(Im)probable stories

Combining Bayesian and explanation-based accounts of rational criminal proof

PhD thesis

to obtain the degree of PhD at the University of Groningen on the authority of the Rector Magnificus Prof. C. Wijmenga and in accordance with the decision by the College of Deans.

This thesis will be defended in public on

Monday 27 February at 16:15 hours

by

Hylke Jellema

born on 21 August 1989
in Sneek
Supervisors
Prof. H. Prakken
Prof. A.R. Mackor

Assessment Committee
Prof. A. Amaya
Prof. C. Dahlman
Prof. H.B. Verheij
‘A main cause of philosophical disease — a one-sided diet: one nourishes one’s thinking with only one kind of example.’

— Ludwig Wittgenstein
# Table of contents

ACKNOWLEDGEMENTS .................................................................................................................. 8

I. INTRODUCTION ........................................................................................................................... 14
   1. Research aim and questions ................................................................................................. 16
   2. The structure of this thesis ................................................................................................. 20
   3. Remarks on methodology ................................................................................................. 23

II. BAYESIAN AND EXPLANATION-BASED ACCOUNTS OF RATIONAL CRIMINAL PROOF ......................................................................................................................... 28
   1. Truth and rationality in criminal proof ............................................................................. 28
   2. Bayesianism ...................................................................................................................... 33
      2.1 The basics of Bayes ..................................................................................................... 34
      2.2 Bayesian models in the courtroom ............................................................................. 39
      2.3 The Bayesian theory of rationality ............................................................................. 47
      2.4 The worries of subjective and evidential probability ................................................. 55
   3. (Inference to the best) explanation-based approaches .................................................. 65
      3.1 Abduction and inference to the best explanation ....................................................... 65
      3.2 Explanation-based models .......................................................................................... 69
      3.3 The irrationality-triviality dilemma .......................................................................... 78
   4. Towards a productive partnership .................................................................................... 85
   5. Conclusion ......................................................................................................................... 93

III. REASONABLE DOUBT AS INference TO THE BEST EXPLANATION .................................. 96
   1. Introduction ...................................................................................................................... 96
   2. Preliminaries: The Simonshaven case and reasonable doubt ....................................... 99
      2.1 A killing in a forest .................................................................................................... 99
      2.2 Proof of guilt beyond a reasonable doubt ................................................................. 101
      2.3 Two explanation-based interpretations of the BARD standard ................................ 105
   3. Why relative plausibility matters ................................................................................... 110
      3.1 Explanatory competition ............................................................................................ 110
      3.2 Relative plausibility and probable truth .................................................................... 113
      3.3 Relative judgment in practice ................................................................................... 116
   4. Why absolute plausibility also matters ......................................................................... 118
      4.1 Plausible innocence explanations ............................................................................. 119
      4.2 The bad lot problem .................................................................................................. 121
   5. Conclusion ...................................................................................................................... 127
IV. IMPLAUSIBLE, INCREDIBLE AND HIGHLY IMPROBABLE STORIES

1. Introduction

2. The Venray murder

3. Rejecting stories with and without evidence

4. Implausible explanations fail to become probable

5. Incredible explanations are told by an unreliable storyteller

6. Highly improbable explanations are obviously false

7. Conclusion

V. THE VALUES OF PREDICTION

1. Introduction

2. On predictions and evidential strength

   2.1 Predictions and accommodations

   2.2 Evidential strength and risky predictions

3. Predicting witnesses

4. Fudged scenarios

5. Fudged evidence, confirmation bias and the argument from choice

6. Explaining predictions

7. Case Study: The murder of Hae Min Lee

   7.1 The case

   7.2 Predicting versus non-predicting witnesses

   7.3 Route talk and severe tests

   7.4 Fudging and the call record

8. Conclusion

VI. EYEWITNESS EVALUATION THROUGH INFERENCE TO THE BEST EXPLANATION

1. Introduction

2. Testimonial inference to the best explanation

   2.1 Sketching the account

   2.2 Storytelling about testimony

3. Being friends with Bayesians

   3.1 Bayesian rationality

   3.2 Heuristic compatibilism

4. Comparing prescriptive accounts
<table>
<thead>
<tr>
<th>APPENDIX II: SAMENVATTING</th>
<th>.................................................................</th>
<th>376</th>
</tr>
</thead>
<tbody>
<tr>
<td>APPENDIX III: SUMMARY FOR THE GENERAL PUBLIC</td>
<td>..................................................................................</td>
<td>392</td>
</tr>
<tr>
<td>APPENDIX IV: PUBLIEKSSAMENVATTING</td>
<td>..................................................................................</td>
<td>393</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

Doing a PhD is strange. Sometimes it is strange in a good way. Over the past years, I was often amazed that I somehow got paid to think, talk, teach and write about philosophy. At other times the strange in-between state of having not-quite-a-job but also being not-quite-a-student was quite demotivating (especially when combined with COVID lockdowns). So, while I am satisfied with the end-result, my PhD experience was something of a mixed bag. However, one thing that I do not have mixed feelings about is the group of people who helped me along the way. The following persons not only helped make this thesis better, without their aid I would likely not have finished it in the first place.

I would first and foremost like to thank my promotors Henry Prakken and Anne Ruth Mackor. Before I began this PhD, I already knew what you were both like as supervisors as you had supervised my master’s thesis. After I was done with that thesis, I sent you an email in which I expressed my gratitude for your kind and constructive supervision. I observed that not everyone is as lucky with their supervisors (I’d heard some horror stories). I also made a prediction: that we would continue to get along well during my PhD. I am happy to say that this was a successful prediction and I stand by the words I wrote in that email. I am grateful for all the time that you devoted to reading my work carefully and critically over these past years. Admittedly, it was not always easy to feel this gratitude when I received yet another round of extensive comments on my work. However, if not for all your help, I wouldn’t be nearly as satisfied with the end result and it made me a better academic. But what I am perhaps most thankful for is that I could always come to you with any problems I had, whether they were professional or personal, in the knowledge that you would offer me understanding, kind words, and good advice.

A close third in the list of people that I should absolutely not forget to thank is Pauline Westerman. Pauline, you had very little impact on what is written in this thesis. When I showed you one of my first chapters, you refused to read it,
exclaiming with disgust that there was a formula on the second page. Nonetheless, you contributed to the successful completion of my PhD in countless other ways. Your first contribution was convincing me to apply for the Research Master of Law, which turned out to be a great choice. After that, you continued to be a mentor to me, and to many of my fellow PhDs as the director of the Graduate School of Law (and these days as the ‘PhD (grand)mother’). You also helped me (and the other PhDs) immensely by organizing online meetings during the pandemic, where we talked about hobbies and other happy things. Those meetings were bright spots during a dull and depressing time. Finally I want to thank you very much for painting the wonderful picture on the cover of this book.

I would also like to express my gratitude to the members of my reading committee: Amalia Amaya, Christian Dahlman and Bart Verheij. Thank you for taking the time to read this thesis, for your helpful feedback and for allowing me to defend it.

Speaking of the defense, I am very happy that Pepa Mellema and Stefan Sleeuw agreed to be my paranymphs. Historically, paranymphs acted as backup for the PhD candidate. The candidate could ask them for advice or to take over the answering of questions. If this would still be the role of the paranymphs, I would have also asked the two of you. After all, you are both not only great friends, but also brilliant philosophers who were kind enough to offer constructive comments on several of the chapters in this thesis.

Many thanks to everyone else who has given feedback on my work or who has been willing to talk to me about my ideas during the writing of this thesis. Outside of those mentioned elsewhere in these acknowledgements, these include at least Floris Bex, Job de Grefte, Leah Henderson, Jelte Hielkema, Peter van Koppen, David Lagnado, Sophie Machavariani, Michael Pardo, Jan-Willem Romeijn, as well as the members of the evidence group, the North Sea group, the epistemology reading group and the audiences at the various workshops and conferences where I have spoken.
This PhD was funded by NWO. I appreciate the opportunity they offered me to pursue my research interests. I’d also like to express my gratitude to the people helped me during the grant-writing process, including Alma Erenstein, Michel Vols, Jeroen ten Voorde and the members of the CWB.

Two persons whom I cannot thank enough for all the help they have offered me (as well as the other PhDs at your faculty) are Anita Kram and Marjolijn Both, from the graduate school.

I began this PhD in a depressing office located in a depressing corner of a depressing building. Luckily the IT law department was kind enough to make room for a stray philosopher in search of a better working space. Special thanks to Aline Klingenberg for making this possible. I would also like to thank my two former office mates (and current friends), Anne Kamphorst and Loes Marinissen. I would not like to thank Tim van Zuijlen. I trained very hard for the 4-mile run of 2018 we did together. It’s impolite to then outrun me with ease while you were suffering from a severe hangover. ;)

About one and a half years ago I ran into Michel Vols on the street. That same day I had had a successful job interview for a position outside of academia and I was pondering whether to accept it. Michel suggested that instead of leaving academia, I should apply at the department of General Law and Legal Skills. I did, and I have not regretted my choice for a single moment. Our ‘sexy sectie’ is filled with fantastic people and I hope to stay here for many years to come. First of all, thank you and cheers to the department as a whole for all the laughs and drinks. Special thanks to Michel for being there that day. I’d also like to express my gratitude to Berend Roorda, for his sunny disposition. Finally, I want to thank all of my office mates who used to and continue to distract me from work (in a good way).

Many thanks to my fellow law PhDs for the teas, beers, walks in the park, shared suffering, deep discussions and silly jokes. There are too many of you to thank individually, but I’d at least like to mention Nischa Vreeling, Mark Beuker, Max de
Vries, Astrid Voorwinden, Kostia Gorobets, Stefan van Tongeren, Martha Buit, Nina Mileva and the members of the PhD council.

I am grateful to all my students. Thank you for all the smart questions and for letting me ramble at you.

I would also like to thank all my friends (including those already mentioned in various places above). I want to begin with my oldest group of friends. A long time ago I lived in a filthy, run-down student flat. During those five years I encountered far too many mice, cockroaches and bedbugs. Luckily, I also met the people of the #Roomboteramandelkrans group: Ad, Annique, Lidwien, Rogier and Xander, with whom I had dinner together almost every night. Since moving out we’ve continued our tradition of having post-Christmas holidays together, where we book a cabin, listen to the Top 2000, play boardgames and eat too many oliebollen during New Year’s eve. Thank you all for this. Special thanks to Rogier for his continuous threats to hit me with a frying pan if I worked too hard these past years.

Another person whom I met in this flat was Marie-Anne. Life is never boring with you around. Thank you for the vacations, theatre plays, coop storytelling, rat-cuddling, good tv shows, bad jokes, silly games, serious moments, Sinterklaas surprises, for trying to teach me how to pronounce ‘th’, and for all of the other good memories.

Melissa, I’m grateful for all the 60-minute coffee dates. I hope that in five years I’ll be reading your PhD thesis. Best of luck!

To my philosophy friends, Stefan, Jesse, Justin, Kritika, Cesar and Caroline: thank you for watching horror movies, discussing an eclectic array of books, cooking, eating, having sleepovers, playing games and, of course, talking about philosophy together.
Many thanks to the bouldering folks. I’m happy to have found a sport that I don’t actively hate. It’s all the more fun to do it together. Special thanks to Dan, for deciding to move to Groningen. It’s great to have you around.

Thanks to everyone with whom I’ve shared a stage in in terrible theatre plays, including Aswin, Maaike, Nadja and Yining.

Ineke, I’m grateful for your attempts to get me to appreciate art. And for everything else of course.

There is also one non-human animal I’d like to thank: my dog Tara. There is nothing that can cheer me up as much as coming home after a long day of work and being greeted by a whole-body wag.

Oan myn heit en mem, Jan en Ant: tige tank foar al jimmë help en leefde. Ik ha in soad gelok bân mei sokke geweldige âlders.

Last of all, I want to thank Sofia. We met shortly before I started my PhD and I could not have gotten through these past years without your love and support. Adequately thanking you for all the wonderful times we had would take far too many words. So let me say one that captures it all: ‘Joepie!’
I. INTRODUCTION

In criminal trials, facts matter. Judges, jurors, lawyers and investigators all regularly face the question: ‘What happened?’ For instance, did the defendant commit manslaughter? Was the witness really present at the crime scene? Is the suspect the source of this blood stain? Such questions can never be answered with certainty; there is always a degree of doubt. Nonetheless, given the available evidence – such as DNA traces and eyewitness statements – the judge or jury may consider the facts of the case proven to a sufficient degree of certainty. When a fact is proven it is, as far as the legal system is concerned, true. The defendant can only be judged guilty and a sanction can only be imposed on them when the facts of the case are proven.

Legal systems typically guide and constrain the proof process through proof rules – for instance by dictating what types of evidence are admissible and by formulating proof standards. There are numerous legal systems, each with their own set of proof rules. However, what many of them share is that the judge or jury who decides on the facts (called the fact-finder) has a great deal of discretion in their evidential reasoning. In other words, the law gives hardly any direction on how fact-finders ought to draw conclusions from the evidence. On the one hand, this is a strength of these systems, as it allows the fact-finder to consider the facts proven in those cases where they feel that this is warranted. On the other hand, this much freedom can also be dangerous. After all, as history has shown over and over again, humans regularly draw mistaken conclusions when dealing with complex sets of evidence. A situation where ‘anything goes’ is therefore undesirable. First, it means that the fact-finder lacks any guidance on how to proceed with the proof process. Second, it leaves us

---

1 These three questions exemplify three kinds of hypotheses common in legal proof: offense, activity, and source level hypotheses. At the offence level, the question is whether the defendant is guilty of a particular offence. At the activity level, hypotheses describe what happened and what those involved did or did not do. Finally, source level hypotheses describe the source of the traces, without specifying how the traces got there (Urbaniak & Di Bello, 2021).

2 For a further explanation of this point, see chapter II, section 1.
empty-handed when we want to criticize a particular decision that we feel is erroneous.

What we require is an account of rational proof, which tells us how to determine what conclusions we may justifiably draw from a given set of evidence. For this we can turn to philosophy. Philosophers have thought extensively about rational evidential reasoning in a variety of contexts, such as daily life, medicine, science and the law. When it comes to rational criminal proof, the two most popular accounts are the Bayesian and (inference to the best) explanation-based.\(^3\) Briefly put, on the Bayesian account we understand and analyze criminal proof in terms of Bayesian probability theory – where beliefs in hypotheses are modeled as probabilities between 1 (expressing complete certainty that the hypothesis is true) and 0 (expressing complete certainty that the hypothesis is false). In contrast, explanation-based approaches are less quantitative. The ‘explanationist’ sees legal proof as a competition between potential explanations of the evidence, often in the form of stories told by defense and prosecution.

Both models have their strengths and weaknesses. Because of this, many have wondered how the two relate to one another and whether they can be fruitfully be combined. In this thesis I argue that a partnership between the two is not only possible, but also desirable. I propose that, currently, neither framework offers a satisfactory account of rational criminal proof. As I explore in further depth in chapter II, both lack a sufficiently clear connection to the aim of truth-finding. For Bayesianism the worry is that, at least some versions of the account, tie what is ‘rational’ insufficiently to the strength of the available evidence. On such versions of the account even patently unreasonable factual decisions may count as ‘rational’

---

\(^3\) These two are not the only approaches to rational proof. For instance, there is the argumentative approach, on which legal proof is viewed as an exchange of arguments for and against a given decision. This approach dates back to Wigmore (1931). In the Anglo-Saxon world this approach was rediscovered by the *New Evidence Scholars* (e.g., Anderson et al., 2005) and was further developed by AI-researchers (e.g., Verheij, 2000; Bex et al., 2003; Prakken, 2004). Another example are non-Bayesian probabilistic approaches, such as belief-function based accounts (Clermont, 2015; Meester, 2020).
and reasonable factual decisions may be judged ‘irrational’. Other, more plausible versions of Bayesianism require that the probabilities put into the model are reasonable in light of the available evidence – so-called ‘evidential probabilities’. However, it is unclear how we should determine what those probabilities are. Explanation-based accounts can complement the Bayesian framework on this point as explanation-based reasoning helps make sense of which evidence is relevant and what conclusions this evidence supports. However, the explanationist account faces its own difficulties. In particular, the process of explanatory comparison has not been spelled out in much detail. A more detailed picture is needed because, some have argued, on some of the more straightforward readings of such accounts, explanation-based thinking will sometimes lead to irrational conclusions. As a result, it is unclear whether, and under what conditions we can accept an explanation as probably true. This is where Bayesian probability theory comes in, which offers a useful language for making precise when explanation-based reasoning leads to accurate outcomes.

1. Research aim and questions

To rephrase the above, in this thesis I aim to show both that the Bayesian and explanation-based account can be combined, and that it is desirable to do so. I therefore suggest that the two can therefore form a productive partnership. By proposing such a partnership, I offer an answer to a well-known conceptual puzzle for explanationists posed to them by Bayesian epistemologists. Briefly put, the puzzle is this:

If explanationism diverges from Bayesianism, it is irrational. But if it does not diverge, it is just Bayesianism in disguise and, therefore, trivial.5

I accept the first horn of this dilemma and reject the second. Explanation-based

---

4 For example, on some interpretations of these accounts we might have to accept a bad guilt explanation as true if the defense failed to provide an alternative explanation.
5 I explain this puzzle in more depth in chapter II, section 2 and further.
thinking can track Bayesian norms without thereby being trivial. It fills in a crucial gap in the Bayesian account, by helping make sense of the evidence while also leading to probably true outcomes (when properly spelled out). The central questions that I seek to answer in this thesis are therefore as follows:

(I) How does explanation-based reasoning help make sense of the available evidence?

(II) When are we justified in accepting an explanation as probably true?

We can divide both questions into a number of subquestions. The first research question relates to how explanation-based thinking helps make sense of the evidence. In other words, how does it aid us in selecting and interpreting the facts of the case? So far most explanationists have focused on how explanation-based thinking helps make sense of evidence that relates directly to the question whether the defendant committed the alleged acts (also known as the ultimate hypothesis). Such evidence includes DNA evidence and eyewitness testimony. In this thesis I add to this discussion by looking at two further kinds of evidence. The first kind is ‘reliability evidence’, which is evidence about the reliability of our ‘regular’ evidence. For instance, how reliable is a particular eyewitness? How we answer this question will influence how much significance we attach to their testimony. This leads to the follow subquestion:

Question I.I: How does explanation-based thinking help us make sense of reliability evidence?

The second kind of evidence that explanation-based thinking helps us make sense of is evidence about whether our set of information (including our evidence,  

---

I am aware of only two exceptions, both of whom have written about their ideas mostly in Dutch. The first is Van Koppen (2011), whose scenario theory includes the evaluation of ‘sub-scenarios’, which can be scenarios about an item of evidence, such as DNA trace evidence or eyewitness testimony. See section 3.2 of chapter II for a further discussion of van Koppen’s theory. The second exception is Rassin (2001), who talks about using scenarios to evaluate witness reliability. See also chapter VI.
hypotheses, arguments) is sufficiently complete. This is an important question in criminal proof. This results in the second subquestion:

Question I.II: How does explanation-based thinking help us reason about the completeness of our set of information?

The second main research question is about when we may accept a given explanation as probably true. The subquestions for this research question arise in response to the standard explanationist account of how we answer this question. On this standard account we are justified in accepting an explanation as probably true only if it is sufficiently ‘plausible’ and if there are no sufficiently ‘plausible’ competing explanations (as it is typically spelled out, plausibility refers to the degree that an explanation displays explanatory virtues, such as internal coherence and fits with the evidence and our background beliefs) (e.g., Pennington & Hastie 1992, 190–199; Pardo & Allen, 2008, 230; Bex 2011, 91–92; Allen & Pardo, 2019). As I will argue in this thesis, this threshold-based idea is problematic. A plausible explanation without plausible competitors is not always probably true. Conversely, an explanation can be probably true without these requirements being met. Whether an explanation is probably true not only relates to whether it and its competitors meet a plausibility threshold, but also depends on other aspects of explanation-based reasoning. For instance, one reason why the threshold idea is problematic is because there can be situations in which one explanation is very plausible and where there is an alternative explanation which only barely meets the threshold for being plausible. The former explanation can then be probable despite having a plausible competitor. Conversely, an explanation may be only barely make the threshold for being plausible, while having a competitor that almost-but-not-quite makes this threshold. The former explanation may then be improbable despite being plausible and having no plausible competitors, especially if there are many almost-plausible alternatives. A more detailed account of how plausibility and probability is therefore needed. This leads me to the first subquestion:
Question II.I: How should the plausibility of explanations influence judgments about their probability?

Another relevant aspect of explanation-based reasoning is the source of an explanation. Explanations may, for instance, be formulated by witnesses, investigators, prosecution attorneys, defense attorneys, or the defendant. Suppose that a witness offers a highly plausible account of what happened but that we have reasons to doubt the veracity of the witness. In that situation the explanation can be improbable despite being plausible. This leads to the second subquestion:

Question II.II: How should the reliability of an explanation’s source influence that explanation’s probability?

Source reliability is therefore not just something that we determine through explanation-based thinking, it is also something that we need to evaluate for explanatory comparison. The same goes for the completeness of our set of information. As I said above, explanation-based thinking can help us determine whether our the information that we have at our disposal is (sufficiently) exhaustive. However, this is also a question that we need to answer when we want to draw justified conclusions about the probability of our explanations. In particular, explanation-based thinking typically involves a small set of quite detailed accounts of what happened. Because of this, it is possible that we overlook one or more plausible alternative explanations. Furthermore, it is not only the set of explanations that is typically limited, so is the available evidence, the arguments that are considered during the trial and so on. Whether we are justified in accepting an explanation as probably true will depend in part on the completeness of our set of information. If this set is insufficiently complete, we may not be justified in believing our best explanation to be probably true. Conversely, if we are sufficiently sure that there are no undiscovered alternative explanations, relevant evidence that we do not possess and so on, then this can raise the probability of our conceived explanations. This brings me to the third and final subquestion:
Question II.III: How should the completeness of our set of information influence judgments about the probability of explanations?

2. The structure of this thesis

This thesis consists of nine chapters (including the introduction and the conclusion). The research questions mentioned above run as threads through these chapters. In the next chapter, chapter II, I set the stage for the rest of this thesis. I begin by reflecting on the notion of rational criminal proof. I then turn to the Bayesian and explanation-based accounts. My discussion of these accounts includes what I see as the most important problems that both accounts face. These problems motivate my compatibilist view. In the final part of this chapter, I sketch the general picture of how Bayesian and explanation-based approaches may go together.

In the chapters that follow, I expand on this general picture by focusing on several underexplored areas of criminal proof where explanation-based thinking provides an especially useful perspective or where its rationality is in question. Each of these chapters is based on a (published or unpublished) journal article and can therefore be read more or less separately from the others. However, even though each chapter stands by itself, they were revised to form a coherent whole and follow a logical progression: in chapter III, I introduce the idea that explanation-based choice involves both an evaluation of the comparative strengths of the scenarios that we have and an assessment of whether there may be plausible scenarios that we did not consider or evidence which we do not possess, which could have overturned the scenarios under consideration. In the following chapters I delve more deeply into both requirements. In particular, chapters III to VI are in part about the question how we choose between the scenarios that we have and chapters VII and VIII relate to the questions of unconceived alternatives and unpossessed evidence.

Chapter III is based on Jellema, H. (2020), The reasonable doubt standard as inference to the best explanation, *Synthese*, 199(1), 949-973. In this chapter I discuss how
proponents of explanation-based accounts should understand the ‘proof beyond a reasonable doubt’ standard. I distinguish between how they usually understand this standard and an ‘inference to the best explanation’ (IBE) interpretation that has so far only been hinted at in the literature. I argue for this latter interpretation using a Bayesian analysis. My main argument is that an IBE-based interpretation deals better with situations of barely plausible, but unfalsified explanations. Furthermore, I improve on this IBE-interpretation by adding a ‘robustness’ condition, which stipulates that we should be sufficiently confident that no further unconceived alternatives and missing evidence exist that could have overturned the case for guilt.7

Chapter IV is based on Jellema (2019). Case comment: responding to the implausible, incredible and highly improbable stories defendants tell: a Bayesian interpretation of the Venray murder ruling. Law, Probability & Risk, 18(2-3), 201-211. In this chapter I deal with the question when we may rationally reject implausible, but unfalsified alternative scenarios offered by the defendant. I show how Bayesian probability theory can help us make sense of the otherwise nebulous Venray ruling by the Dutch Supreme Court, which deals with how courts should respond to such scenarios.8

Chapter V is an updated version of Jellema, H. (2021). The values of prediction in criminal cases. The International Journal of Evidence & Proof, 25(2), 163-179. This chapter discusses the matter of unfalsified scenarios in more depth. It starts from a well-known idea in the philosophy of science that explanations are more strongly confirmed if they survive falsification attempts by testing their predictions than if they only explain known data. I ask whether this is also the case in criminal law and whether there is therefore any merit to the intuition that unfalsifiable scenarios, which do not yield testable predictions are thereby less credible. I draw upon arguments from the philosophy of science on predictivism to offer a precise, Bayesian

7 This chapter therefore deals with research questions I.I and I.III.
8 This chapter further deals with question I.I and starts the discussion of question I.II.
account of why predictions are valuable in criminal proof.9

Chapter VI is based on Jellema, H. (2022a). Eyewitness evaluation through inference to the best explanation. *Synthese.* This chapter discusses another aspect of unfalsifiable scenarios – namely that of evaluating the reliability of eyewitnesses who may put forward such an explanation. When we cannot verify or falsify their scenario, we must sometimes assess how credible the witness themselves is. I argue that such eyewitness evaluation can plausibly be understood as an instance of inference to the best explanation. This approach is complementary to Bayesian accounts of eyewitness evaluation. I argue that the Bayesian account sets the normative standard for rational eyewitness evaluation, but inference to the best explanation provides a practical method of evaluation. I also relate my account to existing frameworks for eyewitness evaluation developed by psychologists.10

In the final two chapters, I turn to the question of evidence and scenarios that we do not have.11 In chapter VII, I explore the matter of unconceived alternative scenarios. I argue that the possibility that there are unconceived, plausible alternatives should sometimes lead to reasonable doubt in criminal cases on both explanation-based and Bayesian accounts. I offer an account of how we may assess whether such alternatives are likely to exist. likely exist. This chapter is based on Jellema, H. (2022b). *Reasonable Doubt from Unconceived Alternatives.* *Erkenntnis.*

Chapter VIII is based on Jellema, H. (submitted). *Reasonable doubt, robust evidential probability and the unknown.* In this chapter, I connect the question whether there are unconceived alternative scenarios, whether there exists unpossessed exculpatory evidence and so on to the question when the fact-finder may justifiably believe that their belief in the defendant’s guilt is supported by the

---

9 This chapter deals with the reliability of the source of explanations and therefore concerns research questions I.II and II.I.
10 This chapter continues the discussion of source reliability and therefore further addresses research questions I.II and II.I. The lion’s share of my discussion of question II.I can be found in this chapter.
11 Together these chapters answer questions I.III and II.II.
evidence. Both criteria concern reasoning about unpossessed information. As a consequence, both suffer from the same problem, namely that it is unclear how we should take unpossessed information into account when making a factual decision. I explore both why and when unpossessed information leads to reasonable doubt, thereby offering a novel interpretation of requirements of robustness and evidential probability.

Chapter nine, finally, offers a conclusion. I answer the research questions of this thesis, reflect on what this thesis has contributed to the study of rational criminal proof and I outline some possible future research directions.

3. Remarks on methodology

Before we move on to the next chapter, I want to briefly remark on some of my methodological commitments. This thesis is a work of analytic epistemology. Legal epistemologists view legal proof as being primarily about obtaining accurate, justified factual conclusions. This is not the only way to look at the criminal law system. We could, for instance, reasonably view it as a conflict-resolving mechanism, an extension of class struggle, a method of preventing arbitrary uses of force by a state or as a way of reducing future crime. Each perspective leads to different criteria for when a criminal law system functions well. On the epistemic view, we evaluate how well the system of legal proof functions in terms of its propensity to produce accurate (i.e., truthful) and justified (i.e., well-supported) factual conclusions.

Over the past decades, the epistemic approach has gained a great deal of interest. This is to a large degree due to miscarriages of justice, where innocent people were convicted and guilty people were acquitted. How may we prevent such mistakes in the future? This brings me to a second methodological point. If one of the aims

---

12 Examples include the Schiedammer parkmurder and the case of Lucia de Berk in the Netherlands (Gill et al., 2018; Posthumus, 2005), Sally Clark in the United Kingdom (Nobles & Schiff, 2005) and Thomas Quick in Sweden (Stridbeck, 2020).
of legal epistemology is to contribute to real-life decision-making, then we have to have some idea how the analyses produced in this field relate to realistic cases. Such cases are usually ambiguous and complex. Yet philosophers often favor simple thought experiments in which the intuitions are clear. Should you throw the switch to redirect the trolley? Does a color-scientist, raised in a black and white room learn anything when seeing the color red for the first time? Legal evidence scholarship also has its fair share of such thought experiments. The most prominent of these are cases relating to *naked statistical evidence* – which feature situations in which a defendant’s guilt is highly probable, but in which we would not want to convict. These cases are constructed so that there is only a single unambiguous item of statistical evidence. Such thought experiments are therefore far-removed from what criminal cases look like in practice. As a consequence, some authors suggest that such ‘weird’ hypotheticals are not especially useful for thinking about rational criminal proof (Allen, 2020; Fratantonio, 2021). I also believe that examples which are overly stylized confuse more than they clarify. That is why I use as realistic examples as much as possible. Nonetheless, note that these examples are still simplified compared to actual cases. Real cases often involve hundreds if not thousands of pages of evidence and arguments. Additionally, when describing such cases, it is difficult to avoid bias, especially when we know the outcome of the case. When I discuss real-life cases I am therefore not arguing that this case ought to have been decided a certain way. Furthermore, by appealing to such cases, I do not want to use our shared intuitions about such cases as evidence for the *correctness* of my ideas. Instead, I intend such examples to be indicative of their *usefulness*: to show that, and illustrate *how* these ideas can be applied to real-life cases. That ideas *can* be applied

---

13 See, however, Spellman (2020) for a defense of the merit of using thought experiments in legal proof scholarship.

14 Using case studies as examples to make a philosophical point is generally problematic (Pitt, 2001, 373). If we choose the case because it illustrates a general point, then we manipulate the data to fit that point. This provides weak support at best. However, starting out with one or more case-studies and then asking what general point it supports is also problematic because it is unreasonable to generalize from one or even a handful of cases.
helpfully to real-life cases is, I believe, the ultimate litmus test for any theory of rational proof.

Another methodological point concerns how this thesis relates to existing legal systems. Some evidence scholars, when publishing in international journals, write about what ‘the law’ says about a particular matter or how a concept like ‘the presumption of innocence’ ought to be interpreted. Such authors are often referring to American federal law (or, sometimes, English law). This has to do with the fact that these are the predominant English-language legal systems and that English also happens to be the contemporary academic lingua franca. Though I engage with many ideas from such thinkers in the coming pages, I do not seek to challenge them on how their national law works or on how it ought to work. Instead, my approach is more akin to that of epistemologists, who often focus on questions such as ‘what is knowledge?’ or ‘what does it mean to understand?’, not with ‘what is knowledge in this country?’ or ‘what does it mean to understand in this culture?’. When I talk about criminal proof, I am similarly not concerned with how specific courts apply particular proof rules. The aim of this thesis is normative, not to describe or to offer a rational reconstruction of specific legal practices. As I mentioned above, and as I will further explain in the next chapter, the study of rational criminal proof begins precisely where legal rules tend to stop – at the question how we draw justified, probably true conclusions from the evidence. The assumption that I, and many other evidence scholars make is that when we answer this question we do not need to consider the particulars of procedural criminal law of various countries.

As a final point, the study of rational criminal proof typically draws on more fields than only epistemology (Park & Saks, 2006). For instance, in this thesis we will encounter ideas from psychology, mathematics, forensic science and legal scholarship. However, there is one field from which I borrow the most in terms of ideas,

---

15 The one exception being chapter V, where I discuss the interpretation of a Dutch court ruling.
and that is the philosophy of science. Criminal evidence scholars have not paid much attention to ideas from this field (or at least not enough, I believe). Yet scientific and legal proof are in many ways similar. For instance, both contexts involve discovering, testing and justifying hypotheses about what happened in a particular case (Schum, 1994). In both we want to minimize factual errors (Burch & Furman, 2019). Finally, the two most influential contemporary accounts of rational scientific proof are Bayesianism and inference to the best explanation – which are also the main accounts in legal proof scholarship. So, it is perhaps unsurprising that ideas from one context can be fruitfully applied to another. However, this is not to say that the analogy between the two fields is one-on-one. For instance, much of science involves repeatable experiments, usually involving quantitative evidence. In contrast, criminal-law requires reasoning about a single case, often using evidence that cannot so easily be put in terms of numbers. When importing ideas from one context to the other, we must be mindful of these differences.

Now that I have explained the questions and methodology of this thesis, it is time to bring the two main characters of this thesis on the stage: Bayesianism and explanationism.
II. BAYESIAN AND EXPLANATION-BASED ACCOUNTS OF RATIONAL CRIMINAL PROOF

This chapter reviews the current work on the Bayesian and the explanation-based frameworks in the law as well as the relationship between them. Bayesianism and explanationism are the dominant accounts of rational proof in legal evidence scholarship. Before I turn to the two models, I first offer some general reflections on the idea of rational proof in criminal cases.¹⁶

1. Truth and rationality in criminal proof

When judges or juries decide whether to convict or acquit a defendant, they first have to come to a conclusion about whether the prosecution has succeeded in

¹⁶ I do not, however, want to spend much time on a wider discussion of legal proof scholarship. Various authors have already written extensively on this history and the central questions that the field has dealt with over the past century or so. I do not want to retread their work here. For those readers who are interested in more comprehensive treatments of the history and key questions of the study of rational legal proof, I can recommend the following works:

proving the facts of the case. These facts underly the legal conclusion of guilt or innocence. For instance, did the defendant kill the victim? This inferential process is called ‘fact-finding’ and it is an essential part of decision-making in criminal cases. In many legal systems, fact-finders have a great deal of freedom in how they go about this task. Their freedom is especially pronounced in continental European systems, where the guiding idea is generally that fact-finders should be constrained as little as possible in how they draw conclusions from the evidence (Twining, 1994, 194-197). Traditionally, the evidence law of Anglo-American legal systems was more detailed than the European, setting stricter limitations on how fact-finders ought to go about drawing factual conclusions. However, over the past century, these systems have moved towards their European counterparts under the influence of an abolitionist wave (Stein, 2005, 109-11). They now contain fewer detailed evidence rules and more discretionary guidelines. As Stein (2005, 109) puts it, it is now “common sense, not common law that functions as [the fact-finder’s] principal guide.”

That fact-finders enjoy freedom from legal constraints does not mean that we can say nothing about how they ought to go about their task, nor that we should say nothing. After all, not every factual decision is correct; jurors and judges can make mistakes. This means that we should have the tools to instruct fact-finders on how to avoid such mistakes and for criticizing their decision when they do make mistakes. It is therefore not enough to tell jurors or judges to ‘use their common sense’. For instance, consider the following example of a judicial error:

Walter McMillian

Walter stands accused of murder. The key evidence against him is the testimony of a man called Ralph. Ralph claims that he was at a gas station when Walter forced him at gunpoint to get in his truck and drive him to a nearby cleaners. He claimed that he could not drive himself because his arm was hurt. When they arrived at the cleaners, Walter went inside. Ralph waited
for a long time. While waiting, he briefly drove elsewhere to buy cigarettes and then returned again. Finally, Walter got back in the car and said that he had killed a store clerk. They drove back to the gas station so that Ralph could retrieve his car. Before letting him go, Walter threatened to kill him if he ever told anyone.

Ralph’s story was highly implausible. After all, why would Walter confess to killing someone and then just let Ralph go? Why did Walter need Ralph to drive when he already drove himself to the gas station? Why did Ralph not just leave when he had the chance? Moreover, Ralph’s testimony was contradicted by that of more than a dozen people from Walter’s church. They testified that they were at a charity fish-fry at Walter’s house on the day of the murder and that Walter had been there all day. This was confirmed by a police officer who had bought fish there that day and had noted Walter’s presence in her logbook. Additionally, there was ample reason to doubt Ralph’s credibility. He had accused two other people of the murder before pointing to Walter. Furthermore, before arresting Walter, the police had Ralph go into a shop where Walter was. But Ralph could not identify him among the shoppers. Nonetheless, the jury found Walter guilty and he would spend six years on death row.

The above example is a simplified version of the real Walter McMillian case, as described in Bryan Stevenson’s 2015 book *Just Mercy*. I will come back to this case several times in this chapter. The reason why I introduce it here is that while the jury’s verdict was quite obviously an error, it did not break any rules of criminal procedure in coming to this verdict. After all, the jury was free in what conclusion to draw from the evidence. The law therefore does not always offer us the right tools for criticizing even the most obvious cases of injustice. However, though the jury did not break any procedural rules when coming to its decision, we may nonetheless criticize their verdict on the ground that it was patently irrational given the evidence in the case. This distinction, between *the law of evidence* and *the rationality of
legal proof dates back to Wigmore (1931, 3), who called the latter ‘ratiocinative process of contentious persuasion’. The former consists of those rules that we might find in a law book or in a court ruling about what kinds of evidence are (im) permissible at trial, how evidence ought to be handled and so on. The latter pertains to how we go ought to, epistemically speaking, go about drawing factual conclusions from a given body of evidence.

What determines whether an inference is rational? On epistemic approaches to rational proof\(^\text{17}\) – such as the Bayesian and explanation-based – whether a proposition is proven depends primarily on if we may justifiably believe that this proposition is true.\(^\text{18}\) Numerous legal scholars have underlined the importance of truth-finding in legal trials (e.g., Nance, 2007, 163; Cleiren, 2008; Park and Saks, 2006, 1030–1031; Thompson, 1991; Ho, 2008; Dworkin, 1985a; Goldman, 2001; Stein, 2005; Allen & Stein, 2013, 567; De Keijser, 2017; Dubelaar, 2018). The United States Supreme Court summarized this idea succinctly when it stated that “[t]he basic purpose of a trial is the determination of truth.”\(^\text{19}\) However, while criminal proof is about discovering the truth\(^\text{20}\), a caveat is in order. In legal terms, the facts

---

\(^{17}\) I use the term ‘epistemic approaches’ to refer to accounts that connect evidential reasoning to the goal of truth-finding. This includes the Bayesian and explanation-based approaches. In contrast, Ho (2021) distinguishes between probabilistic approaches – such as the Bayesian – which cast proof beyond a reasonable doubt in terms of a probabilistic threshold, and epistemic approaches which construe the reasonable doubt rule in terms of justified belief or knowledge.

\(^{18}\) Admittedly, ‘truth’ is a loaded word among philosophers. However, I follow Haack’s (2004, 45) suggestion that we should not get too hung up on what exactly this means. She writes: “[N]o elaborately articulated theory of truth is needed (…) Someone who is trying to find out whether the butler did it, for example, wants to end up believing that the butler did it if the butler did it; that the butler didn’t do it if the butler didn’t do it; and that it’s more complicated than that if it’s more complicated than that.”


\(^{20}\) Of course truth-finding is not the only goal of criminal trials. There are other, non-epistemic values, such as protecting the defendant against arbitrary uses of power by the state, victim rehabilitation and efficient decision-making. Such values can constrain our possibilities for finding the truth. For example, we cannot always use the best evidence available as specific items of evidence may be inadmissible due to considerations of procedural fairness. In some types of trials, the aim of accuracy may even take a backseat to
mentioned in the indictment are either proven or they are not. But we can never establish with certainty whether these facts actually occurred. These are, after all, events that have already happened, which can no longer be directly observed. Hence, criminal proof is always decision-making under uncertainty – we can only reach conclusions about what probably happened. We are therefore after probable truth. So, on epistemic accounts of criminal proof, whether a specific type of inference is rational should be determined primarily by whether such an inference leads to conclusions which are probably true. This is also the fundamental assumption of this thesis.

Epistemic frameworks offer an account of how we go about determining what probably happened and, hence, what is proven. However, one theme in this thesis is that an adequate theory of rational proof should do more than tell us how to draw probably true inferences given a set of evidence. Such evidence and its meaning are not simply given. The act of gathering and interpreting evidence is also a process which can be rational or irrational. As Haack (2012, 206) points out, epistemic theorizing can be helpful precisely for those cases where the evidence is “especially complex, ambiguous, or emotionally disturbing — as it often is in legal contexts”. In other words, a theory of rational criminal proof worth its salt helps us deal with hard cases where ‘common sense’ is not enough, by offering ways of making sense of the evidence and by distinguishing between valid and invalid inferences.21

21 Another important thing that accounts of rational proof can help us do is debias fact-finders. Since the 1960’s an extensive literature has developed studying various types of biases - subconscious, systematic reasoning errors that humans are prone to. For instance, the most infamous of these in legal settings is confirmation bias - “the unconscious tendency to seek out, select, and interpret new information in ways that validate one’s pre-existing beliefs, hopes, or expectations” (Nickerson, 1998). This is one of the leading causes of judicial errors (Gross et al., 2004). One crucial insight of psychologists such as
Having set out what a theory of rational proof ought to do, let us now turn to the first of the two theories that I focus on: the Bayesian account of rational criminal proof.

2. Bayesianism

Bayesianism is an umbrella term for a varied collection of ideas, arguments and frameworks. What they share is the centrality of Bayesian inference to the analysis of evidential reasoning. I will first explain the basic tenets of Bayesianism. Then I turn to the two most important ideas that people refer to when talking about the Bayesian approach to legal evidence. The first is that probabilistic tools – which express the strength of evidence in probabilistic terms – are helpful at trial and elsewhere to support the drawing of conclusions from the evidence. The second idea is that how a rational agent ought to reason about whether facts are proven can best be understood in Bayesian terms. On this view Bayesianism is a theory of epistemic rationality. It is therefore a philosophical account, not a practical one about how evidence should be presented at trial and elsewhere. These two ideas are often conflated. However, as I will argue, the two can be seen as separate enterprises and it is important to distinguish between them as the arguments for and against each idea are distinct as well.

In this thesis I focus on the second idea, that Bayesianism should be considered the primary normative framework for rational criminal proof. I argue that, as it stands, Tversky and Kahneman (1974) and their successors is that reasoning mistakes such as confirmation bias are predictable, rather than random. Therefore, frameworks that offer rational ways of thinking may steer fact-finders away from subconsciously making such systematic mistakes, for instance by explicitly having them consider alternative possibilities.

Amaya (2006, 179-182) offers a more fine-grained distinction between the various uses of the term Bayesianism: as a prescription for action, as a tool for the rational reconstruction of the decision-making process, as an educational tool, as an analytical tool or as a normative model. These finer distinctions map onto the categorization offered here.
the theory has serious flaws. The most defensible version of the Bayesian account interprets the relevant probabilities as ‘evidential’. These probabilities express the degree of belief that is reasonable to hold in the face of the available evidence. However, Bayesianism itself offers no story about how we determine what these probabilities are. Without such a story it is unclear under which conditions Bayesianism leads to probably true outcomes.23

2.1 The basics of Bayes

Bayesianism involves three commitments (Bird, 2017). The first commitment is that agents have “degrees of belief” or “credences” with respect to any proposition. These express how strongly the agent believes the given proposition to be true. A credence of 1 indicates absolute certainty that the hypothesis is true, whereas a credence of 0 means that the agent is completely confident that the hypothesis is false. The second commitment of Bayesianism is that these degrees of belief can be modeled as a probability distribution, satisfying the standard axioms of the mathematical theory of probability. These are most popularly expressed in the form of the Kolmogorov axioms which, stated informally, are:

1. The probability of any event is equal to or greater than 0.
2. The probability that at least one of all the possible outcomes of a process will occur is 1.
3. If A and B are mutually exclusive outcomes, then the probability of either of them happening is the sum of the probability of A happening and the probability of B happening.

According to Bayesians, for an agent to be rational at any given time their personal probability function must obey these axioms. In other words, it must be a probability

23 As I will argue, this problem creates space for combining Bayesianism with explanationism.
The third and final commitment of Bayesianism is that an agent must update their degrees of belief upon acquiring new information by way of Bayes’ theorem, which is a logical consequence of the axioms of probability theory:

\[ P(E | H) \times P_{\text{old}}(H) \]

\[ P_{\text{new}}(H) = \frac{P(H | E)}{P(E)} \]

\[ P(E) \]

\( P(H)_{\text{new}} \) is one’s belief in \( H \) after taking to account evidence \( E \). The formula equates this to \( P(H | E) \), or the posterior probability of the hypothesis \( H \). This is the probability of the hypothesis conditional on — i.e., taking into account — the evidence \( E \). In other words, how probable is the hypothesis given the evidence that we have? On the right-hand side of the formula we find three terms. The first, \( P(E | H) \) is called the likelihood of the evidence conditional on the hypothesis. If we suppose that \( H \) is true, how likely is it that we observe fact \( E \)? Second, there is \( P(H)_{\text{old}} \) or the prior probability of the hypothesis. How probable do we consider the hypothesis to be, before taking evidence \( E \) into account? Finally, there is \( P(E) \), also called the marginal likelihood, which expresses how likely the evidence is, regardless of whether the hypothesis is true. More precisely, when we consider two mutually exclusive and exhaustive hypotheses, this marginal likelihood can be further decomposed into:

\[ P(E) = P(H) \times P(E | H) + P(\neg H)P \times (E | \neg H) \]

When we consider multiple mutually exclusive hypotheses, we can write the marginal probability as follows:

\[ P(E) = P(H_1) \times P(E | H_1) + \ldots + P(H_n) \times P(E | H_n) + P(H_c) \times P(E | H_c) \]

Here \( H_1, \ldots, H_n \) are our conceived hypotheses and \( H_c \) is the catch-all hypothesis, which expresses that ‘none of our conceived hypotheses is true’. We add the catch-all when hypotheses \( H_1, \ldots, H_n \) are not exhaustive. By adding \( H_c \) the set of of
hypotheses exhausts the probability space.\textsuperscript{24}

This version of Bayes’ theorem is quite demanding, since determining $P(E)$ requires one to consider the likelihood and prior of all of one’s conceived hypotheses and that of the catch-all, which is a composite of all alternatives to the hypotheses under consideration. In legal settings a simpler version of Bayes’ formula is more commonly used, called the ‘odds’ version:

\[
\frac{P(H_1 | E)}{P(H_2 | E)} = \frac{P(E | H_1)}{P(E | H_2)} \times \frac{P(H_1)}{P(H_2)}
\]

This equation expresses the relative posterior probability of two hypotheses – i.e., how much more probable is $H_1$ than $H_2$, conditional on the evidence. However, note that we can only calculate the posterior probability of an individual hypothesis from this ratio if the two hypotheses – $H_1$ and $H_2$ – are mutually exclusive and exhaustive. To give an example of what such Bayesian updating entails in the case of the odds version, suppose that our hypothesis is “John stabbed Mary”. Bayesianism presumes that we have some initial probability for both this hypothesis and its negation (‘it is not the case that John stabbed Mary’). Imagine further that we receive some evidence for this hypothesis, for example a witness saying that “John stabbed Mary”. This will typically increase our posterior odds compared to the prior odds, because the likelihood of the witness offering this statement will be much higher if we presume that John killed Mary than if we presume John did not kill her.

Bayesianism interprets the relevant probabilities as degrees of belief or credences. Such an interpretation is sometimes called ‘subjective’ or ‘epistemic’, as it means

\textsuperscript{24} However, determining the likelihood and prior of the catch-all in is problematic. For a discussion of this point, see chapter VII.
that the probabilities express how (un)certain an agent is about the truth of a proposition (Mellor, 2004; Gillies, 2000, 53). In other words, on this interpretation probability is in the mind. We may distinguish this from interpretations on which probability is a feature of the world, which are sometimes called ‘objective’. However, in this thesis I follow Mellor (2004), who argues that ‘physical’ is a more apt name, as epistemic interpretations also fall on a scale of purely subjective to more ‘objective’ (depending on how much space there is for different levels of credence about a proposition to count as ‘rational’).\textsuperscript{25} I will therefore continue to use Mellor’s terminology of epistemic versus physical interpretations of probability.

Physical interpretations can be divided into frequency and propensity interpretations. The frequency interpretation defines the probability of an outcome as the frequency with which that outcome appears in a long series of similar events, such as the number of heads in a set of 1000 coin flips. According to the propensity theory, probability is a tendency inherent in a set of repeatable outcomes. For example, it is a property of a fair coin to land on heads as often as on tails (in the long run). We can therefore assign the probability of heads coming up next as 0.5, even if we have never flipped this coin so far and therefore do not have frequentistic data. However, as many have noted in legal proof, we are not concerned with a series of repeated or repeatable outcomes. Rather, we are concerned with unique events – with what happened in a particular case. This is not to say we never have frequentist data in legal proof – we may, for instance, know the random match probability of a piece of DNA evidence. However, even in the case of DNA evidence, we are in itself not interested in the fact that the DNA profile has been shown to occur with a particular frequency in the population. Rather, we are interested in the probability that this defendant’s DNA would match if he were not the source (Redmayne, 2003, 278). Furthermore, for other evidence, such as eyewitness data.

\footnotesize{\textsuperscript{25} See section 2.4 of this chapter for more on different meanings of the word ‘subjective’.}
evidence, such frequencies are typically not available.  

So, though frequencies may inform our probability judgments, the resulting probabilities are epistemic rather than physical. This adoption of the epistemic interpretation is widespread among legal probabilists (e.g., Cullison, 1969; Finkelstein & Fairley, 1970; Nance, 2016; Robertson & Vignaux, 1993; Kaplan, 1967). As Urbaniak and Di Bello (2021) recently summarized it: “Most legal probabilists agree that the probabilities ascribed to statements that are disputed in a trial—such as ‘The defendant is the source of the crime traces’ or ‘The defendant was at the crime scene when the crime was committed’—should be understood as evidence-based degrees of belief.” The resulting account of legal proof is described by Spottswood (2014, 48). As he explains it, Bayesianism views legal proof as “a series of internal probability adjustments propagating through a vast network of inferential connections. [A] fact-finder identifies a number of key factual questions that are being disputed in a case early in the fact-finding process. Each new piece of evidence either increases or decreases the degree to which the fact-finder believes that each key fact in question is demonstrated or refuted. [...] Then, after all the evidence has been presented, they will choose the winner and loser by applying legal rules to those probable facts.” This picture of fact-finders updating their probabilities is (likely) not intended to be descriptive. After all, people often do not conform to the rules of probability theory when reasoning (Saks & Thompson, 2003). Many therefore regard Bayesianism as an inadequate descriptive theory. Rather, Bayesianism can best be seen as an account of how people ought to reason.

There are two ways in which we can interpret the Bayesian ‘ought’: in a prescriptive

---

26 Additionally, even if they were available, assuming that such frequencies are equivalent to the probability that a witness is telling the truth would lead to the wrong conclusions. See chapter VI for a discussion of this.

27 The concept of probabilities being ‘evidence-based’ is problematic and requires further spelling out. I will say more about the problems surrounding this idea and will offer an account of what this phrase means in sections 2.3 and 2.4 of this chapter as well as in in chapter VIII.
or a normative way. Prescriptive accounts are intended to aid decision-making in practice. On this view Bayesianism is the idea that (Bayesian) probabilistic tools are useful for analyzing and presenting the evidence during the process of legal decision-making. For instance, some suggest that having fact-finders explicitly use Bayes’ theorem during their decision-making helps avoid reasoning errors. Normative accounts are theoretical in nature and describe what it means to draw rational conclusions from the evidence. This version of Bayesianism is a philosophical theory of rational legal proof. To give an indication of what this means, one important idea within this theory is that the main requirement for proof of guilt beyond a reasonable doubt is for the fact-finder to have a high degree of belief, conditional on the evidence, that the defendant committed the alleged acts. These two notions are often conflated, but should not be, as they are distinct ideas and we can embrace one idea while rejecting the other. When we conflate them, we also conflate the arguments for and against each, thereby muddling what Bayesians defend and what Bayesioskeptics attack. In contrast, once we separate them, the benefits and weaknesses of each idea come into much clearer focus. In the next two subsections I explain each idea in turn.

2.2 Bayesian models in the courtroom

The idea that probabilistic tools may be used for analyzing and presenting evidence was popularized during a wave of legal proof that began in the 1980’s and which has become known as the ‘New Evidence Scholarship’ (Lempert, 1986, 440; Welch, 2020). This generation of evidence scholars focused mainly on the scope, limitations, and appropriateness of explicitly using mathematical reasoning about probabilities at trial (cf. Di Bello, 2013, ch 3). Their interest in probabilistic modeling at trial was primarily driven by two factors. First, developments in forensic science over the foregoing decades, in particular in the area of DNA evidence, led to an increase in the use of statistical evidence at trial. This raised the question how such
statistical evidence can be combined with other, non-statistical evidence in order to determine what conclusions the evidence as a whole supports in a given case. Second, the discussion on the use of probabilistic tools was spurred by erroneous convictions following from the *misuse* of statistical models. Examples of such judicial errors include the cases of Sally Clark in the United Kingdom and Lucia de Berk in the Netherlands (cf. Barker, 2017; Dawid et al., 2011; Fenton, 2011; Gill et al., 2018; Schneps & Colmez, 2013). Incorrect uses of statistical evidence led many to ask when such models are useful, how we may help legal decision makers to understand these models and – more generally – how we can help them to avoid the kinds of statistical reasoning errors that led to such false convictions.

There are various ways in which probability theory can support evidential reasoning (cf. Tillers, 2011). For instance, it can be used as an educational method when teaching evidence theory (Nance, 2001, 1607-10). Another, more popular usage – and the one that I will focus on here – is as a tool for presenting the evidence in courtroom settings (Sjerps & Berger, 2013; Fenton, Neil & Berger, 2016). For example, fact-finders may apply Bayes’ theorem to calculate the impact of certain items of evidence – such as a DNA trace – on a given hypothesis – for instance, that the defendant was at the crime scene. A key benefit of using probabilistic tools for this purpose is that they may help detect and avoid probabilistic fallacies. For example, consider the following example, which is a simplified version of the British Sally Clark case (cf. Lagnado, 2021, ch. 9):

*Sally Clark*

Sally was a British lawyer. She had two children, the first born in 1996 and

---

28 In the actual case, there was additional medical evidence which the prosecution used to substantiate the claim that Sally killed her children – such as bruises and blood in their lungs. This evidence ultimately turned out to be both misleading and incomplete (Lagnado, 2021, ch. 9). For current purposes I focus on the statistical evidence in the case, which was also the main point of contention during the subsequent appeals process.
the second in 1998. Both died from sudden infant death syndrome (SIDS). This diagnosis is given when a baby dies before its first birthday and no cause of death can be established. Sally was charged with killing both her children. The key evidence in the case was statistical in nature. The prosecutor argued that the probability of two children dying from SIDS was one in 73 million. They obtained this figure by multiplying the probability of a single occurrence of SIDS (1 in 8,543). According to the prosecution, two cases of SIDS in a single family were therefore so improbable that Sally must have killed them. She was convicted of murdering her children and sentenced to life imprisonment.

The prosecution’s argument given above rests on two well-known probabilistic fallacies. The first is that they assumed that the two occurrences of SIDS were unrelated events whose probabilities could straightforwardly be multiplied. However, this ignores the possibility of genetic and environmental factors making a family more prone to SIDS. In other words, the two events are statistically dependent – and the actual probability of two children dying in this way was much lower (1 in 300,000 according to the defense). Second, this number only expresses the probability of two children dying given that the cause was SIDS, \( P(\text{both siblings died} \mid \text{SIDS}) \). However, the prosecution confused this with the probability that the cause of death was SIDS given that the children died, \( P(\text{SIDS} \mid \text{both siblings died}) \). This confusion is called the prosecutor’s fallacy – a well-known and common probabilistic error. To calculate \( P(\text{SIDS} \mid \text{both siblings died}) \) we also need to look at the probability that the cause was not SIDS – i.e., that Sally killed her children. This requires comparing the prior probability that a mother would kill both her children to the aforementioned prior probability of SIDS.\(^9\) Because double infanticide is extremely rare, the resulting probability of murder would be much lower than the

\(^9\) In this case the conditional probabilities of a double death given SIDS or given Sally killing her children are both 1, and can therefore be left out of consideration. However, in other cases, we might also need to take these probabilities into account.
When we apply Bayes’ formula, we have to take into account both the difference between $P(E|H)$ and $P(H|E)$ and the prior probability of both the hypothesis (such as that the children died from SIDS) and its negation. So, by explicitly using this formula, we can avoid mistakes such as the prosecutor’s fallacy. Additionally, various authors suggest that by reasoning in Bayesian terms, fact-finders may avoid other biases and fallacies, such as confirmation bias and the fallacy of the false dichotomy (Berger & Aben, 2010a; b; c; Rissinger, 2013; Schweizer, 2019; Dahlman, 2020b; Fenton & Lagnado, 2021; Prakken, 2021).

One difficulty for the probability theory in court is that laypeople often have difficulty understanding and performing probabilistic calculations (Dartnall & Goodman-Delahunty, 2006; Amaya, 2006, 166-7). Some have suggested that it might therefore be useful in some cases to provide jurors with a lecture or a tutorial on probability (Koehler, 1992, 31; Prakken, 2018; 2020). However, it is doubtful whether this is enough. For instance, Fenton, Neil & Berger (2016, 16) cite a remark by an unnamed, eminent UK judge that:

No matter how many times Bayesian statisticians try to explain to me what the prosecutor’s fallacy is I still do not understand it and nor do I understand why there is a fallacy.

The judge made this remark despite guidance to UK judges and lawyers on how to avoid the prosecutor’s fallacy. The authors note that, anecdotally, the prosecutor’s fallacy and other probabilistic fallacy continue to be made regularly by both expert

---

30 According to Hill (2004) the actual probability of Clark killing her children was somewhere in the vicinity of 10%.

31 Confirmation bias is the tendency to (a) predominantly search for information in line with the main hypothesis, (b) ignore alternative interpretations, and (c) interpret ambiguous information in line with the main hypothesis (Ask and Granhag, 2005; Fahsing and Ask 2016; Rassin, 2010; Sunde and Dror, 2019).

32 This is the fallacy where a limited number of options are incorrectly presented as being mutually exclusive to one another, or as being the only options that exist, in a situation where that isn't the case.
witnesses and lawyers. Similarly, despite Dutch judges receiving tutorials on probabilistic reasoning, they appear to continue to frequently make the prosecutor’s fallacy when interpreting likelihood ratios of DNA evidence (de Keijser & Elffers, 2012; Prakken, 2018; 2020). Furthermore, even if fact-finders could perform simple probabilistic calculations themselves, another difficulty is that more complex calculations can be needed, especially when dealing with multiple items of evidence. As we just saw with the case of Sally Clark, we often cannot proceed by calculating the impact of each item of evidence one at a time as there may be interdependencies. The more dependencies there are between various items of evidence, the more calculations are needed to determine the probability of any hypothesis. This is known as a ‘combinatorial explosion’ (Callen, 1981, 10-15; Allen & Pardo, 2019b, 38; Amaya, 2015, 83-86). Most contemporary proponents of Bayesian modeling therefore no longer recommend that calculations should be done by fact-finders themselves. Instead, they suggest the use of Bayesian networks, which can be constructed by experts (Fenton, Neil & Lagnado, 2013; Taroni et al., 2014). A Bayesian network is a graphical representation of the probabilities and probabilistic connections between a set of propositions (Nielsen & Jensen, 2009). For instance, consider the following example of such a network:
Here the relevant piece of evidence is that the defendant who was previously poor, suddenly obtained a great amount of money. We have two potential causes of this evidence. The prosecution claims that the defendant obtained this money through fraud. The defense on the other hand posits that he inherited this money from a deceased relative. The table next to each node – the *node probability table* or *conditional probability table* – gives the conditional probabilities of each hypothesis given the hypotheses that this hypothesis is probabilistically dependent on. For instance, if we know that the defendant received an inheritance, then this reduces the probability that they committed fraud – after all, the financial incentive would be much smaller. It is possible that the defendant both received an inheritance and committed fraud, but it is quite unlikely. Bayesian networks can be used to calculate the probability of a hypothesis given a set of evidence. For example, in the above network we might be interested in the probability that the defendant committed fraud given that they suddenly became rich: \( P(\text{FRAUD} \mid \text{SUDDEN RICHNESS}) \). To
this end, we enter the information that SUDDEN RICHNESS = T in the network. The relevant calculations can then be performed by software designed for that purpose, such as GeNlE, SamlAm or AgenARisk.\footnote{See, respectively, www.bayesfusion.com, reasoning.cs.ucla.edu/samiam/ and www.agenarisk.com.}

Bayesian networks have been extensively studied as a tool for aiding probabilistic reasoning in legal situations (Dawid & Evett, 1997; Taroni et al., 2006; Taroni et al., 2006; Vlek et al., 2013; Vlek, 2016). They are typically used to analyze a specific part of a case that has to do with forensic evidence such as DNA evidence or to the modus operandi of several crimes (Evett & Weir, 1998; De Zoete et al., 2015; Fenton, Neil & Berger, 2016). Indeed, even skeptics of Bayesianism tend to concede that Bayesian modeling can be useful when presenting evidence that is statistical in nature (e.g., Stein, 1996; Allen, 1997, 258). However, some Bayesians go further and suggest that these networks might be used to model \emph{entire cases}. One benefit of this idea is that a single model can help combine statistical and non-statistical evidence and that it can allow fact-finders to understand the evidential impact of specific assumptions about the (relations between the) evidence (Donnelly, 1997, 304-5; Fienberg, 1997, 312). These benefits were illustrated by Kadane and Schum (1996) who modeled the American ‘Sacco and Vanzetti’ murder case using Bayesian probability theory to understand the relationships and impact of the evidence in the case. Recent work has focused on how the construction of such larger Bayesian networks can be standardized and thereby made easier (Lagnado et al., 2013; Fenton et al., 2013; Vlek et al., 2013; 2014).

Despite the enthusiasm of its proponents, courts in various countries have responded negatively to the idea of using Bayesian probability theory to analyze entire cases. For instance, the English Court of Appeal ruled in the 2010 R v T case that Bayes’ theorem should not be used in evaluating evidence, except for DNA and
‘possibly other areas where there is a firm statistical base’.\textsuperscript{34} In the United States, the California Supreme Court, took up a comparable position in the People v. Collins decision, saying that “[m]athematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not [be allowed to] cast a spell over him.”\textsuperscript{35} Dutch courts have similarly responded negatively, with one court suggesting that whether guilt has been proven is not ‘a question that is suitable for probabilistic calculation’.\textsuperscript{36}

There are a number of explanations that we could give for the negative responses of courts towards Bayesian modeling. One worry is where the numbers that we enter into these networks come from (Jensen & Nielsen, 2007; Amaya, 2015, 83, 85; Vlek, 2016, 24; Prakken & Meester, 2017; Allen & Pardo, 2019b). For many types of evidence, such as eyewitness evidence, it is not clear what we should base the relevant numbers on. For other types of evidence we do have frequentist data to rely on, but many numbers would be needed and these figures would have to be provided by domain-experts based on their expertise. This would make the building of a Bayesian network highly labor-intensive. Furthermore, even if we have relatively objective data, such as the likelihood ratio of a DNA trace sample, this often leaves out things such as rates of laboratory errors and crime scene errors. This may lead to misinterpretations of how objective this ratio actually is (Butler, Kline, & Coble, 2018; Thompson, 2009). As the England and Wales Court of Appeal put it in the R v Adams case: “the apparently objective numerical figures conceal the element of judgment on which it entirely depends”.\textsuperscript{37} Some Bayesians therefore suggest that, rather than using networks as a tool for calculating the ‘correct’ probability of the ultimate hypothesis, we can best understand them as a method for

\textsuperscript{34} England and Wales Court of Appeal, Criminal Division. (2010) R v T. 2439. For a discussion of this ruling, see the special issue of Law, Probability and Risk on the R v T case (2012, Vol. 4, No. 2).

\textsuperscript{35} California Supreme Court. (March 11th, 1968). People v. Collins, 68 Cal.2d 319 (), 497


\textsuperscript{37} England and Wales Court of Appeal, Criminal Division. (1996). R v Adams. 2 Cr App R 467.
exploring how various assumptions about the strength of the evidence and the connections between the different items of evidence lead to different factual conclusions (e.g., Fenton, Neil & Lagnado, 2013; Fenton & Neil, 2020; Fenton et al., 2020). An expert could, for instance, use a Bayesian network to explore which evidence is the most telling in the case. The aim is then to elucidate the structure of various evidential arguments. For the proponents of this view, where the numbers come from is less worrisome as these numbers do not represent the ‘actual’ strength of the evidence. However, further empirical research will have to show to what degree Bayesian networks help people understand the structure of the evidence.

To summarize the above, using Bayesian probability theory to analyze evidence may help avoid fallacies and biases. This idea has led to a great deal of work on how to make such methods easier to use. Nonetheless, courts remain hesitant to allow the use Bayesian modeling outside of limited areas. Some Bayesians respond that courts are misguided in their skepticism, since Bayesianism is ‘only common sense’ or ‘rational’ (cf. Fenton & Lagnado, 2021). However, such remarks hint at a different conception of Bayesian probability – namely as a standard of rationality. This view is distinct from the idea that we should use probabilistic tools at trial. I now turn to the normative Bayesian account of rational proof.

2.3 The Bayesian theory of rationality

Bayesianism is not only influential in legal proof but also in other areas such as medical diagnosis or judgments of risk (Hardman, 2009). But the field where Bayesianism has had the biggest impact is arguably is in the study of scientific reasoning (Earman, 1992; Howson & Urbach, 2006; Talbott, 2016). Yet epistemologists and philosophers of science who defend Bayesianism are typically not concerned with the specific tools that scientists or others should use when drawing factual decisions. For them, Bayesianism is a philosophical theory which provides norms that govern our degrees of beliefs, including how we ought to change these degrees of
belief in response to a varying body of evidence (Lin, 2022). Meeting these norms does not need to involve probabilistic calculations. Similarly, Bayesianism is sometimes used in legal proof scholarship to refer to an epistemic theory of evidential reasoning. In this section I examine this version of Bayesianism in the context of criminal proof.

Bayesians in the domain of legal evidence often do not distinguish between the normative theory of Bayesianism and the prescriptive suggestion that probability theory is a helpful tool for decision-making. However, at least part of the discussion about legal Bayesianism is philosophical in nature. According to this normative version of Bayesianism, we can best model how a rational fact-finder should reason and when particular propositions are proven by means of Bayesian probability theory.38 For example, Bayesians have extensively discussed how we may best conceptualize the legal standards of proof (Kaplan, 1967; Kaye, 1988; Goldman, 2001; Cheng, 2013; Schwartz & Sober, 2017; Puddifoot, 2018; Hahn & Hartmann, 2020).

Another instance of this strand of work are Bayesian models of how assessments of the reliability of witnesses, experts and other sources of information ought to impact our inferences (Friedman, 1992; Fields, 2013; Dahlman & Wahlberg, 2015; Merdes, Von Sydow & Hahn, 2020). Other topics to which Bayesian analyses have been applied include the relevance of non-discriminating evidence (Fenton et al., 2014), when responsibility may legally be attributed to a person (Chockler et al., 2015) and how we may prove discrimination (Bonchi et al., 2017). While such work can be developed further into prescriptive models that support decision-making, this is not their primary aim. Their goal is to help understand the nature of rational proof.

38 The Bayesian account of rational proof is the most popular subspecies of the position called ‘legal probabilism’. According to legal probabilism, legal proof “is to be governed by the mathematical principles of standard probability theory, and the decision criterion in juridical fact-finding is to be modeled with probabilistic tools” (Stefaniak & Urbaniak, 2018). In this thesis I will not discuss other, less influential forms of legal probabilism, such as those that use belief functions (Clermont, 2017; Nance, 2019a).
To get a better grasp on what the (normative) Bayesian theory of rational proof amounts to, it is helpful to look at some of the objections that have been raised against it. I want to focus on two in particular here: the problem of naked statistical evidence and the prior problem. The problem of naked statistical evidence is a so-called proof paradox, which targets the Bayesian understanding of the proof standards (i.e., the standards that set out when the guilt of the defendant is proven). According to a (naïve) probabilistic interpretation of the proof standard, we may convict the defendant if and only if the probability of their guilt is sufficiently high. The problem of naked statistical evidence is usually presented in the form of thought experiments in which the probability of guilt meets this threshold for conviction, but where many people nonetheless do not want to convict. A famous example is that of the prisoners in a yard:

One hundred prisoners are in the prison yard. Camera footage tells us that ninety-nine prisoners jointly kill a guard, but one of them did not partake in the killing, though the camera recording is not clear enough to tell us which one. Can we now convict any (or all) of these prisoners for the killing?

For many people, the intuitive answer is ‘no’, even though the probability of each individual prisoner killing the guard is 99%. This presents a problem for the probabilistic interpretation of the proof standard, as it would lead to conviction in such a thought experiment. Probabilists have mostly responded to this intuition by adding further conditions on conviction (e.g., Nunn, 2015; Krauss, 2019; Di Bello, 2019b; Dahlman, 2020a; Ross, 2020; Enoch & Spectre, 2021; Littlejohn, 2021; Amaya (2015, ch. 2); Urbaniak & Di Bello (2021); Mackor, Jellema & Van Koppen, 2021).

Various other objections have been raised against the theory. I will leave most of these aside in my thesis, because they are not germane to my goals. For an overview of such arguments, see Amaya (2015, ch. 2); Urbaniak & Di Bello (2021); Mackor, Jellema & Van Koppen, 2021). I will have more to say about such proof standards in later chapters (especially chapter III). However, I will not engage much with the matter of naked statistical evidence, as I neither intend to defend the probabilistic conception of proof, nor have anything substantial to add to the extensive existing literature on this topic.
The second worry for the Bayesian account that I want to look at is the prior problem. It begins with the observation that, in order for Bayesian inference to proceed we have to begin with a prior probability of more than zero. If we began the trial with a prior belief in the defendant’s guilt of zero, this would indicate complete certainty about the defendant’s innocence. No new evidence could then make any difference. Whatever the likelihood ratio of the evidence, multiplying it by the prior probability of zero will end up with a posterior probability of zero. So, guilt can never be proven, regardless of the strength of the evidence. However, suppose that we assign a non-zero value to the hypothesis that the defendant is guilty. According to many, this would conflict with the presumption of innocence – a procedural protection afforded to defendants in many countries (Tribe, 1971, 1368-72; Posner, 1999, 1514; Amaya, 2015, 85). After all, it would mean that we are at least somewhat convinced that the defendant is guilty at the outset of the trial. Admittedly, some question whether this is the proper way of interpreting the presumption of innocence. Others have offered proposals for setting the prior probability of guilt in a way that is consistent with the presumption of innocence. However, no generally acceptable solution to this problem has been found (Dahlman, 2018; Fenton et al., 2019).

The naked statistical evidence problem targets the view that whether guilt has been proven depends solely on the Bayesian posterior probability of guilt. The suggestion that the presumption of innocence and the Bayesian prior are in conflict is an

---

41 Another proof paradox is the conjunction paradox – which relates to the question whether, on the probabilistic interpretation, each element of a crime has to be proven to a specified degree of probability or the conjunction of these elements (Cohen, 1977). See Schwartz & Sober (2017), Spottswood (2016) and Wittlin (2019) for recent discussions of this paradox. For discussions of the proof paradoxes in general, see Ross (2020) and Spottswood (2021).

42 For instance, Robertson and Vignaux (1993) suggest that this presumption is about requiring a high posterior probability of guilt. Another response is given by Picinali (2014) who argues that the presumption of innocence is a matter of procedural rights.
objection to the idea that rational proof occurs through probabilistic updating, where we update our prior probability in guilt by applying Bayes’ theorem upon receiving new evidence. On the view that is being objected to, Bayesianism therefore defines the standard of proof that has to be met and how we ought to go about meeting this standard. Bayesians have responded to worries such as that of naked statistical evidence and the problem of the prior by slightly amending their theories, but retain the idea that Bayesianism offers an adequate description of the essence of criminal proof – namely that we aim for a high degree of belief by conditionalizing a prior belief on the evidence.

The (normative) Bayesian account has been highly influential and widely discussed in legal proof scholarship (Ho, 2015). As said earlier, Bayesianism is also a leading account of rational evidential reasoning in other evidential contexts. We can partially explain Bayesianism’s broad appeal by pointing at the fact that the theory provides an elegant, precise and versatile framework for understanding evidential reasoning. Additionally, for many it matches various fundamental intuitions about evidence and evidential strength, as highlighted in the famous adage by Laplace (2012) that it is but ‘formalised common sense’. Finally, Bayesianism’s normative appeal also derives from philosophical arguments. These arguments purport to show that deviating from the rules of probability theory is epistemically irrational. It is worth considering these philosophical arguments in more depth, as they establish a prima facie connection between Bayesianism and the goal of truth-finding which I discussed in section 2.1 of this chapter.

The best-known argument for the epistemic rationality of Bayesian is the Dutch Book Argument ( Hájek, 2009; Vineberg, 2016). This argument draws an analogy between the degrees of belief that a person has in a given set of propositions and

---

43 However, this updating does not necessarily have to involve assigning numbers to hypotheses. As Spottswood (2019, 78-9) argues, “we can talk about the notion that jurors have varying levels of credence, and even quantify such levels mathematically in our theorizing, without making the silly assumption that jurors typically attach explicit numbers to their own levels of confidence throughout the process.”
hypothesis that this person therefore ought to accept about the truth-value of these propositions. What the argument establishes is that someone who deviates from the prescripts of Bayesian probability theory is vulnerable to a Dutch book—a series of bets which, if accepted, would mean that this person is certain to lose money. Conversely, someone who does follow these prescripts is not exposed to a Dutch book.\footnote{To give an example, a person who assigns a higher probability to a conjunction of two events than to these events individually commits the conjunction fallacy. A well-known example of this is a person believing it to be more probable that Linda is a bank teller and a feminist, than that she is a feminist (Tversky and Kahneman, 1983). As Hahn (2014, 3) explains, anyone who commits this fallacy with respect to propositions A and B, should agree to a series of bets on the truth of A and B where they would lose money, no matter the actual truth-value of these propositions, result in them losing money.} The conclusion of such Dutch Book Arguments is that it is rational to be a Bayesian. Admittedly, the betting analogy may seem strange, at first. After all, it would be deeply problematic if judges or juries would accept any bet about whether the defendant is guilty—this is not the type of thing that one ought to bet on.\footnote{See also Acree (2021, 65) who argues for this point with respect to scientific inquiry. Acree also points out that it seems strange to bet on the truth of unobservable phenomena, in part because we could never settle such a bet. The same is true for legal settings—we can never know whether the defendant truly did commit the alleged acts.} However, we need not think of Dutch books in terms of people’s actual betting dispositions. We can also view them as hypothetical bets—i.e., the bets that one would accept if one were betting (Eriksson & Hájek, 2007). In fact, we can frame any kind of decision as such a hypothetical bet. As Ramsey (1926, 85) puts it:

[A]ll our lives we are in a sense betting. Whenever we go to the station we are betting that a train will really run, and if we had not a sufficient degree of belief in this we should decline the bet and stay at home.

Similarly, when a decision-maker considers how probable it is that the defendant committed a crime, they are, in a sense, betting on that person’s guilt, with the stakes being the positive utilities of convicting a guilty person (or acquitting an innocent person) and the negative utilities of convicting an innocent person (or acquitting someone who is guilty).
Some legal scholars suggest that the Dutch Book argument supports Bayesianism as the correct standard of rationality for legal evidence. For instance, Fenton and Lagnado (2021) invoke this argument as key support for the Bayesian approach in law. However, though it is influential, there are various reasons to be dissatisfied with the Dutch Book Argument as a foundation for Bayesianism. First, deviating from Bayesian prescripts may also have advantages, such as getting to the truth efficiently (Gigerenzer, 2004; Gigerenzer & Engel, 2006; Douven, 2013; 2021). Such advantages may offset the risk of being exposed to a Dutch book. Second, it is unclear what the betting analogy shows, exactly. The connection between an agent’s degrees of belief and bets that this agent would accept is problematic (Eriksson & Hájek, 2007). For instance, a person may be morally opposed to betting or them betting on something may influence their dispositions because they want to win the bet. In this sense actual betting dispositions are neither good ways to define nor to measure the strength of someone’s beliefs. As said, we might also think of betting dispositions in hypothetical terms, as bets that an agent would (or should) accept in a counterfactual world. However, some have doubted whether there is a measurable, non-circular way of spelling out this idea of a hypothetical bet (Eriksson & Hájek, 2007). Additionally, we may ask what the pragmatic norm of avoiding a sure loss has to do with epistemic norms of rationality. After all, we are not concerned with (hypothetical) monetary gain but with having accurate beliefs (Briggs, 2015). Finally, betting is a predictive activity, whereas legal proof is about postdiction – proof of past event. It is therefore unclear whether the Dutch book argument relies on a meaningful analogy (Hamer, 2004).\footnote{See Sleeuw (2020) for a general discussion on when analogical arguments are (un)sound.}

Due to the problems surrounding the Dutch book argument, many contemporary Bayesian epistemologists attempt to justify the framework more directly, by invoking considerations of accuracy (e.g., De Finetti, 1974; Rosenkrantz, 1992; Joyce,
In effect, such authors aim to show that following the prescripts of Bayesianism maximizes the accuracy of one’s beliefs – one’s epistemic utility – in all possible worlds compared to alternative approaches. These results provide a more promising normative justification for Bayesianism compared to the Dutch Book Argument, as they offer a clear connection with the aim of drawing accurate conclusions. The most influential proponent of this approach with respect to legal proof is Alvin Goldman (1999, ch. 9), who showed that following the prescripts of Bayesianism is a practice that is likely to help us end up closer to the truth – at least in areas where our degrees of belief match available ‘objective’ chances.

The upshot of the above arguments is that, if one’s aim is to maximize the accuracy of one’s beliefs, it is more rational for a fact-finder’s beliefs to conform to the basic tenets of Bayesianism than to not conform to them. For instance, it is better to update one’s probabilities in conformity with Bayes’ theorem when receiving new evidence than to update it in a different way. This is an important result, as it connects Bayesianism to the goal of accurate fact-finding. However, there is reason to doubt whether this connection is strong enough to ground the Bayesian account. I now want to turn to what I consider to be the greatest difficulty for Bayesianism, namely whether the account does not connect what is rational too much to the fact-finder’s beliefs, thereby breaking the link with the aim of obtaining accurate beliefs. I want to suggest that a way of solving this worry that has recently become popular, by interpreting the relevant probabilities as ‘evidential’ probabilities, leads to another problem. This problem, in turn, opens up the possibility of combining Bayesianism with the explanation-based account of rational criminal proof. But before I turn to these issues, let me briefly note that the normative version of Bayesian is distinct from the prescriptive version that I discussed in the previous subsection.

47 This ‘accuracy-first epistemology’ is an ongoing project by epistemologists (Douven & Schupbach, 2015; Pettigrew, 2016).

48 However, in the next subsection I argue that the notion of credences matching ‘objective’ probabilities is a problematic idea within the context of criminal proof.
We can be committed to the idea that rational proof is best understood in Bayesian terms, without assuming that the process of proof can be improved by using probability theory in decision-making contexts. Similarly, we can use probabilistic tools without being committed to the Bayesian view of rationality. For instance, as many have noted, the term ‘Bayesian network’ is a misnomer. Such networks are not necessarily Bayesian, because they are neutral with respect to whether the probabilities therein are epistemic or are, for instance, frequentist in nature (Korb & Nicholson, 2010, 22). Because the two notions of Bayesianism are distinct, arguments against one idea are not necessarily problematic for the other view. For instance, proponents of Bayesian rationality do not need to be disturbed by the rejection by courts of Bayesian modeling, as long as there are other ways of getting people to follow their precepts. Conversely, proponents of Bayesian modeling at trial have no need to engage with the philosophical arguments against Bayesianism such as those that I turn to now.

2.4 The worries of subjective and evidential probability

According to most proponents of Bayesianism, the relevant probabilities in legal proof represent the fact-finder’s subjective degrees of belief – or ‘epistemic probabilities’ as they are sometimes called. This interpretation avoids many problems with physical interpretations. For instance, it makes the required numbers easier to get by – we only have to ask how probable the fact-finder considers specific hypotheses to be. However, some have suggested that this interpretation leads to a worry, namely that the account is ‘too subjective’. According to its skeptics, Bayesianism ”allow[s] people to render opinions (...) on what is sometimes no better than a guess.” (Risinger, 2013, 71). Similarly, Allen (2017, 138) is concerned about “uncontrolled subjectivity” and that probabilities “could be any numbers at all, and they need not be constrained in any way by the quality of the evidence” (Allen & Pardo, 2019b, 10). In other words, if the relevant probabilities represent the beliefs
of the fact-finder, what prevents ‘garbage in, garbage out’, where the process of Bayesian updating happens correctly, but is based on irrational beliefs, leading to irrational outcomes?

While the above is a well-known objection to the Bayesian account of rational proof, not all authors are clear what they mean by it. To understand the worry, it is helpful to first make a distinction between two meanings of the term ‘subjective’. I take this distinction from Ian Hacking (2001, 131). As Hacking points out, the word ‘subjective’ may firstly mean that a statement expresses one’s personal opinion (“I believe that…”). I call this ‘subjectivity-as-belief’. Second, we may mean that statement is not supported by the available evidence (“that is just your subjective opinion!”). I call this ‘subjectivity-as-unfounded-belief’. As said, most Bayesians presume that the relevant probabilities are degrees of belief (e.g., Kaplan, 1967; Cullison, 1969; Finkelstein & Fairley, 1970; Robertson & Vignaux, 1993; Bird, 2017; Schweizer, 2019; Hunt & Mostyn, 2020). This is subjectivity-as-belief. However, what the above critics seem to be talking about is subjectivity in the second sense – where the fact-finder’s degree of belief may be considered rational even if it is not supported by the evidence. But why should subjectivity-as-unfounded-belief be especially worrisome for the Bayesian? True, people who follow the prescripts of the Bayesian account can end up with irrational beliefs. However, this is also the case for competing frameworks. For instance, an agent may reason in terms of competing explanations, as the explanation-based account dictates, and nonetheless end up drawing inferences that are wholly unsupported by the evidence. Even in frequentist statistics there can be different ways of modeling a given set of evidence and, as a result, subjective biases can sneak in (Acree, 2021, 273-4).

One way to understand the above worry is that Bayesianism offers especially much room for such irrational beliefs. This objection targets a specific form of the account, sometimes called ‘pure subjective Bayesianism’ (Lutz, 2020). According to

---

49 See section 3 of this chapter for a discussion of such accounts.
this version of the framework, the only requirements for an agent to be rational are that their credences form a probability function (i.e., that they meet the Kolmogorov axioms mentioned in section 2.1 of this chapter) and are updated in accordance with Bayes’ theorem. To put it another way, as long as the fact-finder’s credences are probabilistically coherent, they are rational in a Bayesian sense. Obviously, on such an account, the relevant probabilities can be almost anything and do not need to correspond to the strength of the evidence. For example, consider the case of Walter McMillian again. Imagine a juror in this case who starts out by assigning a high probability, e.g., 0.99, to the proposition that Walter is the killer. Suppose that this juror considers the numerous witnesses who testify that Walter is innocent as completely unreliable. Furthermore, he overlooks the various gaps in the story of the key witness. As a result of these beliefs, he believes that Walter almost certainly killed the victim. Such a juror’s belief is clearly irrational given the evidence, but it does not necessarily violate any of the aforementioned constraints of probability theory.

The arguments mentioned in the previous subsection, such as the Dutch book argument, support the idea that having coherent degrees of belief is conducive to obtaining accurate beliefs. All else being equal, someone whose degrees of belief

\[50\] The question how the priors ought to be set is a well-known problem for the Bayesian account. Because these probabilities concern one-time events there is no obvious, frequentist basis on which to base our priors (Friedman, 1997, 276). Probability theory sets no constraints on what the prior should be (Allen & Pardo, 2019b; Amaya, 2015, 82). So, nothing prohibits the factfinder to set a prior degree of belief in the defendant's guilt to 0.99, or even to 1. Some Bayesians reply to this that such extreme priors get “washed out” by incoming evidence as long as they are not 0 or 1. This means that the priors of individuals converge on the same probability - as long as individuals use the ‘correct’ or at least intersubjectively agreed upon likelihoods for the evidence (Edwards, Lindman, & Savage, 1963; Hawthorne, 1994; Goldman, 2001). So, it does not matter that individuals can start with completely different priors, as they will end up in the same place. However, this response is unconvincing. Depending on the situation, this “washing out” can take a very long time and a great amount of evidence whereas in legal settings time, resources and available evidence are limited (Amaya, 2015, 83; Godfrey-Smith, 2003, 209-10). Furthermore, there is often neither a ‘correct’ likelihood for a given piece of evidence in criminal cases, nor a reason to presume that people will always agree on these likelihoods (Godfrey-Smith, 2003, 209-10; Jellema, 2022a).
form a probability function and who updates these degrees of belief in line with Bayes’ theorem is more likely to hold accurate beliefs than someone whose degrees of belief do not conform to these requirements. However, the connection between being rational in a Bayesian sense and the aim of obtaining accurate beliefs is weak at best if these are the only requirements.\textsuperscript{51} As we saw just now in the example of the Walter McMillian case, to be rational in a (pure subjective) Bayesian sense can go hand in hand with even the most irrational of decisions. Additionally, according to some, pure subjective Bayesianism is ‘too subjective’ in another sense: it grounds what is rational in the coherence of the degrees of belief of the fact-finder. But this arguably gets things the wrong way around; pure subjective Bayesianism presumes what it should provide. To give an example of this point, as I will discuss at various points in this thesis, Bayesians typically understand the proof of guilt beyond a reasonable doubt standard in terms of a probability threshold of, for instance, at least, 95\% (often alongside other requirements).\textsuperscript{52} However, if the relevant probabilities are the fact-finder’s degrees of belief then what the proof standard would effectively boil down to is that, in order for the defendant to be found guilty, the fact-finder needs to strongly believe that the defendant is guilty. However, as Laudan (2006, 79) argues, this is not actually a proof standard because such a standard should tell the fact-finder \textit{when} they may believe that the defendant is guilty. But the (pure subjective) Bayesian interpretation presumes this belief on the part of the fact-finder. So, the Bayesian account is “in danger of putting the cart before the horse” (Cox, 2000, 323).

The above worries target pure subjective Bayesianism. However, Bayesians do not need to be committed to this version of the theory. While they may claim that the minimal coherence requirements just mentioned are \textit{necessary} for one’s beliefs to be rational, they do not have to assume that these criteria are \textit{sufficient}. As Hunt &

\textsuperscript{51} See section 1 of this chapter for a discussion about the aim of obtaining accurate beliefs, or ‘truth-finding’.

\textsuperscript{52} I discuss the proof standard in particular in chapters III, VII and VIII.
Mostyn (2020) point out, the Bayesian can propose that while (pure subjective) Bayesianism is not a complete account of rational proof on its own, it is at least part of one. On this view, the Bayesian coherence constraints are essential for legal proof, but they need to be amended with additional requirements for rational degrees of belief. One way of spelling out this idea is known as ‘objective Bayesianism’. Objective Bayesianism is still ‘subjective’ in the sense of subjective-as-belief; it considers the relevant probabilities to be the agent’s degrees of belief. However, it is ‘objective’ in the sense that it adds additional constraints on rational credences, thereby trying to limit the space for subjectivity-as-unfounded-belief. Some suggest that this is a more promising account of rational legal proof than ‘pure subjective Bayesianism’ (Di Bello, 2019a). Objective Bayesians have suggested several additional requirements for one’s credences to count as rational. For instance, two commonly offered constraints are: (a) The principle of indifference, according to which one should accord equal probability to the propositions under consideration if one’s evidence does not discriminate between them (Williamson, 2010), (b) The Principal Principle, according to which one ought to set one’s credences to conform with known objective chances – e.g., frequencies (Lewis, 1980). For example, if we know that the objective chance of some event is 0.5, then we should set our degree of belief to 0.5. However, though it may seem promising at first, objective Bayesianism does not get the Bayesian out of the woods. The conditions proposed by objective

This is also the basis for my later proposal to wed the Bayesian and explanation-based account (see section 4 of this chapter). However, honesty compels me to mention that I am (subjectively) not fully convinced that the Bayesian coherence constraints are a necessary requirement for an agent to count as rational. For instance, consider another juror in the McMillian case who considers the defense’s witnesses reliable and who does not overlook the gaps in the prosecution’s scenario. As a result the juror believes that Walter is innocent. However, if we were to ask her to assign probabilities to different hypotheses, we would find out that her credences do not satisfy the axioms of probability. She often assigns numbers smaller or greater than 1 to the conjunction of hypotheses and their negations, she does not always update in accordance with Bayes’ rule when receiving new information and so on. Is she irrational? While she would surely be more rational if she conformed to the Bayesian prescripts, I would hesitate to call her irrational. After all, her belief is sensible, as are her reasons for holding that belief.
Bayesians do not map well onto legal proof. For instance, the principal principle is only relevant in cases where we know the objective chance of some event, which is rarely the case in criminal trials. The principle of indifference is similarly problematic. For example, it would have us set the prior of guilt to 0.5 at the start of the trial. Though this suggestion has been defended by Posner (1999), most see it as patently unacceptable, as it clashes with the presumption of innocence (see the discussion in the previous subsection). Opponents of Posner’s idea argue that the prior should be much lower. The other objective constraints that have been proposed in the literature are similarly non-applicable or problematic. Additionally, it is not clear that the objective Bayesian constraints help us in explaining why decisions such as that in the case of Walter McMillian are irrational. Even with these constraints, irrational judgments such as that in the Walter McMillian case could be rational from a Bayesian perspective.

The problem with the objective Bayesian constraints is that they do not adequately capture the idea that the agent’s credences should be well-supported by the evidence. While there have been various objective Bayesian authors who offer further suggestions for connecting degrees of belief to the available evidence, their proposals have tended to be similar to the aforementioned principal principle in that they stipulate how our degrees of belief ought to match up with ‘objective’ (i.e., ‘physical’) chances. For example, De Finetti (1974, xv) wrote that our beliefs should have a “reasonable relationship to any kind of relevant objective data”. Such proposals run into the same problem as the Principal Principle: the required data is usually not available in legal proof. Instead, we need a way of spelling out how one’s evidence ought to constrain one’s degrees of belief which also takes into account non-statistical evidence such as eyewitness testimony to prevent the Bayesian account from being overly permissive in what it counts as rational.

Recently, a number of evidence scholars have suggested variations on the Bayesian

---

54 See Williamson (2010) for an explanation of the various principles.
account which incorporate this idea. These authors suggest that, to count as rational, an agent’s degrees of belief should be “reasonable,” or “evidential” probabilities (Nance, 2016; Wittlin, 2019; Spottswood, 2019; Hedden & Colyvan, 2019). On such accounts the relevant probabilities are not the degrees of belief of the fact-finder. Rather they express the probability that is reasonable to hold in the face of the available evidence – i.e., the degree of belief that a rational agent would have given this evidence – regardless of whether the fact-finder actually holds this credence (Hacking, 2001, 131-3). As Williamson (2002, 209), arguably the most prominent defender of this type of Bayesianism in epistemology, puts it, when we ask about the evidential probability of a hypothesis “[w]e are not asking what objective physical chance or frequency of truth h has. (…) Equally, in asking how probable h is on present evidence, we are not asking about anyone’s actual degree of belief in h. Present evidence may tell strongly against h, even though everyone is irrationally certain of h.” As Nance (2016, 47) points out, on evidential Bayesianism, a fact-finder’s degree of belief in a hypothesis is rational to the extent that it matches the evidential probability of that hypothesis.55

Evidential Bayesianism is, I believe, the most plausible version of legal Bayesianism. As mentioned above, pure subjective Bayesianism not only offers too little connection between what counts as rational and the strength of the evidence, it also gets the order of explanation the wrong way around: an account of why a given belief is (ir)rational should not be grounded in that belief itself. Rather, this account should offer a reason for the agent to hold (or not hold) a given belief. Evidential Bayesianism’s response to both questions is to ground evidential rationality in what is reasonable to believe given the evidence. So, when we ask ‘did the defendant probably commit the alleged acts?’ we are after the evidential probability of that hypothesis. However, despite its advantages, evidential Bayesianism faces a significant

55 In Nance’s (2016, 47) own words: “one’s subjective, credal probability that a certain proposition is true ought to be set equal to the epistemic probability of that proposition, determined in light of (relative to) the evidence that one has for it.”
problem of its own. In particular, the worry is that it is highly unclear what we mean by the term ‘evidential probability’. In other words, what does it mean to say that the available evidence justifies the belief that a given hypothesis is probably true?

Admittedly, to lack a clear definition of something does not mean that the concept is useless. Indeed, as Timothy Williamson (2002, 211) argues, we should deliberately avoid giving a precise definition of this concept:

What, then, are probabilities on evidence? We should resist demands for an operational definition; such demands are as damaging in the philosophy of science as they are in science itself. Sometimes the best policy is to go ahead and theorize with a vague but powerful notion.

According to Williamson (2002, 209), even if we cannot give a precise definition of what evidential probability is, we may perfectly intelligibly ask: “how probable is h on present evidence?” On this view we therefore regard evidential probability as a primitive – a concept which cannot be spelled out in terms of more fundamental concepts, but which is clear enough in context. However, while I agree with Williamson that we may, in some situations, reasonably ask what conclusions the evidence supports, this suggestion is unsatisfactory. Evidential Bayesianism seeks to solve the problem of pure subjective Bayesianism. But without a clearer notion of how we should determine what our evidence supports, we risk reverting back to the latter. After all, as said, we can understand ‘evidential probability’ as the probability that is reasonable to have given the evidence. However, who is to determine this? Or, as Acree (2021, 273) puts it, if such a probability “is the appraisal made by a wise person, or an appraisal from the fullest context of present human knowledge, then who is to say what this appraisal shall be? If it is, for instance, up to the [fact-finder], then we are right back to pure subjective probability.” Evidential probability

---

56 This is similar to authors such as Keynes and Jeffreys who write on the logical interpretation of probabilities and who view logical entailment by evidence as a primitive and do not spell it out further in terms of fundamental concepts (Acree, 2021, 245).
is then just whatever conclusions the fact-finder believes are justified given the evidence. To put it differently, what question do we answer when we try to determine whether our belief is evidentially calibrated? If this question is simply ‘does the evidence convince me?’, then we are back at the purely subjective version of Bayesianism, which, as said, lacks a clear connection to the aim of truth-finding as the numbers can be almost anything. So, we need a more precise account of evidential Bayesianism and this account should have a clear connection to the aim of obtaining accurate beliefs. Additionally, it should offer guidance on how to determine what the evidential probability in a given situation is. Is it possible to provide such an account? Existing attempts to do so do not seem to hold much promise. For instance, Redmayne (2003) surveys various ideas that could be used for such further development when he discusses a proposal by Goldman (1999) about ‘objective’ probabilities in legal contexts. Redmayne concludes that they are all either too limited in scope to capture the evidential richness of legal proof, or so vague that they are uninformative. Nonetheless, I believe that a more informative account of evidential probability is possible. I will develop this account over the course of this thesis.

While I will expand on this topic later, let me already offer a brief sketch of what my account of evidential probability will look like. Evidential probability is a probability assessment that is reasonable given the available evidence. However, in legal cases, what the relevant evidence is and what this evidence supports is not simply given. The relevant information needs to be gathered, selected and interpreted. Only once we have done so can we intelligibly ask what this evidence entails – i.e.,

---

57 Some of the aforementioned authors do make brief suggestions of what evidential probability could be. For instance, Hedden & Colyvan (2019) suggest that we could determine the evidential probability of a hypothesis by considering how well that hypothesis explains the evidence and how intrinsically plausible it is prior to investigation. Nonetheless, such proposals are at best only hints of how the concept could be further developed.

58 I first return to the concept of evidential probability in chapter VI, where I discuss it in the context of eyewitness evaluation. However, the most in-depth discussion can be found in chapter VIII.
what the evidential probability of our hypotheses is. It is at that point that Bayesianism becomes relevant, as it provides a clear and arguably rational calculus for drawing probabilistic conclusions from the evidence. Rational legal proof is therefore also about adequately gathering the evidence, then selecting what is (most) relevant and determining exactly in what way this evidence is relevant.

I propose that we have assurances that our probability indeed aligns with the evidence to the extent that we have made sense of the evidence well. Admittedly, giving a precise and complete definition of what it means to make sense of the evidence well is a task that goes beyond my abilities. However, I want to suggest that we can plausibly think of it in terms of ‘seeing how the facts hang together’ (van Oorschot, 2014). A large part of making sense of the evidence well is therefore seeing all relevant connections between the evidence and hypothesis and between the items of evidence and between the different parts of the explanation (i.e., how well the explanation hangs together). It also mean seeing relevant alternative explanations. To give a simple example, suppose that a suspect testifies multiple times about where she was on the day of the alleged crime and that there are inconsistencies between these testimonies. What does this prove? We will have to decide whether these inconsistencies are indicative of the witness lying, or whether there is an alternative explanation for them (for instance that they are honest mistakes).

In the above example of the eyewitness, we are concerned with the probability relating to a single item of evidence, namely that of eyewitness testimony. I will say more about how to interpret the evidence relating to eyewitness reliability by means in chapter VII. Then, in chapter VIII, I turn to the most important probability that has to be estimated in a criminal case. This is the posterior probability of the ultimate hypothesis – that the defendant committed the alleged acts. In that chapter I connect the idea of evidential probability to another worry for the Bayesian, namely that our set of evidence may itself be defective – i.e., incomplete or biased. Even if our probability is reasonable in the light of our evidence, this probability as a whole may not be reasonable given what we know about the completeness of our set of
evidence. Both requirements are about the possibility of missed information. What the fact-finder needs to be assured of in order to reasonably believe that their belief is calibrated to the evidence is that they have not missed any alternative explanations for the evidence, connections between the items of evidence or arguments against their conclusions. Or so I will suggest. As I will argue in this thesis, explanation-based thinking helps us gain such assurance in various ways. But before I turn to this point in further detail, let me first explain the explanation-based accounts themselves.

3. (Inference to the best) explanation-based approaches

In the previous section I discussed the Bayesian account – one of the two frameworks around which this thesis centers. In this section I turn to the second account, namely explanation-based approaches. I begin with explaining the common basis of these proposals. Then, I turn to the most important types of explanation-based proposals. Finally, I discuss the problems associated with explanation-based accounts.

3.1 Abduction and inference to the best explanation

Explanation-based models are a family of approaches to rational proof that revolve around what is known as abductive reasoning, a term coined by Charles Sanders Peirce (1974). On abductive reasoning, we begin with some empirical fact – for instance a person’s dead body – and then ask what could explain this fact. For example, we might hypothesize that this person was killed. Abductive reasoning is therefore explanatory in nature, where we observe a state of affairs and try to explain how this state could have come about.

59 Parts of this section are based on Mackor, Jellema & Van Koppen (2021).
The term abduction has two meanings. The first is commonly known as ‘creative abduction’. On this conception, abduction is the process by which explanatory hypotheses are generated. Peirce (1974, 5.189) offered the following syllogistic form such inference:

\[
\begin{align*}
\text{The surprising fact E is observed;} \\
\text{But if hypothesis H were true, E would be a matter of course,} \\
\text{-----} \\
\text{Thus, there is reason to suspect that H might be true.}
\end{align*}
\]

In contrast to deductive and inductive inference, the conclusion of an abductive argument is not that the explanation is true or even probably true, but only that is “possibly” or “plausibly true” (Schum, 2001, 308). Numerous authors have argued that creative abductive reasoning should be the starting point of any investigation, whether scientific or criminal (Eco & Seobok, 1983; Carson, 2009; Fraser-Mackenzie, Bucht, & Dror, 2013; Innes, 2003; Keppens & Schafer, 2004; Lipton, 2007; Simon, 2012). In contrast, on the second, stronger, notion of abduction, the term refers not only to the processes of generation of hypotheses, but also comprises the process whereby such hypotheses are evaluated. This is commonly known as inference to the best explanation (IBE) (Harman, 1965; Douven, 2011). On IBE, we accept a hypothesis H because it is the best explanation of the evidence (Psillos, 2004, 83). This type of inference therefore proceeds according to the following syllogism:
D is a collection of data

H would, if true, explain D

No other hypothesis can explain D as well as H.

Therefore, H is probably true.

In other words, IBE relies on the idea of `self-evidencing' explanations, where the fact that something could, if true, provide a satisfactory explanation is a reason for believing that this explanation is actually true.

The most detailed and widely discussed account of IBE is Peter Lipton’s (2004). According to him, IBE is a two-stage process. First, we generate a number of possible explanations of a fact. Second, we choose the best of these explanations. According to Lipton (2004, 60), how good an explanation is depends on its ‘loveliness’ – i.e., how much understanding the explanation would offer us of the facts if it were true. So, as Lipton (2004, 184) puts it, the governing idea of the IBE model is that “explanatory considerations are a guide to inference”. Such loveliness, or ‘explanatory goodness’ or ‘plausibility’ is usually further spelled out in terms of explanatory virtues. For instance, Josephson (2001, 1627) mentions the following desirable features of explanatory hypotheses: consistency, simplicity, explanatory power, predictive power, precision, specificity, and theoretical promise. Others propose yet other criteria. Pardo & Allen (2008) propose that “all other things being equal [an explanation is] better to the extent that it is consistent, simpler, explains more and different types of facts (consilience), better accords with background beliefs (coherence), is less ad hoc, and so on; and is worse to the extent that it betrays these criteria.”

Our initial hypotheses will often be too coarsely grained to determine whether they are plausible in the given context and will need to be made more specific by considering what the most plausible version of the events is. As Walton (2007, 128)
points out, this process of filling in the missing bits is also an instance of inference to the best explanation. Though there are countless ways of filling in the gaps in a story, we choose the version that is the most plausible in the given context. Sufficiently specific candidate explanations lead to predictions that can be tested – they allow us to ask what else we would expect to observe if the explanation were true (Schum, 2001).

IBE’s structure has much in common with legal fact-finding. Abduction involves reasoning from observed to unobserved possibilities. Similarly, legal trials are devoted to reconstructing a historical truth from evidence. Philosophers of science therefore sometimes use criminal proof as an example for their discussions of IBE. For example, Harman (1965, 89) suggests that “when a detective puts the evidence together and decides that it must have been the butler, he is reasoning that no other explanation which accounts for all the facts is plausible enough or simple enough to be accepted.” It is therefore also unsurprising that a variety of legal scholars have applied IBE to process of proof at trial. These approaches conceptualize the fact-finding process as a choice among possible hypotheses, which are usually (though not necessarily) presented by the parties at trial, which (causally) explain the most important evidence and may produce divergent legal consequences.

We will examine the different explanation-based accounts in the following subsection. But first, let me remark that while these approaches are often referred to as employing “inference to the best explanation”, this does not mean that they endorse the syllogism mentioned above. Indeed, as Laudan (2007) argues, criminal proof cannot be inference to the best explanation in the above sense, as criminal fact-finders should not always choose to accept the best of the available explanations as true. Such an approach evaluates the case for guilt strictly in terms of how

---

60 See chapter V for a further discussion on how predictions follow from scenarios and what the value of checking on such predictions is.

61 Some explanations imply that the defendant committed the alleged crime. If such guilt explanations are proven, this can lead to conviction. Other explanations are consistent with the innocence of the defendant (I call these innocence explanations or alternative explanations).
plausible the scenario offered by the prosecution is compared to its competitors. However, in criminal trials the prosecution has to prove that the case for guilt is strong in an absolute sense, not merely stronger than any alternatives that have been presented at trial. For example, suppose that the defense offered no alternative explanations. This does not mean that guilt can be proven based on a weak explanation offered by the prosecution (which would be the best available explanation). While I argue against Laudan’s view in chapter III, and suggest that we can feasibly think of criminal proof as IBE (if properly spelled out), the point stands that not all abductive approaches to legal proof employ inference to the best explanation in the above sense. I therefore use the term ‘explanation-based’ as an umbrella-term for abductive accounts of legal proof.

3.2 Explanation-based models

In this section I briefly explain the four most prominent explanation-based accounts: (i) the story model, (ii) the scenario-approach, (iii) the relative plausibility theory, and (iv) the theory of inference to the most coherent explanation. These accounts share the view that the process of legal proof is – or should – be centered around competing, contrasting explanations of the disputed facts. After discussing these accounts, I reflect on some of their differences and similarities.

3.2.1 The story model

The story model is a descriptive, psychological theory about how fact-finders process evidence at trial according to which fact-finders come to factual decisions by constructing and evaluating one or more stories explaining “what happened”. The development of the story model began with Bennett and Feldman (1981), who found that jurors depend on the construction of such stories when having to make a decision. In Bennett and Feldman’s model, a story is organized around a central
action for which the rest of the story acts as a context.

The story model was further developed by Pennington & Hastie (1986; 1991; 1992; 1993). Using experiments with mock jurors, they found that these jurors explained their decisions by referencing stories they had constructed. These stories combined (some of) the evidence in the case with assumptions that they used to fill in temporal or causal gaps in their narratives. On Pennington & Hastie’s account, stories consist of elements, which are called episodes. Episodes consist of a chronological sequence of specific elements, such as an initiating event, a psychological response, sometimes a goal, an action and a consequence (Pennington & Hastie, 1993, 197). Take, for example, the following episode: a husband has an argument with his wife (initiating event), which makes him angry (psychological response). Because he intends to hurt her (goal), he beats his wife (action), which causes her death (consequence).

According to Pennington & Hastie (1993, 194) stories are constructed based on several types of information:

(i) Case-specific information acquired during the trial, such as the testimony of a witness;

(ii) Knowledge about events similar to those that are under dispute, for instance information about similar crimes in the media; and

(iii) Generic expectations about the structure of stories, such as the knowledge that most human actions (like a murder) are motivated by the achievement of certain goals (for instance, to inherit a substantial amount of money).

If more than one story is constructed, one has to be chosen by comparing them.

---

62 For more recent discussions of the model by these authors, see Hastie, Penrod & Pennington (2013) and Hastie (2019).
Pennington and Hastie (1993, 198-9) mention three ‘certainty principles’ that fact-finders use to assess how confident they are about the truth of stories: coverage, coherence and uniqueness. First, coverage refers to the extent to which the story explains the evidence in the case. Second, a coherent story is one that is consistent with the available evidence, is free from internal contradictions, fits with the fact-finder’s background beliefs about the world and does not lack any part that we would expect it to have. Finally, a story is unique if it is the only coherent story that can account for the evidence. If there is more than one coherent story, confidence in each of these stories will diminish.

After constructing such stories, fact-finders match them with legal categories. For example, if there is a story of one man killing another, the decision maker can choose between the verdict categories of first-degree murder, second-degree murder, manslaughter and self-defense. The fact-finder chooses the category that best fits with their chosen story of “what happened”.63

3.2.2 The scenario approach

The scenario approach is a normative account of how fact-finders ought to reason, developed originally for the Dutch criminal law context.64 It is based on Pennington & Hastie’s descriptive model. The development of the model began with Crombag et al. (1991) and Wagenaar et al.’s (1993) Anchored Narratives Theory. In their work, they discuss examples of mistakes people make when reasoning with evidence and proof. According to this theory, fact-finders reach their decisions based on the

---

63 According to Tuzet (2003) the process of fitting the proven facts into a legal category can also best be seen as inference to the best explanation.

64 The scenario theory has recently been further developed by others. Bex (2011) has developed a framework that combines the scenario approach with argumentation theory, which allows for the support and critique of scenarios with arguments. Vlek (2016) suggests ways in which the insights from the scenario theory can be used in Bayesian modeling. Scholten (2022) incorporates the scenario theory in her model of how courts should justify their fact-finding decisions.
plausibility of the prosecution’s account of what happened and whether this account is ‘anchored’ in the available evidence through common sense generalizations. The authors suggest that erroneous convictions can be prevented by the use of a more systematic approach to reasoning with evidence. The scenario theory was further developed by Van Koppen (2011). On his account judges should not only take into consideration whether the prosecution scenario is well-grounded, but they should also test this scenario by looking for falsifying evidence, and the scenario should be compared to alternatives which support the defendant’s innocence. On the scenario-approach it is especially important to look for evidence that discriminates between the prosecution scenario and its alternatives. The comparison between scenarios happens based on three criteria, which are derived from, but also slightly differ from those of Pennington & Hastie (1993). According to Mackor & Van Koppen (2021) how good a scenario is, depends on how much of the evidence it explains (coverage), how many alternative scenarios there are that explain the evidence in the case (uniqueness) and it should be coherent. This last criterion includes consistency (absence of internal contradictions), plausibility (does the scenario fit with our background beliefs about the world) and completeness (does the scenario have the elements that we would expect it to have, such as a motive, action, actor). Additionally, Mackor & Van Koppen (2021) add the requirement of ‘robustness’. According to their definition a scenario is robust when it has been confirmed in multiple and independent ways.

---

65 See Van Koppen & Mackor (2020) and Mackor & Van Koppen (2021) for English-language discussions as well as a further elaboration of the theory.

66 See chapter VIII of this thesis for an extensive discussion of robustness.
3.3.3 The Relative Plausibility Theory

Over the past decades Allen and Pardo developed the *relative plausibility theory*. The primary goal of this theory is to explain the different facets of juridical proof of the American legal system by means of a single framework. On this theory, juridical proof involves determining the comparative plausibility of competing explanations (Allen & Leiter, 2001, 1527-1528). In their own words: “The primary message of relative plausibility is that from beginning to end the legal system pushes the parties to provide competing explanations, and these explanations structure the decision that is subsequently made” (Allen & Pardo (2019b, 4). They suggest that both proof rules and legal culture encourage or even require participants in the legal system to provide and compare competing explanations. This view, they argue, is supported by case law, which provides numerous examples of trials which reduce to the comparison of competing stories” (Allen & Pardo, 2019a).67

According to Allen and Pardo (2019b, n86), in criminal law these explanations can be stories, but they need not be. Explanations can also be ‘disjunctive’ and ‘general’. Disjunctive explanations are made up of mutually exclusive events – ‘this happened or that’, while general explanations can be of the form ‘something (else) happened’ (Pardo, 2013, 598-9). Such explanations are evaluated based on their plausibility. The authors contrast this notion with that of (Bayesian) probability – which is a quantitative concept, whereas plausibility is a qualitative concept (Allen & Pardo, 2019b).

According to Allen and Pardo, an explanation is more plausible when ‘it is consistent, simpler, explains more and different types of facts (consilience), better accords with background beliefs (coherence), is less ad hoc, and so on; and is worse to extent [sic] it betrays these criteria’ (Pardo & Allen, 2008, 230).68 The goal of assessing the plausibility of competing explanations is to reach probabilistic

---

67 For some of these examples from case law, see Allen (1991, 414-7; 1993, 633-40) and Allen & Leiter (2001, 1528-33).
68 Pardo and Allen derive these criteria from Thagard (1978).
conclusions. Allen and Pardo write that ‘[e]xplanatory criteria guide inferences and judgments about likelihood’ (Allen & Pardo, 2019b, 17) and that ‘the better the explanation, the more likely true’ (Pardo & Allen, 2008, 9, n45).

The relative plausibility theory focuses primarily on the American proof standards and how to interpret these in terms of explanations. For criminal cases, this is the beyond a reasonable doubt standard. According to Pardo and Allen (2008, 238) guilt is proven beyond a reasonable doubt only if there is a sufficiently plausible explanation that implies guilt and there is no plausible explanation which is consistent with innocence. According to Allen and Pardo (2019b, 2) their theory not only accurately describes such features of legal proof, but the theory is also normatively appropriate in the light of the goals of the legal system.69

3.3.4 Amaya’s theory of inference to the most coherent explanation

Amaya’s (2009; 2013; 2015) theory of inference to the most coherent explanation is based on Thagard’s theory of explanatory coherence (Thagard, 2000; 2004). According to Amaya, we choose an explanation through inference to the best explanation. First, we generate a set of hypotheses and evidence. After that we construct alternative theories, which we then refine and revise. Following this, we evaluate the coherence of these theories. Finally, we choose the most coherent explanation. If this explanation implies the defendant’s guilt, it’s degree of justification has to satisfy the applicable legal standard of proof, in the context of criminal law the beyond a reasonable doubt standard (Amaya, 2013, 13).

On Amaya’s account, explanatory coherence is the satisfaction of a set of positive and negative constraints – coherence and incoherence relations – among a set of elements of hypotheses and evidence. She derives these constraints from Thagard (2000, 43). They include, for example, the notion that ‘similar hypotheses that

69 I criticize this account in chapter III.
explain similar pieces of evidence cohere’ and that ‘contradictory propositions are incoherent with each other.’ Jointly these constraints determine how coherent an explanation is with the other available hypotheses and with the admissible evidence.

Another component of Amaya’s theory is a responsibilist epistemology (Amaya, 2013, 24). On a responsibilist epistemology, justification is not analyzed exclusively in terms of evidential support, but also in terms of what a factfinder has done or failed to do, more in particular in terms of how thorough or robust the investigations have been (Amaya, 2009, 154). This includes the duty to actively search for alternative hypotheses, to gather additional evidence about uncertain propositions and to believe all and only propositions that are supported by available evidence (Amaya, 2013, 25-26. For a further discussion of Amaya’s responsibilism, see chapter VII.

3.3.5 Similarities and differences

The explanation-based approaches discussed above share a number of similarities, but there are also points of difference. One important difference is their aim. Pennington & Hastie’s story model is purely descriptive. Allen and Pardo (2019b, 17, n86) suggest that the relative plausibility model differs from the story model because “whereas the story model is an empirical description of juror reasoning, relative plausibility theory is an explanation of the ‘standards of proof and other features of the proof process.’” However, they sometimes also present the account as a description of what occurs at trial or as a normative theory of how fact-finders ought to reason (Pardo & Allen, 2008, 238). Amaya’s theory is normative, in the sense that it is a theory of rational evidential reasoning, which does not seek to describe or explain any proof rules. Finally, the scenario theory is both normative – in the sense that it aims to offer a standard for behavior – and prescriptive – because it gives concrete guidance for fact-finders on how to reason in practice.

Another difference is what the models mean by the term explanation. Pennington
& Hastie’s model and the scenario theory take explanations to be stories. In contrast, Allen and Pardo claim that the explanations offered by parties can, but need not be stories (Allen & Pardo, 2019b, 3, n7; 13, n86). Similarly, Amaya (2021, 236) suggests that narrative accounts are “unduly restrictive in scope, in that they are applicable only to those cases that bear a narrative structure. Because of that, a turn to an explanationist approach to legal proof, according to which fact-finding involves the comparison and selection of the best explanation of the available evidence, rather than the best narrative, seems preferable.” However, even though Allen and Pardo (2019b) suggest that explanations are not always stories, they also argue that their account draws support from the empirical evidence from the story model, thereby fitting better with how people reason naturally.

This brings us to an area in which the explanation-based models are similar, namely that they are sensitive to human cognitive capacities and to the ways in which legal fact-finders actually reason (Redmayne, 2019, 71-5). This is sometimes called a naturalized view on the nature of rationality, as opposed to an ideal view – which does not take human cognitive capacities into account (Nance, 2001). One way in which explanation-based approaches respect human cognitive capacities is that they do not assume that fact-finders and others consider the entirety of the probability space. Rather they have us focus on a select number of explanations which are often quite detailed, which are then (ideally) examined closely. Other possibilities are ignored. So, the explanation-based account has us narrow our view for the sake of tractability (but at the danger of overlooking relevant alternatives). A closely related similarity is that all aforementioned explanation-based models view evidential reasoning as holistic – i.e., they have us evaluate the mass of evidence as a whole (Damaška, 1990, 91; Spottswood, 2014). We may contrast this holism with atomism, according to which fact-finders assess the evidentiary strength of each item of evidence and then integrate their individual assessments according to some general.

70 See chapter VII for an account of when we may justifiably believe that these ignored possibilities do not contain any plausible alternative explanation.
rule to arrive at a conclusion (Twining, 2006, 309; Schweizer, 2013).71

A final similarity is that, apart from Pennington & Hastie, the authors of these models all claim to employ inference to the best explanation (e.g., Pardo & Allen, 2008; Amaya, 2009; Van Koppen & Mackor, 2020; Mackor & Van Koppen, 2021). It is not always clear what these authors mean by this term (see also chapter III of this thesis). For authors such as Allen and Pardo it does not mean choosing the (comparatively) best explanation as true. Instead, what they appear to mean is that the primary way of evaluating explanations is by comparing them to their alternatives.

The explanation-based models all share a benefit which follows from their emphasis on comparative evaluation: they may help avoid confirmation bias (Mackor & Van Koppen, 2021; Fahsing et al., 2021). Confirmation bias is a common cause of judicial errors. It is the tendency to (a) predominantly search for information in line with the main hypothesis, (b) ignore alternative interpretations, and (c) interpret ambiguous information in line with the main hypothesis (Ask and Granhag, 2005; Fahsing and Ask 2016; Rassin, 2010; Sunde and Dror, 2019).72 Nickerson (1998, 211) suggested in his comprehensive work on confirmation bias that “reasoning might be improved by training people to think of alternative hypotheses early in the hypothesis-evaluation process.” By having people generate and compare alternative explanations of an event, confirmation bias may be avoided. Fahsing et al. (2021) review the psychological work on this idea and argue that the data supports “consider the alternative strategies” being an effective way of combating confirmation bias. Similarly, they argue that “the ability to generate and evaluate alternative explanations...

---

71 Atomism was traditionally associated with Bayesian updating. Indeed, according to Allen & Pardo (2019b) Bayesianism suffers in part from the problem of computational complexity because it is atomistic, whereas explanation-based approaches avoid this problem because they are holistic. However, it is doubtful whether we should view more recent Bayesian Networks as atomistic. After all, such networks also evaluate the mass of evidence jointly and the strength of each piece of evidence is determined in part by its dependency relations with other evidence and hypotheses.

72 See chapter V for more on this.
hypotheses is among the core defining features of an expert detective” and that it is “that the ability to adequately generate relevant hypotheses is beneficial to the outcome of actual criminal investigations” (Fahsin et al., 2021, 11).

3.3 The irrationality-triviality dilemma

Though explanation-based approaches have several benefits, they are not without their problems. In this section I focus on what I perceive as the most pressing of these problems. The worry is that – depending on how we spell out such accounts – the requirements for rationality are either too vague and lenient to offer an informative account of rational proof, or that such accounts lead to irrational decisions.

One reason for thinking that explanation-based accounts may lead to irrationality is the well-known observation that storytelling can be used to sneak in irrelevant considerations, hidden prejudices and stereotypes, and to subvert legal principles such as: “judge the act not the actor; consider only relevant evidence; argue from evidence; do not invent or suppress facts, keep separate questions of fact, value, law, and disposition” (Twining, 1999). In other words, if we focus on who tells the most convincing story, we may overlook weaknesses in the evidence. For instance, experiments by Bennett and Feldman (1981) revealed that a complete and well-structured scenario was believed more readily by test subjects than an incomplete or unstructured scenario, even if the latter really happened and the former was false. This led them to conclude that one should be wary of a good scenario pushing out a true scenario, which was later also emphasized by Anderson, Schum & Twining

73 For instance, one issue is that such models are qualitative in a world where quantitative evidence is increasingly important. Yet, as Fienberg (1997, 312) puts it: “Without a probabilistic framework it is nigh impossible sensibly to situate the statistical evidence presented at trial, yet all agree that the evidence is central.” It is not clear how we can rationally incorporate statistical evidence within scenario-based reasoning. This problem mirrors that of Bayesians who want probabilistic models to be used in court and have difficulties incorporating non-statistical evidence in a way that is acceptable to courts.
(2005) and Griffin (2012). One mechanism that may contribute to a good story pushing out a true story are ‘coherence shifts’, where evidence that contradicts one’s preferred story is ignored and the strength of supporting evidence is inflated (Simon, 1998; 2004; Schweizer, 2013). The result of this process is that even when the evidence gives little support to a decision, the fact-finder has a high degree of confidence in having made the correct decision (Simon et al. 2004, 819).

Findings such as these pose a challenge for explanation-based models of proof. What we need is some way to show how story comparison may be done in a rational way. One of the most important questions concerning the rationality of explanatory comparison is by what criteria we ought to judge the quality of explanations (Schum, 2001, 468; 2002). For instance, Allen and Pardo (2019b) suggest that their relative plausibility theory is less susceptible to the worry of subjectivity than Bayesian accounts, as their notion of plausibility is more ‘objective’ than that of probability. After all this notion refers to the quality of the evidence and explanations, not merely to the fact-finder’s degrees of belief. However, this presumes that the explanationist has a well-founded and clear notion of when an explanation is good. But this is not the case. The various models discussed above each come with their own notion of what makes an explanation good. However, such characterizations tend to be quite vague. Furthermore, which of these varying suggestions is the ‘correct’ one? The authors of these models have not extensively engaged with the question why we should use the criteria that they propose to judge explanations. Why is a story with gaps, for instance, less likely to be true? Why is a simpler explanation more likely to be true?

74 See section 2.4 of this chapter for a discussion of the problem of subjectivity for Bayesianism.

75 Or, as Laudan (2006, 82), puts it, explanation-based accounts “tell the juror what features he should be looking for in the evidence at hand that would justify him in convicting the accused. They do not ask the juror to engage in a meta-analysis of his own degree of confidence in the guilt or innocence of the accused.”

76 Lipton (2004, 144) calls this ‘Voltaire’s objection’. He uses the term ‘loveliness’ for the
Allen and Pardo (2019a,b), Amaya (2009) and Mackor & Van Koppen (2021) base their list of criteria for a good explanation – such as coherence and fit with background knowledge – on work from the philosophy of science. We could therefore look at how philosophers of science justify their choice of explanatory criteria and follow their strategies. For example, some philosophers of science have sought to show that specific explanatory virtues are truth-conducive (Douglas, 2009; Thagard, 1978). However, showing a reliable connection between these criteria and the probability of explanations has proven to be difficult. Take the notion of coherence as an example. Epistemologists have given ‘impossibility results’ for coherence, i.e., mathematical proofs that suggest that more coherent theories are not necessarily more probable than less coherent theories (e.g., Bovens & Hartmann, 2003; Olsson, 2005). So, despite intense efforts to show truth-conduciveness, this strategy has, so far, not succeeded. A second approach used by philosophers of science, is to look at historical instances of successful explanatory choice and to see which explanatory virtues guided those choices. However, other scholars voice skepticism about this strategy. They argue that the historical record is mixed and that, for instance, more complex theories regularly prevail over more simpler ones. According to these authors no clear set of virtues follows from this strategy (Lipton, 2004, 60; Cabrera, 2017). I am therefore doubtful whether either strategy would be helpful for the legal explanationist.

We are therefore left without a notion of explanatory goodness with a clear connection to obtaining accurate beliefs. Furthermore, even if we were to establish a list of criteria, we lack any suggestion for combining how well an explanation scores amount of understanding that a scientific theory would offer if it were true. But, he asks, why should the loveliest theory be the most likely to be true?

77 Though see Dahlman & Mackor (2019) for a suggestion on how coherence and probability are connected.

78 As Sober (2015, 46-7) points out, those who pursue such a track-record approach often display confirmation bias, focusing on those grand, unifying theories “that turned out to be (more or less) successful. However, we often fail to consider the many dozens of theories that might have excelled with respect to the explanatory virtues, but in the end were abandoned for being not so successful.”
on various criteria into a single judgment. Allen and Pardo (2019, 19) acknowledge that ‘there is no a priori way to combine or rank the various explanatory criteria in a way that provides a general recipe for determining case outcomes’. Yet without a clear answer to the question what makes an explanation good (and how this notion of goodness links up with the aim of accuracy) explanation-based approaches are also open to a worry of unconstrained subjectivity, just like Bayesian approaches. After all, the judgment of whether the explanations are good enough then depends almost exclusively on the fact-finder’s intuition. As Wittlin (2019, 187) puts it:

> Allen and Pardo have eliminated numbers from the probabilistic calculation—it becomes a qualitative assessment instead of a quantitative one—but they have provided little reason to believe that judicial conclusions about sufficiency will be accurate under relative plausibility and not under a probabilistic framework. Both require personal, subjective assessment.

To give an example of this worry, according to various authors how good an explanation is, depends in part on how well it fits with our background beliefs concerning the world. But who is to say what the ‘correct’ background beliefs are? The most straightforward answer is that it is the fact-finder who decides. However, the fact-finder may have wildly unreasonable background beliefs.

The worry that explanation-based approaches may be too permissive with respect to irrational decisions also extends to the phase after we have determined how good each of the competing explanations is. In this phase the question becomes how we should choose which (if any) of these explanations to accept as true. Most authors mentioned above claim that their model uses Inference to the best explanation. IBE is, at its core, a comparative notion. It presupposes that we compare the quality of different available explanations. However, more is needed for a criminal conviction. For instance, while the defendant may offer his own version of the facts, he does not have a duty to do so (Lempert, 1986; Clermont, 2015, 359). This can lead to a situation in which the prosecution’s story is the only available explanation. But
suppose that this explanation is highly implausible. Obviously, the defendant should not be convicted based on such a weak case for guilt. Yet on a literal reading, IBE does lead to conviction in such cases, as the prosecution’s story is the best available explanation.

One response to the above is to claim that proving beyond a reasonable doubt in fact means that the prosecution’s scenario must be demonstrated to be better than any alternative reasonable scenario – not just with those that the defendant came up with (Wagenaar, et al., 1993; Van Koppen, 2011). However, though this is a step into the right direction, it does not get the explanationist completely out of the woods. Such an account runs into the so-called bad lot problem, which is possibly the best-known argument against IBE in philosophy of science (Van Fraassen, 1989). It has also been raised against IBE in legal contexts (Schum, 2001, 468-9; Amaya, 2009). The problem is that on IBE, we should choose the best explanation as true but that this may lead to choosing a poor explanation if all the explanations that we are able to come up with are poor (Niiniluoto, 1999, 447). Even if we compare our scenario to all reasonable, available alternatives, the prosecution’s scenario may still be both bad and the best available – for instance if we failed to come up with any alternatives.

A plausible response to such worries is that we should not take the name ‘inference to the best explanation’ too literally. For example, according to Allen and Pardo’s theory, a defendant’s guilt can only be proven beyond a reasonable doubt if there is a plausible explanation implying his guilt and no plausible explanation implying his innocence (more on this in the next chapter). That means that factfinders do not necessarily have to look at how well explanations perform compared to one another (Pardo & Allen, 2008). Nonetheless, I see at least three problems with the above version of IBE. I explore one of them in chapter III, namely that by dropping the comparative requirement, Allen and Pardo’s account runs into difficulty in situations where we have a barely plausible alternative scenario, but where the prosecution’s scenario is so much better that it is very likely true. If we were to follow
the rule that we only convict when there are no plausible alternatives, we should acquit. However, in such situations this means acquitting even though the guilt-explanation is probably true.

A second worry, which is closely related to the bad lot problem, is that of *unconceived alternatives*. Even if we currently have a plausible guilt scenario and no plausible innocence scenarios, this may be only because we have failed to conceive of relevant alternative explanations. In such a situation, inferring that the guilt scenario is true is unwarranted. For instance, Amaya (2009, 154-155) states that ‘we need to have some reason to believe that the set of hypotheses from which we have inferred to the best is “good enough”. I delve into the worry of unconceived alternatives in depth in chapter VII, where I argue that in such situations, we should have a reasonable doubt.’

A third worry is that of disjunctive explanations. This is a general objection against IBE, which boils down to the observation “that a particular hypothesis can be the best explanation of a given set of evidence even though the disjunction of its rivals is more likely to be true” (McCain & Poston, 2019). This objection has been raised in particular against the theory of Allen and Pardo (e.g., Clermont, 2017; Nance, 2019b), but if valid, it would also hold against other explanation-based approaches, such as Amaya’s and Van Koppen’s. In essence, the argument is that while one explanation may be much better than any competitor, it might still be less probable than the disjunction of all competitors. Or, alternatively, that no alternative to the guilt scenario may be plausible on its own, but that the disjunction of alternative scenarios may jointly be sufficiently plausible to raise a reasonable doubt. One possible response is that the requirement that the best explanation has to be sufficiently good on its own is not met when it is swamped by alternatives (McCain & Poston, 2019). Another solution, given by Allen and Pardo, is to straightforwardly say that their theory allows comparisons between an explanation and the disjunction of its

---

79 However, I also argue that this problem is equally worrisome for Bayesianism.
competitors (e.g., Pardo & Allen, 2008; Allen & Pardo, 2019b, 26-29). However, this second route runs into the worry that explanation-based approaches are simply Bayesianism in disguise (Nance, 2016, 80-82). After all, the Bayesian account has us compare a hypothesis to its negation. Similarly, we could argue that under Allen and Pardo’s proposal, we compare the guilt scenario to its negation (which is made up of the disjunction of innocence scenarios) (Nance, 2016, 80; Clermont, 2015, 359; Schwarts & Sober, 2017, 653-4). If this is the case, are we not simply doing Bayesian inference? Allen and Pardo (2019b, 26-29) have argued that their account is still distinctive, because it relies on explanatory, not on probabilistic considerations. However, as we saw above, it is not fully clear which explanatory considerations we should adopt.

This brings us to the upshot of this subsection. Explanation-based approaches face a number of difficulties. Many of these counterarguments begin by taking a face-value reading of what the proponents of explanation-based accounts propose and show that this would lead to situations where the account tells us to convict even when the defendant’s guilt is not especially probable (or vice versa). An obvious response for explanationists to many of these worries is to amend their account, so that it is no longer subject to this criticism. However, in doing so they run into the danger of their account being little more than Bayesianism in disguise – where we compare a hypothesis with its negation to determine whether we believe that the former is sufficiently probable given the evidence in the case. This dilemma – triviality or irrationality – mirrors a well-known objection against IBE within the philosophy of science. This objection has led to a research program by philosophers of science on how IBE and Bayesianism may go together, without the former being a trivial restatement of the latter (Douven, 2011, section 4). As I discuss next, this research program also holds promise for wedding Bayesianism and explanation-based accounts in the law.
4. Towards a productive partnership

In the previous two sections, I set out the details of the Bayesian and explanation-based approaches (as well as some of their strengths and weaknesses). These two accounts are sometimes presented as opposed to one another (e.g., Allen & Pardo, 2019b). However, in recent years, more and more attention has gone out to how the two might be combined. For instance, some propose Bayesian modeling approaches that incorporate insights from explanation-based approaches (e.g., Cheng, 2013; Vlek et al., 2013; Urbaniak, 2018; Dahlman, 2020b; Verheij, 2020). I will leave this strand of work aside in this thesis and instead focus on a question that has received much less attention in the legal context, namely, how explanation-based approaches might be wedded with a Bayesian epistemology (as the latter term is used in philosophy).

Several legal scholars have briefly noted that a combination of the two approaches might be possible when discussing Allen and Pardo’s relative plausibility theory. For instance, both Friedman (2001, 2045) and Pardo and Allen (2008, 253) posit that the relative plausibility theory and Bayesianism need not be incompatible. As Pardo and Allen (2008, 253) put it “there is no reason to see explanatory and Bayesian approaches as necessarily incompatible.” Recently Biedermann & Vuille (2019, 18–20), Gelbach (2019, 169) and Welch (2020) made similar compatibilist suggestions. However, remarks on how the two can go together have been brief. In contrast to the legal domain, philosophers of science have extensively discussed the possibility of combining Bayesianism and explanationism, the two main accounts of rational scientific proof (cf. Niiniluoto, 2004; Glass, 2012). It is helpful to consider the debate in the philosophy of science in more depth as it offers useful lessons for legal evidence scholarship.

The compatibilism debate in the philosophy of science has mostly focused on a rather specific question, namely whether explanatory virtues – such as simplicity, consilience and scope – are relevant for how probable a scientific theory is (Douven,
Discussions within this debate usually start with Van Fraassen’s (1989, 169) claim that they are not. On Van Fraassen’s view, IBE is the position that more explanatory hypotheses gain a confirmatory ‘boost’ compared to less explanatory ones. In other words, being a better explanation leads to being assigned a higher probability. He argues that, if this is the case, IBE is incoherent as those who use this updating rule are subject to a Dutch book. In contrast, if there is no such boost, IBE is equivalent to Bayesianism and therefore trivial.

While Van Fraassen’s reconstruction of IBE has been widely criticized, his remarks make the fundamental problem clear. As Lipton (2004) points out, if IBE is to be an informative doctrine, it must not simply tell us to choose the likeliest explanation. The notion of a ‘good’ explanation should somehow be distinct from being a probable explanation. Furthermore, this distinct concept of explanatory goodness must play at least some role in rational inference. If we fail to establish this, IBE is not an independent theory of rational scientific proof. Interestingly, Gijsbers (2011, 39) uses a legal example to explain this worry:

When the detective tries to solve the murder – the naïve defender of IBE may say – surely he infers to that hypothesis that best explains the available evidence? True. But does he infer to that hypothesis because [...] it would be the most satisfying explanation if it were true? Or is the explanation accepted as that which best explains the available evidence because the evidence makes it likelier than any of its competitors? In the latter case, explanatory considerations do no epistemic work at all, and IBE does not take place.

I cite this remark in part to illustrate a contrast between the debate in the philosophy of science and in legal evidence scholarship. Legal explanationists are not all too concerned with the precise role that explanatory virtues play in inference. Rather they propose a particular style of drawing conclusions from the evidence, namely explanatory comparison, where we compare a small number of explanations of

---

80 See section 2.3 of this chapter for a discussion of the Dutch book argument.
the evidence and choose one (or none) as true. Gijsbers does not specify what the hypothesis under consideration is in the above example. However, we might imagine that the detective has some specific version of what happened in mind, which can include a location, time, murder weapon and perhaps even a motive. Furthermore, the above quote does not describe on what grounds the detective comes to their conclusions. We could envision that they conclude that the suspect killed the victim because this is plausible and because there is no plausible alternative explanation. If so, the detective is engaged in explanatory comparison and the objection falls flat (at least with respect to what legal explanationists propose). However, while this objection may not be damaging for explanation-based accounts of rational legal proof, it mirrors the irrationality-triviality problem discussed above, which is worrying for the legal explanationist.

The challenge for the defenders of explanation-based accounts is similar to that of the philosophers of science who wants to defend IBE. First, defenders of IBE have to show that the account does not diverge from Bayesianism or risk being deemed irrational. Second, they need to show that this account is not trivial, but that it plays an essential role in rational criminal proof. On these counts the legal explanationist can learn from the philosophy of science. In particular, after Van Fraassen various philosophers of science suggested ways in which explanatory virtues can have a place within the Bayesian framework, without IBE thereby becoming trivial. We can roughly distinguish three kinds of compatibilist frameworks (Dellsén, 2018, 1746). First, some argue that IBE and Bayesianism are different ways of describing what is essentially the same form of reasoning (Niiniluoto, 1999; Henderson, 2014; 2017). For instance, on Henderson’s account – which she dubs ‘emergent compatibilism’ – a preference for explanatory hypotheses grows out of independently-motivated objectivist Bayesian constraints on probabilities. This means that “IBE would ‘emerge’ without the Bayesian doing anything differently from what she

---

81 See Williamson (2010) and section 2.4 of this chapter for a more in-depth explanation of objective Bayesianism.
would anyway” (Henderson, 2014, 689). To give an example, according to Henderson, simpler hypotheses are favored by (objective) Bayesianism, which would fit with the explanationist claim that simplicity is one of the explanatory virtues to use in IBE. Second, some argue that explanatory considerations help to determine the ‘objectively correct’ prior probabilities from which Bayesian agents ought to reason (Weisberg, 2009; Huemer, 2009; Climenhaga, 2017a, b; Poston, 2014). Third, and most influentially, various authors propose that IBE might serve as a heuristic to approximate correct Bayesian reasoning. On this approach, IBE complements Bayesianism by providing a rule of inference that is appropriate for non-ideal agents and yet enables these agents to approximate the probabilities that Bayesian reasoning would have them assign to hypotheses (Okasha, 2000; McGrew, 2003; Lipton, 2004; Dellsén, 2018).\textsuperscript{82} Lipton (2004, 120) notes that humans have considerable difficulty in thinking with probabilities and are liable to error, whereas explanatory considerations are more natural for them. So, it would be rational to employ a heuristic in place of Bayesian reasoning.\textsuperscript{83}

Could such suggestions work in legal proof scholarship? Some believe that they can. For example, Hedden & Colyvan (2019) have made a proposal for how explanation-based thinking can fit into the Bayesian framework. They suggest that we can determine the Bayesian evidential probability through explanationist means as follows: the prior probability may express the intrinsic plausibility of a hypothesis prior to investigation and the conditional probabilities can be determined by explanatory connections. The conditional probability of a hypothesis H on evidence E will then be high if it provides a good explanation of E. On this view, we

\textsuperscript{82} As Tversky and Kahneman write, heuristics are principles that people rely on “which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors” (Tversky & Kahneman, 1974, 112).

\textsuperscript{83} One way to formulate this heuristic is that if explanation A displays more explanatory virtues than explanation B, we should judge A to be more probable than B (Bird, 2018, 5). However, as said, in legal proof scholarship the concern is less with the precise role that explanatory virtues play.
determine the evidential probability of \( H \) by starting with the prior and condition-
alizing on those conditional probabilities. Much like in the compatibilist accounts
mentioned above, we would then use explanatory considerations to determine the
probabilities needed for Bayesian inference. However, though this is an interesting
idea, it falls short on both aspects of what we want from a compatibilist perspective:
it leaves unanswered whether, and under which conditions, explanatory thinking
leads to probably true outcomes and it fails to spell out what the added value of an
explanationist perspective is. So, a more specific proposal is needed which answers
both questions.

This brings me to my own account. I begin from the ‘heuristic’ conception of com-
patibilism mentioned above, on which explanation-based thinking is a useful but
imperfect way of achieving Bayesian probabilistic judgments. Some legal scholars
have already hinted at such an account. For instance, Allen and Pardo (2019a, 21)
quote Timothy Williamson’s (2016, 267) remark that:

Inference to the best explanation may be a good heuristic to use when—as
often happens—probabilities are hard to estimate, especially the Bayesian
prior probabilities of theories. In such cases, inference to the best explana-
tion may be the closest we can get to probabilistic epistemology in practice.

An important question for the heuristic account is what kind of probability it is a
heuristic for (Dellsén, 2018, 1746). Like Hedden & Colyvan (2019) I assume that
we are after evidential probabilities. As I wrote earlier in this chapter, to determine
what the evidential probability of a statement is requires ‘making sense of the evi-
dence’. In other words, we have to select and interpret the facts in part by coming
up with potential explanations of this evidence.\(^8\) This is one of the most important
aspects of rational criminal proof. However, Bayesianism is silent on how we
should go about this task. This is (perhaps unsurprisingly) where explanation-based
thinking enters the picture. As has been widely recognized in legal proof scholarship

\(^{8}\) See section 2.4 of this chapter.
and the philosophy of science alike, explanation-based thinking is often what determines what evidence is relevant and so on. For example, Nance (2016, 84) writes the following:

One main motivating concern of those who press the explanatory approach is that [probabilistic accounts] focus on the end product of deliberation, rather than the process of arriving there, giving no direction to jurors as to how to go about assessing the evidence in the case.

In the philosophy of science, the main advocate of this benefit of explanation-based thinking has been Lipton (2004). He argued that Bayesianism does not tell us what evidence to conditionalize on and that IBE can complement the Bayesian account on this point, by helping us determine what facts are relevant (Lipton, 2004, 116). When we think in terms of possible explanations we engage in an act of imagination; we envision potential explanations of the available evidence. Once we have done this, we can ask further questions, which help us determine what evidence is relevant and how we ought to interpret this evidence. For instance, ‘if this scenario were true, what other evidence would we expect to find?’ ‘What would we not expect to find?’ ‘What other evidence does it explain?’ ‘What is left unexplained?’ ‘Can we make this scenario more specific?’ ‘What does a given piece of evidence tell us about the imagined sequence of events – is it plausible or not?’

According to Lipton (2004, 116) another benefit of explanation-based thinking is it tells us how to generate novel hypotheses. We can do so through creative abduction, by asking ‘what else could have explained these facts?’ In legal proof, this last advantage of explanationism is similarly stressed by Nance (2016, 87) who writes that “Bayesianism takes the hypotheses as given but provides an illuminating structure that incorporates explanatory analysis. However, it does not provide the hypotheses to be considered.” By answering these questions, we make sense of the evidence. So, explanation-based thinking plays an essential role in rational proof.

85 I spell out this point in further detail in chapters V and VI.
complementing the Bayesian framework. The latter can only get off the ground when we have applied the former.

Some defenders of explanation-based accounts argue that if Bayesianism relies on explanationism then the former has no utility at all; that it is explanationism that is doing all the work. For instance, Allen argues that it is probabilistic reasoning that has then been absorbed by explanatory reasoning. He writes: “if a reformulated view of subjective Bayesianism and a new addition to the epistemologist’s lexicon, —evidential probability—incorporate the lessons of explanatory theory, then they have lost any independent explanatory power” (Allen, 2020, 6). Similarly, Brennan-Marquez (2019) suggests that “[t]he problem with (...) incorporating [explanationism into Bayesianism] — is not that it makes subjective probabilism wrong per se; it’s that it turns subjective probabilism into a species of explanationism, thus draining probabilism of descriptive power on its own terms.” The point that these authors make is that if Bayesianism is only a framework for expressing the conclusions that we draw from explanation-based thinking, then it is not an independently valuable theory of rational proof. However, I believe that this conclusion is too hasty. As I explained in section 3.3, explanation-based accounts come with their own problems. In particular, it is not fully clear under which conditions explanatory choice leads to probable truth.

A straightforward interpretation of how some authors present explanation-based thinking may sometimes lead us to irrational outcomes, for instance if we are in a bad lot situation. How should we choose between the competing explanations if we want to avoid such irrationality? Proposals by explanationists have, so far, been tantalizingly brief. As I will argue, Bayesian probability theory provides the promise of more clarity in this regard as it is (among other things) a precise and versatile framework for exploring how and when explanatory reasoning leads to probably
true outcomes.\textsuperscript{86}

This last point brings me to the question why explanation-based thinking is a heuristic for Bayesian inference. In other words, why does it track Bayesianism only imperfectly? The reason for this is that, as I will argue at various points in this thesis, when we engage in explanation-based thinking there is a trade-off between the aim of making sense of the evidence and that of reaching probably true conclusions. When we try to make sense of complex sets of evidence, we typically do not want to consider mutually exclusive and exhaustive hypotheses, as these are too general in nature to provide guidance in what the evidence tells us. What evidence is relevant and so on often only follows when we are considering detailed scenarios. For example, the hypothesis that ‘the suspect killed the victim at time X, at place Y’ helps us look for trace evidence and investigators can ask those who were at that place at that time what they saw. In contrast, the hypothesis that ‘the witness killed the victim’ provides none of these benefits.

However, there is a tension between the need to consider specific explanations and the goal of reaching probably true conclusions. As various philosophers of science have pointed out informative (and therefore detailed) hypotheses will almost always be less probable than more general, less detailed versions of themselves (Cabrera, 2017; 2018). For instance, the hypothesis that ‘the suspect killed the victim’ implies the hypothesis that ‘the suspect killed the victim at time X, at place Y’. So, if the latter is true, so is the former. But the converse is not always true; the suspect may have killed the victim at another place or time. A consequence of this point is that the more specific our explanations, the smaller the space of possibilities that we can explore and the lower the probability of our conclusions. When we engage in explanation-based thinking we usually deliberately limit ourselves to only a few of the

\textsuperscript{86} This is in line with the remark by Pardo and Allen (2008, 253) put it: ”Explanatory considerations (…) are inherent and fundamental; to the extent Bayesian perspectives can clarify and approve on those considerations, they prove their worth. To the extent they do not, they do not.”
possible scenarios of what happened, for the sake of making sense of the evidence. However, this does mean that we can never be certain that our conclusions are the same as those of a perfectly rational agent with the same evidence; in contrast with this mythical agent, we may have overlooked something. Hence, we can best think of explanation-based thinking as an imperfect, but useful and tractable method of arriving at evidential probabilities.

5. Conclusion

Reaching accurate conclusions is a fundamental goal of criminal trials. Any theory of rationality worth its salt should have a solid connection to this aim. However, the two most popular accounts of rational criminal proof both face serious problems in this regard. For the Bayesian account of rational proof (which we must distinguish from the use of probabilistic tools in court) the worry is that it requires a way to connect rational degrees of belief to the strength of the available evidence. The most common way of doing so by contemporary legal Bayesians is in terms of ‘evidential probabilities’. However, we need a clearer account of this term.

In contrast, a major problem for explanation-based accounts is that it is unclear how one ought to choose between the competing explanations if one wants to obtain probably true beliefs. Critics argue that IBE-based approaches may therefore fall victim to particular problems – such as the bad lot problem or the problem of disjunctive explanations – or become equivalent to Bayesianism. Due to these worries proponents of explanation-based approaches face the objection that their account is either trivial or irrational.

I suggested that the worries for either account open up the possibility for a productive partnership. To spell out what this partnership looks like, we can draw inspiration from the philosophy of science. In particular, we may view explanation-based thinking as a heuristic for optimal Bayesian inference. The value of such thinking is that it helps make sense of the evidence and thereby to reach a Bayesian posterior
evidential probability. In the rest of this thesis, I further develop this picture. I begin with the question how explanationists should understand the proof of guilt beyond a reasonable doubt standard if they want to draw probably true conclusions.
Chapter abstract

Explanation-based accounts of rational legal proof are often criticized for being underdeveloped. One question in need of further attention is how the explanationist should understand the proof of guilt beyond a reasonable doubt (BARD) standard. Most explanationists who discuss this matter explicitly or tacitly accept the `No plausible alternative' (NPA) account of proof BARD. In contrast, I defend an IBE-based interpretation, on which guilt is only established BARD if (i) the best guilt explanation in a case is substantially more plausible than any innocence explanation, and (ii) there is no good reason to presume that we have overlooked evidence or alternative explanations that could realistically have exonerated the defendant. This account has at least three key advantages over existing work. First, it is a comparative account, which I argue is better suited for arriving at accurate verdicts than the non-comparative NPA account. Second, my account is not susceptible to various arguments that have been raised against IBE-based frameworks of rational criminal proof. Third, my account provides helpful guidance for fact-finders when having to deal with unfalsifiable but barely plausible alternative scenarios. I illustrate these points using the Dutch Simonshaven manslaughter case.

1. Introduction

Explanationists view criminal trials as a competition between scenarios. To give an example of such a competition, in this chapter I discuss the Simonshaven case in which a woman was killed in a forest. The prosecution argued that the killer was the woman's husband. The defendant offered the alternative scenario that a
madman jumped out of the bushes and killed her. Imagine that you are in the judge's robe and have to choose between these scenarios – 'her husband killed the victim' and 'an unknown madman killed the victim'. According to the explanationist account, you should ask yourself which of these options best explains the evidence in the case (i.e., which is the most plausible). Now assume that the husband killing her is the best available explanation. Should you then convict him because it is proven that he is the culprit? To put it differently, if fact-finders have to choose between several scenarios – some better than others – what determines which they can accept as proven (if any)? This question has not been worked out in sufficient detail within explanation-based frameworks. How we answer it will depend on how we interpret the relevant proof standard, which tells us when we may deem facts to be legally proven.

In this chapter I discuss how those who defend explanation-based approaches should understand the beyond a reasonable doubt (BARD) proof standard in criminal law. This standard is used in common law countries, though my conclusions here hold for many other legal systems too. The BARD standard is notoriously vague. Even though various interpretations of this standard exist, according to Laudan they are all problematic (Laudan, 2006, 35-47). He claims that these interpretations seem to create more confusion and potential for error than they prevent. In particular, he argues that they are too unclear to be meaningfully applied in real cases, too reliant on the subjective hunches of the person who applies it or insufficiently connected to the goal of avoiding false convictions (Laudan 2006, 87). My aim in this chapter is not just to propose an interpretation that fits best within the explanationist framework, but also to argue that this interpretation overcomes Laudan's criticisms of other, non-explanationist interpretations. The interpretation that I propose is as follows:

Guilt is only established BARD if (i) the best guilt explanation in a case is substantially more plausible than any innocence explanation, and (ii) there is no
good reason to presume that we have overlooked evidence or alternative explanations that could realistically have exonerated the defendant.

My interpretation is an expanded version of Thagard's (2003, 2008) proposal that we could interpret the BARD standard as a form of inference to the best explanation (IBE). This is a comparative account of the BARD standard: we determine whether guilt is proven by looking at how well our guilt explanation performs compared to competing innocence explanations (Thagard, 2003; Thagard, 2008). In contrast, many explanationists explicitly or tacitly adopt the non-comparative no plausible alternative (NPA) interpretation of the standard (Allen & Leiter, 2001; Josephson, 2001; Pardo & Allen, 2008; Allen, 2010; Bex & Walton, 2012). According to NPA, proving guilt BARD requires that we have a sufficiently plausible explanation implying guilt and no sufficiently plausible explanation implying innocence. As I will argue in section 3, such a non-comparative interpretation is undesirable on epistemic grounds. In particular, a defendant should only be convicted if the probability that he is guilty is sufficiently high in order to minimize false convictions. However, a non-comparative account such as NPA sometimes leads to factual conclusions which are insufficiently probable. The reason for this is that the relative plausibility of explanations influences their probability. So, we need an account that captures this comparative dimension.

Many object to a comparative approach to fact finding in criminal trials. According to such critics, criminal law is about whether the case for guilt is good enough, not whether it is better than the case for innocence. However, as I will argue in section 4, IBE does not imply that we should only care about the relative performance of explanations. First, we can look at the absolute performance of explanations to efficiently reach comparative conclusions. Second, and more importantly, IBE does not imply that we can convict a defendant based on an implausible guilt explanation, even when that explanation is better than any innocence explanation. If all our explanations are implausible then we are in a so-called bad lot situation. As I will argue, in bad lot situations we have reason to presume that a better investigation
could have exonerated the defendant. The comparative case for guilt then lacks robustness and we should therefore not convict the defendant. As I argue in chapters VII and VIII, the idea that robustness is a requirement for proof of guilt BARD is a plausible assumption for any account of the BARD standard. So, the proposed IBE-based account overcomes the bad lot problem without the need for ad hoc modifications.

To illustrate how my account can be applied to criminal practice I use the Dutch Simonshaven case as a running example.\(^87\) I further introduce this case in section 2. I also discuss explanationism, NPA, and IBE in more detail in that section. In section 3 I discuss why explanationists need a comparative account of the BARD standard, which NPA does not provide. Finally, in section 4, I argue that my comparative account nonetheless squares with the non-comparative nature of criminal law.

### 2. Preliminaries: The Simonshaven case and reasonable doubt

In this section I set the stage for my argument. First, I discuss the Simonshaven case, which I will use as an example throughout my chapter. After that I introduce the BARD standard in more detail. Finally, I turn to the IBE and NPA interpretations of this standard.

#### 2.1 A killing in a forest

The Simonshaven case took place in the Netherlands in 2009. The case's name comes from a village near the forest where the alleged events occurred.\(^88\) I will

---

\(^87\) This case is also discussed in the special issue of *Topics on cognitive science, titled ‘Models of Rational Proof in Criminal Law’*. See Prakken, Bex & Mackor (2020).

\(^88\) The Hague Appellate Court. (18th February 2015) ECLI:NL:GHDHA:2015:282. In the Netherlands fact finding is done by judges, rather than juries. Furthermore, the Dutch legal system lacks the BARD standard. It has the conviction raisonnée standard, which stipulates that conviction is permitted only if the fact-finder is (reasonably) convinced of the
provide more details of the case throughout this chapter, where needed to illustrate my ideas. The simplified case is as follows:

A couple recently stopped living together due to relationship issues. Nonetheless, they sometimes take walks together in a nearby forest. One day, witnesses see them walk into this forest. About an hour later the husband calls his son in a panicked and confused state. The son races to the forest where he finds his dad who says that someone attacked them. The son then calls the police who find the wife's body lying on a nearby path in the forest in a pool of blood. There are severe wounds on her face. The husband is also injured. He has minor facial injuries. When questioned, the husband claims that an attacker suddenly jumped out of the bushes and knocked him out. He claims to remember only vague details about the attacker, such as that he was light skinned and of average height. After he is marked as a suspect by the police, he calls upon his right to remain silent and makes no further statements. Forensic investigation shows that a gun was fired near the victim's face and that her wounds were caused by being hit with a blunt object, possibly a gun. The police also find shell casings near her body. One of these casings contains DNA-evidence that belongs to neither the victim nor her husband. However, despite intensive searching, no gun is found near the crime-scene. During the trial the prosecution argues that the suspect could have dumped the gun in a nearby river. The children of the defendant testify that he was jealous because the victim had recently started dating a new lover. They also testify that their father owned several guns. Apart from the victim's husband, no other suspect is identified.

The fact-finder (in this case a judge) is now faced with the question which of the following competing explanations to believe (if any):

defendant's guilt. The account that I propose in this chapter can also be extended to this standard because it can be read as an account of justified belief in the guilt of a defendant. My aim is not to draw conclusions about the actual case. The above is only a summary that focuses on the most important items of evidence.
**The husband scenario:** The suspect walked into the forest with the victim. They either got into an argument and the suspect killed the victim in anger, or the suspect had planned to kill her there. The suspect fired a gun in the vicinity of the victim and then beat her to death using his handgun. He then drove to a nearby river to dump his gun after which he drove back and called his son and claimed someone had attacked them.

**The madman scenario:** The suspect walked into the forest with the victim. During the walk an unknown perpetrator randomly attacked them, first knocking the husband unconscious. The unknown man then fired a gun in the vicinity of the victim. Then he beat her to death using the gun and fled.

### 2.2 Proof of guilt beyond a reasonable doubt

The judge in the Simonshaven case faced a choice between the competing 'husband' and 'madman' scenarios. The husband explanation implies that the defendant committed the alleged criminal act (killing his wife). If such a guilt explanation is proven, this can lead to conviction. The 'madman' explanation implies the innocence of the defendant – it is an innocence explanation. If such an explanation is sufficiently convincing it may lead to a reasonable doubt and therefore to acquittal.

In this chapter I ask when the explanationist should consider the events described in the guilt explanation proven beyond a reasonable doubt and when not. When we use the term reasonable doubt, we usually refer to common law legal systems, such as the American and the English. These systems use ‘proof of guilt beyond a reasonable doubt’ as their proof standard for criminal trials. However, when I talk about reasonable doubt in this chapter (and in this thesis in general) I am not talking about a specific doctrinal interpretation used by particular courts for two reasons. First, it is notoriously unclear how common law courts interpret this standard. For instance, the United States Supreme Court has refused to define the term reasonable doubt, stating that “[a]ttempts to explain the term ‘reasonable doubt’ do not
result in making it any clearer to the minds of the jury.” (Strong, 1992, 447). To the extent that other common-law courts have defined the standard, they do so in varying and (often deliberately) vague ways. That is why Laudan (2006, 295-296) calls the standard “obscure, incoherent, and muddled”.90 Second, even in systems that do not have the reasonable doubt standard, such as European civil law systems, courts should not convict defendants if they have serious doubts about their guilt.91 My conclusions pertain to those systems too.

When I ask how the explanationist should understand the reasonable doubt standard, I mean this in an epistemic sense. On an epistemic view of criminal proof the principal goal of trials is drawing accurate factual conclusions and, to the extent that factual errors are made, distributing those errors fairly (e.g., Dworkin, 1985b; Goldman, 2002; Stein, 2005; Pardo, 2021). With respect to this second point, false convictions are generally considered to be much worse than false acquittals (Epps, 2015). This is captured in Blackstone’s well-known maxim that “it is better that ten guilty persons escape than that one innocent suffer” (Blackstone, 1962).92 That is why many authors have suggested that we should only consider guilt proven if it is very probable – e.g., at least 90%, 95% or even 99% – that the defendant committed the alleged acts (Connolly, 1987; Dane, 1985; Dhami, 2008; Lippke, 2010; Laudan, 2011; Walen, 2015). The idea is that the higher the proof standard, the more

---

90 See also Picinali (2015) who argues that people often ask for clarification about this standard and in the absence of a definition, they find it difficult to apply and that this confusion about how we should understand the standard is also widespread among academics and courts.

91 For example, in the Netherlands various scholars have argued that the Dutch proof standard of criminal proof stipulated in article 338 Dutch code of criminal procedure – which merely says that the court should be convinced based on the admissible evidence that the defendant perpetrated the alleged acts – is best read as the reasonable doubt standard (Ter Haar & Meijer, 2018, 7.4; Nijboer et al., 2017, 73-74). Furthermore, in the new Dutch Code of Criminal Procedure, which is currently in preparation, the reasonable doubt standard is explicitly adopted in the legal proof system (4.3.2.1(2) New Dutch Code of Criminal Procedure).

92 Note that this ratio should not be taken too literally; various other ratios have been suggested.
difficult it is to convict someone, the fewer false convictions there will be. However, the cost of this is that more cases in which the defendant quite probably, but not sufficiently probably, will be acquitted, raising the number of false acquittals.\(^\text{93}\)

This idea – that we are after a sufficiently high probability – underlies the Bayesian account of proof beyond a reasonable doubt. On the Bayesian account, proof beyond a reasonable doubt requires (at least) that the posterior probability of guilt meets some threshold (Gardiner, 2019).\(^\text{94}\)

As explained in the previous chapter, the Bayesian and explanation-based accounts differ on many points. However, with respect to the proof standard, they share the assumption that the defendant should only be convicted if it is very probable that they perpetrated the alleged acts. As Allen and Pardo (2018, 1580) put it: “The explanatory account shares the same ends or goals as probabilistic approaches, which have to do with various policy judgments about the likelihood of disputed facts and allocating the risk of error between the parties.” So, what I am after in this chapter is to explore how we should understand the beyond a reasonable doubt standard to ensure that a choice between competing explanations leads to probably true outcomes.

For explanationists, whether an explanation is proven depends primarily on the \textit{plausibility} of that explanation and that of its competitors (Pardo and Allen, 2008; Allen, 2010; Josephson 2001). As discussed in the previous chapter, a plausible explanation is one that explains the evidence (sufficiently) well. Plausibility is usually further specified as a set of explanatory virtues, with different authors suggesting different virtues. For instance, in the story model of legal proof, how good an explanation is depends on how much of the evidence the explanation accounts for, whether it contains internal contradictions and problematic gaps, whether it has all

\(^{93}\) For instance, a standard of 0.9 will result in an expected false conviction rate of at most 10%. However, the exact ratio of the two types of error will depend on the probabilities assigned to defendants and how they are distributed (Allen, 2014).

\(^{94}\) Some argue that this interpretation is problematic, e.g., because it leads to counterintuitive outcomes in cases of naked statistical evidence or that it is too subjective (Ho, 2015). I will leave these considerations aside in this chapter.
its parts and how well it fits with our knowledge of the world (Pennington and Hastie 1992, 190–199; Bex 2011, 91–92). Pardo and Allen (2008, 230) propose that “all other things being equal [an explanation is] better to the extent that it is consistent, simpler, explains more and different types of facts (consilience), better accords with background beliefs (coherence), is less ad hoc, and so on; and is worse to the extent that it betrays these criteria.” Josephson (2001) and Amaya (2009) propose yet other criteria. I do not want to delve too deeply into the debate about which set of criteria best captures plausibility. Instead, I offer the following working definition, which is an attempt at summarizing the above proposals:

An explanation is more plausible to the extent that it: (a) explains more of the evidence in the case, (b) fits better with the fact-finder’s background beliefs about how the world usually works, and (c) has parts that fit together coherently.

So, when we want to determine whether an explanation is plausible, we ask questions such as how well this particular explanation fits with our background beliefs and how much evidence in the case it explains. Understood this way, plausibility is a qualitative concept that is distinct from the quantitative idea of probability as used by Bayesians. Nonetheless, explanationists assume a close connection between the two concepts. In particular, an underlying assumption of this framework is that plausibility is a guide to the probability of explanations. For instance, as Pardo & Allen (2008, n45) put it “[Explanationism] is, at root, based on the notion that explanatory success tracks likelihood of truth – the better the explanation, the more likely true.” In other words, the more plausible an explanation is – i.e., the more of the above virtues it displays – the more probable it ceteris paribus is as well. In this chapter I assume that there is indeed a ceteris paribus connection between plausibility and probability. In other words, all other things being equal, the more plausible

---

95 This is similar to how philosophers of science such as Lipton (2004) and Thagard (1978) approach the notion of ‘best’ in inference to the best explanation. In contrast, authors such as Douven (2017) and Schupbach (2016) attempt to ‘define explanatory goodness’ in a mathematically precise way.
explanation is, the more probable it is as well. However, as I will argue, there are conditions in which a very plausible explanation is not necessarily very probable. One thing that is important to note for the argument that I develop in section 3 is that when we discover, for instance, that a better explanation exists, our formerly best explanation does not become any less plausible. After all, it does not suddenly explain fewer facts or become more internally incoherent.

2.3 Two explanation-based interpretations of the BARD standard

The explanationist framework instructs fact-finders to look at the competing explanations and to choose between them based on how plausible they are. Choosing an explanation means accepting that explanation as proven. The following scheme expresses this idea (adapted from Laudan, 2007):

(1) \( e_1, e_2, \ldots, e_n \) are the salient facts to be explained.

(2) Hypotheses \( h_1, h_2, \ldots, h_n \) each explain \( e_1, e_2, \ldots, e_n \).

(3) Rival explanations for \( e_1, e_2, \ldots, e_n \) have been earnestly sought out but the search has produced only \( h_1, h_2, \ldots, h_n \).

(4) [Proof standard]

(5) Therefore, \( h_i \) is probably true.

I have left premise 4 – the proof standard – unspecified. Such a proof standard determines on what grounds the fact-finder should choose between the available explanations. For criminal cases this is the BARD standard. Explanationists have paid little attention to how the BARD standard should be interpreted. In the literature we find two (brief) proposals: Thagard (2003, 2008) has proposed a version of the BARD standard based on IBE. In contrast, Pardo and Allen (2008) claim that all proof standards within common law systems (such as the preponderance of

\[96\] I discussed some worries regarding this assumption in section 3.3 of chapter II.
evidence standard), except BARD, are IBE. For BARD they propose the NPA standard (Pardo and Allen, 2008; Allen and Pardo, 2019).

2.3.1 Inference to the best explanation

As explained in the previous chapter, the concept of IBE originated in the philosophy of science and epistemology (Peirce, 1974; Harman, 1965; Thagard, 1978; Lipton, 2004). On IBE we infer the probable truth of an explanation on the grounds that it would, if true, provide a better explanation of one’s evidence than any other available alternative explanation. Taking the name IBE at face-value it therefore leads to the following criterion for accepting an explanation as proven:

\[(4) \text{bi is the best explanation in the set } \{b_1, b_2, \ldots, b_n\}.\]

With respect to criminal law, Thagard has proposed — but not argued — that the BARD standard could be interpreted in terms of IBE (Thagard, 2003, 366; 2008, 141). His version of IBE requires more than simply choosing the best available explanation. He proposes that the BARD standard is only met when the best guilt explanation is “substantially more plausible” than any innocence explanation. An explanation being only slightly better than any of the alternatives fails to yield sufficient certainty for conviction. If, for instance, both the prosecution and the defense offer explanations that are roughly equally good, their explanations “will lack uniqueness, and great uncertainty will result” (Pennington and Hastie 1992, 191). This would conflict with the goal of a just error distribution mentioned earlier. We want a high degree of certainty before conviction. That is why the guilt explanation must be ‘substantially more plausible’.

So, what IBE asks the fact-finder to do is to make a judgment about the relative plausibility of the various explanations (I call this a relative judgment). We can contrast this standard with the following, which does not involve such a relative judgment.
2.3.2 No plausible alternative

NPA originates with Pardo and Allen (2008, 238-9), who explicitly contrast it with IBE:

In criminal cases, rather than inferring the best explanation from the potential ones, fact-finders [should infer] the defendants innocence whenever there is a sufficiently plausible explanation of the evidence consistent with innocence (and ought to convict when there is no plausible explanation consistent with innocence assuming there is a plausible explanation consistent with guilt).


(4) \( h_i \) is a sufficiently plausible guilt explanation and there are no sufficiently plausible innocence explanations in the set \( \{ h_1, h_2, \ldots, h_n \} \).

How should we interpret this criterion and how does it differ from IBE? As I mentioned above, IBE depends on a relative judgment of the plausibility of explanations. Although Allen and Pardo do not explicitly state how NPA should be interpreted, I follow the same interpretation that others, such as Sullivan (2018), Ho (2019) and Kolflaath (2019) give, namely that NPA is instead about judging explanations based on their absolute plausibility. This means judging explanations on an absolute scale, independently of one another. As Kolflaath (2019, 5–6) puts it:

[NPA means] that fact-finders should decide first whether there is a plausible explanation consistent with guilt, and then – if they take this to be the case –

---

97 “Guilt has been established beyond a reasonable doubt when there is no plausible alternative explanation for the data that does not imply the guilt of the defendant.”

98 “[I]n criminal cases the prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence.”

99 “[A] plausible explanation consistent with innocence creates a reasonable doubt.”

100 It is similar to the no plausible alternative theory of knowledge, which states that to know proposition P, we need to rule out all relevant alternatives to P (Lawlor, 2013).
decide whether there is also a plausible explanation consistent with innocence. Thus, [NPA seems] to presuppose that the first decision can be made independently of the second. And although the need for the second decision depends upon the first decision, whether or not there is a plausible explanation consistent with innocence would not, as far as I can tell, depend upon whether there is a plausible explanation consistent with guilt.

In other words, fact-finders have to check whether explanations individually meet some standard of plausibility, rather than how well they perform relative to one another. When an explanation is insufficiently plausible on its own – i.e., displays the various explanatory virtues to an insufficient degree – it is eliminated. I will call reasoning about how well an explanation performs on its own *absolute judgment.*\(^{101}\) So, while NPA is comparative in the sense that it instructs fact-finders to consider and compare the possible alternative explanations on each side, it is *not* comparative in the sense that it takes into account the relative plausibility of explanations (Sullivan 2018, 5; Ho, 2019).\(^{102}\)

Let me give a toy example to clarify. One aspect that makes an explanation plausible is how much of the evidence it explains. Suppose that an explanation explains 7 items of evidence. This is a conclusion about how well the explanation performs on its own – i.e., how plausible it is in an absolute sense. In contrast, a conclusion about relative plausibility could, for instance, be `explanation A explains 2 more

---

\(^{101}\) These two types of judgments—relative and absolute—are equivalent to two modes of reasoning distinguished in cognitive psychology. ‘Selective processing’ where something is judged in isolation, whereas ‘comparative processing’ means judging something in the context of the available alternatives (e.g., Sanbonmatsu et al. 1998, 2011; Trope and Liberman, 1996). On NPA, the judgments about whether there is a ‘sufficiently’ plausible’ guilt and whether there is a sufficiently plausible innocence explanation can be made in isolation, which therefore corresponds to selective processing.

\(^{102}\) Though Allen and Pardo’s call their overall account the *relative plausibility theory* (Allen and Pardo 2019), their definition of the BARD standard is not about relative plausibility. As Allen and Pardo (2019, 31) themselves remark: “[The BARD standard is] ‘comparative’ in a different sense, however. It is comparative in the sense that determining whether the threshold is met will involve considering and comparing the possible alternative explanations on each side.”
pieces of evidence than explanation B.' Within this toy example, if we employ IBE we would look at how many more pieces of evidence explanation A explains than explanation B when we decide which explanation to accept. When we employ NPA, we first look at whether explanation A explains a 'sufficient' amount of evidence, say 5, and then we look whether explanation B explains a 'sufficient amount of evidence'.

One important thing to note is that an absolute judgment can be turned in to a relative judgment. For instance, based on the fact that explanation A explains 7 pieces of evidence and explanation B 5, we can reach the relative conclusion that A explains two more pieces of evidence than B. As I will argue in section 4.1, this is one of the reasons why absolute plausibility still plays a key role within IBE – because it is a useful method for reaching relative conclusions. However, a relative judgment about plausibility cannot be turned into an absolute one. For instance, in this toy example we cannot infer that A explains at least 5 pieces of evidence from the fact that it explains 2 pieces of evidence more than B (unless we know how much B explains). This in turn means that because NPA is about how explanations perform in an absolute sense, judgments about relative plausibility serve no function within the framework. In other words, a fact-finder never has any reason for asking how much better one explanation performs relative to the other if she is after conclusions on an absolute scale.

The fact that NPA does not take relative plausibility into account makes it problematic. Both relative and absolute judgment are important for accurate fact finding. In the next two sections I deal with each form of reasoning in turn. I begin with relative judgment and argue that NPA cannot account for this without losing what makes it distinctive (section 3). After that I show how IBE can account for the value of absolute judgments, while still being about relative conclusions at heart (section 4).
3. Why relative plausibility matters

On IBE proof of guilt BARD is about relative judgments of plausibility, whereas on NPA it is about absolute judgment. As I will argue in this section, relative judgment is (at least sometimes) necessary for accurate conclusions. In particular, sometimes the plausibility of one explanation going up can mean the probability of a competing explanation going down. The goal of the BARD standard is ultimately probabilistic; fact-finders should only accept a guilt explanation as proven if it is sufficiently probable. So, in order for plausibility to adequately track probability we should also take into account the relative plausibility of explanations. NPA fails to account for this (or, if it does, collapses into IBE). However, before looking at this argument, let me first say more about how explanations compete both in terms of probability and plausibility.

3.1 Explanatory competition

The relative plausibility of explanations is important to accurately assess their probability when these explanations compete. What do we mean by competing explanations? One sense in which, for instance, the `husband' and the `madman' explanation in the Simonshaven case compete is that they attempt to explain the same facts (e.g., the victim's dead body). However, not all explanations that explain the same facts compete. Events can have multiple causes, which do not have to conflict in any way. As I use the term here, two explanations – A and B – compete when the following condition is met: if evidence raises the probability of explanation A, the probability of explanation B is thereby lowered. In other words, the confirmation of explanation A implies the disconfirmation of explanation B. This is also called explaining away: when an event has multiple possible causes, and the confirmation of one cause reduces the need to invoke alternative causes (Kjaerulff and

103 See Schupbach and Glass (2017) for an extensive discussion on this topic.
Madsen 2008). The probabilistics of explaining away can best be shown in a Bayesian network. For instance, Hepler et al. (2007) give the following idiom\textsuperscript{104}:

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{idiom_for_competing-causes.png}
\caption{Idiom for competing causes}
\end{figure}

This idiom can be understood as follows: The top nodes are potential causes of the event in the bottom node. For instance, in Hepler et al.’s (2007) example, the event is ‘the defendant confesses to the crime’, and the potential causes are ‘the defendant is guilty’ and ‘the defendant was coerced by an interrogating official.’ However, the causes could also be `madman' and `husband' and the event `dead body'. What matters is that there is a path between the two causes through the event that they explain. This expresses the explaining away effect that I just mentioned, where the probability of one explanation influences that of its competitors.\textsuperscript{105}

In such a network, evidence can change the probability of an explanation in two ways. First, some evidence directly confirms or disconfirms an explanation. Second, evidence can (dis)confirm an explanation by (dis)confirming one of its competitors. For instance, take the piece of evidence from the Simonshaven case that the victim had a new lover, which implies that the defendant had a motive for killing her. We can include this evidence in the Bayesian network as follows:

\textsuperscript{104} Modeling mutually exclusive hypotheses requires a more sophisticated approach (Fenton et al., 2013). However, the general idea remains the same.

\textsuperscript{105} Assuming we know that the event occurred. For instance, that we are sure that there is indeed a dead body.
There is a direct probabilistic link between the `husband' explanation and the newly introduced evidence. In this case it means that the piece of evidence increases the probability of the explanation that the husband was the killer. This piece of evidence directly confirms the explanation. Yet because the two explanations compete, this piece of evidence also indirectly disconfirms the `madman' explanation.

What does all this have to do with the plausibility of explanations? Criteria of plausibility – such as how many pieces of evidence an explanation explains – are intended to be a guide to the probability of these explanations. However, the two ways in which evidence can probabilistically (dis)confirm an explanation are reflected differently in terms of their plausibility. Evidence that directly confirms an explanation makes this explanation more plausible in an absolute sense. For example, when we introduce the evidence that the defendant had a motive for the killing, the `husband' explanation now explains one more piece of evidence. It becomes more plausible as a result, regardless of what this evidence does to its competitors. Conversely, evidence that is indirectly relevant influences the relative plausibility of explanations. It disconfirms one explanation probabilistically, by directly supporting a competing explanation. For example, the motive evidence makes the `madman' explanation less probable, by making the `husband' explanation more plausible. However, the evidence of motive leaves the plausibility of the `madman' explanation the same, as it does not make this explanation incoherent, does not make it fit
less with our background beliefs and this motive evidence is not in conflict with the madman explanation. To see why, imagine that we knew with certainty that the husband did not kill his wife. Now suppose that we learned that the defendant had a motive for killing her. Since we know for sure that he did not kill his wife, this does not give us any reason to disbelieve the explanation that a madman killed her. Conversely, if we did not know with certainty that the `husband' explanation is false, this evidence would be relevant. After all, it makes the husband explanation more plausible. So, the relevance of evidence about motive to the `madman' explanation can only be determined by looking at both explanations in relation to one another – i.e., to their relative plausibility.

3.2 Relative plausibility and probable truth

The above is relevant for the discussion between IBE and NPA because a great difference in relative plausibility can mean that the two principles lead to different recommendations about whether the fact-finder should convict or acquit due to the resulting probabilities of these explanations. In particular, in some cases we have a highly plausible guilt explanation that explains a great deal of evidence, making it much more plausible than any innocence explanation. IBE then dictates conviction. However, if there is also an innocence explanation that is sufficiently plausible when looked at on its own, then NPA dictates acquittal. Yet, in such cases, the evidence for the guilt explanation might indirectly disconfirm the innocence explanations to such a great degree that they become highly improbable. For instance, if we believe that the defendant in the Simonshaven case had motive, means and opportunity, then this supports the explanation that he killed his wife. It thereby disconfirms the competing `madman' explanation.

The more plausible we make the `husband' explanation, the less probable the `madman' explanation becomes. Add enough evidence supporting the `husband' explanation and it (indirectly) disconfirms the `madman' explanation to such a great extent that no reasonable fact-finder could believe it. For example, one piece of
evidence that I have not mentioned is that multiple witnesses testified that the defendant had threatened the victim several times in the past. First, the son of the defendant said that the defendant had repeatedly told the victim “If anything happens I'm taking you with me”. Second, the brother of the victim reported that the defendant told the victim “if you leave me, I'll kill you.” We can imagine even more of such evidence for the ‘husband explanation’ being found, which would increase the plausibility of the ‘husband’ explanation, thereby lowering the probability of the ‘madman’ explanation.

So, whether the defendant can be convicted – at least sometimes – depends on the relative plausibility of explanations. Fact-finders should therefore be attentive to the relative plausibility of explanations, especially when the difference in plausibility is large. However, NPA does not take relative plausibility into account. Instead, it directs fact-finders to look at how well explanations perform on their own. It therefore ignores the probabilistic influence of how explanations perform relative to one another. Because we only want to accept an explanation as proven if it is sufficiently probable we should, ceteris paribus, prefer IBE’s conclusion over that of NPA, where the two differ. After all, IBE does pay attention to the relative plausibility of explanations.

One move that proponents of NPA can make in response to this, is to suggest that their account should be interpreted differently; that fact-finders should look at both absolute and relative plausibility. This is, for instance, what Josephson (2001, 1626) proposes. He puts forward the NPA definition of BARD but also says that what makes an explanation plausible depends both on how decisively the leading hypothesis surpasses the alternatives and how well it stands by itself, independently of the alternatives. Presumably, on such an approach, fact-finders should convict when the best guilt explanation is substantially more plausible than the best innocence explanation.

---

106 “Guilt has been established beyond a reasonable doubt when there is no plausible alternative explanation for the data that does not imply the guilt of the defendant” (Josephson 2001, 1642).
explanation, even if the best innocence explanation is plausible when looked at on its own. However, such an interpretation would go against how plausibility is usually defined (as the degree to which an explanation displays certain virtues). Furthermore, such an interpretation would be equivalent to what I propose here. It would then instruct fact-finders to primarily look at the absolute plausibility of explanations, but also that relative plausibility can trump considerations of absolute plausibility to avoid (probabilistically) unwanted outcomes. As I will argue in section 4, on IBE it is also important to check how well an explanation stands by itself. In fact, as I will suggest, it is often useful to primarily look at how well explanations perform in an absolute sense. On this interpretation, the two approaches would then give fact-finders identical instructions. Furthermore, this amended form of NPA would be mostly uninformative about how fact-finders should reason or when guilt can be proven in particular cases. It would amount to little more than a reiteration of the explanationist framework – which tells fact-finders to look at competing explanations of the evidence.

In contrast, on IBE, the goal of the fact-finder is always to reach relative conclusions: defendants can only be convicted if the case for guilt is much stronger than the case for innocence. Absolute judgment can serve such a relative conclusion in at least two ways. First, it helps to reach such relative conclusions efficiently (section 4.1). Second, it tells us something about the possibility of unconsidered explanations and undiscovered evidence that could have overturned the (relative) case for guilt (section 4.2). So, my account of IBE explains why fact-finders should look at both relative and absolute plausibility and how these two aspects interrelate. Furthermore, it provides meaningful guidance to fact-finders on how to reason about both relative and absolute plausibility as I will show below.

\footnote{Indeed, despite offering the NPA definition of the BARD standard, Josephson refers to his account as ‘inference to the best explanation’.}
3.3 Relative judgment in practice

As said, if they only look at explanations on their own, fact-finders risk making inaccurate decisions. In fact, psychological data suggests that judging explanations in isolation may lead to biased conclusions more often than looking at how explanations perform relative to one another (Sanbonmatsu et al., 1998). For instance, people often fail to ask whether evidence that supports one explanation also supports alternative explanations. This can cause individuals to overestimate the probability of potential causes of an event when making absolute judgments (Sanbonmatsu et al., 1993). So, relative judgment can be a method of avoiding biased conclusions.

Apart from its benefits in avoiding biased conclusions, relative judgment can also be strategically preferable to absolute judgment. Parties may want to make their case by focusing their arguments on the relative plausibility of explanations when it is easier to do so than to argue for an absolute conclusion. I am thinking in particular about cases where the defense gives an unfalsifiable ad hoc explanation. For instance, a guilty defendant may give an untrue, vague explanation that only explains the evidence that has been presented to him. Giving direct arguments against this explanation can then be difficult. Yet the prosecution can still make a case by giving reasons for the guilt explanation. For example, the madman explanation is low on details. It does not contain many specifics about what the attacker looked like or how the attack occurred. This makes it difficult to argue that the explanation is implausible based on the evidence. Furthermore, given that it could have happened, a fact-finder might be hard pressed to provide reasons why it conflicts so much with our background beliefs that it can be rejected. In other words, it might be hard to find convincing reasons against this explanation that do not invoke how it performs compared to the alternative explanation. If the fact-finder then wants to convict, a strategy based on relative judgment may then be more prudent than one based on absolute judgment. In other words, a defendant does not have to be acquitted only because it is difficult to offer reasons why his story is implausible. He
can also be convicted if the case for his guilt is substantially more plausible than the case for his innocence (as long as we are sufficiently certain that we have considered all relevant evidence and discovered all relevant evidence, see next section).

Yet is it possible for a fact-finder to reason meaningfully about whether one explanation is ‘substantially more plausible’ than another, as IBE dictates? Laudan objects to this. He believes that jurors will not be able to easily reach consensus about whether the prosecution’s case is substantially more plausible than the defendant’s. According to him, IBE therefore seems “an invitation to confusion and disparate interpretations of the [BARD] standard” (Laudan, 2007, 300). I agree with Laudan that people can have different intuitions about whether an explanation is substantially more plausible than another. However, this is not a worry unique to IBE. The same worry applies to NPA (when are explanations ‘sufficiently good’?) and to probabilism (when is the case for guilt ‘sufficiently probable’?). What matters is that fact-finders can meaningfully reason about the relative plausibility of explanations.

Let me use the example of the Simonshaven case to show that this type of reasoning is not nearly as difficult as Laudan makes it out to be. Again, I am not interested in drawing conclusions about how fact-finders should have reasoned in the actual case. Instead, my point is that rational fact-finders could reason in relative terms. An argument for guilt based on relative plausibility might go as follows:

Guilt is proven in this case, because the ‘husband’ explanation is substantially more plausible than the ‘madman’ explanation. To begin, partners killing their wives is a much more common occurrence than random attackers. After all, while random attackers are rare, when a woman is killed, it is often her (ex)partner who did it. Furthermore, the couple was walking together in the forest at the time of the killing, so the husband had the opportunity. Also, he had a motive to kill her – he was jealous because of her new lover – whereas the attacker would have to have attacked without a clear motive. Finally, it is strange that he

---

called his son, but not the emergency services, when his wife was just killed. Catching the killer would require alerting the authorities as soon as possible. These two facts do not seem to fit well together. Yet the 'husband' explanation does not presume such irregular behavior. The explanation that the husband did it therefore fits much better with the evidence and our general understanding of the world and is more internally coherent, than the explanation that there was a random unknown attacker.

Note that the fact-finder is not establishing that the explanations are (in)sufficiently good when looked at on their own. When looked at on its own, the random attacker explanation might not be unbelievable. After all, it is something that could have happened and that would explain the evidence in the case. On the fact-finder’s reasoning above, it is the relative plausibility that makes the difference. The reasons the fact-finder gives here – such as husbands killing their wives being more common than madmen killing people – establish that one explanation is more plausible than the other. Whether fact-finders can easily reach consensus about the validity of such a relative argument is an empirical matter. However, in contrast with Laudan, I see no reason why relative judgment is inherently problematic or even more problematic than other kinds of reasoning in the context of criminal trials.

4. Why absolute plausibility also matters

So far, I have argued that our interpretation of the BARD standard should be attentive to the relative plausibility of explanations. In other words, the case for guilt being substantially better than the case for innocence is sometimes a good reason to consider the guilt of the defendant proven. Yet this does not mean that this is all we should care about.

For instance, the main reason why Pardo & Allen (2008) reject an IBE-based interpretation of the BARD standard is because it does not fit with the structure of criminal proof. After all, we do not want to convict based on an implausible (and
therefore, all other things being equal, improbable) case for guilt, even when this is much better than the case for innocence. Can the comparative interpretation account for this, without *ad hoc* modifications? I argue that it can. In particular, on IBE we should also look at how plausible both the innocence and guilt explanations are in an absolute sense for two reasons. First, we can look at the absolute performance of explanations to efficiently reach comparative conclusions (4.1). Second, IBE does not imply that we can convict a defendant based on an implausible guilt explanation (4.2). If all our explanations are implausible then we have reason to presume that a better investigation could have exonerated the defendant. The comparative case for guilt then lacks *robustness* and we should therefore not convict the defendant. That robustness is a requirement for proof of guilt BARD is an independently plausible assumption for *any* account of the BARD standard, as I explain further in chapters VII and VIII of this thesis.

4.1 Plausible innocence explanations

Proof of guilt BARD often involves eliminating sources of doubt that come in the form of innocence explanations. This idea is found in interpretations of the standard throughout English case law and other jurisdictions that employ the standard, going back at least as far as English evidence treatises of the nineteenth century (Ho, 2008, 156-61). For instance, an influential formulation of the BARD standard by Wills includes the formulation that “Every other reasonable supposition by which the facts may be explained consistently with the hypothesis of innocence must (...) be rigorously examined and successively eliminated” (Wills et al., 1905, 312). Similarly, Josephson writes that how confident we should be in the truth of an explanation largely comes “from ruling out or discrediting alternative explanations” (Josephson, 2001, 1641-2). Yet at first sight it may seem as if IBE cannot account for the importance of such successive elimination – i.e., eliminating explanations when looking at those explanations on their own – given that IBE is about how explanations perform relative to each other. However, this view is mistaken.
For instance, Bird (2007) describes a special version of IBE that he calls ‘Inference to the only explanation’. This is a form of IBE from the philosophy of science that is similar to NPA. Bird describes situations where scientists eliminate all but one of the available explanations because all but one are not good enough. He argues that this often occurs in science by careful experimental design. This is still IBE, because the remaining explanation is then also the best explanation. After all, it is the only good explanation left. So, elimination through absolute judgment can be a way of inferring to the best explanation.

Such successive elimination is not only a way of getting to the best explanation, it is often more efficient than relative judgment. Psychologists have found extensive evidence suggesting that people often do not use “comparative processing”—i.e., judging something by comparing it to its rivals—“because it is cognitively demanding” (Sanbonmatsu et al. 2011, 369). Such a relative comparison requires you to consider features of both the object you are comparing and the alternatives you are comparing it to and the extra step of weighing how these features perform relative to one another. Looking at how explanations do on their own is often simpler; it only requires checking whether each explanation individually meets the desired plausibility threshold. Finding a single piece of evidence that contradicts the explanation may, for instance, suffice for elimination. This often makes it desirable to look at explanations on their own, rather than to how explanations perform relative to one another.

Of course, absolute judgment does not take into account the explaining away effects that I discussed in the previous section. However, this is only problematic when the difference in plausibility between the explanations is large, but the innocence explanation is still (somewhat) plausible. Such cases involve a highly plausible guilt explanation that outperforms any barely plausible innocence explanations. In those situations, absolute judgment alone is insufficient and we need to take relative plausibility into account as well.
Yet there is a converse problem for relative judgment in cases where the best guilt explanation is not plausible but is still substantially more plausible than any innocence explanation. In such cases it would be a relative judgment that leads us in the wrong direction, because it would require us to accept a bad explanation. This is the bad lot problem that I turn to now.

4.2 The bad lot problem

The bad lot problem is possibly the best-known argument against IBE in scientific contexts (Van Fraassen, 1989). It has also been raised against IBE in legal contexts. For instance, Amaya calls this “the most serious problem that a model of IBE for law has to face” (Amaya, 2009, n13). The bad lot problem arises when a bad explanation is much better than any others only because those alternative explanations are even worse. For instance, suppose that no innocence scenario was formulated, but the guilt scenario was also implausible. In such a case, the guilt scenario would be much more plausible than any alternative scenario. Because IBE looks at how explanations perform relative to one another it seems that it would then dictate conviction. However, in criminal trials, the onus is on the prosecution to prove the defendant’s by producing a sufficiently plausible guilt scenario. So, conviction would then be unwarranted, contrary to what (a naïve version of) IBE would dictate.

I will say more about the bad lot problem in chapter VII. Nonetheless, I also want to briefly address the issue here. For the purpose of the current chapter, it is important to distinguish two versions of the bad lot problem. First, in the philosophy of science, the bad lot problem is sometimes presented as a general worry about any truth-claim based on IBE. After all, even if our explanation is much more plausible than any alternative, an even better explanation may exist (at least in theory)
even if there is no reason to suspect so at the moment. Ribeiro (2018, 32) mentions this version of the bad lot problem as a potential difficulty for IBE in criminal trials. However, this is not the version of the bad lot problem that I am concerned with here. Instead, the bad lot problem that I want to discuss concerns those cases where all our available scenarios are implausible. In such bad lot situations, we should not want to choose the best explanation as true.

Let me use an aspect of the Simonshaven case that I have not mentioned so far as an example of what such a bad lot situation looks like. At some point in the case, the defense argued that the defendant could not have killed his wife because he had arthritis in his hands. They argued that this made him incapable of inflicting the kind of violence that was brought upon the victim. If the defense could prove this claim, it would be destructive to the case the prosecution was trying to make. Yet suppose, counterfactually, that the explanation that the husband was the killer is the only available explanation, for instance because the defense never offered the random attacker explanation and the suspect called on his right to remain silent from the beginning. In that case we would only have the explanation that the husband killed his wife. This explanation would be implausible because of the arthritis. It would also be the best explanation, because it would be the only explanation.

One solution to this kind of bad lot problem is to amend the definition of IBE by adding the demand that the best explanation should also be sufficiently good on its own. Some philosophers of science have attempted to save IBE from bad lots this way (Lipton 2003, 63; Musgrave 1988, 238–239). The most straightforward way to read such a proposal is that the best explanation should also display explanatory virtues to a sufficient degree (Dellsén, 2021). However, there are at least two problems with this idea. First, this would be an ad hoc workaround, meant only to deal

---

109 This bad lot problem is worrisome for scientific realists in part because history teaches us that many of our best scientific theories have turned out to be false, even though there was no reason to suspect their falsity at the time. This argument is known as pessimistic induction (Laudan, 1981).
with the bad lot problem, which does not fit well with the essence of IBE, which is about relative conclusions. Second, and more importantly, an implausible explanation may not always be improbable. Fact-finders in criminal trials often face explanations that are at least somewhat implausible. After all, criminal cases, especially hard cases, are complex matters. Summaries of the evidence and investigation alone can span hundreds of pages. It would be surprising if all the evidence pointed to one side. Furthermore, criminal cases deal almost by definition with extraordinary circumstances. Explanations may therefore conflict with our background assumptions about how the world typically works. Yet suppose that fact-finders deal with an *a priori* implausible scenario. Furthermore, imagine that they have spent countless hours investigating possible alternative explanations and looking for additional evidence, but this search did not result in any new discoveries. Surely this should have a positive impact on how probable it is that the implausible explanation is true. This is a well-known idea in epistemology called the ‘no alternatives argument’. One important implication of this argument is that an implausible guilt explanation can sometimes become probable when a thorough investigation fails to turn up any alternative explanations or evidence.

As an example, take the bad lot case I sketched above with the only explanation being the ‘husband' explanation, which is implausible if we know that the husband had arthritis. Let us take an extra leap of imagination. Assume that the couple was on a small, deserted island rather than in a forest in the Netherlands. If we know for a fact that (i) someone killed the victim and that (ii) the investigation has shown that no one else could have been on the island, then the ‘husband' explanation suddenly becomes highly probable despite the arthritis.

So, the condition that the guilt explanation must be sufficiently plausible would sometimes lead to acquittal in cases where the defendant’s guilt is highly probable.

---

110 That the probability of an existing explanation can be raised when a thorough investigation fails to turn up alternative explanations or disconfirming evidence has also been formalized within the Bayesian framework (Dawid et al. 2015; Oaksford and Hahn 2004).
Furthermore, we can imagine situations in which the best guilt explanation is plausible and where there are no plausible alternative explanations, but only because investigators have not spent much time looking for alternatives or for acquitting evidence. In those cases, the amended version of IBE would have us acquit, even though the defendant’s guilt may not be probable. So, this amended version of IBE is not a satisfactory solution for the bad lot problem.

The above discussion points to a better way of tackling the bad lot problem. The above inferences are all about the question whether there are alternative unconceived scenarios or unpossessed evidence. For instance, an implausible scenario can be probable if we are confident that there are no other plausible scenarios that can explain the evidence in the case nor that there is evidence which we have missed that contradicts the scenario. Conversely, if we have reasons to suspect that such unpossessed evidence or unconceived alternatives do exist, then a plausible guilt scenario can be improbable.\footnote{As I explore in depth in chapters VII and VIII, in those cases there is reasonable doubt about guilt.} The additional condition that I propose to deal with the bad lot problem is therefore this: inference to the best explanation is only justified if we have convincing reasons to presume that we have not overlooked evidence or alternative explanations that could realistically have exonerated the defendant. If we do not have such reasons then the case for guilt is insufficiently robust. In other words, we then have reasons to believe that evidence exists or existed, or we could have formulated alternative scenarios which could have overturned the case for guilt. Acquittal should then follow.

That proof BARD requires a robust case for guilt is a common idea in the philosophy of criminal law (Logue, 1997; Ho, 2008, 278; Stein, 2005, 88; Dahlman et al., 2015; Urbaniak, 2018; Di Bello, 2013). As I further argue in chapters VII and VIII, it is also a requirement for the Bayesian and for the NPA interpretation of BARD.
In those chapters I also reflect on how we should determine whether the case for guilt is robust. In this chapter I only offer a few brief remarks on this concept.

The case for guilt can be overturned by both evidence and scenarios. First, those who write about robustness typically link this term to the evidence in the case. We may sometimes have reason to believe that the set of evidence in a case is incomplete, misleading or biased. If we have reason to believe that a better investigation could have uncovered further evidence that would have exonerated the defendant, it is irrational to believe in the defendant’s guilt (Ballantyne, 2015). For instance, the defense in the Simonshaven case argued that the case for guilt lacked robustness, because of potential missing evidence. This related to a curious feature that I have not discussed so far. During the investigation a different criminal case became relevant. This case was about a serial killer whose modus operandi loosely fitted with how the madman supposedly attacked the defendant and his wife. This serial killer had been active in the area of the crime. The police investigated this possible connection further, but failed to find any evidence in its favor. For example, the suspected killer did not fit the description the defendant gave of the supposed attacker. Furthermore, his DNA did not match with any DNA found on the crime scene. Finally, there were no phone records that showed that his phone was near the crime scene at the time of the killing. However, the defense argued that the police’s investigation came too late. They claimed that relevant evidence may have been destroyed or lost. Their argument was therefore that the case for guilt lacked robustness because a more timely investigation could have produced relevant evidence. The court of appeal ultimately rejected this claim. However, this is an example of an argument against the robustness of the case for guilt.

Second, an investigation is about more than collecting evidence. It is also a creative effort, where investigators formulate possible explanations of the evidence. This leads to a different way in which the case for guilt can lack robustness, namely when
investigators overlook plausible innocence explanations. When unconsidered innocence explanations are discovered that are better (or at least not much worse) than the best guilt explanation, this guilt explanation is no longer substantially more plausible than any innocence explanation. The requirement that the case for guilt should be robust in this sense is expressed in the condition that is a part of the definitions of both NPA and IBE (see section 2.2): “Rival explanations for \(e_1, e_2, \ldots, e_n\) have been earnestly sought out but the search has produced only \(h_1, h_2, \ldots, h_n\).” Amaya (2009, 154–5) makes a similar point in response to the bad lot problem: “we need to have some reason to believe that the set of hypotheses from which we have inferred to the best is `good enough'.” Likewise, Josephson, (2001) claims that for IBE to work well, it is important to consider the exhaustiveness of our set of explanations.

In a bad lot situation – where all available explanations are implausible – the case for guilt will often lack robustness. If we only have poor scenarios then this gives us a reason to believe that either some better, unconsidered explanation has to exist or that our evidence is insufficient or misleading (or both). Take the arthritis example again. If the husband had arthritis, this would make it unlikely that he was able to inflict the degree of violence that we know killed the victim. Yet we know that someone killed the victim. So, if we have no plausible explanations about how the killing occurred, then it becomes more likely that some other, better explanation for the killing exists, that we have not considered. Alternatively, our evidence that the husband has arthritis might be false. Conversely, we may have a plausible guilt explanation and no plausible innocence explanations, but also reason to suspect that there are further alternatives or evidence which, if we had collected them/come up with them, our guilt scenario would not be especially plausible or would have further plausible alternative explanations.

\[112\] Or scenarios on which the defendant is guilty of a crime, but a less serious one (e.g., manslaughter rather than murder).
To summarize, while IBE is about the relative assessment of explanations, making sure that the guilt explanation is plausible on its own is still of key importance. Insufficiently plausible guilt explanations should usually lead to acquittal, even when they are substantially better than any innocence explanation. This is similar to the solution to the bad lot problem mentioned above, that the best explanation also has to be good enough on its own. However, my account is subtly different. Whether the case for guilt is plausible enough for conviction is part of the larger question, whether the investigation (probably) uncovered all relevant explanations and evidence (including evidence about the reliability of our evidence). This means combining information about the quality of the available explanations with other evidence about the robustness of the case for guilt – in particular evidence about quality of the investigation.

5. Conclusion

Explanationists view criminal trials in terms of competing explanations. These explanations can be more or less plausible. But when is an explanation that implies the defendant’s guilt proven beyond a reasonable doubt? In this chapter I proposed an inference to the best explanation (IBE)-based approach on which guilt is only established BARD if (i) the best guilt explanation in a case is substantially more plausible than any innocence explanation, and (ii) there is no good reason to presume that we have overlooked evidence or alternative explanations that could realistically have exonerated the defendant. This is a comparative account on which we determine whether guilt has been proven by looking at how well our guilt explanation performs compared to any innocence explanations. As I have argued, accurate fact finding – at least sometimes – requires looking at how explanations perform relative to one another. This is because confirming one explanation can lead to the disconfirmation of another. The non-comparative ‘no plausible alternative’ view of
the BARD standard, which many explanationists assume, overlooks the importance of such relative performance.

Of course, we should not only care about the relative performance of explanations. If we only have implausible guilt explanations, we should not convict the defendant, even when these explanations are much more plausible than any innocence explanation. Yet as I have argued, we can amend IBE so that it does not imply that we should. When we only have implausible explanations – i.e., when we are in a bad lot situation – this suggests that we have missed something and that the relative case for guilt would have been overturned if we had come up with these alternative explanations or if we had found further evidence. Our relative conclusion – that the best guilt explanation is substantially better than any innocence explanation – then lacks robustness. That we should not convict if the case for guilt lacks robustness is an independently plausible assumption for any account of the BARD standard. So, the proposed IBE-based account overcomes the bad lot problem without the need for ad hoc modifications. I will say much more about robustness in chapters VII and VIII.

The proposed account connects various strands of criminal fact finding into a single framework. For instance, Josephson, 2001, 1626) claims that the strength of any abductive justification is determined by:

(1) how decisively the leading hypothesis surpasses the alternatives;

(2) how well the hypothesis stands by itself, independently of the alternatives;

(3) how thorough the search was for alternative explanations.

Josephson does not discuss why fact-finders should care about these specific criteria or how they relate to one another. In contrast, my account offers a coherent account that answers these questions. On this account, how decisively our best explanation surpasses its alternatives is the central question for criminal fact-finders. How well the guilt explanation stands by itself relates to the question whether
the case for guilt is robust. Finally, the question of whether the search for alternatives was thorough also partially determines whether the case for guilt is robust.

As mentioned in the introduction, Laudan (2006) argues that the most prominent interpretations of the BARD standard are all problematic. He claims that such interpretations are either too unclear to be applied, too reliant on the subjective hunches of the person who applies it or insufficiently connected to the goal of avoiding false convictions. My interpretation of the BARD standard overcomes such problems. First, as I have shown using the Simonshaven case, IBE can meaningfully be applied to real criminal cases. Second, the interpretation frames the standard in terms of the plausibility of explanations and connects this to the goal of getting at probabilistic conclusions. This means that what counts as rational on this interpretation is getting to conclusions that are probably true, not just the hunches of the fact-finder. Third, the connection to probability theory also connects this interpretation to the goal of avoiding false convictions.
IV. IMPLAUSIBLE, INCREDIBLE AND HIGHLY IMPROBABLE STORIES

Chapter abstract

In criminal trials, defendants often offer alternative explanations of the facts when they plead for their innocence. In its ruling on the Venray murder case, the Dutch Supreme Court dealt with the question when and how courts can reject such alternative explanations. According to the Supreme court, while courts should typically refer to evidence that refutes the explanation, they can also argue that the explanation ‘did not become plausible’ or that it is ‘not credible’. Finally, courts can state that an explanation is so ‘highly improbable’ that it requires no response. However, the Supreme Court did not explain these terms, leading to confusion about how they ought to be interpreted. This chapter offers a Bayesian interpretation according to which these three terms relate to (i) the posterior probability of the explanation, (ii) the credibility of the defendant, and (iii) how obvious it is that the explanation is highly improbable. This interpretation clarifies the Supreme Court’s ruling and ties it to the criminal law system’s goals of error minimization and of producing understandable decisions efficiently.

1. Introduction

In 2010, the Dutch Supreme Court ruled on a case that has since become known as the ‘Venray murder’.113 In this case the defendant was accused of having killed his wife. He denied this accusation, claiming that he found his wife dead when he came home. When the court of appeal ruled on the case, it acquitted the defendant

---

113 Dutch Supreme Court. (March 16th 2010). ECLI:NL:HR:2010:BK3359. The name comes from the town where the victim and her husband lived.
on the grounds that the defendant's scenario explained the evidence just as well as (or even better than) that of the prosecution. However, the defendant offered his explanation only after he was informed of the results of the forensic investigation. So, it was possible that he carefully constructed a false scenario that fitted with the evidence. This raised the question – can courts never reject a defendant’s alternative explanation if there is no evidence that refutes it?

The case went to the Supreme Court, which decreed that while courts should ideally point to evidence that refutes the explanation, they can also reject alternative explanations even when it is not contradicted by the evidence. In particular, the Supreme Court distinguished three grounds for such a rejection. First and second, courts can argue that the explanation ‘did not become plausible’ or that they are ‘not credible’. Third, some explanations are so ‘highly improbable’ that they require no response at all.

This ruling is important, because it set forth a framework for whether and how courts should respond to the alternative scenarios offered by defendants. However, the resulting case law is also nebulous. The Supreme Court did not offer any explanation of the phrases it introduced, nor did it specify how these terms should be applied. As a result, both legal scholars and courts have been struggling to make sense of this ruling (Lettinga, 2015). In this chapter I offer a new interpretation of this ruling. Broadly speaking, I seek to achieve three aims. The first is to clarify the distinction between the terms ‘implausible’, ‘incredible’ and ‘highly improbable’. In colloquial settings, these words all relate to probability and are regularly used interchangeably. To make the distinction between the different terms precise, I use

---

115 Or, the Supreme Court notes, through ‘facts and circumstances derived from the evidence that refute it’ (Dutch Supreme Court. (March 16th 2010). ECLI:NL:HR:2010:BK3359).
116 See also Scholten (2021, 168-176) for another discussion of this ruling.
the language of Bayesian probability theory.\footnote{117 Though note that this is not a general account proposal on how the explanationist notion of ‘plausibility’ fits into the Bayesian account. The term ‘plausibility’ here refers to the Dutch legal term ‘aannemelijk’ and my use of Bayesianism is purely to clarify certain distinctions.} My second aim is to connect the Venray ruling to the aims of criminal fact-finding. In particular, I show how the ruling is grounded in several key goals of the (Dutch) criminal law system: to minimize erroneous decisions, to legitimize decisions by making them understandable to others and to reach such decisions efficiently. Third, I offer suggestions on how courts can judge whether explanations are ‘implausible’, ‘incredible’ or ‘highly improbable.

I begin this chapter by examining the facts and the main legal question of the Venray ruling in more detail (section 2). After that, I first discuss situations in which explanations are refuted by the evidence (section 3). Finally, I turn to what it means for an explanation to be ‘implausible’ (section 4), ‘incredible’ (section 5) or ‘highly improbable’ (section 6).

2. The Venray murder

In the Venray case, a man was accused of stabbing his wife to death. At first, the man called upon his right to remain silent. He only offered an alternative explanation after one and a half years had passed. At that point, he knew the results of the forensic investigations. He sketched a scenario in which he had found his wife dead and hypothesized that criminals might have killed her because of an argument they had with him. As the court of appeal noted, the suspect’s explanation fit with the limited available evidence (blood stains and shoe prints) at least as well as the hypothesis that he killed his wife. In fact, the court remarked, the innocence explanation fit even better with the evidence than the guilt scenario. For instance, no blood was found on the defendant’s clothing. When stabbing someone, it is likely that blood would end up on your clothes and such blood is difficult to remove.
afterwards. However, the court did note that the defendant’s story was somewhat hard to believe, especially because he waited so long to come forward with it. Nonetheless, the court acquitted him. It reasoned that it could only convict him if there was evidence that refuted the alternative explanation or if this explanation was so implausible that it needed no explicit refutation. According to the court of appeal, neither was the case.

The court’s position is understandable if we look at it in the context of Dutch criminal law. First, whenever a defendant offers an alternative explanation, the court can only convict him if it provides an explicit justification for rejecting this explanation in its ruling.\(^{118}\) Second, in the Netherlands the court may only consider it proven that the defendant committed the alleged acts if it is convinced of this fact based on the admissible evidence.\(^{119}\) So, at first sight, when the court has to justify why it is convinced that the defendant committed the alleged acts, it should do so by referring to the admissible evidence. If the reason why the court finds the facts of the case proven is that it rejects the defendant’s alternative explanation, this would then presumably also require the court to refer to evidence which refutes this alternative explanation. However, the Supreme Court did not share the court of appeal’s view. It ruled that explanations can sometimes be rejected even when there is no evidence that contradicts it.\(^{120}\) In particular, the Supreme Court stated that courts can argue that the defendant’s story ‘did not become plausible’ or that it is ‘not credible’. Finally, some explanations are so ‘highly improbable’ that they require no explicit

\(^{118}\) Article 359(2) Dutch code of criminal procedure. This article states that courts have to respond any time the defendant offers a clear argument in favor of his innocence (In Dutch legal terminology ‘onderbouwd standpunt’, my translation). This includes any alternative explanations of the facts that the defendant offers that are consistent with his innocence.

\(^{119}\) Article 338, Dutch code of criminal procedure.

\(^{120}\) It then referred the case back to another court of appeal. This court of appeal then convicted the defendant of murder. Court of justice of Arnhem. (October 15\(^{th}\) 2012). ECLI:NL:GHARN:2012:BY0075.
justification by the court to be rejected. The following schema summarizes the ruling:

![Diagram showing the Venray ruling schematically]

**Figure 4: The Venray ruling schematically**

So, the Supreme Court’s ruling is about how courts should deal with cases in which the evidence does not ‘refute’ a defendant’s story, but they still wish to reject this alternative explanation. Before moving on to my interpretation of this ruling, I want to discuss what it means for evidence to refute a defendant’s alternative explanation.

### 3. Rejecting stories with and without evidence

When courts reject a defendant’s explanation, they typically do so by referring to evidence that refutes this explanation. This does not mean that the evidence excludes the story, in the sense that the story cannot possibly be true. Even in cases with very strong evidence against an explanation – say multiple, apparently reliable witnesses – there is always a remote possibility that the story is true. For instance, all witnesses could have had a reason to lie. Instead, evidence refutes a story insofar as it makes the story (very) improbable. In Bayesian terms, this means that the
posterior probability of the explanation, $P(H|E)$, should be low. As explained in the previous chapter, we only want to convict the defendant if they are very probably guilty in order to avoid erroneous convictions in order to achieve a just error distribution.

Whether the evidence in a case makes the hypothesis improbable depends on the likelihood ratio. In a case like the Venray murder, with two competing scenarios, the value of this ratio will typically depend on the presence of 'discriminating evidence', i.e., evidence that fits better with one scenario than with the other (Van Koppen, 2011, 52-55). If the likelihood ratio is sufficiently much greater than 1, the probability of one explanation will be high and the probability that the alternative explanation is true will be low. In such a case, the court can point to the discriminating evidence as a reason why it rejects the alternative explanation.

In the Venray case, the evidence did not discriminate (much) between the competing scenarios. Whether it was the husband who killed his wife or someone else, either way, we would expect to find the kind of evidence that was found (such as the shoe prints and the blood stains). In such a situation, the only way for a story to be highly improbable is for its prior probability to be low. Is this how we should read the Supreme Court’s ruling? To put it differently, when courts reject a defendant’s explanation for being implausible, incredible or highly improbable, is this always a judgment about that explanation’s prior probability? And what should we then make of the distinction between these three terms?

In the following sections I will argue that prior probability only plays a key role in one of the three criteria that the Supreme Court mentioned, namely whether the explanation needs to ‘become plausible’. I will look at this criterion next. For the other two criteria we need different concepts, which I discuss in sections 5 and 6 respectively. In section 5 I argue that whether an explanation is ‘incredible’ depends on the credibility of the defendant. Finally, in section 6 I argue that whether an
explanation is ‘highly improbable’ depends on how obvious it is that the explanation is improbable.

4. Implausible explanations fail to become probable

According to the Supreme Court, courts can reject an explanation if it ‘did not become plausible’ during the criminal proceedings. The phrasing of this term suggests that an explanation can initially be implausible, but can become plausible. This raises two questions. First, what does it mean for a scenario to be initially implausible? Second, how can such scenarios become plausible? The most obvious Bayesian interpretation of these phrases is in terms of the prior probability, the likelihood ratio and the posterior probability of the explanation.

Let us first look at what it means for an explanation to be initially implausible. I follow Williamson (2000, 211) and understand this phrase as “the intrinsic plausibility of hypotheses prior to investigation”, which Williamson equates with the prior probability of that hypothesis. What does an explanation with a low prior probability look like? First, it may have parts that do not fit well together. For instance, the explanation might imply that the defendant was in two places at the same time. Alternatively, the defendant may tell a story in which motive and action do not fit well together, such as a story about a robbery where nothing was stolen. Finally, the explanation may consist of a number of independent and individually unlikely events (Lettinga, 2015, 53; Josephson, 2001). Second, an explanation can also have a low prior probability because it does not fit well with our generalizations about how the world typically works. For instance, we may believe that innocent bystanders do not run away from the police, that a suspect cannot cross the city in ten minutes or that the police rarely forge evidence. The more an explanation violates such generalizations, the lower its prior probability is.

121 In Dutch ‘niet aannemelijk geworden’. My translation.
An initially implausible explanation fails to create a reasonable doubt because it is insufficiently probable. However, such a story can be made more probable by introducing discriminating evidence, which is evidence with a sufficiently high likelihood ratio. If the defendant’s alternative scenario explains that evidence sufficiently well compared to the prosecution’s scenario, the prior probability of the alternative explanation is raised and reasonable doubt about the defendant’s guilt results. In this way, an initially implausible story may become plausible.

Conversely, suppose that the evidence presented does not discriminate between the defense and prosecution scenario. Furthermore, imagine that the explanation provided by the prosecution has a high prior probability, while defendant’s alternative explanation has a low prior probability. In such a situation, the prosecution’s scenario may be sufficiently probable based on its prior, and the defendant’s scenario sufficiently improbable, to support proof of guilt beyond a reasonable doubt. This would mean that if no further evidence were adduced, the defendant would most likely lose the case. The defense may then have a burden to introduce new arguments or evidence that would make the court decide in his favor or he risks losing the case. If they fail to do so, the alternative explanation fails to become plausible, and can be rejected by the court by referring to the initial weaknesses in the story which were not mended by supporting it with discriminating evidence.

122 This is known as the tactical burden of proof (Prakken & Sartor, 2009). While the defendant risks being convicted when offering a weak explanation, he is not always the one who carries the burden of producing evidence that will make his story plausible.
5. Incredible explanations are told by an unreliable storyteller

Apart from arguing that the explanation is implausible, the Supreme Court also decreed that courts can reject explanations by arguing that they are ‘incredible’.123 When the term ‘incredible’ is used by Dutch courts, it typically refers to actions within the scenario that the defendant undertook or in how the defendant told the story (Lettinga, 2015). For instance, suppose that the defendant claims that he was a bystander of a murder but that he did not call the emergency services while he did spend time trying to hide possessions of the victim.124 Such a story would be implausible, in the way we just saw: it contains illogical elements and therefore has a low prior probability. However, it would also be incredible. The defendant would not come across as a reliable storyteller. Telling bad stories and lacking credibility as a storyteller often go hand in hand, but not always. Some stories fit well with the evidence and with our background beliefs, perhaps even better than the true explanation but are still improbable due to the lack of credibility of the defendant.

First, an otherwise plausible story may be incredible because it fits poorly with the characteristics and past behavior of the defendant. For instance, if a defendant has made statements in the past that conflict with his current story and he does not have a good explanation for these earlier statements then this lowers his credibility. An example of this comes from the Venray case. After the case went back to the court of appeal, the defendant gave a partial confession. He admitted to attacking his wife, but claimed that there was another person involved who slit her throat. However, the court of appeal argued that they did not believe this partial confession because the defendant lacked credibility due to the contradictory explanations he

123 In Dutch ‘ongeloofwaardig’, my translation.
had given. Instead, the court ruled that the defendant had premeditated killing his wife and convicted him of murder.\footnote{Court of justice Arnhem. (October 15\textsuperscript{th} 2012). ECLI:NL:GHARN:2012:BY0075.}

Second, some stories are vague. For example, a defendant may claim that ‘something else happened’, without providing further details. People tend to find such stories difficult to believe because they lack relevant details (Pennington & Hastie, 1991). However, such a story does not conflict with the evidence or with our background assumptions, nor is it necessarily internally incoherent. However, the fact that it is vague can sometimes be a reason why the story is improbable. Suppose that the defendant claimed he was a bystander of a murder. If he is telling the truth we might expect him to be able to testify to details of what happened. However, if he then offers a vague explanation, we might become suspicious that he is lying by deliberately offering story that his vague enough not to be contradicted by the evidence. In other words, if we can reasonably assume that the defendant \textit{could} tell a more specific story, which better explains the facts, then we have reason to doubt the credibility of his story.\footnote{Not all vague stories are improbable. For instance, our memory and powers of observation are far less reliable than we like to think (Wise et al., 2014). Especially in stressful situations - such as when we are bystanders of a murder or are being interrogated by the police - our memories may fail us. So, a defendant who offers a vague story may simply not remember much of a given situation.}

A third important category of incredible stories are \textit{ad hoc} explanations. An \textit{ad hoc} explanation is an explanation that is made up to fit the available evidence but that is difficult or impossible to falsify. For instance, a guilty defendant can call upon his right to remain silent and only offer an explanation once all the evidence has been presented that is fitted to this evidence (Mackor, 2017). This was what the defendant in the Venray case may have done. Such a story is not necessarily implausible or incoherent. On the contrary, false explanations of criminal evidence are sometimes more coherent (Vredeveldt et al., 2014) and better supported by the evidence (Gunn et al., 2016) than true explanations. This is because they can be
tailored to the known facts. However, if we have good reasons to suspect that the
defendant has fitted his story to the evidence, then this should lower our degree of
belief that he is truthfully reporting on his own experiences.\textsuperscript{127}

The notion of credibility can easily be expressed in Bayesian terms.\textsuperscript{128} Whether a
scenario is credible depends on the follow likelihood ratio:

\[
\frac{\text{Probability(Defendant offers scenario | H)}}{\text{Probability(Defendant offers scenario | } \neg \text{H)}}
\]

Here H could be the guilt scenario and \( \neg \text{H} \) the innocence scenario. In other words,
how likely is it that \emph{this} defendant would offer \emph{this} testimony in \emph{this} way, at \emph{this} mo-
ment under both scenarios? The contents of the testimony may not discriminate
between the two scenarios, but the way in which the defendant offers the testimony
may discriminate, as it may, for instance, fit well with the defendant making up a
false story.

So, the way in which the defendant offers their scenario can count as evidence
about whether that scenario is true. Now, I said before that the Venray case is about
situations in which the defendant’s explanation is \emph{not} refuted by the evidence. How-
ever, as I just argued, an incredible explanation \emph{is} improbable due to evidence, spe-
cifically evidence about the defendant’s credibility. This verbal confusion is due to
fact that the court of appeal (to which the Supreme Court responded) did not con-
sider the fact that the defendant waited so long to offer his alternative explanation
evidence. It claimed it could not reject the defendant’s story because it was not
refuted by the evidence, despite claiming that the story was difficult to believe given

\textsuperscript{127} Note that not every defendant who fits his story to the evidence is lying. For instance,
receiving post-event information may influence our memories subconsciously (Shaw,
2016). Furthermore, new information can remind us that our initial memories were wrong
(“ah, yes, I remember now!”) (Vredevelt, Van Koppen & Granhag, 2014). So, even ve-
racious defendants may offer seemingly \emph{ad hoc} explanations.

\textsuperscript{128} For more on the Bayesian account of eyewitness reliability, see chapter VI.
the defendant’s timing. However, because it can have an impact on how probable we consider the defendant’s alternative scenario, it is also evidence. In its Venray ruling, the Supreme Court recognized this fact.

6. Highly improbable explanations are obviously false

When a court considers an explanation to be implausible or incredible it must generally justify why it does not believe the defendant’s explanation before convicting them. However, according to the Supreme Court, some explanations are so ‘highly improbable’ that courts do not have a duty to respond to them.129

Of the terms that the Supreme Court introduces in its ruling, this one is possibly the most nebulous. At first sight, the term seems to refer to explanations that have a very low (posterior) probability. But this straightforward interpretation faces the difficulty that any alternative explanation that the court rejects is highly improbable. In chapter III, I mentioned that (in practice) Dutch criminal law requires that a defendant’s guilt must be proven beyond a reasonable doubt. This is a high standard for proof. In probabilistic terms, the standard is often taken to mean that the probability of guilt should be high enough (e.g., 95% or 99%). However, this means that the probability of any innocence scenario which fails to create a reasonable doubt will be very low. So, if all rejected alternative explanations are very improbable, what distinguishes those that are ‘highly improbable’ that they need not be addressed? Perhaps some explanations are highly improbable, say less than 0.01%. Yet this still leaves us with the question why courts do not have to respond to such explanations. What makes highly improbable explanations special?

An answer to this question begins with a discussion about why courts usually should justify their decision to reject an alternative explanation. Legal scholars have

129 In Dutch ‘zo onwaarschijnlijk is, dat zij geen uitdrukkelijke weerlegging behoeft’, my translation.
identified several purposes that explaining one’s decision serves. First, explicit justification helps make the decision understandable for its audience, which includes the parties at trial, the legal community and society as a whole (Knigge, 1980; Dreissen, 2007, 392-404). If the audience understands the arguments for the decision, then this makes the court’s decision more legitimate for them. For instance, the explanation of their decision can act as justification for the punishment that follows for the convicted defendant. Second, such an explanation allows courts of appeal, judicial scholars, experts and other interested parties to check whether the decision was correct and to point out possible flaws (Dreissen, 2007; Verbaan, 2013). Furthermore, by making the reasons for the decision understandable, parties might be less inclined to appeal the ruling. This would aid the efficiency of the criminal law system because courts of appeal would have to hear fewer cases (Buruma, 2005). The final reason why judges should justify their decision is that it forces courts to reflect on their arguments for the ruling. This in turn can help them avoid reasoning errors such as subconscious biases (Dreissen, 2007, 392-404; Gommer, 2007). This idea is in line with psychological research that suggests that explaining one’s decision-making process helps people make better decisions (Wilkenfeld & Lombrozo, 2015; Maegherman et al., 2021).

An explanation why the court rejects the defendant’s alternative explanation has similar benefits. In such cases the justification gives both the court and the audience insight into why that explanation is not probable enough to create a reasonable doubt. However, there are cases in which an explanation is not needed. In particular, some stories that defendants tell are so obviously improbable that we would gain little by arguing against them. For example, take a (real) case in which the defendant pleaded that he was not accountable for the child porn on his computer because his mind was controlled by aliens. It seems fair to say that no reasonable audience would consider the ‘alien’ explanation remotely probable. Furthermore, a defendant

---

who offers such an explanation would either be delusional or insincere. So, it is improbable that arguments would sway him. Hence, the court would (most likely) gain little by justifying why it rejects this alternative explanation, with respect to the parties, legal community and general audience’s understanding of it.

What about the benefit of reflecting on one’s reasoning? It seems fair to say that the more difficult it is to see why an explanation is improbable, the more room there is for error. However, when an explanation’s improbability is obvious, the reasoning required to understand its probability does not require much thought. Hence, there is less to be gained by carefully spelling out one’s reasoning to see whether this reasoning is sound. For instance, the court does not have to carefully reflect on whether they might be making an error when they assume that mind controlling aliens do not exist. So, to summarize, there is little gain to justifying why we reject obviously improbable explanations. Yet spelling out such arguments does take time and effort and impedes the efficiency of decision making. The costs of explicit justification may then outweigh the benefits.

That ‘highly improbable’ should be interpreted as ‘obviously improbable’ is also something that has implicitly been noted by Dutch courts. To give an example, in a 1997 traffic law case, the Dutch supreme court overturned the decision by a lower court because it had failed to give a justification for its decision to reject the defendant’s alternative scenario. The supreme court argued that even if the lower court thinks that a defendant’s alternative scenario is improbable, it will sometimes have to offer a justification for this conclusion, because not every improbability is ‘evident’.

Whether the improbability of an explanation is obvious is not something captured by Bayesian probability theory. I am also unaware of any other framework which captures this dimension of proof. Though I do not have a fully worked out answer to what makes an explanation obviously improbable, I want to propose the

---

following tentative answer. When we try to determine how probable an explanation is, we have to see ‘how the facts hang together’ (van Oorschot, 2014). In other words, we have to understand the connections between the explanation and the evidence, but also between different parts of the explanation (its internal coherence).  

Similarly, understanding that an explanation is improbable requires seeing that it conflicts with evidence or with our background assumptions. Whether it is obvious that an explanation is improbable depends on how difficult it is to see such conflicts. The level of difficulty will depend on the information that the audience has and on their capacities. It also depends on the number and complexity of the probabilistic relations between evidence and scenario that the audience has to understand. For instance, in criminal cases the mere description of the evidence can sometimes be hundreds of pages long. Judging whether the evidence makes the explanation unlikely might therefore require seeing how numerous pieces of evidence cohere with one another. Similarly, an explanation can have a very low prior probability because of internal inconsistencies, without this being immediately obvious. Understanding that the explanation is highly improbable might then involve, for instance, creating a time line of the story and seeing that the story does not make sense. For example, we might then find out that the story implies that the defendant is in two places at the same time. Such a story will have a very low probability (perhaps even a probability of 0) but it will not be obviously improbable.

As a final note, whether something is obvious is not always obvious. One potential pitfall is the curse of knowledge, which refers to the difficulty of imagining what it is like for someone else not to know something that you know (Birch & Bloom, 2007). When we know or understand something, we sometimes imagine this to be common knowledge without wondering whether it is (Nickerson, 2001). So, what is obvious to a judge who has preceded over the entire case, who has seen and reflected on the evidence and the arguments, may not be obvious to outsiders.

132 I will expand upon this idea in chapters VI and VIII.
Second, many if not most of us sometimes suffer from *the illusion of knowing* – the idea that we know and understand more than we actually do (Glenberg, Wilkinson & Epstein, 1982). A common experience associated with the illusion of knowing is the feeling of understanding a concept, but then realizing that this is not the case when you try to explain it to someone else. In other words, our lack of knowing may only become apparent once we explain our reasoning (Schwartz, 2013). So, while courts do not have to respond to absurd explanations, they should be wary of their own biases.

7. Conclusion

In the Venray murder case, the Dutch Supreme Court determined on what grounds courts may reject the alternative explanations offered by defendants and when they should justify their decision to do so. In this chapter I offered an interpretation employing Bayesian probability theory.

At the heart of the Supreme Court’s ruling is the idea that courts can reject a defendant’s explanation even in cases where the evidence does not refute this explanation. While rejecting the story by referring to a ‘smoking gun’ (i.e., refuting evidence) may be the ideal, other responses are possible too. The Supreme Court distinguishes three categories. First, some explanations can be rejected because they ‘did not become plausible’. I argued that whether an explanation needs to ‘become plausible’ during the criminal proceedings depends on its inherent plausibility at the time it is offered – its prior probability. If an explanation with a low prior probability does not become probable by means of the evidence, then the explanation fails to create a reasonable doubt. Second, some explanations are ‘incredible’. Whether an explanation offered by a defendant is probable partially depends on evidence about the credibility of the defendant. Finally, some explanations are so ‘highly improbable’ that the court does not have a duty to respond to them. I argued that what distinguishes these explanations from explanations that the court *should*
respond to is that their improbability is *obvious*. When an explanation is obviously improbable, the court would not serve the goals of making its decision understandable by offering a response. A duty to respond would then only reduce the efficiency of the decision process.

To conclude I want to remark that my conclusions are also potentially informative to other legal systems. First, some legal systems have a similar duty to respond to alternative explanations. For instance, German criminal law also requires courts to respond to alternative explanations of the facts (Dreissen, 2007, 319). One of the key goals of this duty is to make the ruling understandable to others (in particular higher courts that have to check whether the reasons for conviction are valid) (Dreissen, 2007, 405). Second, while common law countries do not have a judicial duty to respond to alternative explanations, my analysis might help to further clarify the beyond a reasonable doubt standard of proof. In particular, as Ho (2008, pp. 153-154) points out such countries often have case law which stipulates that a reasonable doubt is not created by explanations that are 'fantastic and unreal', ‘mere conjecture[s]’ (High court of Australia), ‘illusory’ or 'fanciful' (Supreme court of Singapore) or ‘of which there is no evidence and which cannot be reasonably inferred from the evidence’ (British House of Lords). The argument I presented here could help further interpret these terms, by offering ways to understand why certain explanations require no serious consideration.
V. THE VALUES OF PREDICTION

Chapter abstract

An explanation can be confirmed by evidence in two ways. First, it can explain known evidence (also known as accommodation). Second, it may successfully predict novel evidence. Philosophers of science have made much of this distinction, arguing that explanations which lead to successful predictions are – all else being equal – epistemically superior to those that merely explain known data. Is this also the case in criminal law? If so, is there merit to the intuition that unfalsifiable scenarios which do not yield testable predictions, are thereby less credible? In this chapter I draw on arguments from the predictivism debate in the philosophy of science to distinguish three ways in which the predicted evidence can yield stronger support for a scenario than if the same evidence were accommodated. First, witnesses who predict are – all else being equal – more reliable than those who do not because they are less likely to be biased or lying. Second, investigators who only explain known facts run the risk of ‘fudging’ the scenarios that they formulate, where they propose an implausible scenario to explain as many of the facts as possible. Predictions can protect us against this danger. Third, carefully constructed predictions may help investigators to avoid confirmation bias. I express these arguments using the Bayesian likelihood ratio. To show how these arguments play out in real, complex situations, I use a case study of the murder of Hae Min Lee. The upshot of these arguments is that whether a successful prediction yields stronger evidence than accommodated evidence depends on the presence of alternative explanations for the predictive or accommodative success.
1. Introduction

Like scientists, witnesses, investigators and fact-finders in criminal cases both explain known evidence and use the resulting explanations to make new predictions when formulating a scenario. For instance, imagine that a detective suspects someone of being the perpetrator of a criminal act because he was near the scene of the crime. Alternatively, imagine that this detective comes to suspect him for some other reason and only later finds out that this person was near the crime scene. In both cases the detective ends up formulating the scenario that this person was the perpetrator. Furthermore, in both cases she has at least one piece of evidence supporting this scenario – namely that this person was near the crime scene. Yet this piece of evidence comes to support the scenario in different ways. In the first case, the evidence was accommodated – the detective formulated the scenario based on this known piece of evidence. In the second case, the evidence was predicted – the detective did not use it in constructing the scenario.

Philosophers of science have made much of the distinction between prediction and accommodation. In science, successful predictions are often seen as one of the hallmarks of a good theory and scientific theories that make no testable predictions are often seen as defective – even if they explain the known data well (Popper, 1959; Barnes, 2018). Because of this, many philosophers of science argue for predictivism – the thesis that successfully predicted facts provide stronger evidence for a theory than successfully explained known facts (cf. Douglas & Magnus, 2013). In this chapter I argue that the predictivist debate from the philosophy of science can teach us valuable lessons about criminal evidence too.

In criminal law, predictions do not derive from scientific theories but from scenarios. Like scientific theories, scenarios can both explain known evidence and produce further evidence that we might expect to find. Though it may happen less than in science, we also encounter predictivist intuitions regarding the predictions of scenarios. For instance, Josephson (2001) claims that checking whether a crime
scenario’s predictions are confirmed is one of the most important criteria on the basis of which we should assess such scenarios. The necessity of checking predictions is also repeatedly emphasized by Van Koppen (2011, e.g., 76) – who mentions their importance in science as an illustration of why they are important in criminal cases too. A more explicit appeal to predictivism is made by Mackor (2017), who adopts a Lakatosian position on the value of predictions. She draws an analogy between the role of predictions for assessing scientific research programmes and their role in assessing scenario’s. She briefly alludes to the contemporary predictivist debate but she does not discuss the arguments from this debate in detail. Finally, Tuzet (2019) does discuss one argument from the contemporary predictivist debate, namely Peter Lipton’s fudging argument. I discuss this argument in section 4 of this chapter. Based on this argument he suggests that predictivism also holds with respect to legal evidence. Nonetheless, his discussion of this point is brief and in need of further elaboration. So, while at least some authors consider predictions to have a value above and beyond a scenario merely accommodating the known facts, we currently lack a systematic examination of what this value might be. The goal of this chapter is to address this gap. I examine several arguments about the value of prediction developed by philosophers of science and show that analogous arguments can be made with respect to criminal cases. In particular, this chapter develops three arguments for the special value of predicted over accommodated criminal evidence. First, witnesses who predict are – all else being equal – more reliable than those who do not because they are less likely to be biased or lying (section 3). Second, investigators who only accommodate run the risk of ‘fudging’ the scenarios that they formulate: they sacrifice the quality of a scenario to make it fit the facts. Predictions can protect us against this danger (section 4). Third, carefully constructed predictions may help investigators to avoid confirmation bias (section 5). In section 6 I the above together by explaining why, in the case of all three arguments, determining what a particular prediction proves will depend on the competing explanations for this predictive success.
In order to show how the arguments for predictivism play out in real, complex situations, I end this chapter with a case study of the murder of Hae Min Lee (section 7). This murder was the subject of the popular 2014 true crime podcast *Serial*. The reason for choosing this case is that it offers several interesting examples of predictions (or a lack thereof) which prove or fail to support the reliability of the underlying scenario. However, I begin this chapter by analyzing the notions of prediction, accommodation and evidential strength.

2. On predictions and evidential strength

This chapter examines various arguments for predictivism – the thesis that successful predictions yield stronger evidence than successful accommodations. As a preliminary, let’s look at the notions of prediction, accommodation and evidential strength.

2.1 Predictions and accommodations

The term ‘prediction’ has a subtly different meaning in the context of the predictivism debate than how we normally use the term. When we talk about predictions in daily life, we typically mean that someone made a statement about the future. For example, someone might say ‘I predict that it will rain tomorrow’. However, while that is one example of a prediction, not all predictions are like that. First, predictions do not have to be explicitly stated by a person. Second, they do not have to be about the future. Predictions are empirical consequences that derive from hypotheses. In other words, if we assume that a given hypothesis is true, this creates certain expectations about the kind of evidence we should encounter that we would not otherwise encounter. In science, such hypotheses often take the shape of models or theories – which in turn lead to predictions. In criminal trials the relevant hypotheses under consideration are often scenarios. We can think of such scenarios as stories about what happened in a case. For instance, imagine a
murder case where a woman’s dead body is found and where the woman’s husband is the main suspect. The relevant scenario is then ‘the husband killed his wife’. Certain facts will be more likely to be true if the scenario is true – these are the empirical consequences of that scenario. For instance, if the husband killed his wife, this implies that the wife is indeed dead.

Not every empirical consequence of a scenario counts as a prediction as I am using the term here. In the predictivism debate, the term ‘prediction’ is often used as a shorthand for ‘the prediction of a novel fact’. A fact is novel if it was not used to construct the explanation.\footnote{This is called the ‘heuristic’ or ‘use novel’ conception of novelty. The heuristic conception is by far the most widely used conception of novelty in contemporary philosophy of science (Barnes, 2018). Other notions of novelty also exist. The most well-known of these is the temporal conception, which defines a novel fact as one that was unknown to the scientific community at the time when the relevant theory was created. However, the temporal conception is no longer defended in the philosophy of science (Harker, 2008, 431).} For instance, suppose that the husband became a suspect because he repeatedly threatened his wife in the past. The husband’s threats are empirical consequences of the scenario (if we assume that they are more likely to have happened if the husband did indeed kill her than if he did not). But they are not novel, because they were used to construct the hypothesis that he is the killer. Facts that are used in the construction of a hypotheses are called accommodations. In contrast, if investigators did not know about the husband threatening his wife when they formulated the scenario that he killed her, then these threats are a novel fact. The discovery of these threats means that the scenario was confirmed by a successful prediction.

2.2 Evidential strength and risky predictions

There are various ways in which we can spell out the notion of evidential strength. In this chapter I use the Bayesian notion of the likelihood ratio, which is one of the
most common ways of spelling out the evidential strength of evidence. On the likelihood ratio, how much a piece of evidence confirms a hypothesis is directly proportional to how probable it is that we would observe that piece of evidence given the truth of the hypothesis divided by how probable this is given the falsehood of that hypothesis:

\[
P(E | H) \over P(E | \neg H)
\]

To give an example, how probable is it that we would find a defendant’s fingerprints on the murder weapon if he killed the victim and how expected is this if he did not kill her? If predictions provide stronger evidence than accommodations, we should somehow be able to spell this out in terms of the likelihood ratio. This is what I will attempt to do in this chapter.

However, before I turn to the predictivist arguments – there is another term that is worth spelling out for the ensuing discussion, namely ‘riskiness’. A prediction that – if confirmed – would provide strong evidence for a hypothesis (in terms of the likelihood ratio) is a risky prediction. Such a prediction is likely to come true if the scenario from which it follows is true, but unlikely to come true if the scenario is false. For instance, take again the scenario that the husband was the killer. An example of a non-risky prediction resulting from this hypothesis is that we would find the husband’s DNA on the victim. This prediction is likely to come true regardless of whether the husband actually is the killer. After all, partners often have physical contact, even when they are not killing one another. Contrast this with a riskier prediction. For example, suppose that the victim died from being beaten to death. We can then predict that if the husband did indeed kill her, he is likely to have sustained injuries to his fists. This is a risky prediction because such wounds are likely if he killed his wife by beating her to death (such an act often leaves injuries to fists) but unlikely if he did not kill her. After all, most of us do not typically have
this type of injuries on our hands. Of course, such a prediction is not risky if there is a plausible alternative explanation for why he has injuries to his fists that implies his innocence. For instance, if the husband is a professional, bare-knuckled boxer, such injuries are expected, even if he did not kill his wife. Whether a prediction is risky therefore also depends on the specifics of the case and particularly on the competing scenarios involved in it.

In contrast to predictions, accommodations are, by definition, not risky. After all, the definition of riskiness implies that the fact could turn out to be wrong. However, in the case of accommodation we explain facts that we know to be true. Some have tried to ground predictivism in this observation (see Mayo, 1991 for a discussion). But we should be careful not to confuse riskiness with evidential strength. Accommodations can still yield strong evidence for a hypothesis, even if they are not risky. After all, the accommodated fact can also have a high likelihood ratio. Consider the above example again. Suppose that the husband became a suspect in the killing of his wife because his hands were injured. This would be a case of accommodation, but this does not necessarily change how probable the observation of this evidence is given either the truth or falsehood of the hypothesis that the husband is the killer. So, we need a better argument for why predictions yield stronger evidence than accommodations. In the remainder of this chapter I offer three such arguments.
The first benefit of prediction over accommodation concerns the reliability of witnesses. Witness testimony is one of the most important types of criminal evidence. It includes the testimony of the defendant, eyewitnesses, character witnesses and expert witnesses, such as forensic scientists and psychiatrists. The testimony of these witnesses can be supported by predicted and accommodated facts. However, I want to argue that testimony which is confirmed by predictions is – all else being equal – more strongly confirmed than testimony that is confirmed only by accommodations. The reason for this is that successful predictions make it less likely that the witness is telling a false story because they are lying or because they were subconsciously influenced by information received after they observed the events.

To begin with an example, imagine that a bank robbery was committed. Marcy, an eyewitness who was inside the bank during the robbery, testifies that: “The robber was a bald man in a red sweater with a big scar across his face.” Furthermore, suppose that investigators obtain camera footage from the neighborhood surrounding the bank. On the footage a man – Luke – can be seen a short distance from the bank, ten minutes before the robbery took place. Luke is bald, has a big scar across his face and is wearing a red sweater in the footage. That the footage showed someone matching Marcy’s description of the robber close to the bank, shortly before the robbery occurred, seems to provide strong support in combination with her testimony. When cast in terms of the accommodation/prediction distinction, this counts as a successful, risky prediction. After all, Marcy’s testimony was not based on the camera footage. Furthermore, the likelihood ratio of the evidence is high – which makes the prediction risky. To see why, imagine that Marcy’s description of the robber is entirely wrong. In that case it is very unlikely that we would observe a bald man in a red sweater and a scar over his face near the bank during that time. But if her description is accurate, we would expect the robber to have been near the bank around the time that the robbery occurred.
However, now suppose that Marcy accommodated the camera footage. Recall that an accommodated fact is one that is used in the construction of a hypothesis. For instance, the police might have first shown Marcy the footage and she could have filled in the details of her hazy memory based on Luke’s appearance in the camera recordings. If we know that this is the case, we should assign less credibility to Marcy’s testimony than if she had made a risky prediction. In particular, the likelihood ratio of the testimony is then lower, because we would expect Marcy to report that the robber was a bald man with a red sweater and a scar over his face, even if this was not true.

Obviously, whether a witnesses’ testimony was influenced by specific information is usually a question that we seek to answer, not something that we can know with certainty. Nonetheless, the surer we are that certain facts were predicted, the more these facts confirm the testimony. If we are certain that a witness did not use specific information in the construction of her testimony, then this information confirms her testimony more strongly than if we suspect that her testimony only fits with these facts because she fitted her testimony to them.

To give another example, suppose that a suspect is guilty. It can then be in his best interest to call upon his right to remain silent and only offer an alternative scenario at a late stage, when all the evidence has been presented in court. Or he can continuously change his story to fit any counterevidence. In either case, he fits his testimony to the evidence in ways that make it seem well-supported. Because criminal evidence often deteriorates quickly, it may be impossible to test such a scenario by gathering new data. Furthermore, the fact that this scenario is constructed to fit with the evidence can mean that it can be more coherent (Vredeveldt et al., 2014) and better supported by the evidence (Gunn et al., 2016) than a true explanation. After all, the suspect then knows what evidence to accommodate. However, if a

---

134 We saw an example of this in the Venray case, discussed in the last chapter.
suspect offers an explanation that yields testable predictions then this alleviates the worry that he is lying by fitting his story to the available evidence (Jellema, 2019). The above is an example of consciously fitting one’s testimony to the evidence. However, this process can also happen subconsciously. For instance, it is well known that the reliability of witness memories can be negatively influenced by ‘post-event information’ – i.e., information provided to a witness after her observation (Zaragoza et al., 2007). To give an example, police providing information about a potential suspect to a witness is a well-known cause of witnesses identifying the wrong suspect (Wise et al., 2014). Such post-event information then plays a causal role in the construction of the testimony and hence counts as an accommodation – one that lowers the witness’s reliability.

What goes for eyewitnesses also goes for expert witnesses. Experts are prone to make erroneous judgments when they receive too much information about a case. This is known as ‘contextual bias’ (Kassin et al., 2013; Thompson, 2017). Receiving information about a case can lead experts to develop expectations about the outcome of an examination. For instance, fingerprint examiners are less likely to report a match between a latent print from a crime scene and a suspect when they are told that the suspect had a solid alibi (Dror et al., 2006). Similar types of contextual bias occur in several other types of forensic science, such as document examination, bite mark analysis, bloodstain pattern analysis, forensic anthropology and DNA analysis (Thompson, 2017). Doing the initial assessment without knowledge of the target eliminates such potential influences.

Note however that the point is not that explaining known information inherently makes a witness less reliable. In the case of eyewitness testimony, there is at least one type of data that we do want the witness to base their testimony on – namely

---

135 As one Dutch court wrote, “testimony is more credible if it is provided early in the investigation and is then confirmed by further evidence.” District court of Rotterdam. (3d of October, 2019). ECLI:NL:RBROT:2019:7786. My translation.
their own sensory observations during the events that they are reporting on. Similarly, in the case of the expert witness, there will be certain facts that the expert should base her testimony on. However, other information should be excluded from the expert’s judgment. If such information did not play a causal role in the construction of the expert testimony, but does support her judgment, then it is stronger evidence than if it did play such a causal role. Consider a fingerprint examiner who did not know that the suspect had an alibi and who reports a non-match between a print found on the crime scene and that of the suspect. Contrast this to a second expert who did know about the alibi. In the second case, the alibi might have been a subconscious reason for the examiner to come to the conclusion that the fingerprints did not match. So, we should assign a higher reliability to the former expert’s conclusion than to the latter upon learning that the witness had an alibi.

My claim is similar to that of various philosophers of science who argue that predictions tell us something about the competence of a scientist (e.g., Maher, 1988; Kahn et al., 1996; Barnes, 2008). For instance, Kahn et al. (1996) claim that scientists may choose to predict because they have confidence in their abilities to produce true theories – for instance because they know that they have a good intuitive grasp of the problem at hand – whereas those who accommodate may do so because they lack such confidence. Admittedly, some authors regard such arguments with suspicion. For example, as Lipton (2005) points out, in science we are after the evaluation of hypotheses, not the evaluation of scientists. Therefore, even if predicting scientists tend to be more reliable theory-constructors than those who accommodate, the fact that they predicted should ideally not influence how we evaluate their theories. Nonetheless, even if we grant this point in the context of science, when it comes to witness evaluation, we obviously are interested in how reliable they are as sources of testimony. Such witnesses typically testify during the investigation or the trial because fact-finders expect them to have knowledge that is not easily accessible to others. This can either be because the witness is attesting to their personal experiences or because they report on their field of expertise. In such
situations, fact-finders often cannot readily evaluate the credibility of the claims made by the witness. Instead, they have to evaluate their credibility as a person – i.e., how reliable such a witness is.

4. Fudged scenarios

The argument from the previous section established that witnesses who predict are – all else being equal – more reliable than those who do not. However, the scope of this argument is limited. In the philosophy of science, predictivism is the position that we should assign a higher degree of confirmation to scientific theories that successfully predict certain facts than to theories that only explain known facts. This epistemic advantage does not arise because we consider the scientists who formulated the theory to be better, but because we believe that the theory is better supported by the evidence in some sense. The analogous claim in criminal cases would be that a scenario which is confirmed by predictions is – all else being equal – more strongly confirmed than one that only explained known data. Let us turn to an argument that supports this broader claim. This argument is analogous to Peter Lipton’s (2005) ‘fudging’ argument for predictivism from the philosophy of science. It connects accommodation to a specific kind of biased hypothesis construction. According to Lipton, scientists who only explain known evidence when they formulate a hypothesis can be prone to ‘fudging their theories’, i.e., proposing weak hypotheses that explain as much of the evidence as possible. For example, such a theory may lack internal coherence or be overly complex. I propose that fudging is also a danger in criminal cases and that the act of prediction can protect against this danger. But before I argue for this, let us first look at Lipton’s argument in the context of science.

When scientists formulate a hypothesis, they will usually try to make it fit with as much of the available data as possible. This is generally a good approach to hypothesis construction. However, fitting a hypothesis as closely as possible to the existing
data is not always a reliable method. In particular, such a fit may come at the cost of weakening the hypothesis. As Lipton and others, such as Lange (2001) note, explaining known data can occasionally lead to scientific theories that lack coherence, are overly complex or fit poorly with our background assumptions. Such theories explain the evidence well but lack internal plausibility. To illustrate, Lipton gives the example of Ptolemaic astronomy, which had to accommodate contradicting astronomical data by adding more and more complex epicycles to the theory. In such cases, the scientist is overly focused on making sense of the evidence and as a result (subconsciously) makes ‘unnatural’ modifications to her theory. To put it differently, in order to have it fit with the data, she changes the hypothesis in ways that make it less internally coherent, more complex and/or decrease the theory’s fit with our background knowledge.

Like scientists, when criminal investigators formulate a scenario, they typically also want this scenario to fit as closely to the evidence as possible. After all, most people consider a scenario that explains more of the evidence to be better than one that explains less (Pennington & Hastie, 1991; Pardo & Allen, 2008). But the danger of fudging lurks here as well. The goal of fitting the scenario to the data can lead to sacrifices in the quality of that scenario. There are several reasons why such fudging might occur. Lipton mostly refers to situations where the data is varied – which an overly complex theory can easily account for. However, I believe that there are at least two other causes of fudging that are more relevant for criminal investigations.

The first of these causes is the attachment to existing hypotheses in the face of contradicting evidence. For instance, Lipton’s own example of Ptolemaic astronomy is strictly speaking not about heterogeneous data but about unwanted data. The reason that this theory became overly complex is because it had to explain away problematic data that contradicted the theory. Similarly, when investigators become

136 However, some suggest that Lipton’s example is well-chosen. For example, Kuhn (1957, 169) argues that Copernicus’ system is neither simpler nor more accurate (observationally) than the Ptolemaic system.
attached to a specific explanation they sometimes keep amending that explanation
to explain away further, conflicting data – at the cost of making that scenario more
incoherent or fit less well with certain background assumptions about how the
world works. For example, a contradicting witness statement might be explained
away by arguing that this witness is lying for no reason.

Another possible source of fudging is that in criminal cases, some of the data may
be unreliable or irrelevant. Trying too hard to fit one’s scenario to such data can
result in a weaker scenario. Consider a situation in which multiple eyewitnesses
report on a robbery. We know that eyewitnesses commonly misremember many of
the details of any event (Wise et al., 2014). So, the stories that these witnesses tell
will often diverge both from the truth and from each other in many ways. However,
suppose that we want our scenario to match up exactly with the available testimony.
This may require a complex narrative of the robbery. For instance, if witnesses
misreport the precise timing and location of the events, then a narrative that fits
with their testimony is likely to contain illogical jumps in time (see, for instance, the
prosecution scenario in the Hae Min Lee case, in section 6).137

The problem with fudging is that it is sometimes difficult to notice. As Lipton
(2005, 221) points out, evaluating whether a hypothesis was fudged is always not
always straightforward: “[T]his may be to exaggerate scientists’ abilities or equiva-
ently to underestimate the complexity of the factors that determine the degree to
which the hypothesis is supported by data.” Similarly, it is not always straightforward
to determine how good a crime scenario is – especially in hard cases. Is a given
scenario overly complex, does it fit poorly with our background assumptions?
These questions may not be easy to answer. For instance, the case study at the end

137 This second cause is similar to the point made by philosophers of science Hitchcock
and Sober (2004). They link accommodation to the worry of ‘overfitting’ - where a scientist
wants to fit her model too closely to noisy data, thereby making it overly complex. While
they mainly use this in the context of statistical curve-fitting, they argue that the same
notion can be applied to scientific theories in general.
of this chapter illustrates how all involved in a criminal trial may fail to spot fudging and that careful analysis may be needed to bring it to light (section 6.4).

This brings us to the benefit of prediction over accommodation. In science, fudging happens because the scientist has an incentive to fit her theory to the data. As Lipton (2004, 170) puts it: “[t]he scientist knows the answer she must get, and does whatever it takes to get it.” However, this incentive is not present when the scientist makes predictions. After all, in that case she does not know what the outcome of the experiment will be and cannot fit her hypothesis to it. So, if the hypothesis and the evidence fit well together, this is less likely to be due to fudging. The conclusion of this argument is that, all else being equal, a hypothesis that predicted a fact is better supported than a hypothesis that accommodated that same fact. Similarly, in the criminal trial context, fudging may happen because investigators know the evidence that they want to explain and may sacrifice the quality of their scenario to achieve this. Therefore, if we know that a fact was used in the construction of the scenario then there is a possibility that the scenario and the fact only fit due to fudging. But now suppose that we know that a fact was not used in this construction process. The investigator then did not have an incentive to bend the scenario until it fit the fact. And therefore, we can be relatively certain that the fit between the scenario and the fact offers genuine support to that scenario. After all, they did not know all of the evidence in advance that their scenario had to explain.

Of course, whether a fact was predicted is sometimes also difficult to judge. But suppose that we are in a situation where we know with a high degree of certainty that a scenario made successful, risky predictions. In such situations the evidence provides strong support for the scenario. In contrast, if a scenario only accommodates known data, we have more reason to worry about fudging. Therefore, we should assign a higher degree of probability to the first scenario. This is not because predictions are better in some logical sense than accommodations. Rather, they are indicative of another epistemic virtue – namely non-fudging. This is in contrast with the argument from last section, about witness evidence. Given that the
argument from that section holds, there is a logical difference between predicted and accommodated evidence. In other words, predictions should make even a perfectly rational evaluator assign a higher degree of reliability to the witness.

5. Fudged evidence, confirmation bias and the argument from choice

Let’s now turn to the final benefit of prediction over accommodation, one that is similar to the benefit described in the previous section. The fudging argument that I just discussed concerned investigators fitting their scenario to the evidence. But the opposite is also possible – criminal investigators sometimes consciously or subconsciously fit their evidence to a preconceived scenario by selectively gathering or interpreting it – they fudge their evidence. Just like a fudged scenario, this may lead to the illusion of the evidence supporting a scenario well. I want to argue that checking a scenario’s prediction is an important tool in preventing fudged evidence.

Fudged evidence is a well-known problem in criminal cases. Criminal investigators may fit the facts to a preconceived scenario in ways that leads to weak support. In those situations, they are blind to some of the evidence that the scenario should explain, or they misinterpret evidence to make it fit with the scenario. Such fudging may happen intentionally. However, it usually happens subconsciously, in which case we call it confirmation bias – “the unconscious tendency to seek out, select, and interpret new information in ways that validate one’s pre-existing beliefs, hopes, or expectations” (Nickerson, 1998). Confirmation bias is pervasive in all of human affairs. Yet it is especially dangerous in criminal cases, where it is one of the leading causes of judicial errors (Gross et al., 2004). Investigators, judges and jurors are typically not aware of their own confirmation bias (Charlton et al., 2010). In criminal cases, we often encounter such bias in the form of tunnel vision – “a rigid focus on one suspect that leads investigators to seek out and favour inculpatory evidence, while overlooking or discounting any exculpatory evidence that might exist” (Findley & Scott, 2006).
In this chapter, I defend the epistemic superiority of prediction. However, the possibility of fudged evidence might – at first sight – seem to throw a spanner in the works of this project. After all, the fudging argument for predictivism given in section 4 relies on the observation that first gathering the facts can bias the way we then construct a hypothesis based on those facts. When we predict, the process is reversed: we first formulate a hypothesis and then gather the facts. But this provides an incentive to fudge the evidence so that it supports our preconceived hypothesis. In both cases the danger is that what comes first in the process may bias our second step. If this reversed fudging argument holds, it seems to support the notion that accommodation is superior to prediction when it comes to preventing fudged evidence. So, let me begin by addressing this potential objection, before I move on to the value of prediction in counteracting confirmation bias.

Admittedly, there is some truth to the aforementioned argument: if we are sufficiently certain that the accommodating investigator gathers her evidence without any preconceptions, then we can safely assume that she did not fudge the evidence. In such a situation, there is no preconceived hypothesis to fit the evidence to. In contrast, a predicting investigator will have a motive to fudge their evidence much in the same way that an accommodating investigator will have a motive to fudge their scenario. However, the problem with this argument is that it relies on an unrealistic view of accommodation. The process of explaining known data can involve preconceptions just as much as the process of prediction. Even if we are creating a scenario from scratch, preconceptions do not need to take the form of a fully formed scenario. Investigators can also have vague suspicions and implicit biases. This may lead them to ignore some of the facts or to interpret them in a biased way when accommodating. To give an example, investigators may construct their scenario around the testimony of a select group of witnesses who agree on how a certain event took place, while ignoring the testimony of other witnesses, who offer a different version of the events. There is, therefore, no fundamental difference between prediction and accommodation in this regard. In both cases the
danger of confirmation bias lurks. The above argument does not establish that accommodation is, in some sense, superior over prediction. Instead, confirmation bias poses a problem for both predictors and accommodators. Apparent success at explaining known data or at predicting novel facts may be due to the facts having been fitted to the explanation, which could lead us to overvalue how much our evidence supports a scenario. So, regardless of whether we accommodate or predict, we want some assurance that the evidential support for the scenario is genuine and not the result of fudged evidence.

There are different ways to counteract confirmation bias. For instance, investigators can (and arguably should) consider multiple scenarios during a case (O’Brien, 2009; Van Koppen, 2011). This prevents them from becoming overly focused on a single possibility. However, I want to emphasize a different method here. To prevent confirmation bias, investigators can adopt a falsificationist mindset – where they explicitly seek out evidence that might disconfirm their scenario, rather than implicitly trying to confirm that scenario (Nickerson, 1998; Van Koppen, 2011; Maegherman, 2021). When we adopt such a mindset, we ask the question: “suppose that our scenario is false, which evidence would we then expect to find (or not expect to find)?” Obviously we can only ask this question when we have already formulated a scenario. It does not make sense for an accommodator, who gathers evidence without having a scenario, to ask what could prove this scenario wrong. Of course, it is possible that this question can be answered by referring to the available evidence, which was used in the construction of the scenario. If the accommodator has done a thorough and fair search for evidence before constructing their scenario it might be that she has already searched for – but failed to find – any contradicting evidence. However, research suggests that people often prematurely stop looking for further evidence once they have formulated a scenario that explains the known facts well enough (Hoch, 1985). So, investigators should be careful with concluding that there is no contradicting evidence as this conclusion may be the result of bias.
A falsificationist mindset guards against such bias because it forces us to check on the scenario’s risky predictions – which are those predictions that are likely to fail if the scenario is untrue. This role of predictions was famously emphasized by Popper (1959), who considered falsification attempts of theories crucial to the scientific enterprise. Within the predictivism debate Mayo (1991) argues that our intuitions about the value of predictions are best explained by the fact that testing a hypothesis by means of its risky predictions puts it through a ‘severe test’ – i.e., tests that a hypothesis is unlikely to pass if it is false.

This brings us to the final argument for the value of predictions over accommodations, which Lipton (2005) calls the argument from choice. When we accommodate, we have to make the best of whatever information is available. However, this will not always include the most telling evidence – i.e., the evidence with the most discriminating likelihood ratio. In contrast, when we predict, we can pick which predictions to test. One caveat to this argument is that criminal investigators may not always have the opportunity to carefully test a scenario’s predictions (cf. Van Koppen, 2011, ch. 3). After all, criminal cases usually involve limited and deteriorating evidence. This means that we can choose to check the predictions that test our scenario the most severely. If a scenario passes these tests it will thereby be more strongly supported than if it did not. This strong support does not arise because predicted evidence is inherently better; the resulting evidence would be equally strong if it were used to construct an explanation. Rather, when we test predictions, we have more options for choosing what evidence to gather and how to gather this evidence. For example, suppose that there is a murder case. Once we have a detailed scenario of when, where and how the murder took place, we can start to look for witnesses that were in that area at that time and find out whether their testimony matches this scenario.

Of course, as mentioned above, the danger with testing predictions is that investigators may be susceptible to fudging the results. When they test a scenario’s predictions, they might subconsciously choose to look for only confirming evidence
or to misinterpret the results to fit with their preconceived scenario. For instance, they might be inclined to ask witnesses leading questions. Nonetheless, when investigators adopt a falsificationist mindset, when they are cognizant of the possibility of confirmation bias and deliberately seek out evidence that might contradict their scenario, they diminish this worry. So, testing predictions does not inherently prevent confirmation bias. Rather, carefully testing predictions is a vital part of adopting a falsificationist attitude, which in turn counteracts confirmation bias. When our scenario only explains known facts, we run the risk of not testing our scenario as severely as possible.

6. Explaining predictions

I have discussed three different arguments for the value of predictions. The upshot of these arguments is that, all else being equal, predictions yield stronger evidence than accommodations. However, how strong this evidence is exactly and whether ‘all else is equal’ will depend on context. How do we determine what a particular prediction (or lack thereof) supports in a given context? I want to suggest that the above arguments can best be understood within the framework of inference to the best explanation. Suppose, for instance, that a witness’s testimony fails to yield testable predictions. There can be various explanations for this fact. Some of these explanations may be consistent with the witness being a reliable source. For example, the witness may have been interviewed long after the events on which they report took place and as a result additional evidence may have been difficult to gather. Other explanations might – if they were true – imply that the witness is unreliable. For instance, a lack of predictions may be explained well by the witness strategically lying. What the lack of predictions tells us about the reliability of the witness will depend on the relative quality of such explanations. For instance, if the

\[138\] We will see examples of such context-dependency in the next section, where the murder of Hae Min Lee is discussed.
best explanation (by far) for a lack of predictions is that the witness is lying, then this is strong evidence that they are unreliable. In contrast, if there is a reasonable explanation which does not imply that they speak a falsehood, then their lack of predictions may not tell us anything about whether or not they are reliable. When a witness does make a successful, risky prediction, this may make the alternative explanations – which imply unreliability – implausible. For instance, a successful prediction may make it unlikely that the witness is lying. Similarly, if the scenario formulated by investigators yielded successful predictions, then this makes it less plausible that this scenario was constructed through fudging or that the investigation suffered from confirmation bias.

Applying IBE to predictive success is not without precedent in philosophy. The notion that there can be competing explanations for (a lack of) predictions and that the plausibility of these explanations determines what this proves is also central to the No-Miracles argument from the philosophy of science. This is one of the most famous arguments in support of scientific realism – the position that our best scientific theories are (mostly) true and that the entities described in these theories – such as atoms – exist (cf. Van Fraassen, 1980, 39). The argument roughly goes as follows: science is astonishingly successful at explaining known phenomena and in making novel predictions. Such success would be a miracle if our theories were false. Hence, our best theories are (mostly) true. The no-miracles argument is often understood as an instance of inference to the best explanation, where scientific realism is taken to be the best explanation of science’s success. How convincing we should consider the no-miracles argument to be is the topic of ongoing discussion (see Sprenger & Harmann, 2019, ch. 5). For instance, critics have argued that there is no reason to believe that scientific realism is the best explanation of science’s predictive success (Worrall, 2011). Debate continues about how plausible an explanation realism provides of predictive success and whether there are alternative explanations of such success. Similarly, it is possible to argue about what a given prediction means in the context of a specific criminal case. The prosecution and
defense may, for example, disagree about whether a given explanation is plausible or not. Second, they may suggest alternative explanations for (a lack of predictions). Finally, to complicate things, it is possible to argue about whether there was a (lack of) prediction in the first place at all. For instance, suppose that a witness makes an apparently successful, risky prediction, but that we later find out that this is only the case because the police withheld the fact that they shared information with the suspect, thereby making it look as if they successfully predicted the information shared with them.¹³⁹ Such arguments are also inferences to the best explanation concerning the apparent (lack of) predictive success.

To summarize, (a lack of) predictions can be evidence in favor or against the truth of a scenario. Whether it is evidence, and how strong this evidence is, depends on an inference to the best explanation. This is therefore an example of how explanation-based thinking helps determine the probabilities that Bayesianism needs. I will explore this idea – that IBE helps reach conclusions about the strength of evidence – in much greater detail in chapter VI, where I discuss how we may assess the reliability of eyewitness evidence. To conclude the present chapter, I want to look at a case study which exemplifies all three arguments in favor of predictivism.

¹³⁹ See, for instance, the Venray murder discussed in chapter IV.
7. Case Study: The murder of Hae Min Lee

So far, I have illustrated the special value of predictions in criminal cases with simple examples. Let us now consider the three arguments in the context of a real-life case – the murder of Hae Min Lee. The podcast *Serial* (2014) gained fame by exploring the intricacies of this case in its first season. Given the scope and complexity of this case, I invite the reader to listen to this podcast for a more extensive discussion of the investigation and subsequent trial. The goal of this case study is not to reach a verdict on whether the conviction of Adnan Syed – the defendant in this case – was legitimate. Rather, I discuss this case because it provides vivid, practical examples of predicting and non-predicting witnesses (6.2), severe testing by means of checking predictions, (6.3) and fudged scenarios and evidence (6.4). Hence, it shows that the aforementioned arguments occur in real-life situations.

7.1 The case

*Hae Min Lee*, an American high school student, was murdered in early 1999. Her body was found in a park four weeks after she was last seen alive. The cause of death was manual strangulation. *Adnan Syed*, Lee’s ex-boyfriend, was arrested, charged and convicted for her murder. Though the decision was upheld in appeal, critics argued that it was unclear whether the evidence truly supported the prosecution’s scenario. The conviction was primarily based on two items of evidence. First, the testimony of *Jay Wilds*, who claimed that Adnan had murdered Hae. Second, call records of Adnan’s mobile phone.

7.1.1 Jay’s story

Jay was an acquaintance whom Adnan occasionally smoked weed with. He testified multiple times, both in police interviews and in court, that Adnan killed Hae. His story was as follows: They met up to go shopping in the morning. Adnan then told
him that he was going to “kill that bitch”, referring to Hae who had recently started seeing someone else. Afterwards Jay dropped him off at school and kept both Adnan’s car and phone, so that he could pick him up later. In the afternoon, Adnan called Jay from a payphone to come pick him up. When he arrived, Adnan showed him Hae’s body in the trunk of her own car. After the incident, they left Hae’s car elsewhere. Jay then dropped off Adnan at track practice. Later, Adnan called Jay to come pick him up again and after eating together they buried Hae’s body in a nearby park.

Critics questioned Jay’s credibility, especially because he kept changing his story between testimonies. For instance, he gave differing reports about where he first saw Hae’s dead body. At one point he admitted having lied to the police on previous occasions. However, according to the prosecution, Jay’s story was ultimately credible because it was corroborated by the cellphone records.

7.1.2 Adnan’s story

According to Adnan, he did indeed lend his car and phone to Jay to be picked up after track practice. He claims that he was first in the library and then went to track practice during the time period when Hae was most likely killed. After track practice, Jay did pick him up. But according to Adnan, they went to a friend’s house together. In the evening, he went to the mosque alone.

7.1.3 The cellphone records

The call record was a list of over thirty calls that contained the phone numbers that called or were called by Adnan’s cell phone, the time at which these calls took place, the duration of the calls and – most importantly – which cell phone tower the call was routed through.

According to both Jay and Adnan, it was Jay who had the phone on him that fateful day. The prosecution argued that the cell phone records corroborated Jay’s
testimony. First, they gave a rough location of where Adnan’s phone was – which prosecutors claimed matched up with the locations in Jay’s story. They also connected Adnan to the location of the phone because of one particular call. This was the 3:32 p.m. call made to *Nisha*, a friend of Adnan. She was the only person on the call list who did not know Jay. Hence, the reasoning was, Adnan must have been near the phone that afternoon, because Jay wouldn’t have called her since he didn’t know Nisha. This in turn would mean that Adnan was lying when he said that he and Jay were not together that afternoon until Jay picked him up from track practice.

7.2 Predicting versus non-predicting witnesses

As I discussed in section 3, predicting witnesses tend to be more reliable than those who accommodate. This case featured the testimony of two key witnesses: Jay and Adnan. Interestingly, the testimony of Jay was confirmed by a successful prediction, while that of Adnan was not. Let’s consider what that means for the degree of reliability that we should ascribe to them.

Much of Adnan’s scenario was unverifiable – it was low on details and did not lead to any clear, novel predictions that the police could follow up on. In contrast, Jay did make a successful prediction. His story included him and Adnan getting rid of this vehicle together and he successfully predicted the location of Hae’s car. When Jay was first interviewed by the police, he correctly pointed the police to a hill near the city where it was parked. The police discovering the car there is a prediction of a novel fact. Imagine that the police had already found Hae’s vehicle and had told Jay about this. If Jay’s story then included its location, it could be possible that he (deliberately or subconsciously) fitted his story to the information that the police gave him – i.e., that he was accommodating. Based on this successful prediction, the police considered his story believable enough to arrest Adnan.
What does the above mean for the credibility of Jay and Adnan? The answer to this question depends on the explanations for their (lack of) predictive success. For Jay, the prediction takes away one worry that we might have about his story – namely that he is fitting his story to publicly known information. Nonetheless, there are other scenarios, in which Adnan is innocent, where Jay would also have knowledge of the car’s whereabouts. Such scenarios include Jay killing Hae himself or someone else than Adnan killing Hae and telling Jay about it. How much this prediction tells us depends on how plausible these alternative explanations for his knowledge of the car’s location are.

Now let’s consider Adnan, whose testimony did not lead to clear, novel predictions. Is this lack of prediction itself evidence that he is not telling the truth? Again, this depends on the competing explanations for this lack of predictions. A lack of predictions is explained well by him lying. After all, if Adnan is lying, then it is in his best interest to assure that investigators do not discover evidence that contradicts his story. However, there are also plausible explanations for this lack of prediction consistent with him telling the truth. After all, it will often be difficult to produce exonerating, novel evidence for where you were on any particular day, even if you did not do anything illegal that day. This is especially the case when you are first interviewed about that day over a month after the fact, as Adnan was.

Nonetheless, now suppose that Adnan had made a prediction. In that case the likelihood of his testimony would be much lower under the assumption that he was lying. After all, it would thereby have taken away the worry that he was lying by fitting his testimony to the facts provided by the police. By not predicting any novel evidence, Adnan fails to take away this worry. In other words, though his lack of predictions is not necessarily a sign of unreliability, if he had predicted, it would have been a sign of reliability.

7.3 Route talk and severe tests
Let’s turn to the argument from choice, discussed in section 5. According to this argument predictions can sometimes be designed to provide as severe tests as possible, in ways that accommodations cannot. The reasoning is that when we predict, we can first identify what the most telling evidence would be and we can then deliberately search for that evidence. This option may not be available when we accommodate, where we have to make do with whatever evidence is available. Let’s consider an example to illustrate this argument.

In episode 5 of the podcast, the presenters took up a challenge posed to them by Adnan to test a prediction following from the prosecution’s timeline. The prediction in question is about the time it would take to drive from Hae and Adnan’s school to the parking lot where Hae was allegedly murdered. Hae’s last class ended at 2:15 p.m. that day and multiple people remembered seeing her heading to her car afterwards. Furthermore, Adnan supposedly called Jay at 2:36 p.m. to come pick him up after he had killed her. According to the prosecution, in those twenty-one minutes, Hae drove to Best Buy and Adnan strangled her there. The podcast makers wanted to test the prediction – that it was possible to make this drive in twenty-one minutes, and still leave time to strangle someone. They made several attempts, starting at the high school, right after classes ended. They concluded that it was indeed possible, as long as the driver made no errors and there were no delays because of traffic or other sources. Yet Adnan claimed that such conditions were unlikely, because leaving the school after classes ended meant that fifteen-hundred students were leaving the building and driving off the parking lot usually meant having to wait for buses. Furthermore, even without delays, Adnan would have had only three minutes to strangle Hae, put her body in the trunk of his car (in broad daylight) and call Jay from a public pay phone. Hence, the podcast deemed it unlikely that the state’s timeline was correct.

Note that the prediction as I discuss it here is therefore one made by the prosecution scenario, not one made by Adnan. Adnan merely points out that this prediction is implausible.
The relevance of this example is that it illustrates the key role of predictions in severely testing a scenario. Such severe testing involves asking “which facts could prove this scenario wrong?” When we predict, we can choose to focus on the most promising predictions that could realistically falsify the scenario. For instance, in this case the prediction was that it should be possible to drive from the school to the Best Buy in twenty-one minutes, given similar traffic conditions as Hae would have faced that day, while still leaving time for Adnan to strangle her. This is a very specific prediction, which the podcast makers checked on because its importance had been pointed out by Adnan. He suggested that this was a way to show that the scenario was wrong. To put it differently, this predicted fact arguably had a high likelihood ratio – if the prosecution’s scenario was correct, it was very likely to be proven true. But, given Adnan’s remarks, if the scenario was false then this would be one of the facts which would most likely show it.

To see why only accommodating evidence may lead to a scenario not being tested severely, imagine that the podcast makers had restricted themselves to re-checking the information that was used to arrive at the very conclusion that Adnan was the killer. Such information could turn out to contradict the scenario. I will discuss examples of this in the next subsection. However, if there was any information that could prove his innocence, it was likely to be found outside the set of evidence that first led investigators to suspecting Adnan.
7.4 Fudging and the call record

In sections 4 and 5, I discussed two kinds of fudging – fudged scenarios and fudged evidence. A scenario is fudged when it is designed to fit well with the evidence at the cost of its inherent plausibility. Evidence is fudged when it is selectively chosen or interpreted to fit with a preconception, such as a preconceived scenario. The Hae Min Lee case arguably featured instances of both a fudged scenario and fudged evidence, namely with respect to the fit between Jay’s story and the call record. A look at this example will help to get a better grip on what fudging looks like in practice.

To begin with the apparent fit between the call records and Jay’s scenario: Adnan’s mobile phone pinged specific cell towers during the presumed day of the murder, whenever it was used to make calls. Because cellphones normally ping the closest tower, this meant that rough estimates could be made about the phone’s location during specific times of the day. The prosecution claimed that Jay’s story fit the call record perfectly. Furthermore, they argued that Jay could not possibly have known which towers were getting pinged when he told his story. In other words, they claimed that he had made a number of successful, and risky predictions, which offered strong evidence for his story. Furthermore, Adnan’s story did not seem to fit with these records. So, if the prosecution was correct in their assertion that the records and Jay’s story fit perfectly, then it would be very probable that Adnan was guilty.

However, there were several reasons to doubt that the fit between the call record and Jay’s testimony was as close as the prosecution claimed. Take the twenty-one minutes between Hae leaving class and Adnan allegedly calling Jay to come pick him up for example. What I did not mention earlier is that neither Adnan nor Jay claim that the ‘come pick me up’ call took place at the time that the prosecution’s time line said it did – 2:36 p.m. According to Jay the call happened around 3:40 p.m., much later. But this was a problem for the prosecution, because the call record
shows no call from the Best Buy location around that time. So, they concluded that the 2:36 call – the only call that feasibly matched both the time and location – was the ‘come pick me up’ call (Serial, 2014, 199). However, this time line led to the implausible twenty-one-minute window for Hae’s murder. In other words, the prosecution tried to fit their time line to the evidence – Hae being spotted leaving class, Jay’s testimony about Adnan calling him to be picked up, and the call records. But this also made the resulting time line implausible, as it meant that the alleged events – Hae driving to Best Buy, Adnan strangling Hae, Adnan calling Jay – would have to have taken place in an unrealistically brief period of time. The prosecution could have opted for a different, more realistic time line. However, then their scenario would have conflicted with either Jay’s testimony, sightings of Hae leaving class or the call record. For instance, if they had taken Jay’s claim that the ‘come pick me up’ call took place at 3:40 p.m., then there would be a clear disconnect with the call record – which showed no such call from Best Buy at 3:40 p.m. Instead, the prosecution ended up with an implausible time line, though one that – at the very least appeared to – fit well with the evidence. In other words, they fudged their scenario, i.e., sacrificed the plausibility of their scenario to get it to fit with the evidence.

Susan Simpson (2015), a legal associate, pointed out another possible instance of fudging, in her online article “Evidence that Jay’s Story Was Coached to Fit the Cellphone Records”. She focused on one of the changes in Jay’s story between police interviews. At first, Jay claimed that he was at home when Adnan called him to be picked up from track practice. This matched the location data on the phone records. However, he later changed his story and said that he was at a friend’s house. The friend denied this and it also did not match the location data. According to Simpson, this change is best explained by the fact that the police were, at first, working with an inaccurate map of the phone location data. On this map, the tower that Jay’s phone pinged was displayed in a different location than its actual location and Jay’s new story matches that location. So, Simpson argued, the changes in Jay’s
story are likely due to the police coaching Jay to make his story fit with the cell records.

These are both instances of the scenario being changed to fit the phone records. However, there were also signs of the prosecutor fudging the evidence. For example, during the trial the prosecutor only cited four of the fourteen locations that the phone pinged, because only those four matched Jay’s accounts. The pings that conflicted with his account were ignored. Furthermore, prosecutors and investigators may have misinterpreted a crucial piece of evidence, in order to make it fit with their preferred scenario. This piece of evidence was the Nisha call – which tied Adnan to the location of the phone, because it was the only call to a person that only Adnan knew. This call is often treated as the smoking gun, which definitively disproves Adnan’s story and proves Jay’s. However, there are reasons for skepticism. Most importantly, in her description of the call, Nisha says that Adnan and Jay wanted her to come to a video store where Jay worked. But he did not start that job until two weeks after this call. Furthermore, while the phone tower matches the time of Jay’s story, it does not match the location. In fact, none of the calls from around that time match where Jay says that they were at the time. Many have since pointed out that the prosecution did not disprove the possibility that Jay accidentally ‘pocket dialed’ Nisha’s number. The podcast looked into this option extensively and concluded that it was indeed possible (Serial, 2014, 275). According to this theory, Nisha could be confusing this call with a later call later in the month, when Jay already worked at the video store. So, this call may have been (subconsciously) misinterpreted to make it appear as if it strongly supports the prosecution scenario. If that was the case, there would be no call tying Adnan to the phone’s location.

To summarize, there are several reasons to believe that the apparent match between Jay’s testimony and the call records was due to fudging. Yet this fudging was not at all apparent at first. For example, much of it escaped the attention of the defense during the trial. Presumably the prosecution also missed it (assuming that they were
not deliberately fudging). Furthermore, even after a careful examination by the podcast and others, such as Simpson, we cannot say with certainty that the prosecutors and investigators did indeed fudge – there is simply too much unclarity on this issue.

Let me make one final point about fudging that is potentially interesting. In section 5 I mentioned that predictions can play a key role in preventing confirmation bias. In particular, checking on predictions is often a vital part of adopting a falsificationist mindset. However, I also proposed that the phenomenon of confirmation bias can also give rise to an illusion of predictive success. Consider the Nisha call. Jay mentions the call during a police interview, seemingly without knowing that it shows up on the call record. Nisha independently confirmed that she talked to Adnan who briefly put Jay on the phone. This therefore appears to be an instance of a successful, risky prediction by Jay. However, this appearance might be caused by the aforementioned fudging – where Nisha’s call is re-interpreted to fit with Jay’s story. So, Jay’s apparent predictive success may be due to investigators fudging the evidence. In this case, there is therefore an alternative explanation for the apparent predictive success – namely confirmation bias. As said, such confirmation bias can be counteracted by carefully considering what evidence could disconfirm the scenario. For instance, the investigators could have considered whether there was any evidence that contradicted their interpretation of the Nisha call.
8. Conclusion

Predictions are a vital element of science. It is therefore unsurprising that philosophers of science have thought about their value deeply. Though predictions are less central to criminal cases than to the scientific enterprise, they are also worth examining in the legal context. In particular, I distinguished three ways in which predicted facts can provide stronger criminal evidence than accommodations. First, witnesses who predict are – all else being equal – more reliable than those who do not. Second, investigators who only accommodate run the risk of ‘fudging’ their scenario; predictions can protect us against this danger. Third, carefully constructed predictions can be designed to provide severe tests of our criminal scenarios as a useful tool in preventing confirmation bias. I showed how these arguments can feature in real, complex situations by discussing the murder case of Hae Min Lee.

One conclusion that follows from this chapter is therefore that the predictivist arguments developed by philosophers of science can offer valuable lessons for criminal evidence scholars. However, the reverse may also be true – the above discussion might shine some light on the scientific predictivism debate. First, I opted for a pluralist approach on which predictions have multiple advantages over accommodations. There are also philosophers of science who adopt a pluralist view (Barnes, 2008; Douglas & Magnus, 2013). My proposal further shows the value of such a strategy, because it allows for prediction to matter in different ways in different situations. In particular, the three values of prediction that I distinguished are categorically different. For the evaluation of eyewitness evidence, predictions are intrinsically better evidence than accommodations; predicted evidence gets a confirmatory boost compared to accommodated evidence because it is inherently better to predict than to accommodate (when assessing witness reliability). This is known as ‘strong predictivism’ (Barnes, 2018). In contrast, when it comes to fudging, predicted evidence also gets a confirmatory boost, but only because prediction is symptomatic of other virtues (non-fudging). All other things being equal, a perfectly rational evaluator should therefore assign the same degree of confirmation to
a scenario that was supported with predictions as they should to one that was supported with accommodated evidence. Yet given the fact that those involved in criminal cases are not perfect rational beings, a piece of evidence that was predicted should give a confirmatory boost for the relevant scenario. This is known as ‘weak predictivism’. Finally, the argument from choice does not establish that evidence being predicted gives it a confirmatory boost. However, it does suggest that predicted evidence will tend to be better evidence than accommodated evidence because we can carefully design such predictions to yield telling results. This distinction between an inherent epistemic value, a weaker epistemic value and a practical value of prediction can shine new light on the relation between different predictivist arguments in the philosophy of science.

One issue that I have not addressed in this chapter is that testing predictions may not only lead to more probable beliefs, but also to more robust beliefs. As I explain in chapter VIII, if we do not check on a scenario’s predictions, and a scenario only explains known data, then this may not only lead to the conclusion that the scenario is less probable, but also that there is evidence which could have overturned the case for guilt, but which we do not possess. In those cases, our belief that the defendant probably committed the alleged acts lacks robustness. As I discuss in that chapter, it may be preferable to view missing evidence as having an impact on the robustness of our beliefs rather than on our belief in the guilt of the defendant. The reason for this is that we cannot know how evidence that we do not possess would impact our beliefs. However, in the current chapter I looked at arguments for the predictivist thesis – which is usually understood in terms of the relationship between the probability of a hypothesis and the evidence for that hypothesis being predicted (or accommodated). For that reason, I have not dealt with the connection between prediction and robustness here, but leave this discussion for chapter VIII.

Let me end by rephrasing my conclusions. First, witnesses sometimes fit their story to the wrong facts – for instance to the information from the case file rather than to their own observations. That is why we want to make sure that they predict,
rather than accommodate these facts. Second, investigators sometimes fit their story to the facts wrongly, where they propose a weak scenario to explain as many of the facts as possible. Prediction yields less incentive for this kind of overfitting. Third, investigators sometimes fit their facts to the story wrongly – where they only consider the scenario probable because they ignore relevant information. That is why it is useful for them to adopt a falsificationist attitude. This usually means testing a scenario’s risky predictions. For these three reasons, those involved in criminal cases should care about whether evidence was predicted or accommodated. However, what a (lack of) predictions proves will depend on an inference to the best explanation. This chapter therefore illustrated one area where explanation-based thinking helps determine the likelihood ratio of the evidence. I will discuss this idea in further depth in the next chapter.
VI. EYEWITNESS EVALUATION THROUGH INference TO THE 
BEST EXPLANATION

Chapter abstract

Eyewitness testimony is both an important and a notoriously unreliable type of 
criminal evidence. How should investigators, lawyers and fact-finders evaluate 
eyewitness reliability? In this chapter, I argue that Testimonial Inference to the 
Best Explanation (TIBE) is a promising, but underdeveloped prescriptive ac-
count of eyewitness evaluation. On this account, we assess the reliability of eye-
witnesses by comparing different explanations of how the testimony came 
about. This account is compatible with, and complementary to both the Bayes-
ian framework of rational eyewitness evaluation and prescriptive methods for 
eyewitness assessment developed by psychologists. Compared to these frame-
works, the distinctive value of thinking in terms of competing explanations is 
that it helps us select, interpret and draw conclusions from the available evi-
dence about the witness’s reliability.

1. Introduction

Eyewitness testimony is one of the most important kinds of criminal evidence. It 
is also notoriously unreliable and a major source of judicial errors (Innocence Pro-
ject, 2009). In criminal cases, fact-finders therefore regularly face the difficult but 
crucial task of evaluating eyewitness testimony. This sometimes means checking 
whether the witness’s story fits with other established facts of the case. However, 
the veracity of such a story cannot always be verified or falsified directly. In such 
cases evaluators will have to look at whether the statement comes from a reliable 
source. How should they go about doing so? I argue that *Testimonial Inference to the 
Best Explanation* (TIBE) is a useful but underdeveloped account of eyewitness
evaluation. On TIBE, we compare competing explanations of why the particular statement was offered (see section 2 for further exposition).

TIBE is an explanation-based approach to eyewitness evaluation. As said in previous chapters, the benefits of such approaches compared to competing accounts are that they fit with how people reason naturally and that they offer direction to fact-finders as to how to go about assessing the available evidence (Pennington & Hastie, 1993; Allen & Pardo, 2019; Nance, 2016, 84). So far, these approaches have focused mainly on the decision whether the proof standard has been met – for instance whether guilt has been proven beyond a reasonable doubt in a criminal case. However, as some have suggested – and as I show in this chapter – explanatory reasoning can also be a helpful way of thinking about the reliability of the available evidence, such as eyewitness testimony.

I am not the first to argue that we can use inference to the best explanation to assess eyewitness testimony. However, existing work on TIBE in the philosophy of testimony has mostly been limited to brief, descriptive accounts, intended to capture our intuitions about when we may trust the utterances of others in daily life (Lipton, 2007; Malmgren, 2008; Fricker, 2017). In this chapter I show that explanation-based reasoning is also a rational and useful approach to assessing eyewitness reliability. In particular, I develop a prescriptive account that offers guidance to eyewitness evaluators. Furthermore, I argue that, and explore under which conditions, this approach is rational. To this end, I connect my account to its main competitor, the Bayesian accounts of eyewitness reliability (section 3). I argue that TIBE and Bayesianism are compatible.\footnote{I therefore take the opposite view to that of Thagard (2005) who is the only other author, as far as I am aware, who has discussed the relationship between TIBE and Bayesian accounts of eyewitness evaluation.} I follow a line of thought from the philosophy of science, according to which inference to the best explanation is an efficient, but imperfect heuristic for optimal Bayesian reasoning. Because, and to the extent that, TIBE tracks the Bayesian ideal, it is rational. However, I do not suggest that TIBE is
subsumed under Bayesianism. Bayesian probability theory offers a helpful way of making precise why and when TIBE is rational, but no more than that. I argue that the crucial question for an account of rational eyewitness evaluation is how to make sense of the available evidence concerning the witness’s reliability. And on this front, TIBE is doing all the hard work.

To illustrate both what is distinctive about TIBE, and how this approach may be further developed, I compare it with two popular prescriptive accounts: the capacity view and empirically-informed frameworks (section 4). I argue that, though these accounts differ on crucial points, they can also complement one another. First, while these approaches are not inherently explanatory, once we incorporate an explanatory component, we end up with a more context-sensitive, flexible framework on which it is clearer how evaluators ought to come to their ultimate decision. Conversely, these existing frameworks can inform us on some of the details of explanatory inference in practice. Finally, I further explore TIBE’s distinctive value compared to both these prescriptive approaches and to Bayesianism by turning to two specific aspects of eyewitness evaluation where explanatory reasoning is especially helpful, namely the interpretation (section 5) and selection (section 6) of the available evidence about the witness’s reliability.

2. Testimonial inference to the best explanation

TIBE applies the comparative idea of inference to the best explanation to the evaluation of testimony. However, those who have written on TIBE have not offered an extensive characterization of what such explanation-based evaluation looks like in practice. I therefore begin this chapter by filling in some of the details of this account.
2.1 Sketching the account

When we evaluate the reliability of a piece of testimony, we are not concerned with evaluating directly whether the statements in the report are true. Rather, what we evaluate is whether these statements come from a reliable source. This source can be the witness’s utterances, a report about those utterances, or even our own recollection of what the witness or the report said. On TIBE we evaluate the reliability of this source by comparing a select number of explanations of why this particular statement is offered in this particular situation, by this particular source. For instance, suppose that the witness of a robbery was interviewed and the report from this interview states that:

“The robber wore a red shirt.”

This statement contains the story that the witness (or, more precisely, the report) tells. There can be various explanations why the report states that ‘the robber wore a red shirt, e.g.:

i. The witness accurately observed the robber up-close when the alleged events took place and is now sharing their observations.

ii. The witness misremembered the colour of the shirt due to the stress of the situation and now reports on this flawed memory.

iii. The witness is lying about the colour of the witness’s shirt because they are trying to frame someone.

iv. The person who wrote down the witness’s remark misinterpreted what was said.

We may generate relevant explanations by asking: “What could explain this statement in this context?” Which explanations we consider will therefore depend on
the particulars of the case. Certain contexts will feature typical sources of error, such as memory-related biases in stressful situations. Or there may, for instance, be some indication that the witness is lying, which could make us consider this as a possibility. For instance, when interviewing suspects there are various strategies that the suspect may employ to hide the truth, such as lying, twisting the truth or feigning memory loss (Clemens & Grolig, 2019; Rispens, 2021). Depending on how the suspect behaves during the interview, the hypothesis that the witness is using such a strategy can become an explanation that we should consider.

On TIBE we compare two kinds of explanations: Explanation (i) would, if it were true, also imply the probable truth of the fact that the witness is reporting on. This is therefore a truth-telling explanation. Explanations (ii), (iii) and (iv) would, if they were true, not imply that the hypothesis that the witness is reporting on is true. I call these alternative explanations.

On TIBE, if the best explanation implies the truth of what is reported on, we may assign the report a high degree of reliability with respect. Conversely, the greater the number of plausible alternative explanations, and the more plausible these are, the lower the degree of reliability that we should assign to the report. As I will argue later, the fact that it specifies a method of drawing conclusions about the witness’s reliability is one of the main benefits of TIBE compared to alternative frameworks (see section 4). Here are some examples of the kind of explanatory

---

142 As we will see in sections 4 and 6, TIBE’s open-endedness with respect to the explanations that we consider sets it apart from alternative prescriptive approaches and therefore allows for more flexibility.

143 Interviewers may then use strategic interrogative methods to exclude the possibility of the suspect using such methods – e.g., strategically hiding information (Oleszkiewicz & Watson, 2020).

144 Some might object that this is not inference to the best explanation (e.g., Laudan, 2007). After all, we do not infer that the best explanation is true. Instead, we compare the disjunction of available truth-telling explanations to the disjunction of available alternative explanations. However, I believe that we should not be too nitpicky about our use of this term. What I propose may not be IBE in a strict sense, but it is nonetheless explanatory reasoning. To call such reasoning IBE is not uncommon (e.g., Pardo & Allen, 2008).
inferences that I have in mind.

“This testimony comes from a credible source because it makes sense that the witness actually saw what happened and that they are willing to tell us. There are no alternative explanations for their utterance that make sense in this context.’

“This testimony comes from somewhat credible source, because while this statement is well-explained by the witness being truthful, it is also plausible that they are misremembering.’

“This testimony does not come from a credible source, there are various alternative explanations why the report might say this. For instance, it is plausible that the witness is lying.’

The focal point of eyewitness evaluation on this account is therefore the generation and subsequent evaluation of potential explanations for the witness’s utterance. For instance, suppose that, at trial, the prosecution wants to argue for the reliability of a particular witness with respect to their statement that “the robber wore a red shirt”. They could do so by first putting forward a scenario in which the robber wore such a shirt and the witness accurately observed this fact and then arguing that this scenario is plausible in the given case. The defense might then respond by offering an alternative explanation of the witness’s utterance, for instance by arguing that it is plausible that they are lying. It is then up to the decision-maker (i.e., the judge or jury) to determine whether the parties have succeeded in showing that these respective are plausible in the given case.

As I wrote in previous chapters, much ink has been spilled by both philosophers of science and law on what makes an explanation good (or ‘plausible’, or ‘lovely’). In a general sense, an explanation concerning a witness’s statement is good to the extent that – if it were true – it would help us understand what caused the witness to offer this particular story. When does an explanation provide such potential understanding? As explained in chapter II, explanationists are often vague about this
question. I will not offer a precise account of explanatory goodness either. For my current purposes it is sufficient to note that an explanation’s goodness mostly depends on the extent to which it is internally coherent and fits well with the available evidence and our background beliefs about the world. To begin with the latter, our background beliefs concern “the way things normally happen” (Walton, 2007, 128). An example relating to testimony is that ‘humans generally cannot hear soft sounds from a great distance.’ These folk-psychological generalizations can (and should) be partially based upon the extensive research on eyewitness reliability. The relevant evidence may include facts about the witness (e.g., was their eyesight good?) and about the situation (e.g., did it rain?). Furthermore, the content of the witness’s testimony may be relevant. For instance, if someone offers a detailed description of an event, but we also know that they were watching this event from a large distance, then the explanation that they are telling the truth becomes less plausible. In general, a witness who makes a surprising claim should – all else being equal – be regarded with more suspicion than one who reports on run-of-the-mill facts. So, even though TIBE is about the reliability of the witness, rather than the credibility of their claim, judgments about the two may go hand-in-hand.

Not every instance of testimony calls for inference to the best explanation. Thagard (2005) suggests that, in daily life, people do and should accept the testimony of others by default, unless the content of this testimony conflicts with their beliefs. If that is the case, they enter a ‘reflective pathway’ where they evaluate the witness’s claim through inference to the best explanation. As Thagard (2005, 299) points out, such a strategy is epistemically useful, because while we do not want to uncritically accept everything that anyone tells us, if we’d carefully reflect on everything that we’re told, we would be greatly restricted in acquiring new beliefs. I want to suggest something similar for criminal cases. In such cases there is arguably also a tendency by jurors and judges to believe witnesses by default (e.g., Brigham & Bothwell, 2014).

145 For reviews of this research, see, for example, Dubelaar, (2014); Ross et al., (2014); Wise et al., (2014).
In contrast with daily life, this default trust is regularly unwarranted, leading to judicial errors. So, there are many situations in which critical evaluation is needed and where we should engage in TIBE. Nonetheless, this does not mean that eyewitness testimony should always be subjected to scrutiny. Witnesses may report on mundane facts that neither the defense nor the prosecution seek to challenge. In such instances, careful explanatory reasoning about the witness’s reliability would only obstruct the efficiency of the trial. Fact-finders may instead reserve their scrutiny for contentious, surprising statements, and for types of statements where eyewitness errors are common (e.g., an identification of the perpetrator of a violent crime by a bystander).

Neither Thagard, nor anyone else has offered arguments why TIBE is the preferred approach once we enter this ‘reflective pathway’. Yet explanatory reasoning has numerous benefits. For instance, a well-known advantage of thinking in terms of competing explanations is that it may counteract confirmation bias by preventing fact-finders from becoming overly focused on a single possibility (O’Brien, 2009). Confirmation bias also poses a danger when interpreting eyewitness testimony. As mentioned, fact-finders will often trust eyewitness testimony too easily, ignoring any evidence for unreliability. At other times, they may overly focus on evidence contra reliability, leading to a deflated assessment of the witness’s reliability (Pudifoot, 2020). Rassin (1991) suggests that eyewitness evaluators may avoid confirmation bias by considering competing explanations for an eyewitness report. This forces evaluators to explicitly think about both the possibility that the witness is reliable and the possibility that they are not.\footnote{146 Though my account was developed without knowledge of Rassin’s work, his proposal is similar to mine. He also suggests that evaluators should compare competing explanations of a report. His account is aimed at the Dutch legal context and intended to be used by expert eyewitness evaluators. Rassin (1991) writes that his approach and similar approaches have successfully been used by experts in practice. He also notes that such explanatory comparison may lessen the degree to which people tend to draw extreme, unwarranted conclusions (e.g., ‘this witness is completely (un)reliable’). According to}
Rassin, explanatory comparison instead leads to more cautious conclusions such as ‘some alternative explanations need to be further investigated’ or ‘there are no plausible alternatives’. I discuss various other benefits of explanatory thinking throughout this chapter, including in the following section.

2.2 Storytelling about testimony

When we think about why an eyewitness offers their testimony, the explanations that we take into account may be general. However, such general explanations (e.g., ‘the witness is lying’) will often be difficult to evaluate as they lack detail. So, ideally the explanations that we consider should be more specific stories. For instance, Walton (2007, 109-10) has us imagine a situation in which the testimony of two police officers confirms a suspect’s alibi. He writes: “[T]hough it may be conceded that generally police officers in the line of duty do not lie, there may be evidence in this specific case showing that in fact these two police officers did lie. This could be shown by means of a (…) narrative showing the goals of the police officers and the other physical and psychological conditions of the case.”

I am using the word story in a broad sense here. As discussed in section 3 of chapter II, words like ‘narrative’, ‘story’ or ‘scenario’ are typically used in the context of the story model of evidential reasoning. On this model, stories have a specific structure: they contain a central action (e.g., a person killing someone else) and describe a context that make this action understandable. Other elements of a complete story are a description of the scene, a motive, a central actor and resulting consequences. When we describe why a person offers a piece of testimony, our explanation may have such a story structure. However, it may also lack some of the elements of such a structure. Nonetheless, even if a detailed account of some testimony does not have the aforementioned structure, we are – colloquially – telling a narrative about why this person offers this piece of testimony. For example, suppose that we consider the explanation that the witness misremembered. We could then further
specify when and why the distortion in their memory likely occurred by considering the particulars of this specific witness and the context in which they observed the reported events. This leads to a story of why it is plausible that this witness misremembered these events.

As has often been noted, by constructing, critiquing and comparing competing narratives, fact-finders in criminal cases can structure their reasoning about the available evidence (Walton, 2007). Similarly, thinking in terms of competing narratives that explain a witness’s testimony can be of enormous help to evaluators. In particular, the more specific our explanations are, the more specific our expectations become about the kinds of evidence that we would, or would not expect to encounter. For example, it may be difficult to determine how we might falsify or verify the explanation that ‘the witness is lying’. In contrast, suppose that our explanation is that ‘the witness is lying about being at home when the crime took place, because they were actually at the crime scene and are trying to cover up this fact.’ Such a specific explanation will generate predictions that we could test, for example, that we would likely find trace evidence left behind by the witness at the crime scene or that certain other witnesses would likely contradict the witness’s alibi.147

The process of making our explanations more detailed will typically begin by considering what a given explanation would plausibly entail in the given circumstances if it were true. For instance, suppose that these two police officers lied, what would be the most likely reason why they did so? How might they try and hide the fact that they are lying? What are typical tell-tale signs of liars? As Walton (2007, 128) points out, this process of filling in the missing bits is itself an instance of inference to the best explanation. Though there are countless ways of filling in the gaps in a story, we choose the version that is the most plausible in the given context.

147 That it helps us select, structure and interpret the available evidence is a key benefit of TIBE compared to existing frameworks. See sections 4-6.
Truth-telling explanations can also be made more specific. We may start with the general explanation that ‘the witness is accurately reporting on the events that took place’. This can then be turned into a full-fledged story by filling in the details of where the witness was standing when the event took place, how much time they had to observe the event, and so on. During a trial, one of the parties may argue that the witness is reliable by putting forward such a story. The opposing party may then cast doubt on the witness’s reliability by attacking this story. For instance, they can ask critical questions; or, argue that the story does not fit with the facts of the case or with certain background beliefs. They may also sow doubt about the witness’s reliability by presenting a plausible alternative story in which the witness does not speak the truth. It is then up to the decision-maker to decide which story best explains the witness’s utterance. To give an impression of what such an exchange may look like in practice, consider the following example:

A defendant is accused of smuggling drugs by two witnesses. In exchange for getting their sentences for another crime reduced, these witnesses testify that they smuggled drugs across the border together with the defendant many times. The prosecution argues that the witnesses should be believed, as their testimonies cohere with one another and with the other facts of the case. The defense, in contrast, argues that the witnesses are not credible. They sketch an alternative scenario in which the witnesses lied. In this scenario, the agreement between the testimonies is explained as being the result of the witnesses fitting their stories to each other and to the case file. The defense also points out that the testimonies agree with one another to a suspicious degree; they are very detailed given that the events took place years ago – indicating that they were fabricated. The prosecution tries to show that this alternative scenario is not plausible, by arguing that the witnesses also reported on verifiable details that were not in the case file and that their stories also differed in places. The defense, in turn, counters this argument
by showing that these verifiable details were easy to guess, even if the witnesses were lying.\footnote{This example is a highly simplified version of the arguments offered during a criminal trial which took place in the Netherlands in 2015. Hoge Raad. (29 september 2015). ECLI:NL:HR:2015:2842. I take this example from Stevens (2016).}

We can imagine such an exchange continuing for quite a while, with the prosecution trying to show that the alternative scenario is implausible, while the defense tries to argue that it is plausible. As the exchange goes on, the scenario can become increasingly specific as the parties drill down on the specific details that they believe make the explanation of why these witnesses offer their testimony (im)plausible. These can be facts from the case but also, for instance, facts from psychological research on how well people tend to remember specific details of events after a given amount of time. Additionally, the interpretation of available facts can also change. For instance, in this example, the agreement between the witness’s testimonies is used by the prosecution to argue for the reliability of the witnesses, while the defense uses it as an argument for their unreliability. As I will discuss in section 5 and 6, TIBE explains not only how facts may become relevant, but also the phenomenon that facts can be interpreted in competing ways. After the exchange is completed, the fact-finder can decide how plausible each of the competing scenarios is and assign a degree of reliability to the witnesses accordingly.

3. Being friends with Bayesians

Having laid out the basics of TIBE, I now want to turn to its rationality. My account is intended to be prescriptive. It offers evaluators guidance on how to reason in practice about the reliability of a particular eyewitness (e.g., what kind of questions they should ask and how they should answer those questions). We may contrast such an account with descriptive frameworks, which describe how people evaluate eyewitness reliability in practice. Existing accounts of TIBE are mostly of the descriptive
kind: they are intended to capture our intuitions about why we usually, but not always, trust the utterances of others in daily life (Lipton, 2007; Malmgren, 2008; Fricker, 2017). Another type of account is normative. Normative accounts of eyewitness reliability formulate a standard for epistemically rational evaluation. A good prescriptive account helps real-life fact-finders be (more) rational in their evaluations. Such an account therefore lies at the intersection of well-grounded descriptive and normative accounts: it should fit with how people normally reason\(^{149}\), but there should also be good normative reasons to believe that the approach helps us be more accurate.

Little has been said about TIBE’s normative basis. In this section I want to make some headway on this matter by connecting my account to Bayesianism. Apart from being arguably the dominant account of epistemic rationality and legal proof, Bayesianism’s influence also extends the study of witness reliability (Merdes, von Sydow & Hahn, 2020). I focus on the work of Goldman (1999). Goldman is neither the first nor the last person to offer a Bayesian analysis of witness reliability. However, he is, to my knowledge, the author who most extensively discusses why Bayesianism provides the correct normative standard for thinking about this topic. I argue that Goldman’s work also offers a basis for TIBE’s rationality. In particular, I follow a line of thinking from the philosophy of science which casts IBE as a useful heuristic for optimal Bayesian reasoning.\(^{150}\) On my view, Bayesianism describes the goal at which we aim when we engage in rational eyewitness evaluation. reliability – namely to evaluate the likelihood ratio of the testimony in the light of the available evidence. TIBE is rational to the extent that it tracks the Bayesian ideal. However, I also argue that it is not by itself an adequate account of rational eyewitness

\(^{149}\) Though I will not defend TIBE’s descriptive adequacy here, it is at least plausible that it is adequate in this sense. First, it is well-known that how people tend to make sense of criminal evidence through explanatory reasoning (e.g., Pennington & Hastie, 1993). Second, the aforementioned descriptive accounts of TIBE have (to my knowledge) not been met with much resistance.

\(^{150}\) I therefore echo section 4 of chapter II, where I suggested that explanation-based thinking in general can be a heuristic for Bayesianism.
evaluation because it does not tell evaluators much about how to draw conclusions from the available evidence about the witness’s reliability. I will argue that for this we should turn to TIBE.

3.1 Bayesian rationality

Bayesianism is first and foremost an account of how we should change our beliefs upon receiving new evidence. How much receiving a witness report heightens (or lowers) our prior belief in H depends on the likelihood ratio of the testimony:

\[
\frac{\text{Probability(Witness report | H)}}{\text{Probability(Witness report | ¬H)}}
\]

In other words, how much would we expect the witness to give this report, if H is true, and how much would we expect the witness to offer this testimony, if H were false? For instance, is it likely that the witness would say that the robber wore a red shirt if the robber wore such a shirt? What about if he did not wear such a shirt, how likely would it then be that he said this? The higher this ratio of true to false positives, the more reliable we believe the witness to be.151

The likelihood ratio is central to how the Bayesian account expresses eyewitness reliability but there are different ways to interpret the probabilities in this ratio. As explained in chapter II, possible interpretations can be categorized into ‘physical’ (i.e., ‘probability is in the world’) and ‘epistemic’ (i.e., ‘probability is in the head’) (Redmayne, 2003). Which interpretation we choose matters a great deal both for how we should assess this ratio and for Bayesianism’s normative status. In

\[151\text{ Note that, for the sake of simplicity, I presume that the hypothesis H under consideration is the same as the hypothesis that the witness reports on. This is not necessarily the case. For instance, we might consider the evidential impact of the witness's testimony about the colour of the robber’s shirt on the hypothesis that ‘Edward is the robber’.} \]
particular, Goldman (1999, 116-33) offers a mathematical proof which shows that, if our subjective likelihoods of a report match its ‘objective’ likelihoods (to use Goldman’s terminology), then Bayesian updating will always bring our posterior degree of belief in the hypothesis that the witness is reporting on closer to the truth compared to the prior. This result is part of a broader class of veritistic arguments for Bayesianism which purport to show that, if we update our beliefs using accurate likelihoods, our beliefs will (eventually) converge on the truth (Hawthorne, 1994).152 An account of rational eyewitness evaluation that can draw on such arguments has a strong claim to a normative status.

The central difficulty with this argument is determining what the term ‘objective likelihood’ means (Fallis, 2002). There is good reason to doubt that it is even possible to give an adequate account of this notion with respect to eyewitness reliability. For instance, the most well-defined objective interpretation is the frequentist account, where we measure probability as occurrence in a specified population of events. A typical example of this type of probability is the frequency with which a coin lands on heads out of $n$ tosses. Some authors use this interpretation with respect to witness testimony. They suggest that a reliable eyewitness is one that has a tendency to speak the truth on the relevant topic. This tendency could be determined by looking at the witness’s track record – i.e., the number of times the witness has spoken truths versus falsehoods about this topic. However, as Thagard (2005) argues, this view is deeply problematic. First, we almost never have reliable records of a person’s track record. Second, this view leads to the wrong conclusions. For instance, we should not trust a person completely, no matter what other evidence we have, just because they have so far been accurate about some topic.153 Finally,

152 See also chapter II, section 2.3 for a discussion of these arguments.
153 We also cannot save frequentism by interpreting the relevant probabilities as the tendency of witnesses in general to speak the truth under similar circumstances. Though the research on eyewitness reliability is extensive, we lack the required frequency data on eyewitness reliability in such specific sets of circumstances. Additionally, this suggestion faces the well-known reference class problem. For any individual witness there are different ways to
as Coady (1992, pp. 210-11) argues, this interpretation relies on the mistaken view that “[people] have quite general tendencies to lie, whatever the context or subject matter, [and] to make mistakes in abstraction from particular circumstances.” Goldman himself (1999, 117) also rejects such a frequentist interpretation. He goes on to argue that objective probabilities concerning eyewitness reliability nonetheless exist (Goldman, 1999, 118):

[I]n testimony cases it looks as if jurors, for example, work hard at trying to get accurate estimates of such probabilities, which seems to presume objective facts concerning such probabilities. If the witness in fact has very strong incentives to lie about X, this seems to make it objectively quite probable that she would testify to X even if it were false. If the witness has no such incentives, nor any disabilities of perception, memory, or articulation [...] then the objective probability of her testifying to X even if it were false seems to be much lower.

What Goldman seems to be talking about is an evidential probability interpretation, where the probability of a hypothesis is the degree to which our evidence justifies us in believing that hypothesis. Various authors consider evidential interpretation to be the most suitable ‘objective’ interpretation for the context of legal evidence define what counts as a witness under ‘similar circumstances’, without any particular reference class having a claim to being the ‘correct’ one.

Thagard (2005, 308), who defends TIBE, appears to fall into this trap. He suggests that assessing a witness’s credibility “requires a judgment about a person’s disposition to tell the truth on a particular topic.” He illustrates this point using the example of a police officer in the infamous O.J. Simpson case who was accused of being racist and whose testimony was not believed by the jury. According to Thagard, the jury inferred that the police officer had a disposition not to tell the truth in cases where a black person was accused of a crime and decided not to trust him on that basis. But even if we grant that this inference about the witness’s disposition best explains the jury’s decision in that particular instance, this point does not generalize to all instances of eyewitness evaluation. There may also be cases in which a generally reliable witness, who usually tells the truth, has to observe something in poor conditions. It would be strange to say that this person has a disposition to be mistaken.

I discussed this interpretation in-depth in section 2.4 of chapter II.
(Nance, 2016, 47-8; Wittlin, 2019; Spottswood, 2019; Hedden & Colyvan, 2019). However, as argued in chapter II, an objection to the evidential interpretation is that it is vague. Goldman does not spell out what it means for evidence to make something objectively probable and it may actually be impossible to offer a clear definition. For instance, Redmayne (2003) surveys various ideas that Goldman could draw upon for more precision, but concludes that they would all lead to interpretations that are either too limited in scope to capture the evidential richness that we want, or so vague that they are not much more informative than Goldman’s own brief description above. According to Williamson (2002, 212), one of the most prolific proponents of the evidentialist interpretation, such vagueness need not be lethal as long as our concept is clear enough to get ahead with the matters at hand. Goldman (1999, 118) makes a similar suggestion with respect to probabilistic judgments of eyewitness reliability. He points out that we can often intelligibly ask: Does the available evidence make it probable that this witness would testify that X, if X were true? And does the available evidence make it probable that this witness would testify that X, if X were false?

I agree that we can sometimes straightforwardly determine the reliability of a witness. Furthermore, it is easy to imagine situations where a specific evaluator’s assessment of an eyewitness’s reliability will clearly not be reasonable given the evidence. For instance, consider two witnesses. The first is well-known to be honest. They have good vision and an excellent memory. When the crime occurred, they had ample time to observe the perpetrator. Anyone who assigns such a witness a low degree of reliability (without offering further, convincing reasons) is patently irrational, because their belief goes against the evidence. In contrast, consider a witness who is a compulsive liar, who has a motive to lie in this particular instance and whose testimony conflicts with much of what is known. It would be equally irrational for anyone to ascribe a high likelihood ratio to this person. In this way, evidential Bayesianism constrains our evaluations of eyewitness reliability to some agree.
However, not every instance of eyewitness evaluation is straightforward. In the examples that Goldman gives, determining how the evidence influences the eyewitness’s reliability is a matter of applying commonsense generalizations to the facts about the witness. An example of such a generalization is that having an incentive to lie clearly makes one more prone to report falsehoods. It is only common sense that discovering that someone has such an incentive should lower our assessment of their credibility. But as I will set out in detail in sections 5 and 6 of this chapter, things are not always so simple. Evaluators often face the difficult tasks of deciding both which evidence to consider and which generalization to apply to each item of evidence. Bayesianism does not tell us how to go about these tasks. If we adopt the evidential probability interpretation of the likelihood ratio then this is a glaring gap. The slogan ‘look at the evidence’ is not a very informative statement for eyewitness evaluators when it is unclear what the evidence is or in what light we ought to see it. So, the Bayesian account is not a very informative theory of rational eyewitness evaluation by itself.

This then finally brings us back to TIBE. I propose that the Bayesian account clarifies the aim of rational eyewitness evaluators, namely to determine the likelihood ratio. However, it does not give a method for how to achieve this aim. What is needed is a prescriptive account that tells evaluators how to make sense of the available facts about the eyewitness’s reliability. I argue that TIBE succeeds as such an account on two fronts. First, it helps with the two aforementioned tasks: selecting the relevant evidence and interpreting this evidence. Second, as I will now argue, TIBE’s conclusions quite straightforwardly track the Bayesian likelihood ratio, meaning that TIBE leads to rational conclusions from a Bayesian viewpoint. It therefore has a plausible claim to being rational (or, at the very least to being no less rational than Bayesianism). Let us now turn to how TIBE tracks the likelihood ratio.
3.2 Heuristic compatibilism

Though not much has been said about how explanation-based reasoning and Bayesianism relate in the context of eyewitness evaluation, various suggestions about this relationship have been offered in other contexts, especially with respect to scientific theory choice. As I discussed in chapter II, one well-known strand of thought casts inference to the best explanation as a heuristic to approximate correct Bayesian reasoning (Okasha, 2000; Lipton, 2004; McGrew, 2003; Dellsén, 2018). A heuristic is a method of reasoning that is efficient and tends to lead us to the approximately right answer. On this view, IBE complements Bayesianism by offering a rule of inference that is appropriate for non-ideal agents, yet enables these agents to approximate the probabilities that Bayesian reasoning would have them assign to hypotheses (Dellsén, 2018). I suggest that, similarly, TIBE is a heuristic for approximating the Bayesian likelihood ratio. In what follows, I give an argument for why the approach usually tracks this ratio but makes sacrifices with respect to accuracy, for the sake of efficiency and respecting human cognitive limitations. In particular, rather than consider the entire probability space, this approach tells us to evaluate and compare a limited number of salient, well-specified explanations, and to ignore other possibilities.\footnote{This contrast between explanation-based and Bayesian approaches has also been noted by Allen & Pardo (2019, 12).}

To make it more precise under which conditions TIBE tracks the likelihood ratio and when it diverges from this ratio, let us look at recent, Bayesian Network models of eyewitness reliability.\footnote{Such approaches comprise the bulk of contemporary Bayesian work on witness reliability. These networks are used to explore the rationality of what sorts of changes in evidence increase or decrease the credibility of a witness and their statement and to what extent (see Merdes, Von Sydow & Hahn, 2020 for a review).} These approaches start from the idea that the likelihood ratio leaves much information about the reliability of the witness implicit, whereas we might want to represent this information explicitly. For instance, we may want to model the impact of receiving evidence concerning the witness’s reliability. In
these models, the reliability of the witness is therefore represented as a distinct variable, rather being encapsulated in the likelihood ratio. This variable is called ‘reliability node’, usually denoted as REL, which expresses the hypothesis (HYP) that the witness’s report (REP) is accurate (Bovens & Hartmann, 2003, 57; Lagnado, Fenton & Neil, 2013; Merdes, Von Sydow & Hahn, 2020). The idiom looks as follows:

![Diagram of evidence reliability idiom]

*Figure 5: The evidence reliability idiom*

Once we include this variable, we end up with the following formula for calculating the likelihood ratio:

\[
P(\text{REP} | \text{H}) = \frac{P(\text{REP} | \text{H} \& \text{REL})P(\text{REL}) + P(\text{REP} | \text{H} \& \neg \text{REL})(\neg \text{REL})}{P(\text{REP} | \neg \text{H})} = \frac{P(\text{REP} | \neg \text{H} \& \text{REL})P(\text{REL}) + P(\text{REP} | \neg \text{H} \& \neg \text{REL})P(\neg \text{REL})}{P(\text{REP} | \neg \text{H})}
\]

According to Lagnado, Fenton & Neil (2013, 52), by including REL, we make it explicit that there can be “alternative possible causes of a [statement other than it]
being truthful”. In this chapter, I also emphasize the usefulness of explicitly reasoning about competing possible causes of a testimony. Remarks like these therefore highlight the natural fit between Bayesianism and explanation-based reasoning.

As said, TIBE does not always perfectly track the likelihood ratio. It may deviate from this Bayesian formula due to the limited number of explanations that we consider. The above formula helps make it more precise when such deviation occurs. The first limitation on the explanations that we consider follows from the fact that, on TIBE, we want to draw a conclusion about whether the witness is reliable with respect to the facts that they report on. This means that hypothesis H is assumed to be the same as (part of) the content of the witness’s report. For instance, if the witness reports that “the robber wore a red shirt”, then hypothesis H is ‘the robber wore a red shirt’. Though the formula above does not dictate this, it is an assumption that is made by many of the Bayesian authors as well. It is also a useful assumption, because it leads to a simplification of the formula. First, it means that we can set the value of $P(\text{REP} | \text{REL} \& \neg H)$ to 0. On the accounts of Bovens & Hartmann (2003) and Lagnado, Fenton & Neil (2013) a reliable witness is a truth-teller. It is therefore, for instance, impossible that (a) the witness reports that the robber wore a red shirt, (b) that this statement is false, and (c) that the witness is reliable. In addition to this, I propose that, on TIBE, we may also ignore $P(\text{REP} | \neg \text{REL} \& H)$. Admittedly, it is possible for an unreliable witness to accurately report on the true state of the hypothesis. For instance, Bovens & Hartmann (2003) and Lagnado, Fenton & Neil (2013) presume that a fully unreliable witness is a randomizer. Such a randomizer could accidentally report the truth. For example, the witness could be a liar who makes a random statement about the colour of the robber’s shirt and happens to pick the right one. However, such epistemic luck will often (though, admittedly, not always) be highly unlikely and not part of the most salient causes.

---

158 The authors further divide the reliability node up into more specific causes: sensitivity, objectivity and veracity. See section 4.1 of this chapter for more on these three aspects.
If we ignore the aforementioned terms, we can simplify the formula for calculating the likelihood ratio as follows:

\[
\frac{P(\text{REP} \mid H)}{P(\text{REP} \mid \neg H)} = \frac{P(\text{REP} \mid H \& \text{REL}) \cdot P(\text{REL})}{P(\text{REP} \mid \neg H \& \neg \text{REL}) \cdot P(\neg \text{REL})}
\]

My proposal is that TIBE maps onto this simplified formula. To be precise, the disjunction of our truth-telling explanations maps onto the numerator, whereas the disjunction of the alternative explanations maps onto the denominator. Let us focus on the truth-telling explanations first. A truth-telling explanation is a narrative in which the relevant hypothesis is true and the witness accurately reports on it. In other words, this is a situation where both H and REL are true and the witness testifies that ‘H’.

To offer a plausible truth-telling explanation means to offer a *ceteris paribus* reason to assume that both \(P(\text{REP} \mid H \& \text{REL})\) and \(P(\text{REL})\) have a high value. First, as various philosophers of science have noted, all other things being equal, when one hypothesis explains a fact better than a competitor, then the fact is also more likely to occur given this hypothesis than given its alternative (McGrew, 2003; Henderson, 2014). To give a good explanation of a fact is – at least in part – to show that this fact is expected if the explanation is true. Similarly, a narrative that explains the witness’s utterance well will also make this utterance likely. In other words, if we were to presume that a plausible truth-telling explanation is true, then we may assign a high value to \(P(\text{REP} \mid H \& \text{REL})\). However, this is not all there is to

---

159 Friedman (1986, 688), a proponent of the Bayesian account of eyewitness evaluation has similarly suggested that we may ignore such unlikely possibilities. He suggests a situation where a witness who misperceives an event, then misremembers what they misperceived, and ends up telling the truth as an example of a hypothesis that we may safely ignore.
plausibility. Consider the truth-telling explanation that ‘a powerful, all-knowing extraterrestrial mind-controlled the witness so that this witness offered a perfectly accurate statement.’ If we presume that this explanation is true, then obviously this makes it likely that the witness would be accurate in their report. However, such an explanation is also highly implausible, as it conflicts with background beliefs that most of us hold about the world. It is highly unlikely that this actually happened. Conversely, if a scenario in which the witness is reliable is plausible then this is a *ceteris paribus* reason to believe that this scenario actually occurred. To offer such a plausible explanation therefore means to offer a reason to believe that this witness was indeed reliable in this way. For instance, we may sketch a plausible situation in which this particular witness could have accurately observed these particular events, which gives us a *ceteris paribus* reason to presume that this is the correct explanation of the witness’s testimony. Hence, to show that a truth-telling explanation is true also means to show that the value of $P(\text{REL})$ is *ceteris paribus* high.

The more plausible our truth-telling explanations are, and the more of these explanations we have, the higher the value is that we may therefore assign to $P(\text{REP} \mid H \& \text{REL})(\text{REL})$. Similarly, the disjunction of alternative explanations tracks the denominator of the formula. An alternative explanation is a narrative in which the relevant hypothesis is false, but the witness nonetheless reports that it is true because they are unreliable. They might, for example, misremember. To offer plausible alternative explanations is therefore to offer reasons to assign a high value to both terms in the denominator, $P(\text{REP} \mid \neg H \& \neg \text{REL})$ and $P(\neg \text{REL})$. The argument for this is the same as for the numerator. Conversely, to argue that there are no plausible alternative explanations is to argue that we may assign a low value to this part of the formula.

Our explanatory reasoning may however, diverge from optimal Bayesian reasoning. The key feature of TIBE is that the truth-telling and alternative explanations that we consider are specific and small in number. For instance, the prosecution may offer one version of why this witness is telling us what they are and the defense
another. However, this means that TIBE diverges from optimal Bayesian reasoning when we overlook other plausible explanations. In other words, TIBE faces its own version of the well-known bad lot problem for IBE.\footnote{See chapters III and VII for a more in-depth discussion of the bad lot problem.} Briefly put, the bad lot problem is that we are not justified in concluding that the best explanation is (probably) true if we are insufficiently certain that the true explanation is among those that we’ve considered. As I explain in chapter VII, both IBE and Bayesian inference face this problem. Nonetheless, this does not take away that the fewer explanations that we consider and the more specific these explanations are, the bigger the risk that we overlook plausible alternatives. This is a strength of TIBE, because such specific explanations are often easier to evaluate than general ones and because it is unrealistic to presume that evaluators can consider every single possibility. However, it is also a weakness, as it limits the scope of inquiry, thereby making it more likely that we overlook other plausible explanations. When we miss such alternatives, we misjudge the relevant likelihoods. The evaluator must therefore engage in a balancing act between efficiency and accuracy.\footnote{There are specific aspects of criminal cases that arguably make it more likely that we overlook plausible alternatives, see chapter VII. In particular, the likelihood that we overlook such relevant explanations depends not only on the specificity of our conceived alternatives, but also on how good those alternatives are. Additionally, the quality of our investigation and our set of evidence, as well as the complexity of the case at hand also matter. To what extent TIBE helps us accurately assess the likelihood ratio will depend in part on these factors.}

To sum up the above, I have sketched a novel, feasible way in which testimonial inference is compatible with Bayesian accounts of eyewitness evaluation. On my proposal, we aim for a Bayesian norm, but we do so through the efficient but imperfect heuristic of explanation-based reasoning. The upshot of this proposal is that we may use arguments for the normativity of Bayes as a foundation for the rationality of TIBE.\footnote{In chapter II, section 4, I proposed that we may, more generally, view inference to the best explanation in as a heuristic for the Bayesian inference.} I discussed Goldman’s (1999) argument above, which I view as the most prominent of these arguments. However, as argued above, this does not
mean that TIBE is subsumed under Bayesianism. What we want to know from a theory of rational eyewitness evaluation is how we ought to determine what conclusions the evidence supports, which explanatory reasoning helps us do. It is this aspect of TIBE that I discuss in the remaining sections.

4. Comparing prescriptive accounts

Having spelled out TIBE’s normative basis, I now turn to how this account compares to existing work. In particular, I want to juxtapose TIBE with two well-known prescriptive approaches to eyewitness evaluation: (i) the capacity approach, which has been defended by a number of epistemologists, and (ii) empirically-informed methods developed by psychologists. This discussion will help explain what is distinctive about TIBE, but also how it fits with this existing work. In particular, I want to suggest that TIBE can be a helpful addition to these existing accounts and, conversely, that ideas from these existing frameworks can help fill in some of the details of TIBE.

4.1 The capacity approach

The first account that I discuss is what I call the ‘capacity approach’. This is arguably the best-known prescriptive account of eyewitness evaluation in the philosophical literature. On the capacity approach, we evaluate eyewitness reliability by considering whether the witness’s statement came about through an adequate exercise of certain capacities. A wide range of scholars adopt this approach, including a number of Bayesians, who use it to make reasoning about the evidential impact of testimony.

---

163 Lipton (2007, 246-8) similarly contrasts TIBE with a number of alternative accounts from the philosophy of testimony, such as evaluating the witness’s reliability based on their reputation, in order to show that it is a non-trivial approach. However, as far as I am aware, none of the approaches that he discusses are used in contemporary legal practice. In contrast, the accounts that I describe here do find such use.
more tractable (Friedman, 1986; Schum, 1994; Goldman, 1999; Lagnado, Fenton & Neil, 2013). There is some difference in the capacities that various authors distinguish. Schum (1994) gives the most commonly used list, on which an eyewitness’s reliability depends on whether they were (a) observationally sensitive, i.e., their senses functioned correctly, (b) objective, i.e., their memory aligns with what they perceived, and (c) veracious, i.e., they truthfully report what they believe. So, when we assess the likelihood ratio of the witness, we should ask critical questions such as:

1. Is the witness sincere?
2. Did the witness’s memory function properly?
3. Did the witness’s senses function properly?

(Walton, Reed & Macagno, 2008).

The capacity approach can be seen as a kind of inference to the best explanation, where we consider multiple alternative explanations of the witness’s utterance: lying, misremembering, misperceiving. This way of looking at the capacities fits with how various of its proponents present the approach. For instance, Lagnado, Fenton & Neil (2013) utilize the capacity approach in their Bayesian network-based modeling of eyewitness evidence. As said earlier, in their model, eyewitness reliability is expressed through a ‘reliability node’ that represents the “alternative possible causes of a [statement]” (52). By ‘alternative’, they mean other explanations than ‘the statement is true’. They then suggest splitting up this reliability node into separate sincerity, objectivity and sensitivity nodes, because “there are several different ways in which a source of evidence can be unreliable” (56). In other words, they have us consider several alternative explanations of a statement. Similarly, one of their predecessors, Friedman (1986, 668), uses a Bayesian Network-like approach

164 See section 2.2 of chapter II for an explanation of Bayesian networks.
in which the likelihood ratio of the testimony is evaluated by considering “chains of circumstances that might have led to a given declaration.”

Despite the similarity between this account and my account, they differ in at least one respect. On the capacity approach, the aforementioned critical questions are the focal point of eyewitness evaluation. In contrast, on my account, the alternative explanations that we consider are not fixed in number and in generality. Which explanations we consider and how specific these explanations are, depends on the context (see section 2). Furthermore, these explanations do not have to pertain to the capacities of the witness. For instance, someone might consider the possibility that they misinterpreted the witness’s statements. TIBE therefore fits with the idea behind the capacity-approach, but is a more flexible, context-sensitive way of spelling out this underlying idea.

This more flexible approach has both benefits and drawbacks. I will illustrate the benefits of this flexibility in depth in sections 5 and 6. But, briefly put, one downside of the capacity approach is that it is not very informative about how evaluators ought to interpret evidence about the reliability of the witness in a context-sensitive way. A second downside is that the approach can be both overly and underly inclusive in the kinds of evidence and explanations that it has evaluators consider. By interpreting the capacity approach as an instance of TIBE, we avoid these difficulties.

The capacity approach also has its benefits. An important one is that there is strength in simplicity. It gives evaluators a clear set of questions to ask, whereas TIBE makes them do more work in terms of generating and specifying the available explanations. The explanations that we consider on the capacity approach can therefore be a good starting point for TIBE, after which we then make them more specific and ask whether there are also alternative explanations.
4.2 Empirically informed evaluation methods

The second type of evaluation method with which I compare my account is one developed within psychology. Psychologists have extensively studied eyewitness reliability, distinguishing numerous aspects of events, environments and the witnesses themselves that can (positively or negatively) influence eyewitness reliability (Ross, Tredoux & Malpass, 2014; Wise et al., 2014). This has led to a wide range of proposals on how to minimize eyewitness errors, such as better ways to design interviews. Psychologists have also tackled the question how these empirical insights can be used to improve eyewitness evaluation by developing prescriptive, empirically-informed methods that experts can use to assess eyewitness reliability (Griesel & Yuille, 2017). It would go beyond the scope of this chapter to review all of them. Let me just give two examples of influential approaches to illustrate how these approaches differ from, but also fit with my proposal.

Arguably the best-known and most-used empirically-informed method is Criteria-Based Content Analysis (CBCA). The bedrock of this approach is the hypothesis that testimony about statements derived from the memory of real-life experiences differs significantly in content and quality from fabricated or fictitious accounts (Steller, & Köhnken, 1989). The method consists of a checklist that scores testimony on nineteen criteria such as the quantity of details in a testimony, how self-deprecating a witness is about the reliability their statement and references to the witnesses’ own mental state. For most of these criteria, research is available that supports a correlation with (un)reliable testimony (Vrij, 2005). The CBCA method was developed primarily to be used for the evaluation of testimony by children in sexual abuse cases, though some have argued that it could be used for a wider variety of cases.

---

165 The CBCA is admissible evidence in some courts of North America and in several West European countries such as Austria, Germany, Sweden, Switzerland, and the Netherlands (Oberlader, 2019, 15).

166 Admittedly, others have greeted this claim with skepticism (Vrij 2008, 222).
The CBCA is sometimes interpreted as if it provides an algorithm for calculating the witness’s reliability, where a high score on the list automatically corresponds to a high degree of witness reliability. However, such an algorithmic interpretation is problematic (Vrij, 2008, 241). For instance, while meta-analyses do show that a high score on the list correlates with truthful testimony, the correlation is weak and there are many false positives (Vrij, 2005; Vrij, 2008; Amado et al., 2016; Oberlader et al., 2021). Steller & Köhnken (1989, 231) themselves also argued that the occurrence of criteria on the list does not only depend on whether the source is reliable, but also on personal and situational factors which may be unrelated to the reliability of a statement, such as the witness’s age. When interpreting the results of the CBCA, such factors also have to be taken into account. That is why various authors suggest that the CBCA is part of a broader diagnostic process, known as Statement Validity Assessment (SVA). SVA begins with a case file analysis, which gives the expert insight into what may have happened and the issues that are disputed. The second part is a semi-structured interview on the event in question. Third, the CBCA is used to analyze the transcript of this interview. Finally, a Validity Checklist is used to further interpret these results in context. During this analysis, the expert looks at eleven issues that can affect CBCA scores, such as the quality of the interrogation and certain personality characteristics of the witness (e.g., suggestibility) (Vrij 2008, 204).

The way that the above method structures eyewitness evaluation is not inherently explanatory. Nonetheless, it can straightforwardly be interpreted as such. For instance, as Oberlader (2019, 13) summarizes the method: “SVA examines various alternative hypotheses for the development of a statement”. According to her, these hypotheses are derived during the case file analysis (Oberlader, 2019, 13). Explanatory reasoning is similarly present during the later steps of this method. For example, when we interpret the results of the CBCA, we must consider alternative reasons why the statement scores particularly high or low on the checklist. For instance, “a low-quality statement might be given if the event in question was so
simple and short that many criteria could just not occur” (Oberlader, 2019, 14). In such a situation, there is a plausible truth-telling explanation for why this statement was produced which is consistent with a low score on the CBCA. So, TIBE fits well with the underlying idea of the approach. Furthermore, viewing this approach in terms of explanatory comparison again has several benefits. Some of those benefits I mentioned earlier when I discussed the capacity approach. Another benefit relates to a well-known point of criticism of SVA, namely that it lacks a clear method for drawing conclusions from the resulting analyses (Steller & Köhnken, 1989, 231). To draw such conclusions, we may employ inference to the best explanation, where we assign a degree of reliability to the witness based on the plausibility of the truth-telling explanation compared to the alternative explanations (see section 2.1).

Let’s consider another, more recent method, which was developed by Wise, Fisher & Safer (2009). On this method, the evaluator asks whether the interview and identification procedures in the case were fair, unbiased and sufficiently thorough. Furthermore, the evaluators ask what eyewitness factors during the crime are likely to have increased or decreased the accuracy of the eyewitness testimony. The method consists of a list of questions that help evaluator assess these aspects, such as: “Did the interviewers contaminate the eyewitness’s memory of the perpetrator of the crime?”, “Was there reliable, valid corroborating evidence that establishes the veracity of the eyewitness testimony?” and “was the witness intoxicated?” According to its creators, this method offers a comprehensive analytical framework for “identifying and organizing the myriad of disparate factors that affect the accuracy of eyewitness testimony” (Wise, Fisher & Safer, 2009, 472). The authors argue that this method is useful for judges or jurors during the process of evaluation, providing them with a structured method and a way to incorporate the psychological findings on eyewitness reliability into their judgments. They also suggest that the method can similarly be useful for interviewers to optimize the reliability of their
interview and for attorneys to develop ways to defend or attack an eyewitness’s reliability.

Much like with SVA, the explanatory approach that I propose can complement this method. First, this method also lacks a clear way method for drawing conclusions from the resulting analysis. We may use inference to the best explanation as such a method. Furthermore, we may not always want to spend the same amount of time and attention on each aspect on the extensive list. Which aspects we focus on most and how we answer the questions from the list can be based on which explanations for the testimony are most salient. Conversely, both this approach and the SVA can inform how we go about inferring to the best explanation. For instance, the method of Wise et al. (2009) lists a number of questions that we might ask to check how plausible such a hypothesis is. So, (parts of) this method can be helpful for explanatory inference, by pointing to the relevant facts. Similarly, these checklists can also point to specific explanations that we could (or should) consider in a given case.

In summary, explanatory reasoning not only fits well with existing prescriptive approaches, it also complements them. In particular, TIBE offers a flexible, context-sensitive framework for drawing conclusions from the data that we gather using, for instance, psychological checklists. Such context-sensitivity is especially important given that which evidence deserves our attention and how we ought to interpret this evidence varies with the salient explanations for a particular piece of testimony. In the next two sections I consider this point in more depth. In particular, I distinguish two challenges that evaluators face when drawing conclusions from the evidence about an eyewitness’s reliability. The first is that of interpreting the evidence, the second is that of deciding which evidence is relevant. This discussion will further illustrate the benefits of explanatory reasoning about eyewitness reliability.
5. Interpretation and striking agreement

One difficulty that evaluators face is that they have to interpret the evidence about a witness’s reliability. The problem is that the same fact can support contradicting conclusions about a witness’s reliability, depending on how we interpret it. Consider the work of Shapin (1994, 212-238) who discusses a list of familiar and intuitively plausible maxims for the assessment of testimony. His strategy is to find a ‘countermaxim’ for each. For instance, a knowledgeable witness is often good, but can also be bad because they tend to over-interpret what they observe. Similarly, confidently delivered testimony may be a sign that the witness saw the events that they reported on clearly – as opposed to a witness who offers many caveats about their observations – but it may also be indicative of overconfidence or a liar. Shapin argues that we can do this for almost any common-sense generalization about how certain types of facts relate to the reliability of a witness. So, we cannot draw inferences such as: ‘all else being equal, the fact that this witness is knowledgeable always supports them being reliable’, because the opposite inference may sometimes also be warranted. Rather, we have to determine how a particular fact should be interpreted in a particular context. But how do we do so?

To further illustrate this problem, let’s look at an example in more depth, namely the phenomenon of striking agreement. Striking agreement occurs when multiple witnesses agree on a specific, unlikely detail. For instance, imagine that a robbery took place and that there were several witnesses. Suppose that investigators take statements from each of the witnesses. Now compare the following situations:

i. The witnesses all claim that the robber wore a t-shirt and jeans.

ii. The witnesses all claim that the robber wore a clown suit.

People wearing t-shirts and jeans are much more common than clown suits. Therefore, the agreement between the witnesses in the second case is much more striking than in the first. But what conclusion should we draw from such strikingness? One common thought is that if witnesses independently agree on an implausible
hypothesis, then this provides stronger confirmation for this fact than if the witnesses independently agree on a more likely fact. Bovens & Hartmann (2003) offer a Bayesian analysis of this idea. They argue that the likelihood ratio for the statement of the two witnesses in the second case is higher than in the first. The idea is that when witnesses misreport on what happened, some false stories are more likely than others. For instance, imagine that if the robber wore yoga pants and a hoodie, it would not be especially surprising for a witness to mistakenly report that he wore a t-shirt and jeans. People often wear t-shirts and jeans. So, as Bovens & Hartmann (2003, 113) put it:

[T]his is not simply one out of so many false stories. It is the sort of thing that someone is likely to say if she does not know, or does not wish to convey, any details about the person missing in action, but feels compelled to say something.

In contrast, it would be highly unlikely if several witnesses all independently, falsely claimed that the witness wore a clown suit when this was not the case. According to Bovens & Hartmann, such a story is ‘too odd not to be true’.

However, we can also view striking agreement between witnesses in a different light, namely as a sign that the witnesses are unreliable. Even relatively reliable witnesses will make small mistakes. Therefore, we would expect some incongruity between their reports. If their testimonies cohere to a surprising degree, this can lead to a suspicion of conscious or subconscious collusion – i.e., that the witnesses are not independent. We might then assign a lower likelihood ratio to their testimony. Gunn et al. (2016) offer a Bayesian formalization of this argument, where the agreement is considered ‘too good to be true.’ They use the example of a police line-up. If the line-up is sufficiently big and enough witnesses unanimously identify the same person as the perpetrator, then we can be virtually certain that the line-up was biased in some way. This may be “for example, because the suspect is somehow conspicuous, [or because] the staff running the parade direct the witnesses towards
These conflicting ways to interpret this phenomenon pose a challenge for eyewitness evaluators. Suppose that we encounter an instance of striking agreement between multiple witnesses. Should we see this as a sign of truthfulness or of collusion? Is it too good to be true or too odd not to be true? My answer is that, on TIBE, which interpretation of some multi-interpretable fact we should choose depends on which explanation of that fact is the most plausible in context.\textsuperscript{167} In the case of striking agreement we face multiple possible explanations of the witnesses’ utterances. One is the truth-telling explanation, emphasized by Bovens and Hartmann. They quote Lewis (1946, 346) who remarks that: “[O]n any other hypothesis than that of truth-telling, this agreement is very unlikely”. In contrast, Gunn et al. suggest several alternative explanations (e.g., the suspect being conspicuous in the line-up). Once we have these competing explanations, we can further investigate and evaluate them. Were the witnesses indeed independent? Are there signs of collusion? Was the suspect conspicuous in the line-up?

This answer may seem too obvious to mention. Nonetheless, we may contrast it with the two alternative prescriptive accounts that I discussed earlier. These frameworks do not specify how we should go about interpreting the evidence within these frameworks. On the capacity approach, we might ask whether these witnesses are lying or misremembering. But striking agreement can be both evidence for and against lying or misremembering. The problem of interpretation is also a thorny one for empirically informed methods, which presume that we can determine beforehand which factors will count for or against reliability. However, even if a fact \textit{generally} correlates with, for instance, lower reliability, it may nonetheless be reasonable to take it to signal increased reliability in a specific context. So, it seems that these methods need some further mechanism to account for evidential

\textsuperscript{167} Lipton (1998, 30) also makes this point with respect to the aforementioned work by Shapin (1994).
interpretation. TIBE can be of service in that regard. The fact that explanatory reason-
ing offers such a seemingly obvious approach to interpretation only speaks to its naturalness.

The above discussion also shines further light on the relationship between TIBE and Bayesianism. The work of both Bovens & Hartmann and Gunn et al. is part of the extensive Bayesian modeling literature on witness reliability.\textsuperscript{168} Thagard (2005), who argues that TIBE and Bayesianism are incompatible, further suggests that Bayesian modeling is generally unhelpful for studying eyewitness reliability. In particular, he notes that any Bayesian model will require many conditional probabilities and that it is unclear what numbers we should plug in for these probabilities (especially because it is unclear how these probabilities are to be interpreted).\textsuperscript{169} But regardless of whether we grant Thagard’s points about modeling \textit{particular instances} of eyewitness testimony, we do not have to accept his blanket rejection of Bayesian modeling. Because TIBE is compatible with Bayesianism – and the latter provides a helpful way of legitimizing the former – such modeling is a helpful tool for elucidating the epistemic consequences of certain assumptions within our explanatory inferences. For instance, the work on striking agreement is an excellent example of how the underlying assumptions of arguments about the reliability of testimony based on competing explanations can be made explicit and precise.\textsuperscript{170}

\textsuperscript{168} See Merdes, von Sydow & Hahn, (2020) for a review of some of this literature.

\textsuperscript{169} In contrast, Rassin (2014) notes that Bayesian quantification could help make explanation-based eyewitness evaluation more precise. I am skeptical of this suggestion because of the well-known difficulties that many judges and jurors have with interpreting numerically-expressed evidence, especially when this evidence lacks a clear statistical basis.

\textsuperscript{170} See Bovens & Hartmann (2003), chapters 3 and 5, for how various assumptions about the reliability of witnesses and the probability of what they report lead to differences in the impact of their testimony on our beliefs.
6. Relevance

A second problem that evaluators have to deal with is deciding which evidence they should consider when evaluating an eyewitness’s reliability. Bayesians sometimes assume that we should take into account our ‘total evidence’, i.e., everything relevant that we know (Goldman, 1999, 145). But this is often not practically feasible. For any particular eyewitness there are countless facts that could, conceivably, influence how reliable we ought to consider them. Among these are the hundreds of factors which psychological research suggests may influence the reliability of eyewitnesses, such as their stress level during the events, whether they were asked leading questions and how long ago the reported events took place (Wise et al., 2014). Evaluators cannot easily determine whether, and to what extent, each of these countless facts is present in a particular instance. Nor can they readily draw a conclusion about the witness’s reliability if the facts that they know to be present are too numerous. Finally, they have to decide which further evidence to gather. As Lipton (2004, 116) puts it:

In practice, investigators must think about which bits of what they know really bear on their question, and they need also decide which further observations would be particularly relevant.

The capacity approach and empirically informed methods offer valuable guidance on which factors to look at. However, I believe that thinking in terms of possible explanations has additional value here. For instance, the capacity view directs us to look at the evidence about the eyewitness’s capacities (e.g., perception, memory, veracity). But there will typically be numerous facts that are potentially relevant for assessing whether the witness exercised these capacities correctly. On the other hand, some relevant facts may not be directly related to the capacities of the witness. The same goes for empirically-informed methods, where the facts that evaluators look at are determined beforehand. But this also means that if evaluators use, say, the CBCA, then this list will contain various facts that are not especially relevant in
Explanatory thinking can help focus our attention on the most important facts. On TIBE, relevance of facts is dictated by the explanations that we consider. As Lipton (2004, 116) points out, we may discover relevant evidence for a hypothesis by asking what it would explain. I would add to this that we may also consider what the explanation would not explain. In other words, we select the relevant facts by considering what we would expect to see if this explanation were true and what we would not expect to see. What should be especially interesting during witness evaluation are discriminating facts – i.e., facts that confirm one of the available scenarios but disconfirm the other. The more specific and limited our set of explanations, the easier it will be to answer these questions and the more directed our search for evidence will be.

Let’s consider an example that shows both the breadth of facts that can potentially be pertinent, and how explanatory reasoning may guide us in seeing their relevance. The following passage is taken from a lawyer’s plea in a Dutch case, concerning a defendant who allegedly broke into his girlfriend’s house and destroyed some of her belongings (van Oorschot, 2014, 209):

[W]hen we look at the [report] of the interrogation with my client, I see, typed down, in the middle of one of my client’s questions, the phrase, “theft, unqual.”, which arguably stands for “theft, unqualified”. But my client would never express himself this way, nor would other defendants, presumably. So who is speaking here? The police officer or the defendant?

But why should it matter that the police officer phrased the report this way? The language of a report would normally not be especially relevant to the question of its reliability. Nonetheless, such a fact could become germane if we were (implicitly) considering the possibility that the officer did not accurately report what the defendant said. Indeed, that is how van Oorschot (2014, 209-210), from whom I
borrow this example, interprets this passage:

This lawyer (...) challenges the neutrality of the written transcription (...). He suggests that the police officials have been so set on shaping and rewriting his client’s words that it has now become unclear who precisely is speaking (...) how can we be sure that the rest of the client’s statements (...) are truly his – and not added by zealous police officials?

Van Oorschot (2014, 210) also remarks that “the lawyer does not merely offer an interpretation of the different stories present in the file; he also tells us a story about the file”. In other words, the lawyer proposes an alternative explanation of this passage, which competes with the hypothesis that ‘this report accurately reflects what the defendant said’. They draw our attention to a fact that we might not otherwise consider relevant, and shows that it becomes relevant, once seen in the light of this alternative explanation. In fact, this particular datum might have come to the lawyer’s attention precisely because they were considering the alternative explanation that the police was putting words in his client’s mouth.

This example also shows how we may generate novel explanations for a witness’s testimony. As in the case of striking agreement, we encounter a surprising fact – in this instance an unusual phrase in the case file – and ask what might explain it. In other words, we employ abductive reasoning to generate novel explanations. Lipton (2004, 116) points out that this is one of the strengths of inference to the best explanation, compared to Bayesian epistemology. As Lipton and others have noted, Bayesianism has nothing to say about the act of creating hypotheses. Furthermore, the act of generating hypotheses is not an explicit part of either the capacity approach or empirically-informed methods. Nonetheless, hypothesis generation is an important aspect of both evidential reasoning in general and eyewitness evaluation specifically. The discovery of an alternative explanation for some fact will often lower our confidence in our initial explanation and can be a source of reasonable doubt (Jellema, 2022b).
7. Conclusion

In this chapter, I argued that explanation-based thinking can be rational and helpful in assessing the reliability of one important type of evidence, namely eyewitness testimony. On my account such Testimonial inference to the best explanation begins with the question: ‘what could explain this particular piece of testimony given by this source in this context?’ Once we have one or more candidate explanations, it is helpful to make them more specific. We then compare these narratives of the witness’s explanations and arrive at a conclusion about whether – and to what degree – the source of the testimony is reliable. This account falls within a broader explanationist trend in legal evidence. Nance (2016, 84) observes the following about this trend:

One main motivating concern of those who press the explanatory approach is that [probabilistic accounts] focus on the end product of deliberation, rather than the process of arriving there, giving no direction to jurors as to how to go about assessing the evidence in the case.

My argument for testimonial inference to the best explanation rests on a similar idea. I embrace the Bayesian account of what we’re trying to achieve when we evaluate witness reliability – namely to evaluate the likelihood ratio of the testimony in the light of the available evidence. But this Bayesian story does not tell evaluators much about how to draw conclusions from the available evidence about the witness’s reliability. Yet this is precisely what we want from an account of rational eyewitness evaluation. I have outlined how explanation-based reasoning (i) tracks the Bayesian likelihood ratio, and (ii) helps structure, interpret and select the available evidence. These are the two pillars of TIBE’s rationality.

I further explored TIBE’s prescriptive ambitions by comparing it to existing approaches to eyewitness evaluation. I focused on two influential types here: the capacity approach and empirically-informed methods. As I argued, these approaches and TIBE may complement one another. On the one hand, these existing
approaches may inform us of the kinds of considerations that will also be relevant in explanation-based inference. On the other hand, once we incorporate an explanatory component into these approaches, we end up with a more context-sensitive, flexible framework on which it is clearer how evaluators ought to come to their ultimate decision. I further explored this idea – that TIBE can complement other frameworks by telling us how to make sense of the evidence – by focusing on two specific of eyewitness evaluation where explanatory reasoning is especially helpful, namely the interpretation and selection of the evidence about the witness’s reliability.
VII. REASONABLE DOUBT FROM UNCONCEIVED ALTERNATIVES

Chapter abstract

In criminal trials, judges or jurors have to decide whether the facts described in the indictment are proven beyond a reasonable doubt. However, these fact-finders cannot always imagine every relevant sequence of events – there may be unconceived alternatives. The possibility of unconceived alternatives is an overlooked source of reasonable doubt. I argue that fact-finders should not consider the defendant’s guilt proven if they have good reasons to believe that plausible, unconceived scenarios exist. I explore this thesis through the lens of both Bayesian and explanation-based approaches. I draw on ideas from the philosophy of science to show why and when unconceived alternatives lead to reasonable doubt on either account.

1. Introduction

In criminal trials, judges or jurors decide whether the prosecution succeeded in proving the facts described in the indictment. To make this decision, they often compare competing scenarios such as ‘John stabbed Mary to death during an attempt to rob her’. However, it is not always possible to imagine every possible scenario – there can be unconceived alternatives. Most of these alternatives will be not be worth considering – e.g., that it was actually Barack Obama who killed Mary. However, there may also be one or more plausible, overlooked alternatives. To give an example, someone else could have had the motive, means and opportunity to kill Mary.

Fact-finders often treat the limited number of possible scenarios that they consider as an exhaustive set of possibilities. This presumption of exhaustiveness is, I will...
argue, necessary. Without it, we cannot rationally come to a conclusion about whether it is probable that the defendant is guilty. However, there can also be cases where this presumption of exhaustiveness is unjustified, because the decision-maker has good reasons to suspect that plausible alternatives exist. The main contention of this chapter is that, in those cases, reasonable doubt exists about the defendant’s guilt. More precisely, a decision-maker should not consider the alleged acts proven if their evidence supports the belief that plausible, unconceived scenarios in which the defendant did not commit these acts are likely to exist. Conversely, if the decision-maker lacks good reasons for the belief that such alternatives exist, they may presume their set of explanations to be exhaustive.

Parties sometimes invoke the possibility of unconceived alternatives in court as a source of doubt. For instance, Pardo & Allen (2019, 23) cite various scholars who point out that “there are cases where defendants argue along the lines of ‘something else happened’ without offering any specific alternatives.” However, legal evidence scholars have paid little attention to the rationality of such arguments. When and why it is reasonable to doubt the defendant’s guilt based on the possibility of unconceived alternatives is therefore an open question, which I tackle in this chapter.

To make my argument as precise as possible, I cast it in the language of both Bayesian and explanation-based accounts, and their associated interpretations of the proof beyond a reasonable doubt standard. An additional benefit of using these models is that this allows me to connect the ideas that I develop here to existing work on unconceived alternatives from the philosophy of science. Philosophers of science have thought extensively about unconceived alternatives as a source of doubt about our current scientific theories. They regularly use Bayesian or inference to the best explanation-based accounts of rational scientific inference to develop their arguments. As I will show, many of their ideas can be analogously applied to the legal context.
The next section briefly contrasts the interpretations of proof beyond a reasonable doubt associated with these two frameworks (section 2). I then turn to why the possibility of unconceived alternatives leads to reasonable doubt on either interpretation. My argument has four steps: First, for both Bayesians and explanationists, the discovery of previously undiscovered scenarios may lead to reasonable doubt (section 3). Second, if we have good reasons to suspect that we could discover such alternatives, we should take their potential epistemic consequences into account if we want to meet the beyond a reasonable doubt standard. However, third, it is often impossible to anticipate the epistemic consequences that such unconceived alternatives are likely to have if we were to discover them (section 4). Fourth, we therefore cannot consider guilt proven beyond a reasonable doubt if we are not justified in believing that we have considered all plausible scenarios (section 5). In section 6 I explore when we may justifiably presume that there are no further alternatives.

2. Reasonable doubt and rational criminal proof

I already discussed the two frameworks of rational criminal proof as well as the reasonable doubt standard in depth in chapters II and III of this thesis. To lay the groundwork for my argument in this chapter, I first want to reiterate some of the points explained in those chapters. To start, the term reasonable doubt usually refers to the proof standard used in common law systems, such as the American and the English. One way to approach the problem of unconceived alternatives would therefore be to analyze how specific common law courts interpret the reasonable doubt standard. Such a doctrinal analysis might show that, according to these interpretations, the possibility of unconceived alternatives sometimes ought to lead to acquittal. However, I do not believe that such an approach would be fruitful. First, as also discussed in chapter III, it is notoriously unclear how common law courts interpret this standard and even in systems that do not have the reasonable doubt standard, such as many European civil law systems, courts should not
convict defendants if they have serious doubts about their guilt. When I discuss the doubt arising from unconceived alternatives, my conclusions pertain to those systems too. In this chapter I therefore continue using an epistemic approach to analyze the notion of reasonable doubt. On an epistemic view of criminal proof the principal goal of legal trials is drawing accurate factual conclusions and, to the extent that factual errors are made, distributing those errors fairly (e.g., Dworkin, 1985a;b; Goldman, 2002; Stein, 2005). With respect to this last point, false convictions are generally considered to be much worse than false acquittals (Epps, 2015). This is captured in Blackstone’s well-known maxim that “it is better that ten guilty persons escape than that one innocent suffer” (Blackstone, 1962). On the epistemic approach, the key purpose of the reasonable doubt standard is to distribute errors fairly by setting a high standard for conviction. As a result, the probability of false convictions is lowered at the expense of a higher probability of false acquittals.

For both the Bayesian and explanation-based frameworks, the above considerations mean that a defendant can only be convicted if they very probably committed the alleged act. On the Bayesian account, proof beyond a reasonable doubt requires (at least) that the posterior probability of guilt meets some threshold (Gardiner, 2019). In other words, the decision-maker ought to be highly confident that the defendant committed the alleged acts. Though the explanationist approach is distinct from the Bayesian, it shares the commitment that a defendant can only be convicted if the probability of their guilt is sufficiently high. As Allen & Pardo (2018, 1580) put it: “The explanatory account shares the same ends or goals as probabilistic approaches, which have to do with various policy judgments about the likelihood of disputed facts and allocating the risk of error between the parties.” In other words, on both accounts a high probability of the defendant’s guilt is a requirement for proof of guilt beyond a reasonable doubt for the sake of error distribution.

As argued in chapter III, there are, broadly speaking, two ways to conceptualize the beyond a reasonable doubt standard in explanationist terms. The first, and most popular is the following interpretation from Pardo and Allen (2008, pp. 238-9):
In criminal cases [fact-finders should infer] the defendant’s innocence whenever there is a sufficiently plausible explanation of the evidence consistent with innocence (and ought to convict when there is no plausible explanation consistent with innocence assuming there is a plausible explanation consistent with guilt).

In other words, when we determine whether the guilt of the defendant has been proven, we should check whether there is a believable case for guilt and no believable case for innocence. The second explanationist interpretation of the proof beyond a reasonable standard is as inference to the best explanation – where the requirement is that the guilt scenario should be much more plausible than any alternative. I defended this IBE-based interpretation in chapter III. There I also argued that one condition for a justified inference to the best explanation is that we do not have reasons to suspect plausible, unconceived alternatives. In this chapter, I expand on this argument. Furthermore, I also argue that the same holds for Allen & Pardo’s interpretation of the beyond a reasonable doubt standard and for the Bayesian account. For all these interpretations, if a decision-maker has good grounds to suspect plausible, unconceived alternatives, then they cannot justifiably believe that the defendant is very probably guilty. Conviction is then unwarranted on epistemic grounds.

Before I offer my argument, I want to reiterate some technical points from chapter II relating to the Bayesian account and to the explanation-based framework. To begin with the former, one important idea within Bayesianism is that we use Bayes’ formula, to calculate the probability of any hypothesis – including whether the defendant probably committed the alleged acts. The most relevant part of this formula for the discussion ahead is the denominator \( P(E) \), also called the marginal likelihood. This marginal likelihood can be further decomposed into:

\[
P(E) = P(H_1) P(E \mid H_1) + \ldots + P(H_n) P(E \mid H_n) + P(H_c) P(E \mid H_c)
\]
Here $H_1….H_n$ are our conceived hypotheses and $H_c$ is the *catch-all hypothesis* which expresses that ‘none of our conceived hypotheses is true’. When we include the catch-all our set of hypotheses therefore becomes exhaustive as the catch-all therefore conveys the possibility of further, *unconceived* alternatives. So, in order to determine the value of $P(E)$ we need to determine the likelihood and prior probability of this catch-all. As I will explain in section four, this is problematic.

As explained in chapter II, in legal proof it is common to use the odds version of Bayes’s formula. This equation does not contain a catch-all hypothesis. However, the relative probability expressed in this formula can only be converted to an absolute posterior probability of the two hypotheses if the hypotheses under consideration are mutually exclusive and exhaustive. We will end up denying this presumption if we suspect unconceived alternatives.

Explanation-based accounts express the requirements for proof beyond a reasonable doubt in terms of explanations and their plausibility. How plausibility relates to Bayesian probability is the topic of ongoing discussion (see e.g., Biedermann & Vuille, 2018; Allen & Pardo, 2019; Mackor, Jellema & Van Koppen, 2021). For the sake of the following discussion I will use these terms as follows. A plausible explanation is one explains the relevant facts well. Whether it does so depends on the extent to which it displays certain explanatory virtues. For instance, an explanation may be plausible because it offers a coherent, detailed explanation of the most important facts in a case. In contrast, (epistemic) probability is a quantification of how confident an agent is that some hypothesis – for instance a scenario – is true. The two notions are closely related because – all other things being equal –

\[171\]

The relationship between probability and explanatory quality is also an ongoing discussion in the philosophy of science. For instance, Lipton (2004) argues that, that though explanatory quality - or what he calls ‘loveliness’ - can be a guide to the probability of an explanation – its ‘likeliness’ – the two should not be equated. In contrast, other philosophers of science suggest that explanatory goodness may be translated into a probabilistic notion, and that that explanation-based approaches can therefore be fitted into the Bayesian framework see e.g., Douven, 2017).
the more plausible a scenario is, the more probable we should consider it. However, what is important to note for the argument I make in this chapter is that a plausible scenario is not *always* probable. For instance, as I will discuss later, a plausible scenario can be improbable if it has many plausible competitors, while a somewhat implausible scenario may occasionally be quite probable if we are certain that there are no better alternatives.

With these preliminaries out of the way, it is now time to dive into why and when unconceived alternatives lead to reasonable doubt.

3. *Bad lots and reasonable doubt from new alternatives*

In this section, I begin sketching what I consider to be the problem of unconceived alternatives. Few legal scholars have paid attention to this problem. Those who do discuss it, take it to be a narrower problem than it actually is. Specifically, they frame unconceived alternatives as being problematic only for *inference to the best explanation (IBE)*-based models of criminal proof – a subset of explanation-based approaches. This is called the *bad lot problem* (Laudan, 2007; Amaya, 2009; Ribeiro, 2018; Jellema, 2020).\(^\text{172}\) The bad lot objection invokes the possibility of unconceived alternatives to argue that IBE does not reliably lead to accurate outcomes. According to Amaya (2009, 154) it is “the most serious problem that a model of IBE for law has to face.” However, while I agree that the bad lot problem is a major objection to IBE-based approaches, unconceived alternatives are problematic for other major accounts of rational criminal proof as well. On both explanation-based accounts that do not use IBE and on Bayesian accounts, we cannot justifiably come to the conclusion that the defendant is guilty if we have good reasons to suspect plausible unconceived alternatives. It is this broader problem that I want to discuss here.

\(^{172}\) This problem was originally formulated by philosophers of science (Sklar, 1981; Van Fraassen, 1989, 142-3).
Let me begin by explaining the bad lot objection against IBE-based models. Inference to the best explanation is a specific approach to explanatory inference, where we infer that the explanation which best explains the evidence is (probably) the true explanation (Lipton, 2004). The bad lot objection begins with the observation that IBE uses comparative reasoning (that a hypothesis is the best available) to arrive at an absolute verdict (that this hypothesis is true) (Douven, 2017, 9). However, such a conclusion is only warranted if we may presume that the true hypothesis is probably among those considered. If we drop this presumption, we end up with the merely comparative conclusion that one explanation is better than those we have come up with so far. Or, in terms of criminal cases, without this presumption IBE only justifies the conclusion that one scenario is the best out of those that have been presented in court. Any such comparative conclusion is obviously insufficient to ground conviction. For instance, it might even mean basing a conviction on a poor scenario that is probably untrue, simply because this scenario is the best among the bad lot.

The bad lot problem poses a difficulty for IBE-based models of criminal proof, but it does not affect the Bayesian account. Nor does it pose a difficulty for explanation-based models that use the no plausible alternatives interpretation of proof beyond a reasonable doubt. After all, these models do not rely on comparative reasoning in the same way that IBE does. On these models we do not infer from the conclusion that one explanation or hypothesis is the best available that it is probably true. For instance, recall that on many explanation-based accounts, guilt is proven beyond a reasonable doubt if there is a plausible guilt explanation and no plausible innocence.

---

173 Or, as Van Fraassen (1989, 143) puts it: “[T]o believe the best explanation (...) requires a step beyond the comparative judgment that the hypothesis is better than its actual rivals. (...) For me to take it that the best of set X be more likely than not, requires a prior belief that the truth is already more likely to be found in X, than not.”

174 This is not to say that such comparative reasoning never plays a role on these models. As I will argue below, how one hypothesis fares with respect to its competitors often is relevant for the question how probable it is. Nonetheless, this comparison is not an inherent part of these models, and hence these models avoid the narrower bad lot objection, which targets the move from a relative to an absolute conclusion, as discussed above.
explanation. Yet as Pardo & Allen (2007) – the originators of this interpretation – note in response to the bad lot objection, this means that we are not choosing the best explanation of the available ones. Rather, we check whether the prosecution has provided a scenario that is sufficiently plausible to ground a conviction and whether the defense (or anyone else) has produced a scenario that is sufficiently plausible to create a reasonable doubt.\textsuperscript{175}

Still, unconceived alternatives also pose a problem for Allen and Pardo’s interpretation of the reasonable doubt standard, as well as for the Bayesian account. Specifically, within these frameworks, evidential inference also requires the presumption that we have not overlooked plausible alternative scenarios. If this presumption is unjustified (because we have good reasons to believe that plausible unconceived alternatives exist) guilt cannot be proven beyond a reasonable doubt. This is the broader worry of unconceived alternatives.

My argument that this presumption is required begins with the observation that the discovery of previously unconceived alternatives can create a reasonable doubt on both the explanation-based and Bayesian account. To start with the former, consider the two demands of proof beyond a reasonable doubt on this interpretation: (i) There must be a plausible explanation that implies the defendant’s guilt, and (ii) there must not be a plausible explanation that implies his innocence. Suppose that, at some point, both these conditions are met. A subsequent discovery of a previously unconceived alternative could result in the first of these conditions no longer being satisfied if the new alternative might both be plausible and imply the defendant’s innocence. So, on explanation-based accounts, the discovery of a previously unconceived scenario may overturn the case for guilt. What about the Bayesian approach? At first sight, this approach may seem to avoid the problem of newly discovered alternatives. After all, on the Bayesian account, when we consider the

\[\textsuperscript{175} \text{This is similar to the way that Lipton (2004) suggests that IBE can overcome the bad lot problem, namely by adding the demand that the best explanation should also be ‘good enough’.}\]
probability of some hypothesis H, we typically take into account both that hypothesis and its negation. For instance, in their review article on Bayesian inference in legal settings Fenton, Neil & Berger (2016, 53) begin by introducing the basics of the Bayesian approach. They write: “A hypothesis is a statement (...) whose truth value we seek to determine. Examples include: a. ‘Defendant is innocent of the crime charged.’ (...) b. ‘Defendant was the source of DNA found at the crime scene.’ (...) The alternative hypothesis is a statement that is the negation of a hypothesis.” When we cut up the hypothesis space this way, we seemingly avoid the problem of unconceived alternatives. After all, either the defendant is innocent or she is not. Either she is the source of the DNA or she is not. There are no other options to consider as these hypotheses jointly exhaust the hypothesis space.

The problem for the Bayesian is that the above picture is unrealistic. Humans typically do not reason in terms of such general, exhaustive hypotheses, nor may we expect them to. For instance, legal fact-finders typically consider a small set of specific scenarios which helps them make sense of the available evidence (Pennington & Hastie, 1993). One key reason for this is that to adequately consider such mutually exclusive and exhaustive hypotheses would require the fact-finder to consider all of the countless (or even infinite) possibilities in the hypothesis space. This will often be an impossible feat for even the smartest person. Indeed, contemporary Bayesians also frequently note that thinking in terms of mutually exclusive and exhaustive hypotheses is not always realistic. For instance, Urbaniak & Di Bello (2021) point out that considering the negation of a specific hypothesis (e.g., ‘the defendant did not hit the victim in the head’) “can be unhelpful in assessing the evidence.” They suggest that determining the probability of such a negation will involve on drawing up a more specific scenario, which includes details such as whether the defendant was at the crime scene. Several authors have recently developed Bayesian frameworks that reflect this point, by incorporating the notion of scenario-based reasoning (e.g., Cheng, 2013; Vlek et al., 2013; Urbaniak, 2018; Dahlman, 2019).

If we consider only a limited set of scenarios, then the possibility of unconceived
alternatives rears its head. Hence, the Bayesian also faces the possibility of discovering new, previously unconsidered alternatives. Furthermore, just as on explanation-based accounts, the discovery of such alternatives may lead to reasonable doubt for Bayesians. Firstly, it is a generally accepted fact that discovering that there are previously unconsidered alternatives, may change the probabilities that we assign to the already conceived hypotheses. The question how Bayesians should account for this is called ‘the problem of new theories’ (Talbott, 2016).\textsuperscript{176} Secondly, evidence that there are no plausible alternatives can sometimes raise the probability of a hypothesis (Dawid, Hartmann & Sprenger 2015).\textsuperscript{177} Unconceived alternatives therefore pose the same problem for Bayesians as for explanationists: their discovery may overturn the case for guilt (and sometimes discovering that there are no alternatives may be the reason why guilt is proven beyond a reasonable doubt).

The discovery of a new alternative may therefore lead to reasonable doubt on either account. Yet the central thesis of this chapter is stronger than this. If we have reasons to suspect that we could discover a plausible alternative, this should be a cause of reasonable doubt. Of course, we can never be certain that we have exhausted all possibilities, even in the most clear-cut of cases. The situations that I am concerned with are those where we have good reasons to suspect that there are unconceived alternatives – i.e., that there is some relevant possibility that we have overlooked. A defense lawyer could reasonably invoke such grounds to argue that guilt has not been proven beyond a reasonable doubt. I will develop a fuller account of what such reasons are in section 6. For now, let me just mention two examples. First, we may reasonably suspect that there are further unconceived alternatives if we know that the police focused only on a single suspect throughout their investigation and

\textsuperscript{176} This is related to, but not identical to, the well-known problem of ‘old evidence’. The old evidence problem refers to the question how evidence that has been known for some time can provide support for existing theories, when we discover that there is a logical relation between the two (see Talbott, 2016, 6.2 for further discussion).

\textsuperscript{177} In fact, the earlier Bayesian approach on which fact-finders reason about general, exhaustive hypotheses is arguably unrealistic partially because it does not leave room for the epistemic consequences of unconceived alternatives.
did not look for evidence that there could be others. Second, all our currently conceived scenarios might be implausible. For instance, we might discover that both the defense and prosecution scenario contain illogical time leaps. In that case we may also justifiably presume that there has to be some further alternative that we overlooked. This could either be a completely new scenario, or a variation on an old scenario that explains the difficulty away (e.g., a scenario which posits that the evidence which currently makes our scenario implausible is misleading).

In the next two sections I argue that we are not justified in considering guilt proven to a high degree of probability if we have good reasons to suspect plausible unconceived alternatives. We may then, at best, be justified in drawing the conclusion that some scenario fares comparatively well in relation to the others that we have conceived of.178 The reasoning for this is as follows: because discovering plausible alternatives may lead to reasonable doubt, we should include this possible impact in our inferences about whether guilt has been proven beyond a reasonable doubt if we believe that unconceived alternatives probably exist. However, as I will argue in the next section, including this potential impact is often impossible.

4. The difficulty with considering the unconceived

Imagine that we have good reasons to suspect that there are plausible unconceived alternatives. Furthermore, suppose that we want to take this fact into account when deciding whether the defendant’s guilt has been proven beyond a reasonable doubt. The answer is, at first sight, straightforward. Proponents of explanation-based accounts understand evidential inferences in terms of explanations and their degree of plausibility. Bayesian inference concerns hypotheses and their prior probabilities and likelihoods. So, on either account we could treat the possibility that there are further alternatives as an explanation or hypothesis and we assign it a degree of

178 Which is precisely the charge that critics have leveled against IBE by means of the bad lot problem.
plausibility or a prior probability and likelihood. For instance, on the explanation-based account we might end up with the set of scenarios \{'John killed Mary’, ‘William killed Mary’, ‘Something else happened’\}. The last scenario expresses the possibility of unconceived alternatives. We might then assign each scenario a degree of plausibility. If the last scenario turns out to be plausible, we could treat it as an innocence scenario and acquit both John and William on the grounds that there is a plausible alternative that does not imply their guilt. However, sadly, this proposal does not work. The reason for this is that the possibility of unconceived alternatives cannot be evaluated in the same way that we usually evaluate scenarios. To see why, consider an example derived from Dellsén (2017, 37):

Suppose you come home one day to find the front door open and the lock broken. Furniture is overturned, the contents of the shelves are on the floor, and valuables are missing. One explanation is that someone broke in and stole your belongings, making a mess in the hurried process.

Call the scenario mentioned in the final sentence the break-in hypothesis. Now consider the negation of this, the no-break-in hypothesis – that it is not the case that someone broke in, stole your belongings, and made a mess. Or, to put it differently, ‘something else happened’. On the explanationist approach, we should evaluate these explanations in terms of how well they explain the evidence. This is unproblematic for the break-in hypothesis; the explanation that someone broke into your house explains various facts such as why the lock is broken and why the furniture is overturned. The no-break-in hypothesis, in contrast, offers no explanation of these facts. It offers us no understanding of why the lock was broken or why the furniture was overturned.

As Dellsén (2017, 33; 2018, pp. 1758-9) points out, if we engage in inference to the best explanation, we should therefore trivially favor the break-in hypothesis (or really any explanatory hypothesis) over its negation when we ask which explanation
is the best. This even goes for poor explanations. For example, consider the scenario that ‘aliens dancing wildly to ska music turned over the furniture and broke the lock’. If we evaluate explanations in terms of their explanatory quality we should favor the alien scenario over its negation. After all, the latter has no explanatory quality whatsoever – it offers no account of how the mess in the house was created – whereas the former does (however bad an explanation it may be). But if even a bad explanation ranks above the no break in-hypothesis, then no matter how probable it actually is in this case that ‘something else happened’, this possibility will never be accepted as the best explanatory hypothesis (assuming we have formulated at least a single possible explanation). This would, for instance mean that if the alien hypothesis is our only conceived explanation, we would then have to accept it as true on IBE. But, obviously, this does not mean that the alien hypothesis is probably true.

What about explanation-based accounts that do not rely on IBE? In the last section I pointed out that not every explanation-based approach suggests that we should choose the best explanation. Nonetheless, the above point is equally problematic for those explanation-based accounts which state that proof beyond a reasonable doubt requires a plausible guilt explanation and no plausible innocence explanation. No matter how probable it is that we have missed something, the hypothesis that ‘there are unconceived alternatives’ will never be counted as a plausible alternative scenario if we use explanatory reasoning. Unconceived alternatives simply do not show up in regular explanatory inference. For that to happen, we would have to make it more specific what might have happened, if not our conceived scenarios.

---

179 Dellsén makes this point when objecting to Lipton’s (2004) response to the bad lot problem. Lipton’s response is more complex than I have room to explain here, but it ultimately fails because of this point - that we cannot properly compare explanatory hypotheses to negations (Dellsén, 2017, section 2).

180 Contemporary defenders of IBE have modified their accounts to avoid such a conclusion. For instance, Lipton (2004) adds the demand that the best explanation should also be ‘good enough’.
And in doing so, we would simply be offering another alternative explanation – we would not be assigning a degree of plausibility to the unconceived. We cannot compare the unconceived to the conceived precisely because we know nothing about it.

At this point one might suspect that the problem is with explanatory reasoning and that the Bayesian account will therefore fare better. However, as various authors have noted over the years, Bayesianism is ill-equipped to deal with the possibility of unconceived alternatives. To see this, remember how Bayesians usually include the possibility that there is some unconceived alternative, namely by including a ‘catch-all hypothesis’ which expresses the hypothesis that ‘none of our current hypotheses is true’. As mentioned earlier, in the standard formulation of Bayes’ formula, the catch-all hypothesis is included in the denominator, $P(E)$. In the odds-version, we can only draw conclusions about the absolute probabilities of the hypotheses under consideration if these hypotheses are mutually exclusive and exhaustive. To achieve this, we can include the catch-all, which would be in the denominator, where assigning a likelihood, $P(E | \neg H)$ and prior probability $P(\neg H)$ requires assigning a prior probability and likelihood to the catch-all too. For both versions of the formula, because the catch-all is in the denominator, we can only consider $H$ highly probable if the value of the catch-all is low.

In order to know whether $H$ is highly probable or not, we therefore need to determine the likelihood and prior of the catch-all. However, on this point the Bayesian account faces a difficulty. The catch-all is made up of the disjunction of all alternative hypotheses to those in our set of conceived hypotheses $\{H_1…H_n\}$. So, to assign a value to it we need to determine (or at least approximate) the likelihoods and prior probabilities of these alternatives, without knowing anything about them.

For some contexts, defining these values for the catch-all is unproblematic. An example of this might be a card game. In that setting we can accurately determine the probability of the hypothesis ‘the next card will not be a three of hearts or a four of
clovers.’ In criminal contexts we might similarly be able to determine the random-match probability of some DNA sample. However, on the level of scenarios, such values will typically not be so easy to assign. For any given scenario there will be countless alternative scenarios of which we do not know what they look like (after all, they are unconceived). It is difficult to think of a feasible method of determining the likelihood and prior probability of such a set. We simply do not know enough about what the unconceived looks like to assign it a reasonable likelihood and prior probability. As Steele & Stefánsson (2019) put it:

[I]n order for an agent to make sense of a catch-all, she would presumably need to entertain some universal set of possibilities relative to which the catch-all can be defined as the complement of those possibilities she can properly articulate. But it is hard to see how the agent could have access to this universal set of possibilities (…), given that, by assumption, some of these possibilities cannot be articulated.181

For instance, consider the earlier example of the possible break-in again. How should we assign a likelihood of the evidence given the catch-all in that situation? For the break-in hypothesis we could reasonably say that the likelihood is high because a break-in explains the facts well. What about the no-break-in hypothesis? How can we determine how expected the evidence is given the fact that there was no break in, if we don’t know any details about what this no-break-in-situation actually looked like? It seems that there is no meaningful way of doing so. As Salmon (1990, 329) writes: “What is the likelihood of any given piece of evidence with respect to the catch-all? This question strikes me as utterly intractable”. The same goes for the prior probability of such a catch-all hypothesis. As Sklar (1981, 19)

181 Similarly, Bradley (2017, 255) writes: “[G]iven that we don’t know anything about the prospects that we are potentially unaware of, on what basis are we to determine (…) what probability we should assign to the catch-all prospect?”
writes: “[W]e must distribute a priori probabilities over all the alternative hypotheses to be considered. If there is only a finite set of hypotheses we have in mind, this is easy to do (...). But if we must keep in mind the infinite and indeterminate class of all possible hypotheses, known and unknown, how can we even begin to assign a priori probabilities to those few hypotheses (...) we do have in mind (...)?”

That it is difficult to assign any definite prior probability to unknown hypotheses becomes especially evident once we consider what happens when we do discover such a previously unconceived hypothesis. This is a well-known problem for Bayesians, because it seems that the probabilities of our existing hypotheses should change, but there is no new evidence to conditionalize on (Talbott, 2016, section 6.2). The most common way in which Bayesians in the philosophy of science deal with this is by ‘shaving off’ (Earman, 1992; Wenmackers & Romeijn, 2016). Simply put, a newly conceived hypothesis ‘steals’ its new probability from the catch-all; the latter’s probability is lowered so that we can assign a positive probability to the new hypothesis. Suppose that we originally had the set of hypotheses H₁,...,Hₙ and the catch-all Hᶜ. Now suppose that we conceive of the alternative hypothesis Hₙ₊₁. We have to assign this hypothesis some non-zero probability. Because the set {H₁,...,Hₙ, Hᶜ} was mutually exclusive and exhaustive, their probabilities by definition added up to 1. To assign the new hypothesis, Hₙ₊₁ a non-zero probability therefore requires that we lower the probability of the original set. We may do so by assigning the new catch-all, Hᶜ', a lower probability, so that the probabilities of the resulting set add up to 1.¹⁸² However, as philosopher of science Earman (1992, 195) argues, this means that as more and more new hypotheses are discovered and the catch-all becomes smaller, we can only assign them smaller and smaller probabilities when ultimately the point may be reached “where the new theory has such a low initial probability as to stand not much of a fighting chance.” So, suppose that we want to leave open the possibility of discovering many plausible

¹⁸² Note that by doing so we create a new catch-all which differs from the former as it no longer includes the previously unconceived Hₙ₊₁.
alternatives, the prior probability of the catch-all should be very high. Yet this means that the prior probability of our conceived hypotheses has to be very low, regardless of how plausible these hypotheses actually are.

Admittedly, the critique raised by Earman is often more readily applicable to the scientific context than to the legal. Throughout scientific history, there have been numerous instances of novel theories leading to revisions of the probability of the old theories. Some have argued that we therefore have good reasons to continue to expect further plausible theories to be discovered in the future and that it is therefore rational to assign a low probability to our current best theories too (Stanford, 2006). In contrast, in a specific criminal trial, we typically have little reason to expect numerous plausible alternatives to be discovered.\(^{183}\) Nevertheless, the question remains how high the prior of the catch-all should be if we want to leave room for the discovery of plausible alternatives, without knowing exactly how plausible these alternatives are. If we make the prior of the catch-all too low, we may not leave enough room to assign newly discovered, plausible alternatives a high enough probability. If we make the prior too high, it becomes difficult to prove anything, even if our current scenarios are highly plausible. Hence, it seems that there is no reasonable way to assign a prior probability to the catch-all.

To clarify these points, consider the analogous problem of missing evidence.\(^{184}\) For instance, imagine that investigators neglect to interview an important witness. Various authors suggest that, if there is a great deal of missing evidence, the case for guilt may lack the appropriate ‘weight’ or ‘evidential completeness’ (Ho, 2015, section 3.3). But how should we account for the weight of the evidence? One idea is

\(^{183}\) Though this will, of course, depend on the particulars on the case. For instance, in some cases it may be relatively clear how the crime happened and which people might have committed it – for instance in so-called island-cases, where the number of possible culprits is clearly delineated (see Fenton et al., 2019). In other cases, we may expect there to be many – or even infinite - plausible alternatives which vary with respect to the perpetrator, means, motive, location, time and nature of the crime.

\(^{184}\) See also chapter VIII for a more in-depth discussion of this problem.
that worries about weight can be reflected in how probable the evidence makes the ultimate hypothesis – that the defendant committed the alleged acts. We then take the missing evidence into account by estimating what the influence on the relevant probability could be if we did have it. However, as Nance (2008, pp. 633-9) points out, the problem with this is that we do not know what the content of the missing information is. We may sometimes guess this content, for example, if we know that one party repressed evidence, then that evidence likely supported the other side’s case. But, as Nance (2008, 274) writes “that gives little hint of an answer to the question, ‘By how much?’” Similarly, we cannot reasonably draw inferences about whether unconceived alternatives would support the defendant’s guilt or innocence and by how much. We therefore cannot take the impact of these unconceived alternatives into account before we discover them by including a catch-all hypothesis or by adding the scenario that ‘something else happened’. There is simply no way to evaluate the unconceived alternatives in a way that would reflect their potential impact. Any value that we assign to the catch-all does not necessarily correspond to what we would assign to these alternatives were we to discover them.

Admittedly, there are situations in which we may justifiably exclude the catch-all from our inferences, even if our set of alternatives is not exhaustive in a strict sense of the word. For example, Fitelson & Thomason, (2008 26) point out that we may sometimes believe that every unconsidered possibility may be so implausible as to be negligible. We can then presume that the prior probability of the unknown hypotheses is (very) low and that the sum of the prior probabilities of the known theories is therefore (very) high. We are then justified in treating our current set of hypotheses as exhaustive. Indeed, this is arguably a commonplace assumption in criminal law, where fact-finders only look at a select few scenarios, but treat those as if they exhaust the hypothesis space. But the question I want to ask is, what if we cannot justifiably make this presumption – because we have good reasons to suspect plausible unconceived alternatives? In the next section, I argue that we should then have a reasonable doubt about the defendant’s guilt.
5. Reasonable doubt from unconceived alternatives

The upshot from the previous section is that we cannot meaningfully take the possibility of unconceived alternatives into account when we assign plausibility or probability to the available scenarios. Yet suppose that we (a) have good reasons to suspect the existence of unconceived, plausible alternatives, but (b) do not take this possibility into account when evaluating our set of scenarios. In other words, we let go of the presumption that our set of scenarios is exhaustive. If we do so, we can no longer draw conclusions about the absolute probability of the defendant’s guilt, which is what the beyond a reasonable doubt standard requires.

For instance, some epistemologists propose Bayesian accounts that do not rely on an exhaustive set of hypotheses. Salmon (1990) is an example. He suggests that we should only consider conceived alternatives when evaluating the confirmation of a given hypothesis. But, as Rowbottom (2016, 3) points out in response, if we only consider conceived alternatives, this means letting go of the assumption that we are evaluating whether a theory is “truth-like.” Similarly, Wenmackers & Romeijn (2016) propose an ‘open minded’ version of Bayesianism, which drops the assumption “implicit in standard Bayesianism – that the correct empirical hypothesis is among the ones currently under consideration”. However, they admit that their approach “fails to provide us with the required normative guidance” about the absolute confirmation of scientific theories, because it only tells an agent what to believe if she supposes “without committing to it, that the true theory is among those currently under consideration” (Wenmackers & Romeijn, 2016, 1243) In other words, if we drop the presumption of exhaustiveness, we cannot consider the absolute probability of any hypothesis to be high.\(^\text{185}\) Though the authors just mentioned write primarily in the context of the philosophy of science, this point is also applicable to legal proof. If we cannot presume that our set of hypotheses is

\(^{185}\) Nor can we always say that it is low.
exhaustive then we cannot assign a high probability to any hypothesis. The demand for proof beyond a reasonable doubt – that the defendant is very probably guilty – then cannot be met.

What about explanation-based accounts? As mentioned in section 3, the bad lot problem arises for inference to the best explanation when we cannot presume that the true hypothesis is probably amongst those considered – which is precisely the presumption that we are rejecting here.\textsuperscript{186} However, as I also discussed in that section, many explanation-based accounts do not rely on inference to the best explanation. They require that there is a plausible guilt explanation and no plausible innocence explanation. Nonetheless, without the presumption of exhaustiveness, these accounts too only lead to comparative conclusions. We then cannot conclude that there are no plausible alternatives. At best we can infer that no plausible alternatives have been presented in court. But we do not care only about what has been presented in court. We also want to know whether what has been presented in court reflects the actual strength of the case for innocence and guilt.

The point of the beyond a reasonable doubt standard – both on the Bayesian and explanation-based account – is to allow conviction only when the defendant probably committed the alleged acts. But we cannot draw such absolute conclusions while at the same time taking the possibility of unconceived alternatives into account. Some may bite this bullet and suggest adopting a comparative account of rational criminal proof.\textsuperscript{187} As far as I am aware, only Cheng (2013) has made such a proposal. He argues that the reasonable doubt standard could be reconceptualized as a Bayesian likelihood-ratio. On his account fact-finders should compare “a single defense narrative of innocence versus a single prosecution narrative of guilt” and

\textsuperscript{186} In response, some philosophers of science have also offered strictly comparative accounts of inference to the best explanation – where we merely infer that one explanation is more likely to be true than the other available ones (e.g., Kuipers, 2004).
\textsuperscript{187} While Allen & Pardo (2019) call their approach the ‘relative plausibility theory’, their account does not lead to relative conclusions in criminal cases. See section 2.2.
consider the latter proven if it has a sufficiently higher likelihood ratio. However, I am not certain that Cheng indeed intends his account to be merely comparative or whether he sees some connection to the absolute probability of guilt. After all, the price of bullet-biting is high, as it severs the connection between the proof standard and its epistemic aim of fair error distribution.

So, how do we deal with unconceived alternatives without biting this bullet? The solution that several philosophers propose – and that I also endorse for the legal context – is to treat the exhaustiveness of our set of hypotheses as a prerequisite for our Bayesian or explanationist inferences to be justified. For example, as Amaya puts it, in order for IBE to work, our set of explanations should be ‘good enough’. For Bayesianism, the story is the same – the presumption that our set of conceived possibilities is exhaustive can be seen as a prerequisite for Bayesian inference (see e.g., Gillies, 2001).

Note that by prerequisite, I do not mean that we first have to determine whether there are unconceived alternatives before we can engage in Bayesian or explanation-based reasoning. As I will argue in the next section, how plausible our current scenarios are also informs us about whether there are unconceived alternatives. Additionally, as discussed earlier, we should sometimes consider a scenario plausible because we are confident that there are no alternatives. So, while these are two separate questions, they are interrelated and may have to be answered in conjunction with one another. The question of unconceived alternatives therefore does not come prior to whether our conceived scenarios are plausible or probable. Rather, what I mean by prerequisite is that if we cannot justifiably believe that our set of explanations is sufficiently exhaustive then we are also not justified in accepting the conclusions of our Bayesian or explanationist inferences.
6. Justified belief in no unconceived alternatives

Given our human, cognitive limitations, we usually cannot be certain that we have considered every plausible alternative. Nonetheless, we may be justified in presuming that we have. The challenge is then to spell out when this presumption is justified. On my account this depends on how strong our evidence is for the existence of plausible unconceived alternatives. This is therefore an evidentialist account. Before I delve further into what counts as evidence for unconceived alternatives, I first contrast my account with that of Amaya (2009). She is, to my knowledge, the only other author who has tackled the question how we may reason about unconceived alternatives in the context of legal proof. She proposes a responsibilist framework, which she explicitly contrasts with an evidentialist account. In this section I reject her position in favour of an evidentialist account, though I also argue that the two are not very far removed from one another.

6.1 Responsibilism

Amaya discusses the problem of unconceived alternatives when she defends her inference to the best explanation-based account of rational criminal proof from the bad lot problem. In order to avoid bad lots, she proposes that we are only justified in inferring to the best explanation if our set of explanations is ‘good enough’. Without this, she suggests, we may not conclude that the best explanation is likely to be true. Her aim is therefore the same as what I seek to achieve in this chapter, though her solution differs from mine.

Amaya distinguishes between responsibilist and non-responsibilist (or evidential) views of justification. On the first, justification is about what an agent has done (or failed to do) to ensure that her beliefs are true. According to Amaya (2009, 154) on responsibilism, “if one has done all that one can be expected to do for insuring that one’s claim is not defeated by an alternative explanation in the particular case, there is an important sense in which one’s claim may be said to be justified.” On evidentialism,
justification is analyzed purely in terms of evidential support. Amaya does not spell out how exactly she understands this position. However, I take the evidentialist account to mean that whether we are justified about assuming that there are no unconceived alternatives depends on our evidence regarding the existence of such alternatives. Such evidence can include whether investigators did what was expected of them. However, our evidence can also indicate that there are likely no unconceived alternatives even when investigators failed in doing what is expected of them, or it can indicate that there likely are unconceived alternatives despite investigators doing what is expected of them.

According to Amaya, we ought to be responsibilists. This means “complying with some epistemic duties and exercising a number of epistemic virtues in the course of inquiry and deliberation about factual problems in law” (Amaya, 2009, 155). She lists some of these virtues: “open-mindedness in collecting and appraising evidence, perseverance in following a line of inquiry, or readiness to change one’s views in the face of new conflicting evidence” (Amaya, 2009, 155). So, whether we may presume that the set of scenarios is good enough depends on whether investigators and prosecutors have acted in an epistemically virtuous way (and nothing else).

Note that on Amaya’s account, these epistemic duties apply to the prosecution and investigators. We may distinguish her proposal from the responsibilist position advocated by Picanelli (2015) who interprets the beyond a reasonable doubt standard in terms of epistemically responsible behaviour on the part of the jurors. Admittedly, I am not completely certain that this is the correct interpretation of Amaya’s position. In her writing on the bad lot problem, she seems to require actual responsible behavior from the relevant agent, as she writes about IBE being warranted only if “one’s set of alternative explanations has been constructed in an epistemically responsible way” (Amaya, 2009, pp. 154-5). However, in other writing she adopts a counterfactual responsibilist position (Amaya, 2008; 2015). This means that a belief is justified if it could have been the outcome of an epistemically responsible process, even if the actual agent holding the belief did not act epistemically responsible. I am inclined to read her position in the former way not only because it most closely fits with how she presents it herself when talking about the bad lot problem, but also because she is not clear on how we ought to assess whether the belief could have been the result from an epistemically responsible process.
Though Amaya does not argue why we ought to be responsibilists when discussing the bad lot problem, she does offer reasons for this elsewhere: while our evidence may support a belief, this may only be the case because this evidence was gathered in an epistemically irresponsible (e.g., biased, lazy) manner (Amaya, 2008). Such a belief is then not actually justified, even though the evidence supports it. We can imagine a similar justification with respect to unconceived alternatives: a scenario that implies defendant’s guilt may be quite probable given our evidence, but only because we tried insufficiently hard to coming up with alternative scenarios. Additionally, a commitment to responsibilism is generally understandable in the context of criminal law. Epistemic duties on the part of the prosecution are crucial in criminal proof. In particular, the prosecution holds the burden of proving the defendant’s guilt. If a prosecutor or investigator has acted epistemically irresponsible, this should undermine the case that they try to build against the defendant.

However, despite this intuitive plausibility, the responsibilist account is unsatisfactory as a solution to the problem of unconceived alternatives. Prosecutors and investigators acting epistemically responsibly is neither a necessary nor a sufficient condition for being justified that we have uncovered all plausible alternatives. To see why, consider two fictional cases:

*The lazy detective:* Detective A is lazy, biased and stuck in her ways. She displays few if any of the epistemic virtues that we’d desire in an investigator. Nonetheless, in the case she is currently working on – the robbery of a jewelry store – she is lucky. She arrives at the crime scene and the store’s employees have already placed the alleged robber under citizen’s arrest. The suspect immediately confesses. Finally, the employees show the detective camera footage of him threatening the employees and stealing the store’s jewelry. After sloppily taking

The counterfactual reading of her work is therefore more vague than the reading that I adopt here.
statements, the detective goes to get a hamburger. She justifies her lack of a proper investigation by telling herself that ‘he obviously did it, no need to look further.’

The virtuous detective: Detective B, in contrast, is a hard-working investigator who approaches each case with cleverness and objectivity. However, the current case that she is working on – a murder – proves to be especially complex. It features a host of conflicting witness statements, potentially misleading traces and a towering stack of notes about possible leads. The detective does have a primary suspect who seems to have a means, motive and opportunity. She has also tried hard to come up with a plausible alternative explanation, but failed. However, despite her hard work, she still has the feeling that there may have been something that she missed, as she does not yet see all the connections between the different aspects of the case.

I believe that both detectives are right. The lazy detective justifiably believes that she did not miss any plausible alternatives, despite her lack of a virtuous investigation. After all, the case was straightforward. Any scenario in which the defendant is innocent would have to somehow imply that the testimony of both the employees and the defendant himself, as well as the camera evidence are all misleading. There is no reason to believe a scenario that implies such misleadingness could be plausible. Hence, a responsible investigation is not a necessary condition for justifiably believing that one’s set of explanations is exhaustive. In the second case, the investigation was responsible. Nonetheless, the detective is not justified in presuming exhaustiveness; she has good reasons to suspect that she missed some plausible alternatives. So, a virtuous investigation is also not a sufficient condition for the
presumption of exhaustiveness to be justified. Responsibilism is therefore inadequate as a solution to the problem of unconceived alternatives.\textsuperscript{190}

6.2 The evidentialist view

If not responsibilism, then what? I believe that Amaya rejects the evidentialist option too quickly. On my view, whether we may presume that we have considered all plausible alternatives depends on our evidence. That the underlying investigation displayed certain epistemic virtues is one important (if not the most important) type of evidence, but not the only one. For instance, while the lazy detective case featured some evidence in favor of unconceived scenarios existing (the unvirtuous investigation), the evidence as a whole supports the presumption that no plausible alternatives exist. In contrast, in the second case the available evidence does give good reasons to suspect that she may have missed some plausible alternative, even if her investigation was virtuous.

On the evidentialist view we assess whether our evidence indicates that our investigation missed relevant possibilities. Amaya (2009, 155) herself already comes close

\textsuperscript{190} Note that my examples relate to the \textit{detective} holding a justified belief. Nonetheless, the fact-finder – i.e., the judge or jury – will often base their decision on the set of explanations constructed during the preceding investigation by the investigators. They will then be similarly unjustified in holding a belief in exhaustiveness. For the sake of completeness, let me also mention that responsibilists such as Amaya are often concerned with the \textit{justification of a belief} rather than whether a person is \textit{justified in holding that belief} (Baehr, 2009; Cloos, 2015). They note that there are cases where the two diverge – i.e., where a belief is justified, but the agent who has that belief is not justified in holding it. It is for this reason that Amaya develops the counterfactual position mentioned in the previous footnote. I have no objection against this counterfactual position as a standard of justified belief. However, in this chapter I am concerned with when the fact-finder may justifiably believe that her set of explanations is exhaustive. I am not certain whether this is also the question that Amaya (2009) tackles when she discusses the bad lot problem. Nonetheless, whether the fact-finder is justified in presuming exhaustiveness is the appropriate question to when we talk about the problem of unconceived alternatives. After all, it is the fact-finder who has to decide whether the set of scenarios is good enough to convict on. I am interested in the rationality of this decision. The above examples show that responsibilism is an inadequate answer to this question.
to describing such an evidentialist view when she writes: “provided that one has conducted a thorough search for other potential explanations and there is no reason that justifies a further search, then one is justified (in the sense that matters) in accepting as justified the best explanatory hypothesis of the events at trial. However, whether a further search is justified not only depends on whether investigators were virtuous in their conduct during the investigation, but more generally on also on whether we have a reasonable expectation of finding further relevant scenarios. This will depend on our evidence regarding how likely it is that such alternatives exist. So, Amaya’s position may actually be close to the evidentialism that she rejects. However, my evidentialist account is slightly different still. There can also be situations in which we both know that we have likely missed something and that a further search would not be likely to produce anything else. Consider the case of the hard-working detective again. Her previous failures to find alternatives might give her an inductive reason to suspect that similar, future endeavors will also not yield anything – that she will not be able to see all the connections between the evidence and thereby come up with all plausible explanations. Nonetheless, her failure to grasp these connections provide her with a good reason – i.e., good evidence – to suspect that she has missed something.

As I argued before, the question how probable it is that plausible, unconceived alternatives exist is not the same as the question what the impact of actually discovering such possibilities would be. This latter question will often be impossible to answer because it would require us to know the details of scenarios that we have not even come up with (see section 4). In contrast, the first question is one that we can – at least sometimes – meaningfully reason about. Various philosophers of science write about the kinds of evidence that indicate (an absence of) unconceived alternatives. For example, Stanford (2006) invokes the failure of past scientists to conceive of all relevant theories to argue that contemporary scientists have likely also missed such alternatives. Musgrave (1988) and Lipton (2004) note that, in order to assuage the worry of unconceived alternatives, our best scientific theory should
also be ‘satisfactory’ or ‘good enough’. Dawid, Hartmann & Sprenger (2015) suggest that the possibility of unconceived alternatives plausibly depends on the difficulty of the relevant scientific problem, on the cleverness of the scientists and on the available resources to investigate the problem. Similarly, Dellsén (2017) argues that the probability of unconceived alternatives remaining undiscovered depends on the complexity of the domain, how dogmatic and skilled the scientists in that domain are and how long they have been searching for alternatives. Using these ideas as inspiration, I want to propose that at least three types of evidence are relevant for assessing whether we likely uncovered all plausible scenarios:

A. **Quality of the investigation:** All other things being equal, the better the search for alternative scenarios the more reason we have to presume that the investigators have uncovered all relevant possibilities. The quality of such a search depends in part on the amount of time and resources spent on it as well as on the imaginative faculties of the investigators and on the methods used during the investigation. It also depends on the kinds of virtues central to Amaya’s account, such as open-mindedness and perseverance. Whether investigators displayed these virtues can also depend on the nature of the case. For instance, as Amaya (2015, 517) points out, in emotionally disturbing cases, investigators may be more likely to be biased, thereby failing to conceive of plausible alternatives.

B. **Quality of the conceived scenarios:** All other things being equal, the better our conceived scenarios, the more reason we have to presume that no plausible alternatives exist. If our conceived scenarios explain everything adequately, then this gives us a good reason to suspect that there are no relevant alternatives. Conversely, if we only have implausible scenarios, then we have good reasons to suspect that there is a better scenario that we have not conceived of. Furthermore, our existing explanations can be of a high quality in the sense that they are specific. It is sometimes easier to think of alternative explanations when we know precisely what we are seeking an alternative for.
C. Quantity of the information: All other things being equal, having too much or too little information can be a reason to suspect that we have overlooked plausible alternatives. The information in a case includes at least the evidence, explanations and arguments. We may firstly have too much information. For instance, the case may be accompanied by a thick case file, indicating a great deal of potentially relevant evidence, which may also conflict or be incomplete. It may then be difficult to ‘visualize’ what is going on – i.e., to see how all the facts hang together (van Oorschot, 2014). Additionally, much information can create ‘noise’, where irrelevant information drowns out the (more) relevant. The same goes for having too many possible explanations to consider or having to make sense of many competing arguments. Such an abundance of information can make it difficult to judge whether all plausible alternatives were considered. This is, for instance, why the virtuous detective has good reasons to suspect further unconceived alternatives. Second, we can have too little information. For example, if a victim is found in the middle of the forest, with no witnesses, obvious fingerprints or other marks of a possible perpetrator, and if it is not clear whether he was killed or died of natural causes, then it is almost pure guesswork as to what happened. In contrast, if there are multiple witnesses and other evidence that clues us in on what happened, then there is less of a chance that we have overlooked something.

Let me say something about how these criteria related. They are not necessary and sufficient conditions for being justified in believing our set of scenarios to be exhaustive. Instead, they are items of evidence which may jointly justify such a belief. As with any form of evidential reasoning, whether a belief in exhaustiveness is justified will depend on the particulars of the case and may involve weighing these factors against one another. Furthermore, the stakes involved can also be relevant. Cases with higher stakes typically require better evidence for our belief to be

---

191 As I explain in the next chapter, this type of evidence is ‘higher-order evidence’.
7. Conclusion

In criminal cases, reasonable doubts sometimes arise about the defendant’s guilt. For instance, our evidence may be insufficiency strong, the prosecution might fail to present a plausible scenario or the defense can present a plausible alternative scenario. Yet another, mostly overlooked, source of reasonable doubt is the possible existence of plausible, unconsidered alternatives. There are specific situations in which we have good reasons to presume that we might have failed to imagine some relevant alternative scenario. In such cases, we cannot consider the guilt of the defendant proven. In this chapter, I discussed this claim through the lens of the Bayesian and the explanation-based accounts of rational criminal proof.

What this normative, epistemic claim means for the practice of criminal proof will depend on the particulars of the legal system. One option is that the problem of unconsidered alternatives imposes a burden on the prosecution to prove that all relevant possibilities were considered. On the other hand, it could also be up to the defense to raise possibility of unconsidered alternatives in specific situations, to argue that there is a reasonable doubt about the defendant’s guilt. Perhaps the burden to ensure that our set of scenarios is good enough should be placed at the feet of investigators. Alternatively, writing in the context of the American system where the fact-finder is the jury, Nance (2008) suggests that the decision whether the available set of evidence is sufficiently complete for the case to go to trial lies with the judge in the case. Whether we suspect unconsidered alternatives could be part of this decision.

192 For instance, Amaya (2015, 527-8) and Josephson, (2002, 1626) both suggest that the stakes matter for proof beyond a reasonable doubt.
In the next chapter I expand upon the argument made here. I will suggest that not only the possibility of unconceived alternatives, but of other kinds of overlooked information can also lead to a reasonable doubt.
VIII. REASONABLE DOUBT, ROBUST EVIDENTIAL PROBABILITY
AND THE UNKNOWN

Chapter abstract

Most legal evidence scholars agree that proof of guilt beyond a reasonable doubt requires the belief that the defendant probably committed the alleged acts. However, they also agree that this is not a sufficient condition, as this belief may be unreasonable. I focus on two popular proposals for additional conditions: (i) that the degree of belief should be robust and (ii) that it should be calibrated to the evidence. Both criteria face difficulties regarding their meaning and utility. I propose a new interpretation which overcomes these difficulties and which combines the two conditions into a single criterion. On this interpretation, both conditions are about the possibility of overlooked, exculpatory information. Proof of guilt beyond a reasonable doubt requires that we are sufficiently certain that we did not miss such information for the sake of achieving accurate beliefs and a just error distribution. How certain we can be that we did not miss anything relevant depends on our higher-order evidence. One important type of higher-order evidence is whether the hypotheses under consideration are sufficiently detailed, as this helps reason about whether any exculpatory information was missed. This point leads to a novel way of combining the Bayesian and explanation-based accounts of rational legal proof.
1. Introduction

There is wide-spread agreement among evidence scholars that proof of guilt beyond a reasonable doubt requires at least that fact-finder believes that the defendant (very) probably committed the alleged acts.\(^{193}\) For example, if proving murder requires the defendant to have killed someone with intent and premeditation then the judge or jury should not convict him unless they believe that he likely did so. As I have mentioned in previous chapters, the idea that we should conceptualize the BARD standard as a probabilistic threshold is an important part of Bayesianism.\(^{194}\)

However, a high degree of belief cannot be all there is to proof of guilt beyond a reasonable doubt. After all, the fact-finder’s belief may be unreasonable. For example, it may be the result of a biased interpretation of the available evidence or it could be based on a set of evidence that is the product of a flawed investigation. Various proposals for further requirements have been put forward. In this chapter, I focus on two popular ones. The first is that the degree of belief should be robust (or resilient, safe or stable) (Ho, 2008, 278; Stein, 2005, 88; Di Bello, 2013; Dahlman et al., 2015; Urbaniak, 2018). Briefly put, on this requirement the fact-finder’s belief should not be easily overturned. I touched upon this requirement in chapter III. The second is that one’s belief should be calibrated to the evidence (that it should be a ‘reasonable’, ‘epistemic’\(^{195}\), or ‘evidential’ probability) (Nance, 2016; Wittlin, 2019; Spottswood, 2019; Hedden & Colyvan, 2019). The idea behind this requirement is that one’s belief should be in line with what the available evidence supports. I introduced this idea in chapter II and discussed it in the context of eyewitness evaluation in chapter VII. In this chapter I expand upon the discussions from the aforementioned chapters. I focus on the probability of the ultimate hypothesis –

\(^{193}\) Many authors put the required degree of belief around 0.9–0.95 (e.g., Laudan, 2011; Walen, 2015; Gardiner, 2019).

\(^{194}\) See chapter III, section 2.2 for a discussion of the Bayesian interpretation of the BARD standard and how it differs from the explanation-based interpretation.

\(^{195}\) Note that I have used the term ‘epistemic probability’ so far as equivalent to a degree-of-belief based, or ‘subjective’ probability interpretation. I mention the term here because some authors also use it to refer to the idea of evidential probability.
that the defendant committed the alleged acts. When is this probability evidential and robust?

Both criteria face difficulties. As I will argue, for the most common interpretation of robustness, the problem is that it is unclear why it is worthwhile to have robust beliefs. For evidential probability the worry is that the notion is so vague that it does not tell us when one’s belief is actually calibrated to the evidence. In this chapter I offer a new interpretation of both terms which overcomes these difficulties. On my account both criteria are about the same issue: the need to take into account the possibility that we have overlooked exculpatory information. Such information includes at least further evidence, alternative scenarios, arguments, interpretations of the evidence or connections between our items of evidence. Of course, we can never be certain that we did not overlook anything, nor can we know whether what we may have overlooked was exculpatory. Nonetheless, we can be justified in presuming that we considered all pertinent information. If this presumption is not justified there is reasonable doubt, or so I shall argue. More precisely, I propose that too great a possibility of missed exculpatory information should lead to a reasonable doubt for the sake of two central goals of criminal trials: error minimization and error distribution. Error minimization means that criminal fact-finding should lead to as few errors as possible. The goal of error distribution is that, to the extent that errors are made (as is unavoidable in a legal system), these errors should mostly be false acquittals, not false convictions.

Whether the presumption that we did not overlook anything is justified depends on our higher-order evidence. Examples of such evidence are the quality of the underlying investigation and the complexity of the case at hand. Another important type of higher-order evidence is how specific the hypotheses under consideration were. More specific hypotheses can be critically tested more easily and give us guidance on what evidence we may have overlooked and on whether there are alternative plausible explanations. So, thinking in terms of well-specified hypotheses helps ensure that the belief in the defendant’s guilt is robust and supported by the available
evidence. This observation leads to a new way in which the Bayesian and explanation-based accounts of rational proof can be combined. In particular, while suggestions on how the two may go together typically focus on how explanation-based reasoning leads to probably true beliefs about the defendant’s guilt, such thinking therefore similarly leads to an increase in our higher-order probability.

I begin this chapter by explaining the concepts of robustness and evidential probability and their main problems and I offer an alternative interpretation of both (section 2). After that, I argue why there should be reasonable doubt if these criteria are not met (section 3). I then turn to the concept of higher-order evidence, give some examples of such evidence and briefly discuss the notion of higher-order probability (section 4). Finally, I explore a novel way of marrying explanation-based reasoning with the Bayesian account of rational criminal proof (section 5).

2. Robust beliefs that respect one’s evidence

As I explained in chapters II and III, the Bayesian framework of rational proof casts legal proof in probabilistic terms and understands the beyond a reasonable doubt standard of proof as a probability threshold. More precisely, according to Bayesianism, we should only consider guilt proven if the posterior probability of the defendant having committed the alleged acts is very high – for instance, at least 90%, 95% or even 99% (Gardiner, 2019). The relevant probability is usually understood as the fact-finder’s degree of belief in the ultimate hypotheses. However, as I discussed in section 2.4 of chapter II, having a high degree of belief in the defendant’s guilt is not sufficient for proof of guilt beyond a reasonable doubt. First, this is too permissive a standard, as the fact-finder’s belief may be irrational. Second, we want our proof standard to offer guidance on when the evidence is sufficiently strong to justify a belief that the defendant probably committed the alleged

\[196\] Meaning the hypothesis that the defendant committed the acts described in the indictment.
acts. A standard that is formulated purely in terms of the fact-finder’s degrees of belief fails to do so as it starts from the premise that the fact-finder already believes that the defendant is probably guilty. As Laudan (2006, 80) writes “a proper [proof standard] does not depend on one’s subjective confidence in a hypothesis; on the contrary, the standard tells us whether our subjective confidence is justified.” So, this interpretation of the proof standard presumes what it is supposed to deliver – it is based on the fact-finder’s degree of belief, but should be telling the fact-finder when they may justifiably hold such a degree of belief. Or, to phrase it differently, such an account puts ‘the cart before the horse’ (Cox, 2000, 323). In response to such worries, evidence scholars have suggested additional requirements for proof of guilt beyond a reasonable doubt. In this section I discuss two such requirements, namely evidential probability and robustness. These two conditions target distinct ways in which the fact-finder’s degree of belief can be unjustified.

2.1 Evidential probability

Recently, a number of evidence scholars have suggested that to count as rational, an agent’s degrees of belief should be ‘calibrated to the evidence’, ‘reasonable’, ‘epistemic’, or ‘evidential’ probabilities (Nance, 2016; Wittlin, 2019; Spottswood, 2019; Hedden & Colyvan, 2019). I already touched upon this idea in chapter II, section 2.4 and in section 3 of chapter VI. To recap some of the discussion there, on this probability interpretation, the high probability required for proof of guilt beyond a reasonable doubt refers to the degree of belief that is reasonable to hold in the face of the available evidence. In other words, it is the degree of belief that a reasonable agent with the same evidence would have – regardless of whether the fact-finder actually holds this credence (Hacking, 2001, 130). For example, suppose that

---

197 To illustrate why the purely subjective Bayesian interpretation is unsuitable as a proof standard, Laudan (2006, 79-80) equates it to telling a mathematician that a theorem is proven if they believe that it is true or telling an epidemiologist that they may consider a causal link between A and B proven if they are highly confident that such a link exists.
multiple witnesses testify that they saw the defendant somewhere else during the
time of the alleged crime, that these testimonies are consistent with the defendant’s
story and that there is no reason to doubt the reliability of these witnesses. Addition-
ally, imagine that there is no incriminating evidence. In such a case it would
obviously be unreasonable to assign a high probability to the hypothesis that the
defendant is guilty. Conversely, imagine that all the evidence points to the defend-
ant’s guilt, it would then be unreasonable not to assign a high probability of guilt,
even if the fact-finder happens to believe that the defendant is innocent.

When does one’s belief conform to the evidence? Most legal evidence scholars who
discuss the notion of evidential probability do not offer an account of when this is
the case.198 As Timothy Williamson (2002, 211), one of the most prominent authors
who discusses evidential probability, points out, such an account may not be
needed. According to Williamson (2002, 209), even if we cannot give a precise ac-
count of what evidential probability is, we may often perfectly intelligibly ask: “how
probable is [a given hypothesis] on present evidence?” He therefore suggests that
we should regard evidential probability as a primitive – a concept which cannot be
spelled out in terms of more fundamental concepts, but which is clear enough in
context. However, while I agree that we may sometimes reasonably ask what con-
clusions the evidence objectively supports and that the fact-finder’s answer to this
question do not need to align with their personal belief. Sometimes they may con-
clude that ‘I believe that the defendant is guilty, but my evidence does not support
this belief’. For instance, they may be aware that their belief is based on inadmissible
evidence. Conversely, they may conclude that guilt is likely given the available evi-
dence, even if they do not personally believe that the defendant is guilty. For

198 One notable exception are Hedden & Colyvan (2019) who suggest that we could deter-
mine the evidential probability of a hypothesis by considering how well that hypothesis
explains the evidence and how intrinsically plausible it is prior to investigation. However,
they do not expand on this brief suggestion. Additionally, their suggestion is made in the
context of a discussion on how the Bayesian may defend their account against explana-
tionist critique and the remark can perhaps best be read as a suggestion on how we could
interpret this concept, not about how we should.
instance, they may recognize that their belief that the defendant is innocent is based
on wishful thinking. Nonetheless, there will also be cases in which the fact-finder
does need more guidance than the slogan ‘just look at the evidence’. For instance,
suppose that two jurors or two judges disagree over whether the evidence supports
the belief that the defendant is guilty. How might they settle this disagreement? Or
imagine that a fact-finder believes that the defendant is probably guilty, but is not
entirely certain about whether this belief conforms to the evidence. How should
they then decide? In both cases we need a more precise account of how the fact-
finder should determine whether their belief conforms to the evidence.199 However,
it is not clear that any existing epistemic framework will do for this purpose. Red-
mayne (2003) surveys various ideas that we could draw upon for this when when
he discusses a proposal by Goldman (1999) about ‘objective’ probabilities in legal
contexts. Redmayne concludes that they are all either too limited in scope to capture
the evidential richness of legal proof, or so vague that they are uninformative.

In this thesis I propose a way of spelling out this idea in terms of whether the fact-
finder has overlooked exculpatory information. This may seem strange suggestion
at first: evidential probability relates to how probable a given hypothesis is in the
light of information that we do have. However, as I will argue in a moment, this
leads to an account which meets the required desiderata. What are those desiderate
for an adequate account of evidential probability? First, this requirement must be
more informative than just the slogan ‘look at the evidence’. Second, it should be
able to capture the evidential richness of legal proof – it must not be restricted to
statistical information.200 Third, the fact-finder must be able to determine whether

---

199 Another reason why we should strive for a more precise account of evidential proba-
bility is that, as outlined in chapter II, this is a crucial notion within the Bayesian account
of rational legal proof. How plausible and informative this account is will therefore depend
in part on how we spell out this interpretation.

200 As noted in section 2.4 of chapter II, to the extent that Bayesians have given precise
formulations of how rational degrees of belief should be connected to the strength of the
evidence, these have been ‘chance-credence principles’, which stipulate that one’s degrees
of belief should align with ‘objective’ data, usually in the form of frequencies.
the condition has been met. For instance, suppose that we understand evidential probability as the degree of belief that a ‘wise’ person would have given the same evidence (Acree, 2021, 273). The fact-finder cannot meaningfully determine what a wise person would believe.\textsuperscript{201} Fourth, we should offer an account of why the condition is epistemically rational (Acree, 2021, 273). In the case of legal proof this means that we should spell out why it is epistemically rational to acquit if the requirement of evidential probability is not met. I take up this last question in section three of this chapter. I first want to offer a novel account of evidential probability that is more clear than a mere slogan, not restricted to statistical information and determinable for the fact-finder.

In chapter II I suggested that evidential probability can be defined as ‘probability that is based on evidence that has been made sense of adequately’. Evidence is never simply given, it has to be discovered or chosen out of the known information and interpreted. Only then can we draw conclusions from it. As I have argued at various points in this thesis, making sense of the evidence is far from a trivial task. What does it mean to perform this task adequately? This question is too large to answer in this chapter or even in this thesis (and I am doubtful that a fully satisfactory answer is possible). Nonetheless, here is my attempt at an answer. When we make sense of the evidence, we see how all the facts hang together (van Oorschot, 2014). A large part of making sense of the evidence well is therefore seeing relevant connections. These connections can be between the evidence and the hypothesis, but also between the items of evidence themselves. For instance, as outlined in chapter VI, if witnesses agree with one another in their testimonies, this will typically have an impact on the combined strength of these testimonies. Furthermore, apart from seeing connections between the evidence, making sense of the evidence at the very least also involves coming up with hypotheses that explain the evidence.

\textsuperscript{201} Unless they take themselves to be that wise person. However, as Acree (2021, 273) points out, this would risk collapsing the notion of evidential probability back into the subjective condition that it was supposed to supplement; evidential probability is then just whatever conclusions the fact-finder believes are justified given the evidence.
For instance, there may be alternative versions of what happened which are consistent with the defendant’s innocence. To give an example of this, suppose that in a murder case there is, apart from the defendant, another person who could have realistically killed the victim. Our evaluations of the evidence can be unreasonable if we overlook such explanations. For instance we may then overvalue the strength of the case for guilt. Additionally, we can be unreasonable because we overlook inconsistencies within the explanations that we have come up with. To give an example, a particular scenario may seem plausible, but only because we have overlooked that it implies that the defendant traveled from point A to point B in an unrealistically short time – the hypothesis is then internally inconsistent. Finally, our evaluations of the evidence become more reasonable to the extent that we do not overlook relevant arguments about, for example, the reliability of an item of evidence. Such arguments include objections to the background beliefs that we use in the evaluation of the evidence and hypotheses. These beliefs can be unreasonable, for instance if they are contradicted by scientific evidence. I want to suggest that to make sense of the evidence adequately involves (at least) conceiving of relevant alternative hypotheses, noticing what evidence is relevant, seeing relevant connections between the items of evidence as well as any inconsistencies within the hypotheses under consideration and becoming aware of pertinent arguments for or against one’s beliefs. When we overlook such information we have not adequately made sense of the evidence.

Of course, even someone who sees all these connections can come to unreasonable assessments of the evidence.\textsuperscript{202} I do not want to suggest that not overlooking relevant information is a sufficient condition for the fact-finder’s beliefs to be

\textsuperscript{202} For example, suppose that a judge holds background beliefs which imply that any exculpatory evidence in favor of a defendant is caused by a massive conspiracy. They may be aware of arguments against their background beliefs, but may consider these arguments to be unconvincing. Such a person might then end up believing that the defendant to be guilty, even if the evidence points to the defendant’s innocence and despite them not overlooking any exculpatory information.
reasonable. However, I do not consider this to be a serious objection to my proposal. Note that what I am concerned with here is providing an interpretation of evidential probability as a condition for proof of guilt beyond a reasonable doubt. Whether guilt has been proven is always a judgment by the fact-finder and, as said, the fact-finder must therefore be able to assess whether this condition has been met in a particular case. To do so they must use the information that they have, interpreted in the light of the background beliefs that they hold. In other words, they must rely on their own judgment and cannot somehow determine whether their beliefs are reasonable according to some absolute standard. What the fact-finder can assess, however, is whether there is a serious possibility that they overlooked something. As I will argue later, this requires the fact-finder to consider their ‘higher-order evidence’. I suggest that if the fact-finder can reasonably believe that they did not overlook any relevant alternative hypotheses, interpretations of the evidence and so on, then they can also reasonably believe that they have made sense of the evidence well.

2.2 Robustness

A second worry for the Bayesian interpretation of proof beyond a reasonable doubt is that, even if our belief lines up with the evidence, this set of evidence may itself be flawed. For example, this set may be too incomplete if investigators, after a lack-luster investigation, failed to gather many crucial items of evidence. Or the set of information may be biased if the investigators suffered from tunnel vision when collecting the evidence. If we had to draw a conclusion based on this evidence, we might, for instance, have to conclude that the defendant’s guilt is proven beyond a reasonable doubt. However, many share the intuition that such a conclusion is unwarranted if it is based on a flawed set of evidence. For this reason, various authors propose that one’s belief should be resilient or robust.
Robustness refers to the stability of one’s degrees of belief. However, I want to suggest that there are actually two meanings of this idea. These two meanings are sometimes mixed up in discussions on robustness in legal evidence scholarship. The first is robustness as sensitivity to new information (i.e., ‘how would our belief change supposing that we were to find further information supporting innocence?’). The second is robustness as the absence of undiscovered exonerating information (i.e., ‘is it realistic to presume that a more thorough search could have discovered information that overturned our belief in guilt?’). Before we can move on with the discussion at hand, we need to disentangle them, as I will argue that the first of these interpretations is problematic and should be rejected while the second is more plausible and should be adopted as a condition for proof of guilt beyond a reasonable doubt.

2.2.1 Robustness as insensitivity to new information

The notion of robustness as insensitivity to new information was introduced to explain how evidential weight is reflected in one’s degrees of belief. Weight refers to how substantial the set of evidence is on which we base a specific belief (Keynes, 1921, 77). For example, suppose that you find a coin. You do not know whether the coin is biased or fair. Imagine that you are considering the probability that the one-hundredth flip of this coin will land on heads. Because you know nothing about the coin, your credence in this hypothesis is 0.5. Now you toss the coin ninety times and find that it does indeed land on heads roughly as often as it lands on tails. As a result, you come to believe that the coin is fair. Your credence that the 100th flip will land on heads therefore remains 0.5. Nonetheless, the two situations are different. In the second case, the credence of 0.5 is based on a greater amount of information. Hence, while your credence has stayed the same, the weight on which that credence is based has increased.
Skyrms (1977) and others have suggested that weight is reflected in the robustness of our beliefs, where robustness is a measure of how much our credence changes in the face of additional data. For example, consider the coin flipping situation again. Suppose that you flip the coin nine times and all nine times it lands on heads. If these were the first nine flips, your belief that the coin is biased towards heads should increase significantly and therefore so should your belief that the hundredth toss will be heads. However, if you have already flipped the coin ninety times, you have a substantial amount of data which suggests that the coin is not biased. Hence, flips ninety-one to ninety-nine all landing on heads should not influence your belief in the outcome of the one-hundredth flip as strongly as in the situation where you did not flip the coin before these nine flips. So, your belief is more robust than in the first case and this is the result of the second probability being based on a weightier set of evidence.203

Some suggest that robust beliefs should be a requirement for proof of guilt beyond a reasonable doubt (Logue, 1997; Ho, 2008; Stein, 2005; Dahlman et al., 2015; Dahlman & Nordgaard, forthcoming; Urbaniak, 2018; Di Bello, 2013).204 However, while it may have some intuitive plausibility, the robustness condition is problematic, as it is unclear why having robust beliefs is epistemically valuable and, hence, why a lack thereof should lead to reasonable doubt. Nance (2016, 270-8) surveys the literature on robustness and finds only two arguments in favor of this position. The first argument is that it may explain our intuitions in cases of ‘naked statistical evidence’ – idealized thought experiments in which a single item of statistical evidence makes it highly probable that the defendant committed an illegal act.205 Many

203 This idea is sometimes expressed in terms of weighty sets of evidence having a smaller confidence interval (Cohen, 1985; Nance, 2016, 254).
204 In this chapter, I focus on a lack of robustness as a source of reasonable doubt. However, note that a solid investigation may also be a requirement for acquittal. For instance, a court may only choose to acquit once a sufficiently thorough investigation has been completed.
205 See chapter II, section 2.3 for an example of such a case, relating to the one-hundred prisoners in a yard.
find it counterintuitive to convict the defendant in such situations, despite the high degree of probability that he committed the alleged acts. Some suggest that this is because this probability lacks robustness; if we would find new evidence, such evidence could easily overturn our belief (Davidson & Pargetter, 1986; 1987). However, this argument is weak at best. Various authors have cast doubt on whether we can draw any conclusions from these intuitions as the cases to which they relate are unrealistic and underdescribed (Allen, 2020; Fratantonio, 2021). Additionally, there are numerous competing accounts that also explain these intuitions (e.g., Nunn, 2015; Krauss, 2019; Di Bello, 2019b; Dahlman, 2020a; Ross, 2020; Enoch & Spectre, 2021; Littlejohn, 2021; Smith, 2021; Dahlman & Pundik, 2021). It is far from clear that a robustness-based account is the best of these explanations.

Logue (1997) gives a second argument for a robustness condition. He suggests that a stable judgment helps to avoid “embarrassment costs” if we discover further evidence that overturns the case. These are costs associated with finding out that the conviction turned out to be erroneous. However, Nance (2016, 272-77) rejects this argument. As he points out, this argument ignores the costs associated with false acquittals. If we were to acquit based on a lack of robustness even when the evidence indicates that the defendant is probably guilty then we similarly run a risk of creating such embarrassment costs, namely if the acquitted would turn out to be guilty afterwards. Hence, Nance writes, “[w]hatever is gained by way of embarrassment costs in terms of the increased stability of positive verdicts can be expected to be more than offset by increased embarrassment costs associated with decreased stability of the additional negative verdicts.”

Apart from a lack of arguments for why robust degrees of belief are epistemically valuable, there are also reasons for why it may, at least sometimes, be bad for one’s beliefs to be robust. First, note that robustness in the sense above refers to the

---

206 Nance (2016, 276) additionally points out that it is unclear how the fact-finder should determine the appropriate level of robustness.
stability of our beliefs in the face of new evidence. However, acquiring more evidence is not the only way to obtain robust beliefs. Another way is by stubbornly refusing to change them in the face of countervailing evidence (Dotson, 2014). While this is obviously epistemically unwarranted behavior, it would meet the robustness condition. Our degrees of belief would not change much in the face of contradicting evidence. The second argument against the value of robustness relates to its connection to the weight of the evidence. I will explain this argument below in greater detail, but, briefly put, the suggestion is that, in legal proof, weight and robustness may be negatively correlated. So, a more weighty set of evidence may correlate with less robust beliefs. Because optimizing the weight of our evidence is epistemically valuable, less robust beliefs will actually tend to indicate greater epistemic value.

To spell out this second argument in greater depth, let us first consider why it is epistemically valuable to optimize the weight of our evidence. To support this claim Nance (2016) refers to a theorem by Horwich (1982, 127-9) in the philosophy of science which shows that as we accrue relevant evidence, our expected error – which is the difference between our credence in a proposition and its truth value – goes down in the long run. Others, such as Allen and Pardo (2007, 134) and Stein (2005, 122-3) note informally that it generally seems to be the case in legal proof that as the amount of evidence increases, factual error decreases.

Weight and robustness can covary – as one goes up or down, so does the other. For instance, as we saw in the coin-flip example above, if we perform a series of independent, identical trials then increasing the weight of the evidence will also increase its robustness. If this is the case then a lack of robustness can indicate a lack of weight. Because increasing weight tends to lead to fewer errors, increasing robustness covaries with something of epistemic value. However, the same is not necessarily true in criminal proof. Criminal investigations are not about conducting repeated, identical trials. Instead, investigators gather a diverse set of evidence. Yet, as Hamer (2012) argues, as the diversity of our evidence increases, the robustness
of our beliefs may decrease. This happens because more investigation often means that we make our hypotheses more specific. We then develop more detailed, fine-grained accounts of what might have happened. Such specific hypotheses are less resilient because it is easier to overturn a specific statement than a general one. For example, consider the following hypotheses:

(i) The defendant was at the crime scene between 9 a.m. and 9:15 a.m.

(ii) The defendant was at the crime scene between 9 a.m. and 5 p.m.

Now suppose that we gather evidence which indicates that the defendant was not at the crime scene before 10 a.m. This falsifies the first hypothesis, but not the second. In the first case we might therefore have to conclude that the evidence does not support the belief that the defendant is (probably) guilty. In the second case the evidence does not lead to this conclusion.

As the repeated use of the words ‘may’ and ‘could’ above indicates, the above argument does not establish that a weightier set of evidence will always lead to less robust beliefs. For instance, the hypothesis that we consider could stay the same as we accrue more evidence. Nonetheless, the argument provides a reason to doubt the connection between robustness and epistemically desirable behavior (improving one’s evidence). In the absence of a convincing argument why robustness is valuable by itself, it is unclear why a belief lacking robustness should lead to reasonable doubt.

2.2.2 Robustness as the absence of exonerating information

The upshot of the above is that robustness as the resistance of our beliefs to change in the light of new evidence is not a plausible condition for proof of guilt beyond a reasonable doubt. However, when we look at some of the authors who defend a robustness condition, it seems that they have a different (and, I suggest, more plausible) concept in mind than resistance to new information. For instance, according
Di Bello (2013, 216) the robustness condition is met when the defense had the opportunity to raise challenges to the prosecution’s narrative of guilt (and took advantage of this opportunity). He suggests that this typically requires the narrative to be ‘complete’ – i.e., it should give a detailed account of what happened, which includes, for instance, the perpetrator, the actions that they undertook and their motive for undertaking this action. A complete narrative helps reach robust beliefs because it allows the defense to more easily ask critical questions.\textsuperscript{207} In other words, according to Di Bello, making our hypotheses specific is an important part of meeting the robustness condition. However, this seems to contradict Hamer’s point that more specific hypotheses tend to be less robust. Does having a more specific hypothesis increase or decrease robustness? The key to resolving this issue is to see that Di Bello is not talking about robustness in the above sense. Whether we critically tested a hypothesis does not matter for how our belief in that hypothesis would change if we were to receive new evidence. Rather, it relates to whether we could realistically expect to find information that would change our belief if we were to search further. This is what critically testing a hypothesis assures us of – that there is no convincing counterargument that we overlooked.\textsuperscript{208}

Other recent defenders of a robustness requirement also (implicitly) employ the idea that robustness refers to the (probable) absence of undiscovered exculpatory information. For instance, Mackor & Van Koppen (2021) suggest that robustness is primarily an assessment of the quality of the search for evidence and possible alternative hypotheses. Similarly, Dahlman & Nordgaard (forthcoming) argue that if important evidence was missed during the investigation, then the case for guilt

\textsuperscript{207} I explore this idea in more detail in section 6.

\textsuperscript{208} To see how the two notions of robustness may come apart, consider a person who is highly diligent in looking for alternative explanations, facts and arguments against their belief, but who did not find any such information. Their belief is robust in the sense that they are justified in believing that they did not miss any information that contradicts their belief. However, also presume that this person’s belief is fickle – if they had found even the slightest scrap of information, they would have immediately revised their belief. Their belief is therefore unrobust in terms of resistance to contradicting information.
lacks robustness. Neither of these suggestions is about the hypothetical situation in which we actually find new information and how the fact-finder’s degree of belief would then change. Regardless of how good our search was, we might imagine a situation in which we gain reliable information of a vast conspiracy against the defendant and our belief in his guilt would (or should) then change. What a thorough search assures us of is that there is no such (currently undiscovered) evidence of a vast conspiracy. It is this notion of robustness, as a justified belief that no exonerating information was missed, that I defend in this chapter.

3. Reasonable doubt from unavailable information

On the interpretation that I proposed, both the robustness and evidential probabili-
ty conditions are about a justified belief that we did not overlook any exculpatory information. However, suppose that we are not very confident that we did not miss anything exculpatory. Why should this lead to a reasonable doubt? In this section I argue that the answer to this question lies in two of the central aims of criminal trials: we want to minimize the number of errors made (error minimization) and, to the extent that we do make such errors, we would prefer them to be false

209 That missing information can be relevant for criminal proof has also been defended by several authors who argue that proof of guilt beyond a reasonable doubt requires that our evidence is ‘sufficiently complete,’ (e.g., Kaye, 1986; Di Bello, 2013). Roughly speaking this condition entails that the prosecution should present any items of evidence that can be reasonably expected of them. More specifically, no items of evidence should be missing that we would expect to have if the prosecution scenario were true and that the most important parts of the scenario should be backed up with evidence (Di Bello, 2013, 208-210). However, though this suggestion is similar to mine, my proposal is broader. Whether we have missed something relates not only to specific, identifiable items of evidence – e.g., particular witnesses not being presented at trial. We may also have reasons to believe that important evidence is missing even if we do not know what this evidence is, for instance if the search for further evidence was lackluster. Furthermore, exculpatory information can, for instance, also take the form of unconceived alternative hypotheses and insights into the connections between the evidence and the hypothesis.
acquittals rather than false convictions (error distribution). Taking into account the possibility that we may have missed exculpatory information serves these aims.

Before I turn to my own account, I want to briefly consider what I believe to be the strongest argument for why possible overlooked information should not lead to a reasonable doubt. This argument will act as the foil for my own position. It combines insights from, among others, the work of Dale Nance (2008; 2016), Larry Laudan (2006) and Amalia Amaya (2009). The argument begins with the observation that, when we are faced with potential missing evidence, a reasonable response is to gather further information (to the extent that this is cost-effectively possible) (Nance, 2008; 2016). In criminal cases this ‘burden of production’ will typically fall on the prosecution, unless the defendant has sole access to the missing evidence (Nance, 2008, 278). To the extent that parties fail to meet this burden, sanctions may be imposed on them, especially if the missing information is the consequence of the prosecution or defense acting in a culpable way – for instance if they kept important evidence behind.210 One such possible sanction is the conclusion that there is reasonable doubt.

Nance writes on missing evidence. As I discussed in section 6 of chapter VII, Amaya (2009, 155) takes up a similar position with respect to the possibility of unconceived alternatives. She argues in favour of a ‘responsibilist’ account on which we may only take our set of explanations to be sufficiently complete if, during the construction of this set, we complied “with some epistemic duties and exercising a number of epistemic virtues in the course of inquiry and deliberation about factual problems in law” (Amaya, 2009, p. 155). Such virtues include, for instance,

210 Admittedly, Nance’s point may not be as normative as I interpret it in this section. His goal is in part to explain how the federal rules of evidence work within the United States of America. Additionally, Nance suggests that within this system, sanctions relating to a party for failing to meet their duty of producing evidence is for the judge, not the jury (i.e., the fact-finder) to decide (Nance, 2016, 241). He does leave open the option that the jury may, in some cases, sanction the party who fails their duties of production. Nonetheless, the conclusion that there is a reasonable doubt due to a lack of weight will only rarely follow on his account.
open-mindedness. So, on her account, whether the possibility of overlooked alternatives can lead to reasonable doubt also depends on the prosecution or investigators acting culpably, by failing to meet their epistemic duties. Such culpable behavior lead to reasonable doubt (or it may lead to further attempts to formulate alternative explanations).

However, suppose that no party acted morally culpable and that no further information can be cost-effectively produced. This, we could argue, leaves the fact-finder with no choice other than to make the best use of whatever information they do have (Nance, 2016, 124-137; Biedermann & Vuille, 2019, 17). Missing information does not give us any reason to change our degree of belief in guilt. After all, we do not know what conclusions the missing information supports.²¹¹ It could be exculpatory but it could also be incriminating. So, according to this argument, if the fact-finder’s information indicates that the defendant is probably guilty, then they should convict. To presume that missing information should always benefit the defendant is to have, what Larry Laudan (2006, 119) calls a ‘pro-defendant bias’. As he points out, the reason why we set the proof standard to a high degree of probability is to distribute the errors fairly. The standard helps us ensure that, to the extent that errors are made, these are mostly false acquittals, not false convictions.²¹² As various authors have suggested, this is the primary aim of the beyond a reasonable doubt standard (Laudan, 2006; Hamer, 2014; Pardo, 2013, 559; Stein, 2005, 138). However, as Laudan (2006, 119-144) argues, many legal scholars have the intuition that other parts of the criminal proof system should also benefit the defendant. For instance, such authors suggest that deficiencies in the production of evidence and in the relevant procedure always benefit the accused, by having these deficiencies lead to acquittal or the case being dropped. However, Laudan argues, this intuition is mistaken. It increases the number of false acquittals beyond the

²¹¹ I made this point extensively with respect to unconceived alternative explanations in chapter VII.
²¹² See chapter III, section 3.3 for a further explanation of this point.
ideal distribution, which is already enshrined in the proof standard, thereby ignoring
the negative utilities associated with that kind of error. Similarly, one could argue,
because we cannot know what the missing information would support, we may not
presume that a possibility of overlooked information should necessarily benefit the
defendant. Hence, the fact-finder should decide based on their available informa-
tion in order to achieve the optimal error distribution.

Though the above argument has some merit, it ultimately fails. In particular, it relies
on what it rejects. The argument presumes that a high proof standard leads to a just
error distribution. However, this is not always true – this claim only holds if the
fact-finder generally assigns high probabilities of guilt to the guilty and low proba-
bilities of guilt to the innocent. If there is no correlation (or a weak correlation)
between the fact-finder’s beliefs about guilt and actual guilt, then a requirement of
a high degree of belief for conviction will not necessarily lead to errors being shifted
in favor of false acquittals. To give a simple (and admittedly silly) example, imagine
a fact-finder who convicts by rolling a 10-sided die. If the die lands on 10, they
convict. If not, they acquit. There is no reason to presume that those cases in which
the fact-finder rolls a 10 will tend to be cases in which the defendant is guilty and
those in which they do not roll a 10 will tend to be cases in which the defendant is
innocent. There is therefore no reason to expect a just error distribution. Further-
more, such a decision-procedure would also lead to an unacceptably high number
of errors, as there would be no necessary correlation between the strength of the
evidence and the decision made. This point relates to a second key goal of criminal
proof, namely error minimization. As various evidence scholars have argued, we want
verdicts to be based on accurate factual beliefs as much as possible – i.e., we want
to minimize the number of errors made (see e.g., Nance, 2007, 163; Ho, 2008;
Dworkin, 1985a;b; Goldman, 2001; Stein, 2005; Allen & Stein, 2013, 567). How-
ever, this condition is satisfied only if the fact-finder is able to discriminate between
the guilty and the innocent; we want them to be a reliable assessor of who is guilty
and who is not.
What is required for the fact-finder’s judgments to be reliable? As Laudan (2006, 73) points out, whether a high proof standard leads to the correct error distribution depends in part on the completeness\textsuperscript{213} of the evidence and the validity of the fact-finder’s inferences from that evidence (Laudan, 2006, 73).\textsuperscript{214} In other words, does our evidence contain the most important facts and does the fact-finder assign the appropriate strength to this evidence? It is not difficult to read the criteria of robustness and evidential probability into this remark. To give a simple example of the underlying idea, consider a fact-finder who always has to draw conclusions from a very incomplete case file and who consistently interprets the evidence that they do have in unreasonable ways. Obviously, we would not expect such a fact-finder to discriminate well between the guilty and the innocent and hence, we would not expect their decisions to yield a just error distribution.

Apart from error distribution, robustness and evidential probability are also key requirements for error minimization. To begin with the notion of evidential probability, Laudan (2006, 79) argues that the beyond a reasonable doubt standard should not be purely a measure of personal belief, because a purely subjective standard has no necessary connection with error distribution. As I already discussed in section 2.4 of chapter II, such a standard allows for conviction based on weak sets of evidence for guilt acquittal in cases where there is strong evidence for guilt as it means that almost any belief can count as rational, even when it does not align with what the evidence supports. Similarly, Allen and Pardo (2019, 9-10) argue that unconstrained subjective beliefs have “no necessary relationship to advancing accurate outcomes” as they could be any number at all and need not be constrained by the evidence. Evidential probabilities ensure that rational degrees of belief are constrained by the evidence. Nance (2008, 270) offers a similar reason for why we are after evidential probabilities in legal proof. He suggests that this is to ensure

\textsuperscript{213} Laudan actually uses the term ‘robustness’. His meaning of this term is the same as how I use it, namely to describe the quality, i.e., completeness of the set of evidence.

\textsuperscript{214} Additionally, it also relies on the distribution of truly innocent and truly guilty defendants who go to trial (Laudan, 2006, 75).
that the relevant probability is “both well-considered and productive of accurate verdicts”.

The requirement of robustness (understood as the probable absence of overlooked exculpatory information) similarly contributes to the reliability of criminal fact-finding. To illustrate, imagine a fact-finder who always assigns the correct strength to the available evidence. However, imagine that their set of evidence is consistently the result of a sloppy investigation, which overlooks most of the relevant facts. We would not expect this fact-finder’s judgments to be especially accurate, as there is no reason to presume that this evidence is indicative of the actual status of the hypothesis that the defendant is guilty. So, as Ho (2008, 167) suggests “if the trier of fact is aware that the available evidence adduced in support of a hypothesis is significantly incomplete, that too much of relevance is as yet hidden from her, that ‘there is a significant chance that there is a better explanation’ for the event in question, she would not be justified in believing that the hypothesis is true.”

In the quote above, Ho mentions the possibility of unconceived alternatives as one of the sources of reasonable doubt. This connects to the discussion in chapter VII, where I argued in favor of unconceived alternative explanations as a source of reasonable doubt. I want to briefly recap part of the argument from that chapter here. As various Bayesians have pointed out, we may only assign a probability to any given hypothesis if we presume that the set of hypotheses that we consider exhausts the probability space. For instance, Salmon (1990) proposes a Bayesian account which does not presume that our set of hypotheses is exhaustive, but on which we only consider our conceived alternatives when evaluating the confirmation of a given hypothesis. But, as Rowbottom (2016, 3) points out in response, if we only consider conceived alternatives, this means letting go of the assumption that we are evaluating whether a theory is “truth-like.” Similarly, Wenmackers & Romeijn (2016) propose an ‘open minded’ version of Bayesianism, which drops the assumption “implicit in standard Bayesianism – that the correct empirical hypothesis is among the ones currently under consideration”. However, they admit that their approach “fails
to provide us with the required normative guidance” about the absolute confirmation of scientific theories, because it only tells an agent what to believe if she supposes “that the true theory is among those currently under consideration” (Wenmackers & Romeijn, 2016, 1243) In other words, unless we presume that our set of hypotheses is exhaustive, we must let go of the assumption that the probability that we assign to a hypothesis is indicative of the actual truth-value of that hypothesis (Jellema, 2022b). What the above discussion suggests is that the same is true of other kinds of missing information. If we cannot presume that we have not missed anything exculpatory, then we also cannot presume that a high degree of belief in guilt is an indicator that the defendant is actually guilty (or, conversely, that a low degree of belief in guilt indicates that the defendant is innocent).

So, to summarize the above, we want robust evidentially calibrated beliefs for the sake of achieving a just error distribution and for minimizing the number of factual errors. If a fact-finder were to act upon a belief while they know that they have likely missed a great deal of relevant information, then the risk of error is too great. And, in high-stakes situations, which criminal cases regularly are, we arguably want to err on the side of caution by opting for the less risky option (Horowitz, 2014; Henderson, 2021, 6-7). Because a false acquittal is less costly than a false conviction, the fact-finder should acquit.

In chapter VII I argued that we should be ‘evidentialists’ with respect to when we can presume that we did not miss any innocence explanations. In other words, we should look at the evidence that we have for the possible existence of unconceived alternative explanations. I want to expand upon this suggestion and propose that, more generally, whether we have good reason to suspect that we missed exculpatory information depends on our ‘higher-order evidence’. I now turn to the meaning of this notion and offer some examples of such evidence as well as discussing the notion of higher-order probability.
4. Higher-order probability and higher-order evidence

Information about whether we missed anything is not evidence in the ordinary sense, as it does not relate directly to the hypothesis under consideration (that the defendant committed the alleged acts). The reason for this is that such information does not sanction a specific change in our degree of belief in the hypothesis. After all, we do not know what the missing information would support. As Hamer (2012, 136) writes, “probabilistically, it is not possible to take account of unavailable information for the simple reason that it is unavailable and its content is unknown.” Instead I want to suggest that we can better think of it as ‘higher-order evidence’. Higher-order evidence is a well-known concept from epistemology. There are various ways of spelling out this idea. For instance, Henderson (2021) mentions several characterizations that epistemologists have given of the concept, including ‘evidence concerning the reliability of our own thinking about some particular matter’ (Christensen, 2016), ‘evidence about what your evidence supports’ (Sliwa & Horowitz, 2015) and evidence that ‘induces doubts that one’s doxastic state is the result of a flawed process’ (Lasonen-Aarnio, 2014). So, roughly speaking, higher-order evidence is evidence about how reliable the conclusions are that we drew from our first-order evidence. Though the question of missing evidence has, as far as I am aware, not been linked to the idea of higher order evidence, the

---

215 As Nance (2008, 274) points out with respect to missing evidence, even if, for instance, “one can say, crudely, that evidence effectively repressed by one side is likely to be unfavorable to them, [this] gives little hint of an answer to the question, ‘By how much?’”, which is needed to make adjustments to our credences. The fact-finder would then have to make highly speculative adjustments.” I made the same point with respect to unconceived alternative hypotheses. See Jellema (2022b) and chapter VII of this thesis.

216 In this sense, higher-order evidence bears some semblance to ‘ancillary evidence’, which is evidence that has a bearing on the probative force of directly relevant evidence. In other words, it is evidence about evidence (Schum, 1999). An example of ancillary evidence is information on the credibility of a witness. The difference between these two concepts is that higher-order evidence relates to the reliability of the fact-finder’s inferences about the evidence, not the reliability of particular items of evidence. Additionally, higher-order evidence (at least as I use the term here) does not have an impact on the probative force of the evidence, but on whether the evidence is sufficiently good to act upon.
latter concept fits closely with what I suggested in the previous section, namely that the possibility of missed exculpatory information can induce doubt about the reliability of the fact-finder’s inferences from that evidence.

In this section I distinguish several types of higher-order evidence, which may jointly justify the belief that we did not miss any exculpatory information. As with any form of evidential reasoning, our evidence can be stronger or weaker. So, how justified we are in presuming that we have not missed anything comes in degrees. Hence, my proposal effectively boils down to the idea that the proof beyond a reasonable doubt requires the fact-finder’s higher-order probability to meet a minimum threshold (cf. Dahlman & Nordgaard, forthcoming for a similar suggestion). Before I move on to exploring the various types of higher-order evidence, let me briefly discuss how this higher-order probability relates to our first-order belief.

4.1 Higher-order probability

There are various ways of interpreting the notion of a higher-order probability, but not all of them are suitable for legal proof. Because I suggested that my account relates to the reliability of the fact-finder’s inferences, I want to distinguish my own account from one recently proposed interpretation in particular, as this account interprets higher-order probability in terms of the reliability of first-order probability assessments.\textsuperscript{217} On this interpretation a higher-order probability expresses the

\textsuperscript{217} For the sake of completeness, let me mention two further interpretations. First, a higher-order probability is sometimes used to express how resistant to change upon receiving new information our first-order probability is (Skyrms, 1977; Steglich-Petersen, 2019). However, as I argued earlier in this chapter, this is not the concept of robustness that I am concerned with in this chapter. Second, in the context of statistical inference a higher-order probability may express our confidence that the first-order probability is the ‘correct’ or ‘true’ probability (Kyburg, 1989). In some contexts, events have a physical, first-order probability between 0 and 1. Our higher-order probability is then a judgment about how strongly we believe that we have correctly estimated the first-order probability. For example, suppose that we are interested in what the probability of heads is when tossing a coin, and we assign a probability to this event. The higher-order probability would
rate at which the fact-finder tends to make errors in their first-order judgments. The higher this error rate, the lower the higher-order probability. Such an error rate can be taken into account within the first-order belief through ‘calibration’, where we lower the first-order probability in proportion to the error rate (Schoenfield, 2015). Malcai & Rivlin (2021) suggest we use this idea of higher-order probability with respect to legal proof. For instance, they propose that sometimes a judge might lower their credence in the guilt of a defendant if they know that they have erroneously convicted innocent defendants in the past. However, this suggestion strikes me as problematic for several reasons. First, this only works for fact-finders who have a track record – i.e., judges, not juries. But even in the case of judges, we rarely, if ever, have reliable data about their error rate. Second, historical data is a poor guide to one’s error rate. Each case is unique and simply because errors were made in the past does not mean that errors will be made at the same rate in future cases. The extent to which we would expect errors will depend on the particulars of the case at hand. Finally, this interpretation does not fit with my proposal here, which is about the possibility of overlooking information. As said, evidence of having overlooked something does not sanction any particular change in our first-order degrees of belief.

I want to propose that we should instead understand the relationship between our first- and higher-order beliefs in terms of a full-belief framework – where we either believe something or not. Within such a framework, higher-order evidence is usually conceptualized as an undercutting defeater (Lasonen-Aarnio, 2014; Henderson, 2021). An undercutting defeater casts doubt on the connection between the first-order evidence and the belief concerning the first-order proposition. As said, our

then reflect the degree to which we are confident that the coin’s propensity to land on heads is indeed 0.5. However, this idea only has meaning if there is a single ‘correct’ probability. Yet legal proof concerns events in the past that either happened or they did not. So any first-order probability other than 0 or 1 cannot be the correct probability. Hence, if our first-order probability is not 1 or 0, our higher-order probability concerning whether we have identified the ‘true’ first-order probability should always be 0. This is therefore not a useful account of higher-order probability in the context of legal proof.
higher-order belief (about whether we are justified in our first-order belief) comes in degrees, but these higher-order credences do not change the level of our first-order credences. Rather, when our higher-order probability is too low, this severs the link between the fact-finder’s belief that the defendant is probably guilty and the conclusion that the evidence supports the defendant’s guilt. In other words, the fact-finder is not justified in assigning a (high) first-order probability to the defendant’s guilt if their higher-order evidence does not justify them in believing that their set of information is (sufficiently) exhaustive. The requirements for proof of guilt beyond a reasonable doubt are then not met.

4.2 Types of higher-order evidence

Having explained the concept of higher-order evidence, I now want to distinguish different types of such evidence that may occur in criminal proof. One important type of higher-order evidence is how open-minded and attentive the fact-finder was when considering the evidence. To give a fictional example, at the start of the 1957 film 12 Angry Men, only one of the jurors, Davis, votes not guilty. Another juror asks him whether he believes the defendant’s story. Davis replies: “I don’t know whether I believe it or not. Maybe I don't. (...) There were eleven votes for guilty. It's not so easy for me to raise my hand and send a boy off to die without talking about it first.” The juror therefore indicates that the issue is not whether or

218 I base the following list in part on work from the philosophy of science, epistemology and legal evidence scholarship on how we may reason about overlooked evidence and unconceived alternatives (e.g., Musgrave, 1988; Amaya, 2009; 2015; Dawid, Hartmann & Sprenger, 2015; Dellsén, 2017; Malcai & Rivlin, 2021; Jellema, 2022b; Dahlman & Nordgaard, forthcoming). It is an expanded version of the types of evidence for unconceived alternative scenarios mentioned in chapter VII.

219 A related, but controversial, type of evidence is evidence concerning the competence of the fact-finder – i.e., their inherent capacity to make sense of the evidence. However, as Malcai & Rivlin (2021, 38) point out, such evidence is usually deemed inadmissible. Hence, I do not include it here.
not he is convinced at that specific moment, but about carefully considering the evidence and potential explanations for this evidence, before arriving at a verdict.

Another, related type of higher-order evidence is discussed by Di Bello (2013, ch. 7.5). He argues that robustness depends in part on the degree to which the defense had an opportunity to level charges against the prosecution’s case and the degree to which they took advantage of this opportunity. In many countries there are specific legal guarantees that ensure that defendants have enough monetary, legal, intellectual, and evidentiary resources to exercise their right to a defense, such as the right to effective council. If such procedural rights are not respected, then this alone may lead to the case being dropped or to an acquittal. Additionally, such a violation of rights may lessen the degree to which the fact-finder believes that they understand what the evidence actually supports and whether nothing was overlooked.

A third type of higher order evidence is the quality of the underlying investigation. All other things being equal, the better the search for evidence, alternative hypotheses, connections between the evidence and weaknesses in the case, the more reason we have to presume that our set of information is good enough to base a reliable belief on. The quality of such a search depends in part on the amount of time and resources spent on it as well as on the imaginative faculties of the investigators. It also depends on epistemic virtues on the part of investigators, such as open-mindedness and perseverance as well as on how methodically they construed their investigation. Whether investigators displayed these virtues can also depend on the nature of the case. For instance, as Amaya (2015, 517) points out, in emotionally disturbing cases, investigators may be more likely to be biased, thereby failing to conceive of plausible alternatives. Additionally, we may have information that important evidence was not collected during the investigation (Dahlman & Nordgaard, forthcoming), for instance, that an important witness was not heard. As Dahlman et al. (2015, 20) point out, what counts as important evidence depends on our generalizations about whether this kind of evidence is “likely to produce information that will change [our first-order belief] than other inquiries.” To give
an example, a witness who was at the crime scene could easily offer testimony that could overturn our beliefs.

The quantity and quality of our set of information also matters in several ways. We may firstly have too much information to adequately make sense of it. For instance, the case may be accompanied by a thick case file, indicating a great deal of potentially relevant evidence, which may also conflict or be incomplete. It may then be difficult to see how all the facts hang together. Additionally, having too much information can create ‘noise’, where irrelevant information drowns out the (more) relevant. The same goes for having too many possible explanations or having to make sense of many competing arguments. Such an abundance of information can make it difficult to judge whether all plausible alternatives were considered and whether all relevant evidence was collected. Second, we can have too little information. For example, if a victim is found in the middle of the forest, with no witnesses, obvious fingerprints or other marks of a possible perpetrator, and if it is not clear whether he was killed or died of natural causes, then it is almost pure guesswork as to what happened (and therefore whether there are alternative explanations) and what other evidence to gather. In contrast, if there are multiple witnesses and other evidence that clues us in on what happened, then there is less of a chance that we have overlooked something. Finally, some evidence, such as statistical information, is known to be difficult to interpret for lawyers (Malcai & Rivlin, 2021, 29). This can increase the chance that relevant arguments are overlooked, as it is more difficult to come up with such arguments, thereby decreasing the fact-finder’s confidence in their own probability assessment.

Another type of higher-order evidence is the quality of our conceived hypotheses. If none of our current alternatives explain the evidence well, then we have good reasons to suspect that either there is a better explanation that we have not conceived of, or that some of the evidence is misleading. For example, consider a

---

220 I also discussed this point in section 4.2 of chapter III.
situation in which none of our current scenarios explain the testimony of multiple witnesses well. This might make us think that we have overlooked a scenario that does explain these testimonies adequately, or that we have failed to realize that one of the witnesses is lying. Also, as Dahlman & Nordgaard (forthcoming) point out, the better the case for guilt, the less likely it is that it will be overturned by newly discovered evidence. If the case for guilt only barely meets the threshold for conviction, then it is more easily overturned.

A final way in which our existing explanations can be of a high quality is that they are well-specified. It is to this form of higher-order evidence that I turn now.

5. Explanation-based thinking, specificity and robust evidential probability

If the hypotheses under consideration during the investigation and subsequent trial were well-specified, this arguably increases the degree to which the fact-finder is justified in assuming that their belief in the defendant’s guilt is robust and evidentially calibrated. The reason why I single out this particular type of evidence is that it leads to a novel answer on how the Bayesian and explanation-based accounts of rational legal proof can be combined. As I explained in chapter II, more and more evidence scholars suggest that these two accounts are compatible and may also complement one another. However, discussions about the relationship between the two mainly focus on how explanation-based thinking can lead to a justified high credence in first-order propositions (e.g., Hedden & Colyvan, 2019; Biedermann & Vuille, 2019, 18–20; Gelbach, 2019, 169; Welch, 2020). What I want to suggest here is that explanation-based thinking can (also) complement the Bayesian account on the level of the fact-finder’s higher-order belief.

To recap some of the discussion from earlier in this thesis, explanation-based accounts cast rational legal proof in terms of competing explanations of the evidence. On this account, whether guilt is proven depends on the plausibility of the available guilt and innocence explanations. As most legal explanationists use the term,
'plausibility' refers to the extent that an explanation exhibits ‘explanatory virtues’, such as internal coherence and fit with background beliefs about the world. One explanatory value that is sometimes mentioned, both by explanationists in the law and in the philosophy of science is ‘specificity’ (or ‘preciseness’) (Thagard, 1978; Pennington & Hastie, 1991; Salmon, 2001; Ylikoski & Kuorikoski, 2010). The virtue of specificity is especially important in criminal proof, where the relevant explanations often take the form of scenarios – i.e., narratives that describe a sequence of events which led, for instance, to the death of the victim (Allen & Pardo, 2019, 13, n86; Mackor & Van Koppen, 2021). We typically want our scenarios to be sufficiently detailed. For instance, according to Bennett & Feldman (1981) they should ideally contain a central action and describe a context that make this action understandable, in the form of a description of the scene, a motive, a central actor and resulting consequences. According to Pennington & Hastie (1993) a complete scenario includes an initiating event, a psychological response to this event, a goal, a resulting action, and consequences. Apart from these details the scenarios under consideration should arguably be (sufficiently) specific about the time and place of the alleged events as well as about how those events took place.

Yet the idea that more detailed hypotheses are preferable to general ones raises a problem for the explanation-based accounts. As explained before, this account shares the idea with Bayesianism that we only want to accept conclusions that are (very) probably true. So, we should only accept an explanation as true if it is very probably correct. However, there is a tension between this aim and the wish to have detailed hypotheses. All else being equal, more detailed hypotheses are less probable than less detailed ones. For example, consider two explanations of why a person died:

\footnote{See chapter III, section 2.2.}
(i) The defendant killed them.

(ii) The defendant killed them with a hammer, around 5 p.m., after they got in a fight over an unpaid loan.

The second hypothesis can never be more probable than the first, as it is a more specific version of the former. Whenever the second is true, so is the first but there are situations where the first is true while the second is not. So, if our aim is to accept only highly probable explanations, why would we want to reason in terms of the latter rather than the former? More generally, on the explanation-based account we reason in terms of a small set of non-exhaustive, specific hypotheses (Allen & Pardo, 2019, 11-2). But why not take the full probability space into consideration? Critics of explanation-based accounts have argued that this makes the account irrational compared to the Bayesian framework of rational legal proof (e.g., Guerra-Pujol, 2018).

This worry brings us back to the argument by Di Bello (2013) that more specific hypotheses make our beliefs more robust, where robustness means the probable absence of exonerating information. This is an interesting argument that has, as far as I am aware, not received much attention in legal proof scholarship. I want to expand upon this argument and explore how explanation-based thinking helps us obtain robust, and evidentially-calibrated beliefs, as it helps us in various ways to ensure that we did not miss anything. Di Bello’s point is that specific scenarios can more easily be subjected to scrutiny as it is easier to ask critical questions of specific details about broad hypotheses. He (216) writes:

[A] more complete narrative is more widely exposed to challenges because

---

222 This argument has also been made in the philosophy of science. As Cabrera (2017) and others point out, while some explanatory virtues, such as simplicity, plausibly correlate with more probable explanations, not everything that makes an explanation ‘good’ can be linked to its probability. One example is the virtue of precision – which concerns how detailed an explanation is. As Popper (1959) already pointed out, a theory with more informational content cannot be more probable than a less informative theory because of the above point.
it says more, it commits itself to more propositions than a less complete narrative which leaves many issues and questions open.

Because it is easier to scrutinize specific hypotheses there is less chance that a flaw in that hypothesis escaped our attention.

Well-specified scenarios also have other benefits. For instance, as Pennington & Hastie (1993) argue, thinking in terms of such scenarios helps fact-finders make sense of complex sets of evidence. One way in which scenario-based thinking helps us make sense of the evidence is that it helps reason about which facts are relevant in the given case. For example, imagine a case in which there is a medical report which states that the defendant has a limp. Such a report is not, by itself, evidence of anything. It is simply one of the countless facts regarding the defendant. However, suppose that the report is evidence in a burglary case. The prosecution scenario in this case may stipulate that the defendant entered the house through the garden. As a result, the report may become pertinent information. For instance, we may then want to look for footprints in the garden and see whether the pattern of these prints indicated that the perpetrator had a limp. If the footprints match someone with a limp, the report supports the guilt of the defendant. So, because the scenario specified how the perpetrator entered the house, investigators could consider what evidence was pertinent. Conversely, if this had not been specified, both the footprints and the fact that these prints matched the limp of the defendant might have been missed.

A sufficiently specific scenario can also help us discover internal contradictions within that scenario. For instance, suppose that the prosecution’s scenario entails that the limping defendant fled on foot but that he could not have traveled the required distance within the allotted time frame given this limp. The discovery of

---

223 This is an example of how a scenario can yield predictions about what evidence we would or would not expect to see (Douglas, 2010; Mackor & Van Koppen, 2021; Jellema, 2021). See also chapter V of this thesis.
this difficulty required a detailed scenario which stipulates which path was taken within what time-frame.

Another benefit of thinking in terms of well-specified scenarios is that it can lead us to consider alternative scenarios. For example, suppose that we search for further footprints in the garden but do not find any. This can make us wonder whether there is an alternative explanation for why no footprints were found there. For instance, the defendant might not have entered through the yard.\textsuperscript{224} To come up with such alternative scenarios it is helpful to know which part of our conceived scenario(s) is implausible, as we may be able to think of a similar scenario without this implausible element. Furthermore, once we have multiple scenarios we can look for \textit{discriminating evidence}, which is evidence that confirms one of the scenarios and disconfirms the other.

The above remarks are only some of the ways in which comparing well-specified scenarios help us make sense of the evidence. The upshot of this is that when we think in terms of multiple detailed scenarios, this makes it easier to determine whether we may have missed something.\textsuperscript{225} If nothing is found, this strongly suggests (all else being equal) that the fact-finder’s degree of belief is robust and evidentially calibrated. What we sacrifice in first-order probability when we make our hypotheses more specific is then offset by a gain in higher-order probability.

That explanation-based thinking helps make sense of the evidence is a well-known idea. For instance, Nance (2016, 84) observes the following about the popularity of explanation-based accounts: “One main motivating concern of those who press the

\textsuperscript{224} In other words, we generate a new scenario by asking ‘how else could we explain this surprising fact?’ This is a well-known advantage of abductive – i.e., explanation-based – thinking over Bayesianism, which says nothing about the generation of new hypotheses (Lipton, 2004, 116).

\textsuperscript{225} Another good example of the various ways in which detailed scenarios help make sense of the evidence can be found in chapter V, section 7, where I discussed the case of Hae Min Lee. As we saw there, by making the exact time-line precise, it became possible to find inconsistencies in the prosecution scenario and to generate alternative hypotheses.
explanatory approach is that [probabilistic accounts] focus on the end product of deliberation, rather than the process of arriving there, giving no direction to jurors as to how to go about assessing the evidence in the case.” What is new in my analysis is the connection between this idea and the notions of robustness, evidential probability, higher-order evidence and higher-order probability.

6. Conclusion

In criminal trials we are after proof of guilt beyond a reasonable doubt. Most agree that this at least requires the fact-finder to believe that the defendant very probably committed the alleged acts. However, a high degree of belief alone is insufficient for meeting this standard. After all, this belief may be unreasonable. In this chapter I focused on two popular suggestions for additional criteria: that the fact-finder’s belief should be an evidential probability and that it should be robust. As I interpreted these terms, both are about the possibility of overlooked exculpatory information. I argued that if we cannot justifiably presume that there is no such overlooked information, then there should be reasonable doubt. Whether this presumption is justified or not will depend on our higher-order evidence. I surveyed several kinds of higher-order criminal evidence. I ended this chapter by looking at a particular type of higher-order evidence, namely the specificity of our hypotheses. This led to a novel way of combining the Bayesian and explanation-based account of rational legal proof.
IX. CONCLUSIONS

When are the facts of a criminal case proven? Lawbooks or case law may say something like that ‘the fact-finder must be convinced by the evidence or that ‘there must not be a reasonable doubt’. But when is it reasonable to be convinced by the evidence that something has happened? And when is it reasonable to doubt? These are not legal but epistemic questions, which are studied within what is varyingly known as legal evidence scholarship, legal epistemology, the philosophy of legal evidence or the study of rational legal proof. In this thesis I contributed to this field by combining the Bayesian and explanation-based accounts, the dominant theories of rational legal proof. The Bayesian account casts legal proof in terms of our degrees of belief in hypotheses, which we update when receiving evidence. For the Bayesian, reasonable doubt primarily arises when the evidence makes it insufficiently probable that the defendant committed the acts described in the indictment. On the explanation-based account, legal proof is a comparison between competing explanations of the evidence. In criminal cases such explanations are often stories that describe a timeline of causally connected events that occurred. Such stories can be plausible if they display certain virtues, such as fit with the evidence, internal coherence and fit with background beliefs. For the explanationist, reasonable doubt results primarily from either the presence of one or more plausible scenarios in which the defendant did not commit the alleged acts, or the absence of a sufficiently plausible scenario in which they did commit these acts.

Is rational criminal proof best understood in terms of stories or in terms of probabilities? More and more evidence scholars suggest that perhaps the answer is ‘both’ and that the two accounts can complement one another. This recent trend is akin to that in the philosophy of science, where Bayesianism and explanationism are similarly the dominant accounts of rational evidential reasoning and where an increasing number of authors defend ‘compatibilism’ – the notion that the two accounts are compatible and complementary. In this thesis I drew on ideas from the
philosophy of science to show how the two most prominent theories of rational
criminal proof can similarly be wed.

One obstacle for this wedding is that it is not always clear what authors mean when
they talk about ‘Bayesianism’ and ‘explanation-based accounts’. To begin with
Bayesianism, there are at least two distinct ideas that legal evidence scholars refer
to by this name. On the one hand, there is the prescriptive idea that people
should (sometimes) use probabilistic tools, such as Bayesian networks, for analyzing
and presenting evidence at trial or in related contexts. On the other hand, Bayes-
ianism may refer to a normative, epistemic theory of rational proof that offers an
account of how we ought to change our degrees of belief in the light of a varying
set of evidence and when the fact-finder may therefore believe that something
probably happened. I focused on this latter notion of Bayesianism. I agreed with
the suggestion of many contemporary Bayesians that the relevant probabilities are
best interpreted as ‘evidential’ degrees of belief, which are those credences that are
reasonable to hold in the face of the evidence that one has. However, the problem
for the Bayesian is that it is unclear what this means. One of the aims of this thesis
was therefore to further clarify the concept of ‘evidential probability’. I suggested
that we can best think of these as the degree of belief of an agent who has made
sense of the evidence well. However, this means that Bayesianism is in need of an
account of what it means to ‘make sense of the evidence’ in the first place. It is
there that we find the space for explanation-based accounts, which tell us how to
select and interpret the available evidence.

Yet explanation-based accounts face a problem of their own. On such accounts,
legal proof occurs (or ought to occur) through explanatory comparison, which
means to look at several possible accounts of what may have happened and to
choose one (or none) of these. However, it is not clear how the fact-finder can
make this choice rationally. To put it another way, when is the fact-finder justified

\[226\] See chapter II, sections 2.2 and 2.3.
in believing that a particular explanation is (probably) true? I argued that Bayesianism can help us answer this question. It offers a precise framework for exploring how and under which conditions explanation-based thinking leads to probably true outcomes. These considerations led me to two questions which are at the heart of this thesis:

(I) How does explanation-based reasoning help make sense of the available evidence?

(II) When are we justified in accepting an explanation as probably true?

I further divided both into a number of subquestions. By answering these questions, we arrive at a combined theory of rational criminal proof. Below, I consider each question in turn.

1. (Un)helpful explanations

Evidence is never simply given. It has to be gathered, selected and interpreted. In other words, we have to see how the facts ‘hang together’. The result of this process will depend in part on which explanations we consider and how detailed we make these explanations. For example, determining whether a scenario is internally coherent may involve spelling out a detailed timeline of the alleged events. As a result, we may discover that the scenario implies that the defendant moved from one place to another in an unrealistically short amount of time and that it is therefore incoherent. Conversely, very general explanations (for instance, ‘the defendant did not kill the victim’) do not offer such guidance and it can be difficult to evaluate how probable these explanations are in the light of the evidence.

So far, most explanationists have focused on how thinking in terms of competing explanations can help us make sense of the evidence for or against the ultimate
hypothesis – that the defendant committed the acts alleged in the indictment. However, as I showed in this thesis, there are at least two other types of evidence that explanation-based thinking can help us make sense of: evidence about the reliability of sources of information and evidence about the completeness of our set of information.

Question I.I: How does explanation-based thinking help us make sense of reliability evidence?

In criminal cases, we not only encounter evidence that directly relates to the defendant’s guilt or innocence but also evidence that tells us something about the reliability of such ‘regular’ evidence. As I will explain below under question II.II, how reliable we consider the sources of our evidence to be can matter greatly for the probability of our conclusions. But how do we assess the reliability of a source? I argued that assessments of reliability will depend on the explanations under consideration. For example, the Dutch Venray case that I discussed in chapter IV, stipulates when courts may reject a defendant’s alternative explanation. One ground on which a court may reject such an alternative scenario is that the defendant themselves lacks credibility as a storyteller. I suggested in this chapter that a court can, for instance, conclude that a defendant lacks credibility because they lied in the past, because the story that they offer is vague or because it is ad hoc – i.e., unfalsifiable because it was offered at a late stage, once all the evidence has already been presented. None of these facts are always signs of unreliability. What conclusions we may attach to them will depend on what best explains these facts. For instance, can the vagueness of the story also be explained by a poor memory? Was there something different in past situations that explained their unreliability then, which is not present in the current situation? Did they not have the opportunity to offer their story sooner? We can only draw conclusions about the defendant’s credibility once

---

227 Exceptions being Rassin (2001) and Van Koppen (2011). See chapter VI and section 3.2 of chapter II. See also footnote 6.
we have answered such questions.

I expanded on the above idea when I discussed the value of predictions in chapter V. An *ad hoc* scenario may be suspicious in part because it was not confirmed by any successful predictions of novel facts, where the scenario only explains known facts and it is not supported by evidence that could not have been used to construct the scenario. For instance, suppose that the alternative explanation fits well with the facts in the case file well, but that the defendant knew about what is in the case file before offering their scenario. They could have constructed a false story based on the facts mentioned therein. Similarly, a scenario developed by criminal investigators may only explain known facts, or it can be supported by information that it successfully predicted. When a scenario only explains known facts, this can, for example, sometimes be an indicator of confirmation bias on the part of the investigators. However, whether a lack of predictions is a reason to be suspicious of a scenario will depend on the plausibility of the competing explanations for this lack of predictive success. For instance, is it a sign of a witness lying or are there other plausible explanations why the witness only explained known facts? Did the investigators not make predictions because they suffered from confirmation bias or were they actually evenhanded and thorough in gathering the evidence prior to formulating their scenario?

In chapter VI, I took a broader look at how explanation-based thinking helps evaluate eyewitness reliability. I further developed a method for eyewitness evaluation, called testimonial inference to the best explanation (TIBE). On this method we compare multiple explanations of why a witness offers a particular scenario. This helps to select and interpret the evidence about the reliability of an eyewitness. For example, consider multiple witnesses who agree about an unlikely detail in their testimony. Depending on what explanations we consider and how plausible we believe these explanations to be, such agreement may count strongly in favor of the

---

228 See also question II.II.
reliability of the agreeing witnesses or it may count strongly against it. For instance, the agreement can sometimes be best explained by the witnesses talking to one another, subconsciously influencing the testimony of the others, or the best explanation may be that they independently observed the same event and are now accurately reporting on it. This approach also complements existing frameworks for eyewitness evaluation developed by psychologists as it adds a way of selecting, interpreting and drawing conclusions from the available evidence about the witness’s reliability to these accounts.

Summing up, which explanations we consider determines what the relevant facts are regarding the reliability of some source. Furthermore, looking at how plausible these explanations are helps to structure and interpret such reliability evidence and to draw conclusions from it.

**Question I.II: How can explanation-based thinking help us reason about the completeness of our set of information?**

Reasoning about the possibility of unconceived explanations and overlooked evidence is an important aspect of criminal proof. I argued in chapter VIII, by thinking in terms of detailed scenarios we lower the chance of overlooking something. For example, it is easier to ask critical questions about detailed scenarios, as this helps detect flaws and inconsistencies in the case for guilt (or innocence). Furthermore, thinking in terms of detailed scenarios points us to the relevant facts, especially if we consider multiple such scenarios as it allows us to look for ‘discriminating evidence’ – evidence which provides support for one scenario and disconfirms the other. This means that it is easier to see what is relevant within the set of facts that we already have, but also what further evidence we should gather. Finally, it is easier to come up with alternative hypotheses to detailed scenarios. Those who fail to think in terms of well-specified hypotheses therefore run a greater risk of
overlooking relevant facts, arguments and hypotheses. So, thinking in terms of well-specified explanations helps us think about whether we overlooked anything.

2. (Im)probable explanations

Explanatory choice means that the fact-finder has to choose between potential explanations of what happened. These explanations can be plausible to varying degrees, depending on how well they explain the evidence, fit with our background beliefs about the world and are internally coherent. Such explanations can be formulated by investigators, by the defense or prosecution attorneys, by witnesses or by the defendant themselves. Typically, not all possibilities are considered but only a select few that are brought up. The probability of an explanation is a judgment by the fact-finder about how likely it is that the explanation is true in the light of the available evidence. Over the course of the past chapters, I examined the relationship between the notions of probability and plausibility, the influence that the reliability of an explanation’s source has on its probability and how the completeness of our set of evidence and explanations should influence our judgments of how probable a given explanation is.

*Question II.I: How should the plausibility of explanations influence judgments about their probability?*

A good story is not always a true story. Nonetheless, as explanationists in both the philosophy of science and in legal evidence scholarship have pointed out, how plausible an explanation is (and how plausible its competitors are) is perhaps the most important determinant for whether we may believe that it is true. In criminal evidence scholarship, explanationists have translated this idea into the following

---

See question II.III below for how the possibility of overlooked information should influence judgments about the probability of the competing explanations.
interpretation of the beyond a reasonable doubt standard of proof: we should accept an explanation that implies the defendant’s guilt as true if it is plausible and if there are no plausible alternative explanations consistent with his innocence. This is called the ‘no plausible alternatives account’ (NPA). On this interpretation, the guilt explanation should therefore meet some minimum threshold for plausibility and no explanation consistent with innocence should meet this threshold.

The plausibility of the competing explanations is indeed an important determinant of their probability. For instance, in the Venray ruling discussed in chapter III, the Supreme Court mentioned a lack of plausibility as one of the main reasons why a defendant’s alternative scenario does not lead to a reasonable doubt. However, as I suggested at various points in this thesis, plausibility is not the only aspect that we should take into account when deciding whether an explanation is probably true. As a result, NPA is inadequate as an interpretation of the beyond a reasonable doubt standard. First, an explanation may be plausible and it may lack plausible alternatives, but we may nonetheless not be justified in believing that it is probably true. For instance, as the Supreme Court also noted in the Venray case, an explanation may be improbable because it comes from an unreliable source (see also subquestion II.II below). Another reason why we may not be able to assign a high probability to an explanation is if there are reasons to suspect that we overlooked exonerating evidence or plausible innocence explanations (see subquestion II.III).

There is a further worry for the threshold-conception of proof beyond a reasonable doubt. Plausibility comes in degrees: some explanations are more plausible than others. This means that there can be situations in which an innocence explanation is only barely plausible, but where the guilt explanation is much more plausible. In chapter III, I used a Bayesian analysis to show that in those situations, the guilt explanation can be probable despite the presence of a plausible innocence explanation. This observation led me to an alternative definition, according to which proof

230 See chapter III for further explanation.
of guilt beyond a reasonable doubt requires that the guilt explanation is substantially more plausible than any innocence explanation. Admittedly, one worry for this account is that it does away with the idea of a threshold altogether. For instance, suppose that the prosecution only managed to produce an implausible guilt explanation and that the defense produced no alternative explanation. Because the prosecution’s explanation has no competitors, do we have to accept it as true? This brings me to the second part of my conception of proof beyond a reasonable doubt: in such a situation, our lack of any plausible explanations is evidence that there may be plausible, unconceived innocence explanations. This is a cause for reasonable doubt, as it means that we cannot justifiably assign a high probability to our best explanation.231

To sum up, while a more plausible explanation will tend to be more probable than a less plausible one, plausibility is not the only factor that we should take into account when coming to probability judgments. Other relevant factors include the reliability of the explanation’s construction process, how well it performs compared to alternative explanations and how certain we are that no relevant information was overlooked.

**Question II.II: How does the reliability of an explanation’s source influence its probability?**

Imagine a defendant who calls upon their right to remain silent until late in the criminal proceedings, when they have come to know the contents of the case file. Suppose that this defendant now offers an alternative explanation which explains all the evidence well, but only because they used their knowledge of the evidence to carefully craft a false story. Or imagine investigators who, when gathering the evidence, suffered from confirmation bias. As a result, all the evidence points toward the guilt explanation. In both cases the explanation may very well be plausible.

---

231 See question II.III for more about how and why the possibility of unconceived alternative scenarios may lead to reasonable doubt.
It may, for instance, explain most of the evidence, be internally coherent and be believable in the light of our background beliefs. Nonetheless, the explanations can be improbable if we have good reasons to believe that they were provided by an unreliable source.

In the above example, we might mistrust the defendant’s scenario because it only comes after the facts were known. In that situation, we may not be able to rule out the possibility that it is a false story tailored to the case file. This is, for instance, the reason why the defendant’s scenario was ultimately rejected in the Venray case. In chapter V, I connected this idea to a discussion in the philosophy of science. There is a widespread intuition among philosophers of science that theories and models which only explain known facts are thereby suspect and that we want our theories to be confirmed by risky predictions – i.e., predictions which would likely fail if the theory were false. I suggested that this intuition holds in criminal evidence too. Witnesses who fail to offer predictions can be less credible as a result of this fact because they may – consciously or subconsciously – have fit their story to these facts. For instance, they may have been influenced by information that they received from the police or that they might have seen in the media. We similarly want scenarios produced by criminal investigators to be confirmed by predictions in order to prevent them proposing weak scenarios out of the desire to explain away contradicting evidence and to prevent confirmation bias when gathering evidence.

To sum up the above, a plausible explanation may not be probable if it follows from an unreliable construction process. When we evaluate how probable an explanation is, we should therefore also take the reliability of this process into account.

*Question II.III: How does the completeness of our set of information influence judgments about the probability of explanations?*
When we engage in explanatory choice, we typically consider only a small number of hypotheses and these hypotheses focus our attention on a small set of the available evidence. However, in doing so, they open up the possibility of deviating from optimal Bayesian reasoning, which takes our reasoning to be about the complete probability space and the full set of the available evidence. After all, we may then overlook alternative explanations or evidence. As I argued in chapters VI, VII and VIII, such suboptimal reasoning is often unavoidable when we have to make sense of a complex set of evidence.\(^{232}\) However, there can be instances in which our set of information is so incomplete that it leads to reasonable doubt. As I argued in chapter VII, if we have likely overlooked a plausible innocence explanation then neither the explanationist nor the Bayesian definition of proof beyond a reasonable doubt is met, as both accounts start from a presumption that our set of hypotheses is exhaustive. More generally, I suggested in chapter VIII that it is not just the possibility of unconceived alternatives, but also the possibility of having overlooked other kinds of information such as exculpatory evidence that can lead to a reasonable doubt. Of course, we can never know whether information that we missed is exculpatory or not. After all, we do not know its content. Rather, the belief that we have not missed anything exculpatory is always a presumption. When the fact-finder cannot justifiably presume that they have not missed exculpatory information, this undermines the connection between their factual judgments and two key goals of criminal trials. First, we want to make as few errors as possible (error minimization). Second, to the extent that we do make mistakes we prefer them to be false acquittals rather than false convictions (error distribution).

Whether the presumption that we did not miss anything relevant is justified will depend on our ‘higher-order evidence’. As I said when I discussed question II.I, the plausibility of explanations is one type of evidence for the possible existence of

\(^{232}\) So, as I proposed in chapter II, explanation-based thinking can best be seen as a heuristic for the Bayesian ideal: it is a useful, efficient way of thinking that may nonetheless sometimes lead to irrational outcomes, namely when we overlook something.
unconceived alternatives. If all our explanations are poor, this is a reason to suspect that there are other explanations out there. The worse our explanations, the stronger the evidence. However, such higher-order evidence also includes the degree to which the fact-finder was open-minded and attentive, whether the defense had an opportunity to level counterarguments against the prosecution’s case, the thoroughness and objectivity of the underlying criminal investigation and the quantity and quality of our evidence.

Whether we missed anything is also relevant to our probabilistic conclusions in another way. Under certain conditions an implausible guilt scenario can be the basis for. If the prosecution’s scenario is (somewhat) implausible, but our higher-order evidence as a whole makes it sufficiently probable that there are no further realistic possibilities, then it can be rational to conclude that the implausible scenario is probably true. The idea behind this is as follows: criminal cases are often about unusual situations, in which people may act in strange ways. This means that true accounts of what happened may sometimes be implausible. However, by investigating such a scenario thoroughly, and by looking for alternatives, we may become confident that this scenario is (by far) the most plausible explanation of what has happened and therefore probably true. In those cases, a scenario which we would otherwise not consider sufficiently plausible may lead to proof of guilt beyond a reasonable doubt.233

In summary, a plausible explanation may not be probable if we have good reason to suspect that we overlooked relevant information, such as exculpatory evidence or alternative explanations. Conversely, an implausible explanation can sometimes be probable if we are confident that we did not miss anything.

233 This is another example of a problem for the NPA account of the beyond a reasonable doubt standard.
3. Towards a productive partnership

In this thesis, I discussed the two most popular accounts of rational criminal proof. Explanation-based accounts begin with an observation about what humans do naturally when trying to make sense of complex evidence: they make up stories of what might have happened and see which of these they find most believable. Explanationists ask how we can ensure that people make this choice as rationally as possible. In contrast, Bayesians start the other way around. They begin with an abstract theory of rationality and try to apply this theory to legal practice. While explanation-based accounts fit with how humans reason naturally, the Bayesian framework goes well with the increased use of quantified evidence, in particular statistical forensic evidence, such as DNA traces. Given their different origins and the difference in the language and concepts that they employ, it is hardly surprising that the two theories have often been thought of as adversaries. Others, such as myself, have suggested that the two accounts are compatible. However, if they are compatible, are they then just different ways of saying the same thing? If so, are explanation-based views just Bayesianism for people who are not adept at mathematics? Or, conversely, does Bayesianism just express intuitions that are already encapsulated in explanation-based thinking but in a way that gives a false sense of precision? In other words, is one of them trivial and the other ‘correct’? In this thesis I offered a way of combining the two theories without relegating either one to the margins. On their own, neither theory offers a complete, or even near-complete account of rational criminal proof. However, the gaps in either theory offer a space for a fruitful partnership between the two.

My account of how the two go together relies on the observation that in both the Bayesian and explanation-based accounts, fact-finders strive for the same thing: to draw probably true conclusions. But what does it mean to say that a conclusion is probably true? I agreed with the evidential Bayesians who suggest that probability statements are about the probability that a rational person would assign to a conclusion given the available evidence. Of course, this raises the issue how we may
determine what a rational person would believe. One of the main contributions of this thesis was the development of an answer to this question. I proposed that a rational person can be understood as a person who made sense of the evidence well. Whether a fact-finder may therefore assume that their beliefs align with the evidence depends on whether they have good reasons to think that they made sense of the evidence adequately. Yet this means that a theory of rational proof worth its salt has to give us some indication of how we make sense of the evidence in the first place. I suggested that to make sense of the evidence means at least to come up with hypotheses, to select the evidence that is relevant and to determine how that evidence relates to the hypotheses under consideration. Bayesianism does not tell us how to go about these tasks, but explanation-based thinking does: we make sense of the evidence by formulating, comparing and choosing between explanations of the evidence.

Of course, to reason in terms of explanations does not automatically get you to justified, probably true conclusions. There are different ways to spell out the notion of explanatory comparison and some of the most straightforward interpretations of this notion lead to patently irrational factual conclusions. This brings us to how Bayesianism complements the explanation-based accounts: it offers a calculus for exploring when explanation-based thinking leads to probably true outcomes. More precisely, I suggested that, under specific conditions, explanatory comparison is a useful but imperfect heuristic for Bayesian inference. When we engage in explanation-based thinking we deliberately limit ourselves to only a few of the possible scenarios of what happened, for the sake of making sense of the evidence (including evidence about whether we missed anything). However, this means that we can never be certain that our conclusions are the same as those of a perfectly rational agent who considered every possibility. We can, at best, be justified in believing that our conclusions are sound. So, when are we justified in accepting a particular explanation as probably true? I proposed that to answer this question judges, juries, criminal investigators and lawyers should look at the plausibility of that scenario.
compared to its competitors, whether the scenarios come from a reliable source and how likely it is that they overlooked alternative stories or other exculpatory information.

4. Future research

While I hope that this thesis has been a helpful addition to the field of legal evidence scholarship, much more remains to be done. I want to mention a few research topics which I believe warrant further attention and which could be further explored by building on ideas developed in this thesis.

**Perpetrator knowledge**: Suspects sometimes confess to crimes that they did not commit, for instance when they are put under severe pressure during a police interview. One of the most important ways of determining whether their confession is genuine is whether they display ‘perpetrator knowledge’ – information that only the person who committed the crime could have known. Additionally, various methods of conducting police interviews are aimed at tricking a non-confessing suspect into letting such information slip (Hartwig et al., 2005). But how do we determine whether something counts as perpetrator knowledge? This question has not received much attention from epistemologists. An analysis in terms of probabilities and possible explanations for the suspect’s utterance could help shine a light on this matter. Such an analysis would draw upon what I have said in this thesis on eyewitness reliability and predictions. After all, in the case of perpetrator knowledge we are assessing the reliability of a piece of testimony by using a type of prediction, as it is a type of knowledge that confirms a scenario strongly because it is not merely an explanation of publicly known facts.

**The decision to investigate further**: Police, prosecutors, defense attorneys and courts are all under constant time pressure and budgetary constraints. This means that not every lead can be pursued, nor is it possible to formulate numerous scenarios. While such constraints dictate much of the practice of criminal proof, not much attention
has been paid, as far as I am aware, to what this means for rational evidential reasoning. In this thesis I did look at the consequences of having to draw factual conclusions from a limited set of hypotheses, using an incomplete set of evidence. However, I did not consider how budgetary and time constraints influenced this matter. Additionally, I focused mainly on the role of the fact-finder (the judge or jury) not on that of the other parties such as investigators. Further work on this important type of constraint on the different parties in the criminal process could help bridge the gap between legal epistemology and criminal law practice.

Fact-finding outside (domestic) criminal law: Fact-finding is a part of every area of the law. For example, assessing what probably happened is a crucial question in administrative law (Tollenaar, Marseille & de Graaf, 2008), private law (Grando, 2009), asylum law (Bex & Bex-Reimert, 2016) and international criminal law (Schot, 2022). One avenue of further research is whether the ideas developed in this thesis can be fruitfully applied to such areas as well. To give an example, the assessment of eyewitness reliability is of crucial importance in both of the aforementioned areas of law.

Risk assessment: The field of risk assessment uses both probabilistic approaches and scenario-thinking. However, it is not fully clear how these two are interrelated (Ramírez & Selin, 2014). The scenarios used in risk assessment differ from those in criminal proof as they concern what might happen, rather than what has happened in the past. Nonetheless, for both types of scenarios we need to assess how plausible they are. Furthermore, an important question in risk-assessment is how to deal with the possibility of unconceived alternatives (Taleb, 2007). Many questions in risk assessment are therefore similar to those in legal proof. Future research could focus on cross-pollination between the assessments of what happened in the past and what might happen in the future.

Translating explanation-based thinking into Bayesian networks: Several factors influence the probability of explanations, including their plausibility. However, the relationship between these factors and the probability of our conclusions is still far from
precise. For instance, various definitions of plausibility have been given but it is not clear why or when a more plausible explanation is more probable than a less plausible one. A feasible way of making these matters more precise is by formalizing explanatory reasoning in terms of Bayesian networks. This is, for example, the project undertaken by Vlek (2016), who offers a method for translating crime scenarios into Bayesian networks. Nonetheless, more work remains to be done in order to show not just that we can translate explanatory reasoning in Bayesian terms, but also to determine which (if any) formalization is correct. One strand of work that may serve as an inspiration for this is that of philosophers of science who have sought to ‘Bayesianize’ Inference to the best explanation (e.g., Douven, 1999; 2017; Glass, 2012; Henderson, 2014; Climenhaga, 2017; Schupbach, 2017).

**Education:** Theories of rational proof help us understand why particular forms of reasoning are fallacious. Such insights can be translated into tools that help practitioners to reason better. For instance, both the Bayesian account (e.g., Fenton & Neil, 2011) and explanation-based accounts (e.g., Rassin, 2018) have been proposed as ways to improve the evidential reasoning of fact-finders, police officers and other relevant parties, by helping them to commit fewer fallacies and to avoid biases. Given that both approaches have their respective strengths and weaknesses it is worth asking whether a combined view might be helpful in teaching people to reason better.
5. Concluding remarks

Someone once told me that researchers have a habit of being drawn to what they themselves have a problematic relationship with.²³⁴ This is certainly true for this thesis. Part of the reason why I am fascinated with the topic of rational reasoning is because I frequently fall into all kinds of cognitive traps myself. It is a good thing that I will (hopefully) never have to decide on the fate of a criminal defendant. Of course, as psychologists have shown over and over again, those who do have to make such decisions are also vulnerable to reasoning mistakes. This sometimes leads to judicial errors with severe consequences. How may we prevent such mistakes? This question has sparked a debate that has been going for decades. One important goal within this debate has been the development of a theory of rational proof that explains why certain factual inferences in criminal trials are mistakes and what it means to reason well. In this thesis I sought to contribute to the development of such a theory.

This thesis was written in the tradition of analytic epistemology. It is therefore neither legal-dogmatic nor empirical in nature: it does not deal with the law-on-paper nor with the-law-out-there of a particular legal system. Rather, it starts from the premise that, regardless of the particulars of one’s legal system, in criminal trials we are after the truth - we want to know what happened. In many legal systems, the law does not have much to say about how we should go about the task of truth-finding. In contrast, epistemologists (as well as their sisters and brothers in the philosophy of science) have written extensively on this matter. In this thesis I drew upon their ideas to make sense of several thorny issues that regularly pop up in criminal cases. One of the things that I therefore hope to have shown is how philosophical ideas can find fruitful soil outside of the philosophy department.

Of course, I do not mean to imply that philosophers should descend to the other sciences to deliver haughty edicts on how things should be done. When engaging

²³⁴ Thank you for this insight, Pauline.
with legal matters, epistemologists should not forget that their work has to be grounded in real-life practice, in order for it to make a real contribution. One part of the analytic tradition that I have therefore tried to leave behind is the tendency to only engage with highly stylized thought experiments. Instead, I used real-world examples as much as I could.\textsuperscript{235} This brings us back to the quote by Ludwig Wittgenstein on one of the first pages of this thesis, about the dangers of a one-sided diet of examples. Another contribution of this thesis has been to illustrate how epistemologists can enrich their diet.

Finally, let me say something about my writing process. To use terminology borrowed from Isaiah Berlin (2013), my approach has been that of the fox, not that of the hedgehog: I did not start out with the aim of providing a unified theory of rational proof. Instead, this thesis is the result of struggling with several related but independent questions, which happened to interest me at the time. For example, I began thinking about the ideas in chapter III when someone made a remark at a workshop devoted to the Simonsbaven case that the court was correct in convicting the defendant because his story (that he and his wife were attacked by a madman) was much less plausible than that of the prosecution. Another person responded by saying that the defendant’s scenario was not completely unbelievable and that it therefore raised a reasonable doubt. This made me wonder about which of these two remarks was correct. Many of the other ideas in this thesis came about in a similar way. This way of writing does have its downsides. For example, it took some effort to knit the resulting collection of ideas into a coherent story. But there is also a major upside to being a fox: it allows you to chase after what catches your fancy, without the burden of being chained to a single idea. And with that, I say goodbye to you, dear reader, to go chase after something new.

\textsuperscript{235} Though, these were, by necessity, still simplified. I tried to preserve their most important details.
LITERATURE


Allen, R. J. (2010). No plausible alternative to a plausible story of guilt as the rule of decision in criminal cases. In J. Cruz (Ed.), *Proof and standards of proof in the law* (pp. 10–27). Northwestern University School of Law.


Simpson, S. (2015). *Serial: Evidence that Jay’s story was coached to fit the cellphone records.* view-fromll2.com/2015/01/13/serial-evidence-that-jays-story-was-coached-to-fit-the-cellphone-records/


Tollenaar, A., Marseille, A. T., & de Graaf, K. J. (2008). Establishing facts by the administrative judge. dx.doi.org/10.2139/ssrn.1294303


APPENDIX I:

SUMMARY

Fact-finding is one of the most important parts of criminal cases. The fact-finder (a judge or a jury, depending on the legal system) has to determine whether it is proven that the defendant committed the acts alleged in the indictment. But when are they justified in considering these facts proven? To answer this question, we need an account of rational criminal proof. Over the past decades, two such accounts have been especially influential: the Bayesian and explanation-based frameworks. The Bayesian account casts legal proof in terms of our degrees of belief in hypotheses, which we update when receiving evidence. On the explanation-based account, legal proof is a comparison between competing explanations of the evidence. These explanations can be (im)plausible to varying degrees, where a plausible explanation is one that explains the evidence well in that it makes this evidence expected, fits with our background beliefs about the world and is internally coherent. In criminal cases such explanations are often ‘scenarios’, which describe a timeline of causally connected events that may have occurred. These two accounts are often presented as rivals. However, some suggest that it may be possible to combine them and that such a combination can be beneficial, as the two can complement one another. The aim of this thesis is to spell out such a compatibilist account.

One obstacle for combining Bayesianism and explanationism is that it is not always clear what authors mean when they talk about these accounts. With respect to Bayesianism, there are at least two distinct ideas that legal evidence scholars refer to by this name. The first is the prescriptive idea that people should (sometimes) use probabilistic tools, such as Bayesian networks, for analyzing and presenting evidence at trial or in related contexts. Second, Bayesianism may refer to a normative, epistemic theory of rational proof that offers an account of how we ought to change our degrees of belief in the light of a varying set of evidence and when the fact-finder may therefore believe that something probably happened. In this thesis, I
focus on this second notion. I argue that a crucial fact-finding question which Bayesianism does not answer is how we should make sense of the available evidence. In other words, how do we determine what facts are relevant and how to interpret these facts? Explanation-based accounts offer an answer to these questions: it depends on the scenarios that we consider. However, explanation-based accounts face a problem of their own. On these accounts, legal proof occurs through explanatory comparison, which means to look at several possible accounts of what may have happened and to choose to accept one (or none) of these as true. However, it is not clear how the fact-finder can make this choice rationally. To put it another way, when is the fact-finder justified in believing that a particular explanation is (probably) true? I propose that we can best answer this question through the lens of Bayesian probability theory.

These considerations lead to the following research questions:

(I) How does explanation-based reasoning help make sense of the available evidence?

I.I: How does explanation-based thinking help us make sense of reliability evidence?

I.II: How does explanation-based thinking help us reason about the completeness of our set of information?

(II) When are we justified in accepting an explanation as probably true?

II.I: How should the plausibility of explanations influence judgments about their probability?

II.II: How should the reliability of an explanation’s source influence that explanation’s probability?

II.III: How should the completeness of our set of information influence judgments about the probability of explanations?
Summary per chapter

Chapter II introduces the Bayesian and explanation-based accounts of rational proof and their main problems, and it sketches a picture of how the two can complement each other. The general proposal that I defend in this thesis is that explanation-based thinking complements Bayesianism, because it helps us select and interpret the available evidence. On the other hand, Bayesianism offers a useful calculus for making precise when explanation-based reasoning leads us to probably true conclusions. I suggest that we can draw lessons from the philosophy of science on the precise connection between the two. Following several philosophers of science, I argue that we can best view explanatory choice as an imperfect, but necessary and useful, heuristic for optimal Bayesian inference. Explanation-based thinking may sometimes lead to outcomes which are not fully rational from a Bayesian perspective, but such thinking is an efficient method of making sense of the evidence.

Chapters III to VI focus on rationally choosing between the available explanations. Chapter III discusses the proof of guilt beyond a reasonable doubt standard. As most explanationists interpret this standard, guilt is only proven if there is a plausible guilt explanation and no plausible innocence explanation. In contrast, I defend an inference to the best explanation-based interpretation, on which guilt is proven if (i) the best guilt explanation in a case is substantially more plausible than any innocence explanation, and (ii) there is no good reason to presume that we have overlooked evidence or alternative explanations that could realistically have exonerated the defendant. Using the Dutch Simonshaven case as an example, I argue that my interpretation is better suited for arriving at accurate verdicts than the non-comparative no plausible alternative account. The latter yields the wrong verdict in cases where there is a barely plausible innocence scenario which is strongly outperformed by a highly plausible guilt scenario. Furthermore, I argue that my account is not susceptible to the main argument against IBE-based interpretations of proof of guilt, namely that they yield the wrong verdict in bad lot situations, where we
only have poor explanations.

Chapter IV concerns the Dutch Supreme Court’s ruling in the Venray murder case. In this ruling, the court established when and how courts can reject alternative explanations offered by the defendant if there is no evidence that refutes these explanations. The Supreme Court distinguishes three such situations. First, some explanations can be rejected because they ‘did not become plausible’. Second, explanations may be ‘incredible’. Finally, some explanations are so ‘highly improbable’ that the court does not have a duty to respond to them. However, the court did not explain these phrases, leading to confusion about how they ought to be interpreted. I use Bayesian probability theory to clarify them. First, if an explanation with a low prior probability does not become probable by means of the evidence, then the explanation fails to create a reasonable doubt. It has therefore failed to become plausible. Second, whether an explanation offered by a defendant is probable partially depends on evidence about the credibility of the defendant. An explanation may, looked at on its own, be plausible. But given how and when the defendant offered the scenario, it may be incredible. Third, what distinguishes ‘highly improbable’ explanations from improbable explanations that the court should respond to is that their improbability is obvious. When an explanation is obviously improbable, the court would not serve the goals of making its decision understandable by offering a response.

Chapter V examines the value of predictions. An explanation can be confirmed by evidence in two ways. First, it can explain known evidence (also known as ‘accommodation’). Second, it may successfully predict novel evidence. Philosophers of science have made much of this distinction, arguing that explanations which lead to successful predictions are – all else being equal – epistemically superior to those that merely explain known data. I draw on arguments from the predictivism debate in the philosophy of science to distinguish three ways in which the predicted evidence can yield stronger support for a scenario than if the same evidence is accommodated, using the Bayesian likelihood ratio as a measure of the strength of
evidence. First, witnesses who predict are – all else being equal – more reliable than those who do not. Witnesses who only explain known facts can be biased by their knowledge of these facts, or they may deliberately offering a false story that fits with the available information. Second, investigators who only explain known facts run the risk of ‘fudging’ the scenarios that they formulate. Fudging means that their desire to explain as much of the known facts as possible means that they propose a scenario which fits with these facts but lacks internal coherence or fits poorly with our background beliefs. In contrast, prediction yields less incentive for this kind of overfitting. Third, carefully constructed predictions may help investigators to avoid confirmation bias, as they allow a ‘falsificationist’ mindset, where investigators explicitly seek out evidence that might disconfirm their scenario, rather than implicitly trying to confirm that scenario. Tying the three arguments together, whether a scenario which does not yield testable predictions is thereby suspect depends on the plausibility of the competing explanations for its lack of predictive success. I illustrate these arguments using the case of the murder of Hae Min Lee.

Chapter VI deals with the question how we should evaluate eyewitness reliability. I develop an account based on the notion of Testimonial Inference to the Best Explanation (TIBE). On this account, we assess the reliability of eyewitnesses by comparing different causal explanations for the testimony that was offered. On my account the explanations that we consider are detailed scenarios about why the witness might be telling this particular story. My account is prescriptive in the sense that it offers a procedure that helps evaluators select and interpret the evidence regarding a particular eyewitness’s reliability. I also show that my account fits well with several existing prescriptive methods for eyewitness assessment developed by psychologists and philosophers, and that it fills in some gaps in these existing accounts. Finally, I argue that my account is rational because TIBE is an imperfect but useful heuristic for the Bayesian likelihood ratio of the witness’s testimony. If we assess this ratio accurately, taking the witness’s statements into account will reliably bring us closer to the truth. To the extent that TIBE tracks this ratio it is
therefore rational. That it only tracks this ratio imperfectly is a result of thinking in terms of detailed scenarios, which means that we may overlook alternative scenarios and evidence. However, such scenario-based thinking is a necessary requirement for making sense of the reliability evidence.

In chapter VII, I discuss the problem of unconceived explanations. In criminal trials, investigators, lawyers, judges and juries cannot always imagine every relevant sequence of events – there may be unconceived alternatives. Drawing on ideas from the philosophy of science, I argue that the possibility of unconceived alternatives is an overlooked source of reasonable doubt. In other words, fact-finders should not consider the defendant’s guilt proven if they have good reasons to believe that plausible, unconceived scenarios exist. I propose that this conclusion holds for both the Bayesian and explanation-based frameworks: on either notion of reasonable doubt, we may only consider guilt proven if we presume that our set of hypotheses is exhaustive. Whether or not we are justified in presuming this will depend on our higher-order evidence. I distinguish several types of such higher order evidence: the quality of our investigation, the plausibility of our conceived scenarios and the quantity of our information.

Chapter VIII expands on the ideas from chapter VII. I begin this chapter with the observation that most legal evidence scholars agree that proof of guilt beyond a reasonable doubt requires the belief that the defendant probably committed the alleged acts. However, most scholars also agree that this is not a sufficient condition, as this belief may be unreasonable. I discuss two popular proposals for additional conditions: (i) that the degree of belief should be robust and (ii) that it should be reasonable given the available evidence (should be an evidential probability). Both criteria face difficulties regarding their meaning and utility. For the most common interpretation of robustness, the problem is that it is unclear why it is worthwhile to have robust beliefs. For evidential probability the worry is that the notion is so vague that it does not tell us when one’s belief is actually reasonable in the face of the available evidence. I propose a new interpretation which overcomes these
difficulties and which combines the two conditions into a single criterion. On my account both criteria are about the same issue: the need to take into account the possibility that we have overlooked exculpatory information. Such information includes at least further evidence, alternative scenarios, arguments, interpretations of the evidence or connections between our items of evidence. Proof of guilt beyond a reasonable doubt requires that we are sufficiently certain that we did not miss such information, for the sake of achieving *accurate beliefs* and a *just error distribution*. Error minimization means that criminal fact-finding should lead to as few errors as possible. The goal of error distribution is that, to the extent that errors are made (as is unavoidable in a legal system), these errors should mostly be false acquittals, not false convictions. Both goals rely on not overlooking relevant information. How certain we can be that we did not miss any such information depends on our ‘higher-order evidence’. In this chapter, I focus on one particular type of higher-order evidence: whether the hypotheses under consideration are sufficiently *detailed*. Thinking in terms of detailed scenarios can help us be more confident that we did not overlook anything relevant. By thinking in terms of such scenarios we are less likely to overlook relevant evidence, alternative scenarios, inconsistencies within a scenario or other arguments against our preferred scenario. This idea point leads to a novel way of combining the Bayesian and explanation-based accounts of rational legal proof: explanation-based thinking (thinking in terms of detailed, competing explanations) helps reaching justified judgments of higher-order probability.
Answers to the research questions

The results from the individual chapters lead to the following conclusions regarding the central research questions of this thesis:

(I) How does explanation-based reasoning help make sense of the available evidence?

Evidence is never simply given. It has to be gathered, selected and interpreted. In other words, we have to see how the facts ‘hang together’. The result of this process will depend in part on which explanations we consider and how detailed we make these explanations. So far, most explanationists focus on how thinking in terms of competing explanations can help us make sense of the evidence for or against the ultimate hypothesis – that the defendant committed the acts alleged in the indictment. However, there are at least two other types of evidence that explanation-based thinking can help us make sense of: evidence about the reliability of sources of information and evidence about the completeness of our set of information.

I.I: How does explanation-based thinking help us make sense of reliability evidence?

Reliability evidence informs us about the reliability of some source of information. What conclusions we draw from such evidence will depend on the explanations under consideration. Explanation-based thinking firstly helps us interpret the available reliability evidence. Whether a given fact indicates the reliability or unreliability of a source of information hinges on the comparative plausibility of the explanations that we take into account. Second, explanation-based thinking allows us to select what facts are relevant evidence regarding reliability in a given situation and helps us structure complex sets of reliability evidence.
I.II: How does explanation-based thinking help us reason about the completeness of our set of information?

Reasoning about the possibility of unconceived explanations and overlooked evidence is an important aspect of criminal proof. By thinking in terms of detailed scenarios we lower the chance of overlooking something. For example, it is easier to ask specific, critical questions about detailed scenarios. Such thinking therefore helps us detect flaws and inconsistencies in the case for guilt (or innocence). Furthermore, thinking in terms of detailed scenarios points us to the relevant facts, especially if we consider multiple such scenarios which allows us to look for ‘discriminating evidence’ – evidence which provides support for one scenario and disconfirms the other. It is then easier to see what is relevant within the set of facts that we already have, but also whether there is any further evidence that we could gather. Finally, it is easier to come up with alternative hypotheses to detailed scenarios. Those who fail to think in terms of well-specified hypotheses therefore run a greater risk of overlooking relevant facts, arguments and hypotheses.

(II) When are we justified in accepting an explanation as probably true?

Explanatory choice means that the fact-finder has to choose between potential explanations of what happened. Whether an explanation is probably true will depend on its plausibility, the plausibility of its competitors, the reliability of its source and whether our set of explanations and evidence is complete.

II.I: How should the plausibility of explanations influence judgments about their probability?

A plausible explanation is one that explains the evidence well in that it makes this evidence expected, fits with our background beliefs about the world and is internally coherent. How plausible an explanation is (and how plausible its competitors are) is perhaps the most important determinant for whether we may believe that it
is probably true. In criminal evidence scholarship, explanationists translate this idea into the following interpretation of the beyond a reasonable doubt standard of proof: we should accept an explanation that implies the defendant’s guilt as true if it is plausible, and if there are no plausible alternative explanations consistent with his innocence. This is called the ‘no plausible alternatives’ (NPA) account. On this interpretation, the guilt explanation should therefore meet some minimum threshold for plausibility and no explanation consistent with innocence should meet this threshold. However, this interpretation is problematic for several reasons. For instance, the guilt explanation can be probable despite the presence of a plausible innocence explanation, if the former is much more plausible than the latter. Furthermore, an implausible guilt scenario can sometimes lead to conviction. In particular, such a scenario can be probable if, after a thorough investigation, we are confident that this scenario is (by far) the most plausible explanation of what has happened and that there are no alternative explanations that we overlooked. So, on my account, what matters for the probability of an explanation is how well it fares compared to its competitors and whether we have reasons to believe that we overlooked alternative explanations.

II.II: How should the reliability of an explanation’s source influence that explanation’s probability?

Another factor that can influence the probability of a scenario is whether its source is reliable. For example, a defendant may wait to offer an alternative scenario until late in the criminal proceedings, when all the evidence has been gathered and presented. This can allow them to come up with a carefully tailored, false story that explains the evidence well. Similarly, criminal investigators may come up with a story that seemingly explains the available evidence well, but only because the evidence has been gathered in a biased way. In both cases the resulting explanation is plausible. Nonetheless, the explanations can be improbable if we have good reasons
to believe that they were provided by an unreliable source. It is not only rational to take such reliability into account, Dutch courts also consider this factor when deciding whether a defendant’s alternative scenario leads to reasonable doubt.

One reason to be confident that the scenario’s source is reliable, is if that scenario was confirmed by successful risky predictions. When investigators or defendants make predictions, this offers less room to bias and to lying than merely explaining known facts.

**II.III: How should the completeness of our set of information influence judgments about the probability of explanations?**

When fact-finders and criminal investigators engage in explanatory choice they typically consider only a small number of detailed scenarios. As a result it is possible that they overlook alternative explanations. It is also possible to overlook other types of relevant information, such as exculpatory evidence. Of course, we can never be certain that we did not miss anything but we can be justified in presuming this. In situations where we are not justified in this presumption, the connection between our factual judgments and two key goals of criminal trials is severed. First, we want to make as few errors as possible (*error minimization*). Second, to the extent that we do make mistakes, we prefer them to be false acquittals rather than false convictions (*error distribution*). Because the connection between our beliefs and these two aims is then severed, we cannot justifiably believe that guilt is proven beyond a reasonable doubt. Whether the presumption that we have not missed anything is justified depends on our ‘higher-order evidence’, which includes, among other things, the plausibility and level of detail of our conceived explanations, the quality of the investigation and the complexity of the case at hand.

**The overarching question: wedding the Bayesian and explanation-based accounts**
This thesis is, in part, a response to a well-known question from the philosophy of science. If Bayesian and inference to the best explanation accounts of scientific proof are compatible, are they then just different ways of saying the same thing? If so, is one of them superfluous? We can ask the same about Bayesian and explanation-based accounts of legal proof. The answer that I defend in this thesis is that neither theory offers a complete, or even near-complete account of rational criminal proof on its own and that the gaps in either theory offer a space for a fruitful partnership between the two.

The main worry for the Bayesian theory begins with the observation that, on the most defensible version of the account, the probabilities that we are after are ‘evidential’: they express what is reasonable to believe about some hypothesis given the available evidence. But this means that Bayesian reasoning can only begin once we already know what relevant evidence is and how we should interpret this evidence. Bayesianism does not tell us how to go about this task of making sense of the evidence. We therefore have to turn to explanation-based accounts. Within these accounts, what is relevant information, and how this information should be interpreted, depends on the explanations we consider. By comparing and choosing one (or none) of these available explanations, we come to factual conclusions. However, the main problem for this account is that it does not offer a clear answer to the question how we should make this choice in a way that will lead us to accurate (i.e., probably true) factual conclusions. Bayesian probability theory allows us to distinguish the conditions under which explanation-based choice does and does not lead to probably true conclusions. By showing how and when explanationism tracks the Bayesian account, we show that it is a rational way of drawing factual conclusions. I argue for such a tracking claim by drawing on a well-known idea from the philosophy of science. According to this idea, explanatory comparison is an imperfect but useful heuristic for Bayesian inference. When we engage in explanation-based thinking we deliberately limit ourselves to only a few of the possible scenarios of what happened and focus our attention on a subset of the available evidence. When
we limit ourselves in this way, we open up the possibility of overlooking alternative explanations and evidence. This is suboptimal from an idealized, Bayesian perspective but necessary to make sense of the evidence.

In conclusion, when we spell it out properly, explanation-based thinking is a rational and useful method for drawing probably true conclusions. However, we can only understand why it is rational and useful by means of the Bayesian account. Both accounts are therefore necessary components of an adequate theory of rational criminal proof.
APPENDIX II: SAMENVATTING

Waarheidsvinding is een van de belangrijkste doelen van strafzaken. De rechter of jury (afhankelijk van het rechtsstelsel waarin we ons bevinden) moet bepalen of bewezen is dat de verdachte de in de tenlastelegging genoemde feiten heeft gepleegd. Maar mogen we deze conclusie trekken? Wetboeken en jurisprudentie bieden nauwelijks handvatten om deze vraag te beantwoorden. In plaats daarvan hebben we een theorie nodig van *rationeel strafrechtelijk bewijs*. De twee meest invloedrijke theorieën van de afgelopen decennia zijn *Bayesiaanse en verklaringsbenaderingen*. Bayesiaanse benaderingen begrijpen rationeel juridisch bewijs in termen van kansen. Deze kansen zijn ‘graden van geloof’ die uitdrukken hoe sterk we overtuigd zijn van de waarheid van een bepaalde hypothese. De Bayesiaanse aanpak vertelt ons hoe we deze graden van geloof moeten bijstellen wanneer we nieuw bewijs verkrijgen. Volgens *verklaringsbenaderingen*, daarentegen, kunnen we juridisch bewijs het beste begrijpen als een strijd tussen concurrerende, causale verklaringen van het bewijsmateriaal. Deze verklaringen kunnen in meer of mindere mate aannemelijk zijn. Een aannemelijke verklaring is een verklaring die het bewijsmateriaal adequaat verklaart, in die zin dat (i) het vinden van dit bewijs logisch zou zijn als de verklaring waar is, (ii) de verklaring past bij onze achtergrondovertuigingen over de wereld en (iii) deze intern coherent is. In strafzaken zijn dergelijke verklaringen vaak ‘scenario’s’. Een scenario is een tijdlijn van (deels) causaal verbonden gebeurtenissen die mogelijk hebben plaatsgevonden.

Bayesiaanse en verklaringsbenaderingen worden vaak gezien als rivalen. Sommige auteurs suggereren echter dat ze compatibel zijn en elkaar kunnen aanvullen. In dit proefschrift werk ik dit idee uit en laat zien hoe een gecombineerd perspectief ons in staat stelt om een reeks lastige vragen rondom strafrechtelijk bewijs te beantwoorden.
Eén obstakel voor het combineren van Bayesianisme en verklaringsbenaderingen is dat het niet altijd duidelijk is wat we precies moeten verstaan onder beide theorieën. Zo zijn er ten minste twee verschillende benaderingen die men onder de noemer ‘Bayesianisme’ schaart. Volgens de eerste aanpak is het soms nuttig om methoden uit de kansrekening, zoals Bayesiaanse netwerken, te gebruiken voor het analyseren en presenteren van bewijsmateriaal tijdens of voorafgaand aan het strafproces. ‘Bayesianisme’ kan echter ook verwijzen naar een normatieve, epistemische theorie van rationeel bewijs. Een dergelijke theorie beschrijft wat rationeel is om te geloven gegeven het beschikbare bewijs. In dit proefschrift richt ik mij op deze tweede vorm van Bayesianisme.

Ik betoog dat een cruciale vraag bij het vaststellen van feiten is hoe we het bewijsmateriaal moeten kiezen en interpreteren. Hier heeft de Bayesiaanse aanpak geen antwoord op. Verklaringsbenaderingen bieden wel een antwoord op deze vraag. Welk bewijs relevant is en hoe we dit moeten interpreteren hangt af van de scenario’s die we overwegen. Verklaringsbenaderingen kennen echter een eigen probleem. Binnen deze benaderingen kiezen we uit mogelijke verklaringen voor het bewijs. Een bewezenverklaring van de feiten vereist dat we één van deze verklaringen voor waar aannemen. Het is echter niet duidelijk hoe we deze keuze rationeel moeten maken. Met andere woorden, wanneer mogen we gerechtvaardigd geloven dat een verklaring (waarschijnlijk) waar is? Ik bepleit dat we deze kwestie op een precieze, productieve manier kunnen onderzoeken door middel van het Bayesiaanse model. Deze overwegingen leiden tot de volgende onderzoeksfragen:

(I) Hoe helpt verklaringsdenken het beschikbare bewijs te begrijpen?

I.I: Hoe helpt verklaringsdenken ons bij het begrijpen van bewijsmateriaal over de betrouwbaarheid van ander bewijs?

I.II: Hoe helpt verklaringsdenken ons te redeneren over de volledigheid van onze set aan verklaringen en bewijs?
(II): Wanneer zijn we gerechtvaardigd om een verklaring als waarschijnlijk waar te aanvaarden?

II.I: Hoe beïnvloedt de aannemelijkheid van een verklaring haar waarschijnlijkheid?

II.II: Hoe beïnvloedt de betrouwbaarheid van de bron van een verklaring haar waarschijnlijkheid?

II.III: Hoe beïnvloedt de volledigheid van het bewijs en de set verklaringen de waarschijnlijkheid van onze verklaringen?

**Samenvatting per hoofdstuk**

**Hoofdstuk II** introduceert de Bayesiaanse aanpak en verklaringsbenaderingen, hun belangrijkste problemen en schetst een beeld van hoe de twee elkaar kunnen aanvullen. In dit proefschrift suggereer ik dat de verklaringsbenaderingen ons helpen het beschikbare bewijsmateriaal te kiezen en te interpreteren. Dit is een vereiste voordat we kunnen gaan denken over of de feiten in de tenlastelegging waarschijnlijk plaats hebben gevonden, en dus bewezenverklaard kunnen worden. Het Bayesianisme biedt op haar beurt een nuttige calculus om precies te bepalen wanneer denken in termen van strijdende verklaringen ons tot waarschijnlijk ware conclusies brengt. Om de relatie tussen de twee theorieën preciezer te maken kunnen we putten uit de wetenschapsfilosofie. In navolging van verschillende wetenschapsfilosofen, stel ik voor ik dat we verklaringsdenken het best kunnen zien als een imperfecte maar nuttige heuristiek voor optimaal, Bayesiaans redeneren. Verklaringsdenken kan soms leiden tot uitkomsten die niet volledig rationeel zijn vanuit een (geïdealiseerd) Bayesiaans perspectief, maar deze manier van denken is een efficiënte en mogelijk zelfs noodzakelijke wijze om het beschikbare bewijs te kiezen en interpreteren.

Hoofdstukken III tot en met VI gaan over de vraag hoe we rationeel kunnen kiezen tussen de beschikbare verklaringen. **Hoofdstuk III** bespreekt het concept ‘gerede twijfel’ (*reasonable doubt*). In veel rechtssystemen stelt de strafrechtelijke
bewijsstandaard dat schuld voorbij redelijke twijfel bewezen dient te zijn. Sommige auteurs bepleiten dat het Nederlandse strafrechtstelsel ook (impliciet) deze standaard gebruikt. De meeste voorstanders verklaringsbenadering leggen het concept van gerede twijfel als volgt uit: schuld is alleen bewezen als er een aannemelijke schuldverklaring is en geen aannemelijke onschuldverklaring. Wanneer niet aan beide eisen is voldaan dan is er volgens deze interpretatie sprake van gerede twijfel. Ik verwerp deze opvatting, en verdedig een interpretatie gestoeld op het idee van *inference to the best explanation*. Volgens mijn interpretatie is schuld bewezen voorbij gerede twijfel wanneer (i) de beste schuldverklaring in een zaak aanzienlijk meer aannemelijk is dan enige onschuldverklaring, en (ii) er geen goede reden is om aan te nemen dat we bewijsmateriaal of alternatieve verklaringen over het hoofd hebben gezien die de verdachte realistisch gezien hadden kunnen vrijpleiten. Mijn belangrijkste argument voor waarom we deze interpretatie moeten kiezen is dat deze beter omgaat met een specifiek soort situaties dan de bovengenoemde, heersende interpretatie. Ik gebruik de Nederlandse *Simonshaven*-zaak om te laten zien dat er soms gevallen zijn waarin een aannemelijk onschuldsscenario bestaat, maar waarbij het schuldscenario vele malen meer aannemelijk is. Volgens de heersende opvatting schept het onschuldscenario in zulke gevallen gerede twijfel. Ik betoog echter dat het schuldscenario in zulke gevallen hoogstwaarschijnlijk waar kan zijn. Als het schuldscenario inderdaad waarschijnlijk waar is, zouden de feiten bewezenverklaard moeten worden. Mijn interpretatie leidt in dit geval tot de juiste conclusie. Ook is mijn interpretatie niet onderhevig aan een van de belangrijkste argumenten tegen *inference to the best explanation*, namelijk het *bad lot* probleem. Dit probleem draait om situaties waarin we alleen maar onaannemelijke verklaringen hebben, waarvan een schuldverklaring de beste is. Volgens sommige versies van *inference to the best explanation* zouden we deze schuldverklaring voor waar aan moeten nemen, ondanks het feit dat deze onaannemelijk is. Het is immers de beste verklaring. Mijn interpretatie vermijdt dit probleem via de tweede eis: dat we geen relevant bewijsmateriaal of alternatieve verklaringen over het hoofd hebben gezien. Wanneer we alleen
onaannemelijke verklaringen hebben, dan geeft dit ons een sterke reden redenen om aan te nemen dat er aannemelijke alternatieve verklaringen zijn die we over het hoofd hebben gezien.

In Hoofdstuk IV staat het arrest van de Hoge Raad in de zogeheten ‘Venrayse moordzaak’ centraal. In dit arrest legt de Hoge Raad uit wanneer en hoe rechters alternatieve verklaringen, aangedragen door de verdachte, mogen verwerpen wanneer er geen bewijs is dat deze verklaringen weerlegt. De Hoge Raad onderscheidt drie van soorten gevallen. Ten eerste kunnen sommige verklaringen worden verworpen omdat ze ‘niet aannemelijk zijn geworden’. Ten tweede kunnen ze ‘ongeloofwaardig’ zijn. Ten slotte zijn sommige verklaringen zo 'hoogst onwaarschijnlijk' dat de rechter er überhaupt niet op hoeft in te gaan. De Hoge Raad lichtte de betekenis van deze termen echter niet toe. In dit hoofdstuk ontwikkel ik een interpretatie van de drie begrippen. Hierbij gebruik ik de Bayesiaanse kansrekening om het onderscheid tussen deze termen precies te maken. Ten eerste, als een verklaring met een lage a priori waarschijnlijkheid niet voldoende waarschijnlijk wordt nadat we het beschikbare bewijs in acht nemen, dan slaagt de verklaring er niet in redelijke twijfel te creëren. Ze is dus niet aannemelijk geworden. Ten tweede hangt de waarschijnlijkheid van het naar voren gebrachte verklaring gedeeltelijk af van de geloofwaardigheid van de verdachte als verteller van het verhaal. Een verklaring kan, op zichzelf beschouwd, aannemelijk zijn. Maar, afhankelijk van hoe betrouwbaar we de verdachte inschatten, kan de verklaring toch ongeloofwaardig zijn. Ten derde, wat ‘hoogst onwaarschijnlijke’ verklaringen onderscheidt van onwaarschijnlijke verklaringen waarop de rechter wel zou moeten reageren, is dat hun onwaarschijnlijkheid vanzelfsprekend is. Als de rechter zou moeten reageren op verklaringen die vanzelfsprekend onzinnig zijn, dan zou dit tijd en moeite kosten en niet helpen de uiteindelijke beslissing begrijpelijk te maken.

Hoofdstuk V onderzoekt de waarde van voorspellingen in het strafrechtelijk bewijs. Bewijs kan een verklaring op twee manieren ondersteunen. Ten eerste kan de verklaring bekend bewijs accommoderen – het bewijs wordt achteraf verklaard. Ten
tweede kan zij nieuw bewijs voorspellen. Wetenschapsfilosofen hebben veel geschriven over dit onderscheid. Zij betogen dat verklaringen die leiden tot succesvolle voorspellingen ceteris paribus waarschijnlijker zijn dan verklaringen die alleen bekende feiten accommoderen. Op basis van argumenten uit het zogenaamde predictivisme debat uit de wetenschapsfilosofie, onderscheid ik drie manieren waarop het voorspellen van strafrechtelijk bewijs meer steun oplevert voor een verklaring dan wanneer dat zelfde bewijs achteraf wordt verklaard. Ik gebruik de Bayesiaanse likelihood ratio als maatstaf voor bewijskracht om deze argumenten precies te maken.

Ten eerste zijn getuigen die succesvolle voorspellingen doen doorgaans betrouwbaarder dan getuigen die alleen achteraf verklaren. Getuigen die alleen bekende feiten verklaren lopen een groter risico bevooroordeeld te zijn. Ook is het bij achteraf verklaren eenvoudiger om een aannemelijke leugen te vertellen dan bij voorspellen. Een tweede voordeel van voorspelling is dat wanneer rechercheurs alleen bekende feiten verklaren, het risico bestaat dat de wens om alle beschikbare informatie te verklaren leidt tot incoherente of anderszins ongelofwaardige scenario’s. Voor- 
spellen biedt minder kans op dit soort overfitting. Ten derde kunnen zorgvuldig opgestelde voorspellingen rechercheurs helpen om confirmation bias te vermijden, omdat ze falsificatiepogingen mogelijk maken, waarbij we zoeken naar bewijs dat de verklaring mogelijk kan ontkrachten. Deze drie argumenten leiden samen tot de conclusie dat we de vraag of een scenario dat alleen bekende informatie verklaart daardoor onwaarschijnlijk is, beantwoorden door te kijken naar hoe aannemelijk de mogelijke verklaringen voor dit gebrek voorspellingen zijn. Ik illustreer dit argument door middel van de Amerikaanse zaak over de dood van Hae Min Lee.

Hoofdstuk VI behandelt de vraag hoe we de betrouwbaarheid van ooggetuigen moeten beoordelen. Ik ontwikkel een aanpak op grond van een idee uit de epistemologie genaamd Testimonial Inference to the Best Explanation (TIBE). Dit idee houdt in dat we de betrouwbaarheid van ooggetuigen beoordelen door verschillende, causale verklaringen over hoe de getuigenis tot stand is gekomen met elkaar te vergelijken. Deze aanpak helpt om het bewijsmateriaal over de betrouwbaarheid van
ooggetuigen te kiezen en te interpreteren. Ook past deze aanpak bij bestaande methoden voor de beoordeling van ooggetuigen die zijn ontwikkeld door psychologen en filosofen. Mijn aanpak vult tevens enkele leemten in deze methoden in. Ten slotte is mijn aanpak rationeel. Ik toon namelijk aan dat TIBE een nuttige maar imperfecte heuristiek voor de Bayesiaanse aanpak van ooggetuiguevaluatie. Volgens deze Bayesiaanse aanpak moeten we, wanneer we ooggetuigen evalueren, de likelihood ratio van de getuigenis inschatten. De likelihood ratio drukt uit hoe waarschijnlijk het is dat de getuige met deze getuigenis komt gegeven dat hij de waarheid spreekt in vergelijking met hoe waarschijnlijk deze getuigenis is als we aannemen dat de getuige (bewust of onbewust) niet de waarheid spreekt. Bayesiaanse filosofen hebben betoogd dat wanneer we deze ratio accuraat inschatten, we altijd de kans om tot ware overtuigingen te komen vergroten, wanneer we deze getuigenis meenemen in ons bewijs. Wanneer TIBE over het algemeen leidt tot gerechtvaardigde inschattingen van deze ratio, is ze rationeel. Ik beargumenteer dat TIBE een imperfecte heuristiek is voor deze ratio. Soms leidt ze namelijk tot oordelen die afwijken van het Bayesiaanse ideaal. Dit komt doordat we binnen TIBE vaak denken in termen van een beperkt aantal gedetailleerde verklaringen voor de getuigenis. Dit is een noodzakelijke voorwaarde om het beschikbare bewijsmateriaal over de betrouwbaarheid van de getuige te kunnen begrijpen. Dit zorgt er echter ook voor dat we mogelijke alternatieve verklaringen over het hoofd zien.

In hoofdstuk VII bespreek ik het probleem dat er tijdens een strafproces soms alternatieve verklaringen over het hoofd worden gezien. We kunnen niet altijd ieder mogelijk, plausibel scenario ontdekken. Op basis van ideeën uit de wetenschapsfilosofie betoog ik dat de mogelijkheid dat we scenario’s hebben gemist een bron van gerede twijfel is. Met andere woorden, we mogen de schuld van de verdachte niet bewezen achten als we goede redenen hebben om aan te nemen dat we aannemelijke onschuldverklaringen over het hoofd hebben gezien. Deze conclusie geldt voor zowel de Bayesiaanse aanpak als voor verklaringsbenaderingen. Binnen beide theorieën mogen we de ten laste gelegde feiten alleen als bewezen beschouwen als
we gerechtvaardigd kunnen aannemen dat we alle relevante verklaringen in acht hebben genomen. Of deze aannames gerechtvaardigd is hangt af van ons ‘hogere orde bewijs’. Ik onderscheid verschillende soorten hogere orde bewijs: de kwaliteit van de opsporing, de aannemelijkheid van de beschikbare verklaringen en de kwantiteit van het bewijs.

**Hoofdstuk VIII** werkt een aantal ideeëen uit hoofdstuk VII verder uit. Ik begin dit hoofdstuk met de observatie dat de meeste filosofen het erover eens zijn dat de feiten in een zaak alleen bewezen zijn als de rechter of jury tot de conclusie komt dat deze hoogstwaarschijnlijk waar zijn – de beschreven gebeurtenissen hebben plaatsgevonden. Tegelijk is er consensus onder filosofen dat dit geen voldoende voorwaarde is voor een bewezenverklaring. De overtuiging dat deze feiten hoogstwaarschijnlijk waar zijn kan immers onredelijk zijn. Ik bespreek twee populaire filosofische voorstellen voor aanvullende voorwaarden: (i) dat de mate van overtuiging *robust* moet zijn (niet eenvoudig te veranderen) en (ii) dat deze overtuiging redelijk moet zijn gezien het beschikbare bewijs (een *evidential probability* moet zijn). Het is onduidelijk wat beide criteria betekenen en waarom het niet voldoen van onze overtuiging aan deze criteria tot gerede twijfel zou moeten leiden. Voor de meest gangbare interpretatie van *robustness* is niet helder waarom het waardevol is om *robuste* overtuigingen te hebben. Het begrip *evidential probability* is zo vaag dat het ons niet vertelt wanneer iemands overtuiging redelijk is in het licht van het beschikbare bewijs. Ik stel een nieuwe interpretatie voor die deze problemen ondervangt en de twee voorwaarden combineert tot één criterium. Beide criteria hebben betrekking op de noodzaak om rekening te houden met de mogelijkheid dat we ontlastende informatie over het hoofd hebben gezien. Dergelijke informatie kan bestaan uit ontlastend bewijsmateriaal, alternatieve scenario’s, onvoorziene argumenten, interpretaties van het bewijsmateriaal of verbanden tussen onze bewijsstukken. Schuld is alleen bewezen wanneer we er voldoende zeker van zijn dat we dergelijke informatie niet over het hoofd hebben gezien. We moeten rekening houden met deze mogelijkheid om ten minste twee redenen. Ten eerste willen we ervoor te
zorgen dat onze overtuigingen accuraat zijn. Ten tweede is het wenselijk dat, voor zo ver er toch gerechtelike dwalingen plaatsvinden, deze vooral bestaan uit onterechte vrijspraken en in mindere mate uit onterechte veroordelingen. Hoe zeker we kunnen zijn dat we geen ontlastende informatie niet over het hoofd hebben gezien, hangt af van ons hogere-orde bewijs. In dit hoofdstuk richt ik me op één bepaald type hogere-orde bewijs, namelijk of de beschikbare verklaringen in een strafzaak voldoende gedetailleerde scenario’s zijn. Denken in termen van gedetailleerde scenario’s kan ons helpen er zekerder van te zijn dat we niets relevants over het hoofd hebben gezien. Wanneer we denken in termen van dergelijke scenario’s wordt het namelijk minder waarschijnlijk dat we alternatief bewijs, alternatieve scenario’s, inconsistenties binnen een scenario of andere argumenten tegen de beschikbare scenario’s over het hoofd zien. Deze observatie leidt tot een nieuwe manier om de Bayesiaanse aanpak en verklaringsbenadering te combineren: verklaringsdenken (denken in termen van gedetailleerde, concurrerende scenario’s) helpt om tot gerechtvaardigde oordelen over hogere-orde waarschijnlijkheid te komen.
Antwoorden op de onderzoeksvragen

De resultaten van de afzonderlijke hoofdstukken leiden tot de volgende conclusies wat betreft de onderzoeksvragen van dit proefschrift:

(I) Hoe helpt verklaringsdenken het beschikbare bewijs te begrijpen?

Bewijs is nooit kant en klaar beschikbaar. Het moet worden verzameld, geselecteerd en geïnterpreteerd. Dit proces hangt voor een belangrijk deel af van de verklaringen die we overwegen en hoe gedetailleerd deze verklaringen zijn. Tot dusver hebben de meeste voorstanders van de verklaringsbenadering zich beziggehouden met de vraag hoe het denken in termen van wedijverende verklaringen ons kan helpen het bewijsmateriaal te begrijpen dat direct betrekking heeft op de vraag of de verdachte de in de tenlastelegging genoemde feiten heeft begaan. Er zijn echter ten minste twee andere soorten bewijs waarbij verklaringsdenken ons kan helpen het te begrijpen: bewijs over de betrouwbaarheid van bewijs (zoals een getuigenverklaring) en bewijs over de volledigheid van onze set verklaringen en bewijs.

I.I: Hoe helpt verklaringsdenken ons bij het begrijpen van bewijsmateriaal over de betrouwbaarheid van ander bewijs?

Denken in termen van mogelijke verklaringen helpt ons ten eerste bij het interpreteren van bewijs over de betrouwbaarheid van een informatiebron. Of een bepaald feit voor of tegen betrouwbaarheid spreekt wordt bepaald door de relatieve aanneemelijkheid voor de causale verklaringen voor dit bewijsstuk die we in ogenschouw nemen. Ten tweede helpt verklaringsdenken ons de keuze maken welke feiten relevant betrouwbaarheidsbewijs zijn in een bepaald geval. Ten slotte helpt zulk denken ons bij het ordenen van het beschikbare bewijs over betrouwbaarheid.

385
I.II: Hoe helpt verklaringsdenken ons te redeneren over de volledigheid van onze set aan verklaringen en bewijs?

Nadenken over of we mogelijk verklaringen en/of bewijs over het hoofd hebben gezien is een belangrijk onderdeel van strafrechtelijk bewijs. Door te denken in termen van gedetailleerde scenario's verlagen we de kans dat we iets relevants vergeten op te merken. Het is bijvoorbeeld gemakkelijker om specifieke, kritische vragen te stellen over gedetailleerde scenario's. Op die manier helpt zulk denken ons ook om fouten en inconsistenties te ontdekken. Ook maakt denken in termen van gedetailleerde scenario's het eenvoudiger om te zien welke feiten relevant bewijs zijn, vooral wanneer we meerdere van dergelijke scenario's beschouwen. Dit stelt ons namelijk in staat te zoeken naar ‘discriminerend bewijs’: bewijs dat het ene scenario meer en het andere minder aannemelijk maakt. Ten slotte is het gemakkelijker om alternatieve hypotheses te bedenken voor gedetailleerde scenario's. Wie niet in termen van gedetailleerde hypotheses denkt, loopt dus een groter risico relevante feiten, argumenten en hypotheses over het hoofd te zien.

(II): Wanneer zijn we gerechtvaardigd om een verklaring als waarschijnlijk waar te aanvaarden?

Of een verklaring waarschijnlijk waar is, hangt af van haar aannemelijkheid, de aannemelijkheid van haar concurrenten, de betrouwbaarheid van haar bron en de volledigheid van onze set verklaringen en bewisstukken.

II.I: Hoe beïnvloedt de aannemelijkheid van een verklaring haar waarschijnlijkheid?

Een verklaring is aannemelijk als zij het bewijsmateriaal goed verklaart, in die zin dat het vinden van dit bewijs logisch zou zijn als de verklaring waar is, die past bij onze achtergrondovertuigingen over de wereld en die intern coherent is. Hoe aannemelijk een verklaring is (en hoe aannemelijk haar concurrenten zijn) is misschien
wel de belangrijkste factor die bepaalt of zij waarschijnlijk waar is. Filosofen die strafrechtelijk bewijs bestuderen hebben dit vertaald naar de volgende interpretatie van de ‘beyond a reasonable doubt’ bewijsstandaard: schuld is alleen bewezen als er een zeer aannemelijke schuldverklaring is en geen aannemelijke onschuldverklaring. Volgens deze interpretatie moet de schuldverklaring dus voldoen aan een bepaalde minimumdrempel voor aannemelijkheid en mag geen enkele verklaring die met onschuld overeenstemt, aan deze drempel voldoen. Er zijn echter meerdere redenen waarom deze interpretatie problematisch is, en er soms toe leidt dat we onwaarschijnlijke verklaringen ten onrechte als waar bestempelen of waarschijnlijke verklaringen als onwaar. Zo kan een schuldverklaring waarschijnlijk waar zijn, ondanks de aanwezigheid van een plausibele onschuldverklaring, wanneer de eerstgenoemde veel aannemelijker is dan laatstgenoemde. Ook kan een ongelooifwaardige schuldverklaring soms tot een veroordeling leiden. Een dergelijk scenario kan waarschijnlijk waar zijn als we er, na een grondig opsporingsonderzoek, van overtuigd zijn dat dit scenario (verreweg) de meest plausibele verklaring is van wat er is gebeurd en dat we geen alternatieve verklaringen over het hoofd hebben gezien. Kortom, voor de beoordeling van de waarschijnlijkheid van een verklaring is het dus onder meer van belang hoe goed zij het doet in vergelijking met haar concurrenten en of wij redenen hebben om aan te nemen dat we alternatieve verklaringen over het hoofd hebben gezien.

II.II: Hoe beïnvloedt de betrouwbaarheid van de bron van een verklaring haar waarschijnlijkheid?

Een andere factor die de waarschijnlijkheid van een scenario kan beïnvloeden is of de bron van deze verklaring betrouwbaar is. Een verdachte kan bijvoorbeeld wachten tot hij of zij met een alternatief scenario komt, tot laat in het strafproces, wanneer al het bewijsmateriaal is verzameld en gepresenteerd. Hierdoor kan de verdachte een op het bewijs afgestemd, maar onwaar verhaal bedenken. Ook rechercheurs kunnen, bijvoorbeeld, met een scenario voor de dag komen dat het
beschikbare bewijsmateriaal goed lijkt te verklaren, maar alleen omdat ze vooral hebben gezocht naar bewijsmateriaal dat dit scenario bevestigt. In beide gevallen is de resulterende verklaring aannemelijk in het licht van het beschikbare bewijs, maar toch onwaarschijnlijk, omdat we goede redenen hebben om aan te nemen dat hun bron – de verhalenverteller die ze naar voren heeft gebracht – niet betrouwbaar is. Willen we tot goede inschattingen komen over hoe waarschijnlijk een bepaalde verklaring is, dan moeten we dus ook de betrouwbaarheid van deze bron in acht nemen. In de praktijk doen Nederlandse rechters dit ook wanneer zij beslissen of het alternatieve scenario van een verdachte tot gerede twijfel leidt.

Eén reden om aan te nemen dat de bron van een scenario betrouwbaar is, is als dit scenario is bevestigd door succesvolle, risicovolle voorspellingen. Wanneer de scenario’s van rechercheurs of verdachten tot zulke voorspellingen leiden, biedt dit minder ruimte voor vooringenomenheid en liegen dan wanneer alleen bekende feiten achteraf worden verklaard.

II.III: Hoe beïnvloedt de volledigheid van het bewijs en de set verklaringen de waarschijnlijkheid van een verklaring?

Tijdens een strafproces en het daaraan voorafgaande opsporingsonderzoek worden meestal maar een beperkt aantal gedetailleerde scenario’s in beschouwing genomen. Daardoor is het mogelijk dat alternatieve verklaringen over het hoofd worden gezien. Hetzelfde kan gebeuren met andere soorten relevante informatie, zoals ontlustend bewijsmateriaal. We kunnen er nooit zeker van zijn dat we niets over het hoofd hebben gezien. Toch kunnen we wel gerechtvaardigd zijn om dit aan te nemen. Wanneer we niet gerechtvaardigd zijn om aan te nemen dat we niets over het hoofd hebben gezien dan leidt dit tot gerede twijfel over de schuld van de verdachte. In zulke situaties bestaat er namelijk geen, of onvoldoende verband tussen onze overtuigingen en twee belangrijke doelstellingen van strafprocessen. Het eerste doel is om zo min mogelijk fouten maken (error minimization). Het tweede doel
is, dat voor zo ver we toch fouten maken, dit bij voorkeur onterechte vrijspraken zijn in plaats van onterechte veroordelingen (*error distribution*). Wanneer er onvoldoende verband bestaat tussen onze overtuigingen en deze twee doelen zijn we niet gerechtvaardigd om de schuld van de verdachte bewezenverklaard te achten. Of we gerechtvaardigd aan mogen nemen dat we niets over het hoofd hebben gezien hangt af van ons ‘hogere-ordebewijs’. Zulk bewijs omvat onder meer de aannemelijkheid en mate van detail van de beschikbare verklaringen, de kwaliteit van het opsporingsonderzoek en de complexiteit van de zaak in kwestie.

**De overkoepelende vraag: een partnerschap tussen de Bayesiaanse aanpak en de verklaringsbenaderingen**

Dit proefschrift is deels een antwoord op een bekende vraag uit de wetenschapsfilosofie: als de Bayesiaanse en *inference to the best explanation* theorieën van rationeel wetenschappelijk bewijs verenigbaar zijn, zijn het dan gewoon verschillende manieren om hetzelfde te zeggen? Zo ja, is één ervan dan overbodig? Dezelfde vraag kunnen we stellen over Bayesiaanse aanpakken en verklaringsbenaderingen van rationeel bewijs in het strafrecht. Mijn antwoord op deze vraag was dat geen van beide theorieën op zichzelf een volledige, of zelfs maar bijna volledige, analyse biedt van rationeel strafrechtelijk bewijs. De lacunes in beide theorieën bieden ruimte om de theorieën te combineren.

Het belangrijkste probleem voor de Bayesiaanse theorie begint met de observatie dat de waarschijnlijkheden die we moeten inschatten uitdrukken in hoeverre het beschikbare bewijsmateriaal een bepaalde hypothese ondersteunt. Maar dit betekent dat we pas Bayesiaans kunnen redeneren als we weten wat het relevante bewijsmateriaal is en hoe we dit moeten interpreteren. De Bayesiaanse aanpak vertelt ons niet hoe we dit moeten aanpakken. De verklaringsbenadering heeft hierop wel een antwoord. Volgens deze benadering bepalen we wat als relevant bewijs telt en hoe we dit bewijs moeten interpreteren aan de hand van de beschikbare
verklaringen in een zaak. Door deze verklaringen te vergelijken en één (of geen) ervan voor waar aan te nemen, komen we tot conclusies over of de feiten in de zaak bewezenverklaard zijn.

Een belangrijk probleem voor verklaringsbenaderingen is echter dat zij geen duidelijk antwoord hebben op de vraag hoe wij deze keuze op een rationele manier moeten maken. Met andere woorden, wanneer mogen we gerechtvaardigd aannemen dat een bepaalde verklaring waarschijnlijk waar is? De Bayesiaanse kansrekening stelt ons in staat te bepalen wanneer verklaringsdenken al dan niet leidt tot waarschijnlijk ware conclusies. Door aan te tonen hoe, en wanneer, de conclusies die we trekken uit verklaringsdenken overeenstemmen met optimaal, Bayesiaans denken, tonen we aan dat zulk denken een rationele manier is om feitelijke conclusies te trekken. In dit proefschrift onderbouw ik deze zogeheten *tracking claim* door voort te borduren op een idee uit de wetenschapsfilosofie. Volgens dit idee is verklaringsdenken een onvolmaakte, maar nuttige heuristiek voor Bayesiaans denken. Wanneer we denken in termen van verklaringen, beperken we ons bewust tot slechts enkele van de mogelijke scenario's van wat er gebeurd is, en richten wij onze aandacht op een subset van het beschikbare bewijsmateriaal. Door deze beperking wordt het mogelijk dat we alternatieve verklaringen en bewijsmateriaal over het hoofd zien. Dit is suboptimaal vanuit een geïdealiseerd Bayesiaans perspectief, maar noodzakelijk om het bewijsmateriaal te kunnen begrijpen.

Kort samengevat betoog ik dat verklaringsdenken een rationele en nuttige methode om feitelijke conclusies te trekken. We kunnen echter alleen door middel van de Bayesiaanse aanpak begrijpen waarom, en wanneer, ze rationeel is. Beide theorieën zijn dus noodzakelijke elementen van een volledige theorie over rationeel strafrechtelijk bewijs.
APPENDIX III:
SUMMARY FOR THE GENERAL PUBLIC

A key question in criminal trials is, ‘may we consider the facts of the case proven?’ Partially in response to miscarriages of justice, philosophers, psychologists and mathematicians have considered how we can answer this question rationally. The two most popular answers are the Bayesian and the explanation-based accounts. Bayesian models cast criminal evidence in terms of probabilities. Explanation-based approaches view the criminal justice process as a comparison between causal explanations of the evidence. Such explanations usually take the form of scenarios – stories about how a crime was committed. The two approaches are often seen as rivals. However, this thesis argues that both perspectives are necessary for a good theory of rational criminal proof. By comparing scenarios, we can, among other things, determine what the key evidence is, how the items of evidence interrelate, and what further evidence to collect. Bayesian probability theory helps us pinpoint when we can and cannot conclude that a scenario is likely to be true. This thesis considers several questions regarding criminal evidence from this combined perspective, such as: can a defendant sometimes be convicted on the basis of an implausible guilt scenario? When can we assume that we are not overlooking scenarios or evidence? Should judges always address implausible innocence scenarios of the accused? When is it necessary to look for new evidence? How do we judge whether an eyewitness is reliable? By combining the two theories, we arrive at new insights on how to rationally reason about these, and other questions surrounding criminal evidence.
APPENDIX IV:
PUBLIEKSSAMENVATTING

Een van de belangrijkste vragen in strafprocessen is, ‘mogen we de feiten waarvan de verdachte beschuldigd wordt bewezenverklaard achten?’ Mede naar aanleiding van verschillende onterechte veroordelingen hebben filosofen, psychologen en wiskundigen nagedacht hoe we deze vraag rationeel kunnen beantwoorden. De twee belangrijkste antwoorden zijn Bayesiaanse en verklarende benaderingen. Bayesiaanse modellen begrijpen strafrechtelijk bewijs in termen van kansen. Verklaringsbenaderingen zien het strafproces als een vergelijking tussen causale verklaringen voor het bewijs. Zulke verklaringen zijn doorgaans scenario’s – verhalen over hoe een strafbaar feit is gepleegd. Deze benaderingen worden vaak beschouwd als rIVALEN. Dit proefschrift bepleit echter dat beide perspectieven nodig zijn voor een goede theorie van rationeel strafrechtelijk bewijs. Door scenario’s met elkaar te vergelijken kunnen we onder meer bepalen wat het belangrijkste bewijs is, hoe verschillende bewijsstukken samenhangen en welk bewijs nog moet worden verzameld. De Bayesiaanse kansrekening helpt ons precies te maken wanneer we mogen concluderen dat een scenario waarschijnlijk waar is en wanneer niet. Dit proefschrift beschouwt verscheidene vraagstukken over strafrechtelijk bewijs vanuit dit gecombineerde perspectief, zoals: kan een verdachte soms veroordeeld worden op grond van een onaannemelijk schuldscenario? Wanneer mogen we aannemen dat we geen scenario’s of bewijs over het hoofd zien? Moeten rechters altijd ingaan op ongelofwaardige onschuldscenario’s van de verdachte? Wanneer is het nodig om te zoeken naar nieuw bewijs? Hoe beoordelen we of een ooggetuige betrouwbaar is? Door de twee theorieën te combineren komen we tot nieuwe inzichten over hoe we rationeel over deze, en andere kwesties rondom strafrechtelijk bewijs kunnen redeneren.