Prefinal manuscript. See the printed version:

Bex, F.J., & Verheij, B. (2020). Stories about Evidence. Bakens in de Rechtspsychologie. Liber amicorum voor Peter van Koppen (Beacons in Legal Psychology. Liber amicorum for Peter van Koppen) (eds. Horselenberg, R., van Koppen, V., & de Keijser, J.), 309-327. Den Haag: Boom Criminologie.

Stories About Evidence

Floris Bex and Bart Verheij

Introduction

Crime stories seem to come in an unbounded variation. This is evident from fictional crime stories and real crimes alike. In fiction, crime stories have been popular since at least Victorian times – with Sherlock Holmes and Jack the Ripper as prototypes of the rational good and the emotional bad – and remain to be adapted to a contemporary perspective as in the bleak Scandinavian crime series filled with social criticism. Actual criminals also keep surprising the public and the experts.

The variation of crime stories does not make the life of investigators easy. If only all criminals would nicely fit a few prototypes, say a typical drug gang-related drive-by shooting or hotel room *crime passionel*, it would be easy to determine what evidence can show what has happened and solve a case. In practice, it is often hard to determine what exactly happened in a crime, and – even if there is a strong hypothesis of who did it and how – it can be hard to uncover proof beyond a reasonable doubt.

A key issue is then how to rationally manage the evidence and hypotheses about what has happened in a crime in such a way that it can be determined whether the belief in a hypothesis is justified or not. When should we believe the prosecutor's story of what has happened, and when has the defence raised a reasonable doubt?

To us it came as a surprise that stories play an important role in the rational analysis of evidential reasoning. We came to this conviction through our meeting with Peter van Koppen – grand reciter of crime stories and expositor of whether to believe them or not. For one of us (Bart), the exchange started in Maastricht when Peter was appointed as professor of legal psychology in the Department of Metajuridica (from 2003 to 2014). But the importance of stories only imbued our understanding of evidence when Peter joined Floris' PhD research project as a supervisor (with Henry Prakken and Bart; from 2005 to 2009). It was in those days that Peter explained to us that we did not understand evidence at all. Evidence is all about stories, and only stories; such was his message. It took us a while to get a feeling for what he meant, and to appreciate the force of Peter's message. Peters emphasis on open discussion 'with our legs on the table' as a scientific method helped us make progress and enjoy the erratic journey that is research.

In fact, we have come to the conviction that stories play a role that is only addressed by stories: guarding the local and global coherence of the events in light of the evidence. This role is neglected by the other two main rationality tools for evidential reasoning: arguments (focusing on managing conflicting information) and probabilities (analyzing evidential strength).

In the following, we discuss three parts of our journey with Peter, his stories about evidence and his perspective on artificial intelligence.

When to Believe a Story: Anchored Narratives

The first part of the journey concerns the theory of anchored narratives as developed by Peter, Hans Crombag and Willem Wagenaar. The theory is inspired by the psychological experiments of Bennett and Feldman and Pennington and Hastie, who showed the danger of story-telling in court: when a story is told well, for instance, when it is told in the actual ordering of events, more test persons believe the story than when it is told in a different order, for instance following the presentation of the evidence. These and similar experiments emphasize that stories can actually pose a risk for the rational handling of evidence in courts.

The important proposal investigated by Peter, Hans and Willem in their anchored narratives theory is that stories are not merely dangerous, they can actually also play a positive role in court. They emphasized properties of stories and their relations to the evidence that could show that a belief in a story was justified. And, importantly, they did not develop their theory as arm-chair researchers in the ivory tower that a university so often is. On the contrary: they prominently took on their role of what would now be called influencers, and applied their research to real crime cases, serving as experts in the media and in court.

The starting point of the anchored narratives theory (ANT), as described in the books *Dubieuze zaken* and *Anchored narratives*, is that decision makers have to decide on both the *quality of the story* or stories presented by the parties and the *anchoring of the stories*. When judging story quality, it is, for instance, considered whether the story is sufficiently concrete, whether there are elements missing and whether it is consistent. Story anchoring is about anchoring the story in our commonsense beliefs about the world around us. Anchoring is a process in which stories are critically evaluated: questions are asked about the story and the commonsense knowledge needed to believe the story, until the decisionmaker is satisfied and the story can be anchored in the world knowledge. For example, say we have a story about a shooter, let us call him Sacco[1], travelling to the town of Braintree just outside Boston by train. In Braintree, Sacco shot a payroll guard and took off with the payroll money. When Sacco is later apprehended, he is carrying a gun. Sacco argues he has this gun because he has a job as a night watchman. In order to believe this, we have to accept that the fact that ‘night watchmen usually carry guns’. We might say this is not really the case, but that only night watchmen at specific organizations carry guns. We can then ask for evidence showing that Sacco was a night watchman at such an organization. Furthermore, we can also question why Sacco has the gun on him outside of work hours. Sacco might then respond with a new story that acts as a ‘sub-story’ to the original story: that he was on his way home from work. This can be anchored in the knowledge that ‘night watchmen who carry guns take them home with them after work’. Again, this may lead to new questions (‘is it normal that night watchmen take their gun home with them?’, etc.), but at some point, we are satisfied and stop asking questions. Figure 1 illustrates how a story is anchored in commonsense knowledge. At the top, there is the main story describing the course of events around a crime. It has substories which themselves also have substories, and the stories are finally grounded, ‘anchored’, in knowledge of the

[1] This name is obviously taken from the famous Sacco and Vanzetti robbery/murder case, see (Kadane & Schum, 1997).

world, typically in the form of generalizations ('night watchmen usually carry guns, who take them home with them after work').

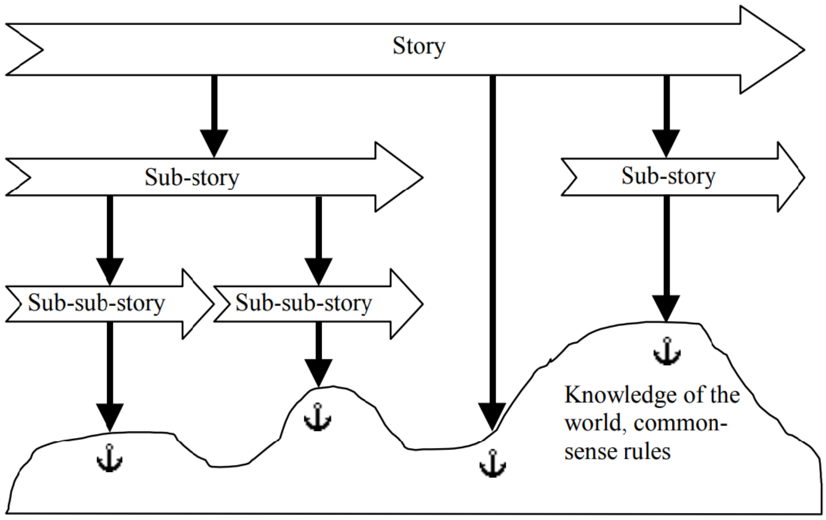

Figure 1. Graphical Illustration of the Anchored Narratives Theory (After *Dubieuze zaken*, p. 72)

ANT was used to establish ten 'universal rules of evidence''(*Anchored narratives*, p. 231 et seq.):

1. The prosecution must present at least one well-shaped narrative.
2. The prosecution must present a limited set of well-shaped narratives.
3. Essential components of the narrative must be anchored.
4. Anchors for different components of the charge should be independent of each other.
5. The trier of fact should give reasons for the decision by specifying the narrative and the accompanying anchoring.
6. A fact-finder's decision as to the level of analysis of the evidence should be explained through an articulation of the general beliefs used as anchors.
7. There should be no competing story with equally good or better anchoring.
8. There should be no falsifications of the indictment's narrative and nested sub-narratives.
9. There should be no anchoring onto obviously false beliefs.
10. The indictment and the verdict should contain the same narrative.

A first attempt at understanding ANT in light of AI & Law research was done at the occasion of the liber amicorum that Peter and Nico Roos prepared for Hans Crombag at the occasion of his valedictory lecture (Van Koppen & Roos, 2000). In that volume, it was considered to what extent the ANT was connected to argumentation as studied in AI & Law (Verheij, 2000). It was concluded that the theories have similarities in the sense that both address the problems of exceptions, conflicts and justification, and that both consider legal decision-making as a process of gradual theory construction, instead of as the straightforward application of rules. Also, differences between the theories are noted. First, the two theories use different primitives as starting point, namely stories and arguments, respectively. Second, the anchored narratives theory is formally less explicit than the theory of argumentation as studied in AI & Law. Third, the approaches have a different setting of use:

whereas anchored narratives theory was used from the start as a tool for evaluating real crime cases, in order to possibly uncover miscarriages of justice, argumentation in AI & Law was studied as a formal, computational model, intended for instance as the foundation of software for developing, diagramming and evaluating arguments.

A second attempt at understanding the anchored narratives theory used argumentation schemes (Walton *et al*., 2008) as a modelling approach. Argumentation schemes can be thought of as semi-formal generalizations of the formal rules of logic, such as modus ponens and classical syllogisms. An example is the following version of an argumentation scheme on expert testimony:

Argumentation scheme for expert testimony
Conclusion.
    There is reason to believe that claim *C* is true.
Conditions.
    Person *P* is an expert on topic *T*.
    Person *P* claims that *C* is true.
    Claim *C* belongs to topic *T*.

As can be seen from the example, argumentation schemes can cover reasoning patterns in a specific, concrete setting (here: expert testimony). Also, they do not guarantee their conclusion when their conditions are satisfied. For instance, in the example the expert can make a mistake or can be lying. Situations that block the use of a scheme are made explicit in the critical questions that come with argumentation schemes. For instance, for the example scheme we could have the following example critical questions:

Critical questions.
    Has the expert not made a mistake when assessing claim *C*?
    Does the expert have a reason to lie about claim *C*?

In sources such as Walton *et al*., (2008), lists of argumentation schemes and their critical questions have been collected.

We have used argumentation schemes to make explicit how (in our interpretation) the anchored narratives theory can be used to determine whether a story is true. Here are two argumentation schemes we proposed (slightly rephrased from Verheij and Bex, 2009):

Argumentation scheme for accepting a story as true
Conclusion
    Story *S* about topic *T* is true.
Conditions
    Story S is good.
    Story S is well-anchored.
Critical questions
    Is there a story S' about topic T (unequal to S) that is good?

Argumentation scheme for assessing story quality
Conclusion

Story S is good.

Conditions

Story S has a central action to which all elements are related.
Story S explains how the central action was performed.
Story S explains why the central action was performed.
Story S is unambiguous.
Story S does not contain contradictions.

(For this scheme no critical questions were listed.)

The first scheme specifies the two mentioned evaluation criteria used in anchored narratives theory: story quality and story anchoring. The second scheme shows how story quality is evaluated, in terms discussed in the original sources of the anchored narratives theory. In Verheij & Bex (2009), this second scheme is later adapted to include the use of story schemes, as inspired by discussions with Peter during Floris' PhD research, in which story schemes play a significant role, as discussed in the next section.

One of the more vocal critics of anchored narratives has been William Twining (1995, 2006). Twining, who, as one of the New Evidence Theorists[2] re-popularized Wigmore's charting method for evidence (Wigmore 1913/1931), argues that in any case, the evidence itself is the most important, and that this evidence should form the basis of the reasoning about what happened in a case. Take as an example the Wigmore chart in Figure 2. Black circles denote evidence brought forth by the prosecution, from which step-by-step inferences can be made to reach the final conclusion 'Sacco was conscious of having been involved in a robbery and shooting' (white circle 155a). Similarly, black diamonds indicate defence evidence, which can be used to infer other conclusions (white diamonds, *e.g.* 'Sacco believed he was being arrested because of his political beliefs', 469), which can in turn be used to counter the original prosecution's inferences.

[2] Together with Terence Anderson, David Schum, Peter Tillers and others.

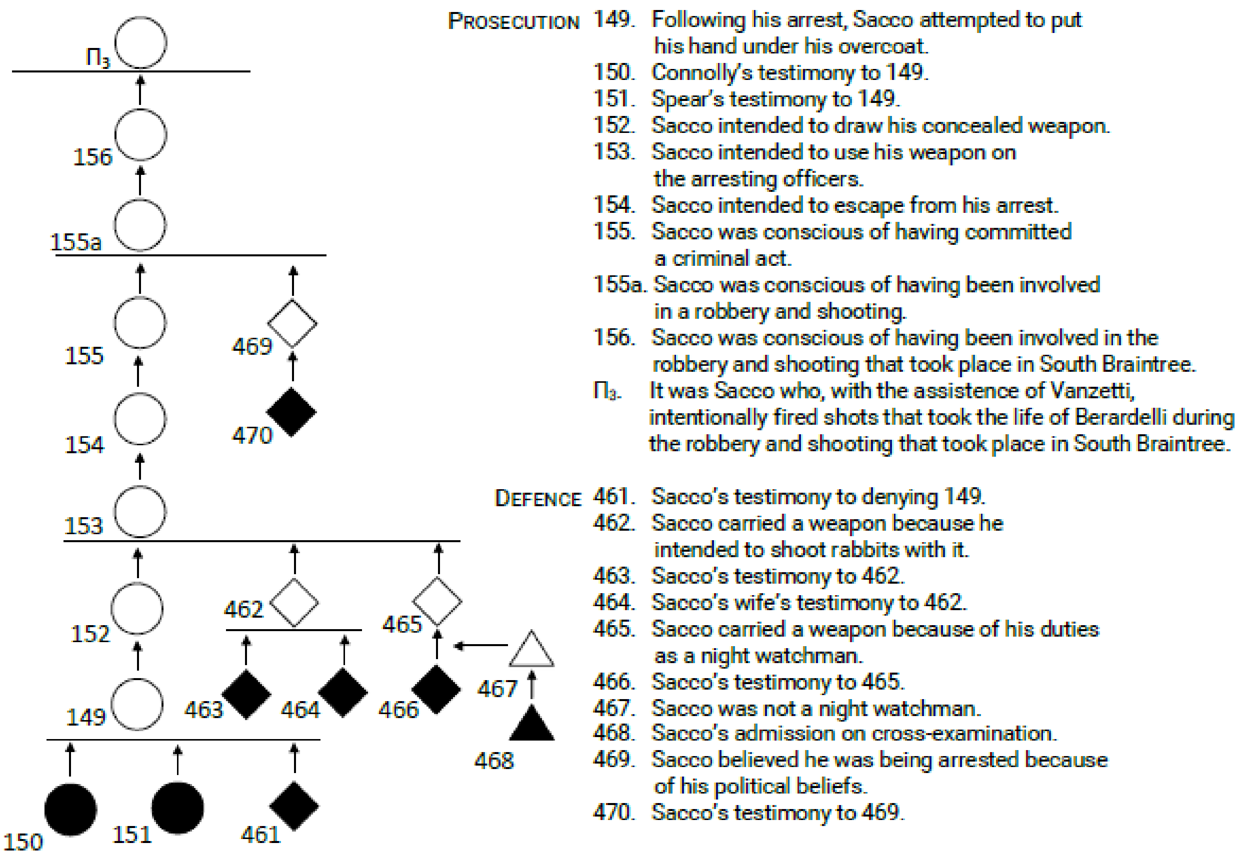

Figure 2. Wigmore Chart or Diagram (Kadane and Schum, 1996)

Twining's main criticism of ANT was that ANT does not consider the concrete, case-specific evidence, but rather seems to consider only general commonsense knowledge to anchor stories, which can contain stereotypes, biases, folk beliefs, and so on (Figure 1). And indeed, in the anchoring process, the search for evidence definitely seems to play a part but the role of the individual pieces of evidence is never really clarified in the description of the anchored narratives theory. At the same time, there does not seem to be deep disagreement between Twining and Crombag, Van Koppen and Wagenaar, since they also argue that a story should be anchored in reality by means of the evidence.

All in all, the anchored narratives theory as developed by Peter and his colleagues provided much inspiration for our thinking about reasoning with evidence, and also led to new ideas for developing a hybrid theory that combined stories with arguments.

Arguments About Stories: A Hybrid Theory

The second part of our journey with Peter took off in early 2005. Henry Prakken, who like Bart worked on logical theories of legal argument (see *e.g.* the influential ASPIC+ model, Prakken 2010), had been awarded an NWO grant 'Making sense of evidence: software support for crime investigations'. He asked Bart and Peter to join as supervisors in an interdisciplinary project performed by Floris for his PhD at both the Law Faculty and the AI Department in Groningen.

The project had as one of its goals to connect the Wigmorean approach, which renders arguments and counterarguments based on evidence in a case as diagrams, with the story-based approach of ANT. Furthermore, the project aimed to incorporate ideas from Artificial Intelligence and Data Visualization into rational theories on evidential reasoning. A clear link already exists between evidential reasoning using arguments and the subfield of AI called computational argumentation, the latter containing work on argument diagramming (Reed *et al*., 2007; Verheij, 2005), logics for legal and commonsense reasoning (Pollock, 1987; Prakken, 1997, Bex *et al*., 2003), and the critical evaluation of arguments in a dialectical process (Dung, 1995). Bex and his colleagues (2003) showed that there are standard argument schemes for evidential reasoning, such as the scheme from expert opinion (see section 2), the scheme from witness testimony, and so on. Using such schemes as inference rules, we can derive conclusions from the evidence, thus building logical arguments, viz.

1.	Eyewitness 1 says ‘I saw Sacco in Boston’	(Evidence)
2.	Sacco was in Boston	(1, Witness Testimony Scheme)
3.	Sacco was not in Braintree	(2)

Here, the Witness Testimony Scheme is used to infer from the evidence that Sacco was in Boston. This scheme is similar to the expert opinion scheme – we have reason to believe the claim is true because an eyewitness says so. However, we can ask the critical question ‘Does the eyewitness misremember?’. Say that answers this question affirmatively because we have evidence that the witness had seen a photo of Sacco (*e.g*., in the newspaper) before being interrogated. We can then build the following argument, which acts as a counterargument to the above.

1.	Witness 1 had seen a photo of Sacco	(Evidence)
2.	Witness 2 misremembers seeing Sacco in Boston *counterargument to*	(1, Misremember Scheme, *Witness Testimony*)

The Misremember Scheme mentioned in this argument is, like the other argument schemes, also a generalization, something like ‘if people see photographs of suspects before they are interrogated, they may mistakenly think they saw the suspect in person’. Note how the dialectical process of asking and answering critical questions in argumentation is similar to the anchoring process as described in the section entitled When to Believe a Story: Anchored Narratives.

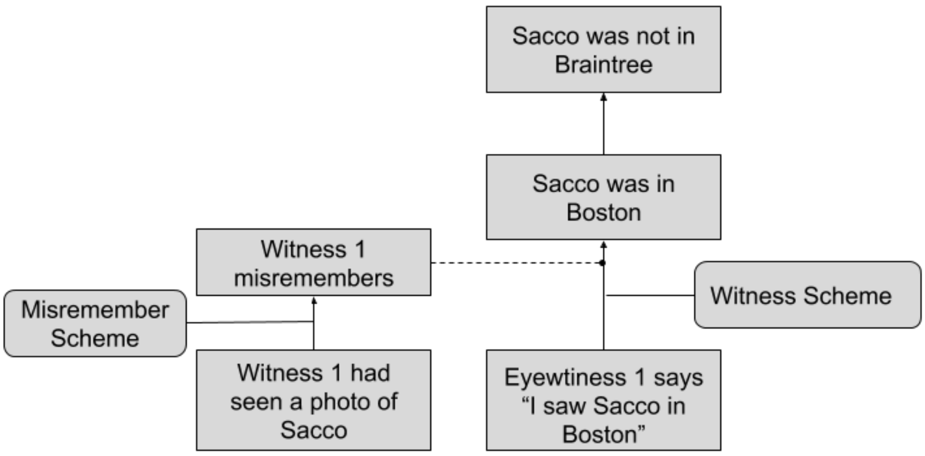

Figure 3. Arguments in a Diagram.
Note Normal arrows denote a 'support' relation and the dashed arrow an 'attack' relation.

The arguments can also be rendered as a semi-formal diagram (Figure 3). Note the similarity between such a diagram and the Wigmore diagram in Figure 2.[3]

Where the link between arguments and AI was thus easily made, the link between stories and AI was less obvious. Looking for more formal work on stories in AI, we quickly found *scripts* – general stories or explanation patterns that were a popular way of representing episodic knowledge in AI (Schank & Abelson, 1977). Such a script is very similar to Pennington and Hastie's typical story structure, also mentioned in ANT: every (good) story has a context, some motive and goals for the main character, a central action and clear consequences of that action. While this is a quite generic 'intentional action' scheme, other more specific story schemes are also possible. Take, for example, the 'murder' scheme (Bex & Verheij, 2012).

Story scheme for murder

1. *Anomaly that the scheme explains*: person y is dead.
2. *Central action of the scheme*: person x kills person y.
3. *Other relevant information*: the motive m, the time of the killing t, the place of the killing p, the weapon w.
4. *Pattern of actions*: person x has a motive m to kill person y – person x kills person y (at time t) (at place p) (with weapon w) – person y is dead.
5. *More specific kinds of murder*: assassination (*e.g.*, liquidation), felony murder (*e.g.*, robbery murder), killing of one's spouse.

Such crime story schemes are thus general scenarios detailing how crimes typically take place (Cornish, 1994). Note that, like argumentation schemes, we can ask critical questions for a crime scheme.

[3] The links between such diagramming methods and AI models of computational argumentation are further explored in Bex *et al*., 2003; Reed *et al*., 2007; Verheij, 2005.

Critical questions

1. Is the motive a legitimate motive for the action?
2. Can the murder be induced by some other motive?
3. Can the action have the stated consequences?

The question was how to incorporate story schemes and stories into a more formaliszd, AI-based reasoning framework. The above story scheme already mentions an ‘anomaly to be explained’, which leads us to diagnostic expert systems,[4] in which some evidence or observations are explained by a hypothesis. This is very similar to explaining the evidence in a legal case by telling a hypothetical story. Take, for example, the fact that was mentioned before, namely that Sacco reached for a gun when he was apprehended. This can be explained by the following story:

Sacco travelled to the town of Braintree just outside Boston by train. In Braintree, Sacco shot a payroll guard and took off with the payroll money. When Sacco was later apprehended, he was still carrying the gun, which he wanted to use to prevent the police from arresting him.

However, it can also be explained by the following story:

Sacco was a night watchman who always carried a gun. When he was arrested, he instinctively reached for his gun because he thought he was being discriminated against because of his political beliefs.

As can be seen, the first story conforms to the ‘murder/robbery’ story scheme, whereas the second story recounts an unfortunate incident or mix-up. The question is then: how do we compare these stories and choose the ‘best’ story? Bex proposes a hybrid theory of stories and arguments, in which alternative hypothetical stories about ‘what happened’ in a case are constructed and discussed in a dialectical process of argument and counterargument. The hybrid theory allows for both alternative explanations and arguments to reason about these explanations. For example, arguments can be used to support an explanation with evidence or to reason about the plausibility of the explanations. Figure 4 shows a simplified example of how arguments can be used to support and attack a story using evidence, and how an alternative explanation can be given. Here are two arguments that support the main story – that Sacco shot and robbed the guard –based on the evidence of eyewitness 2 and the police report. Note how the story is ‘anchored’ in evidence. Also note that, while the generalizations underlying the arguments and the stories have not been rendered in the figure, they are represented by the arrows. Finally, note it is also possible to attack stories and arguments: one argument based on witness testimony 12 attacks the story, but is itself attacked by another argument.

[4] Often used for medical or system diagnosis, see *e.g.* Buchannan & Shortliffe, 1984.

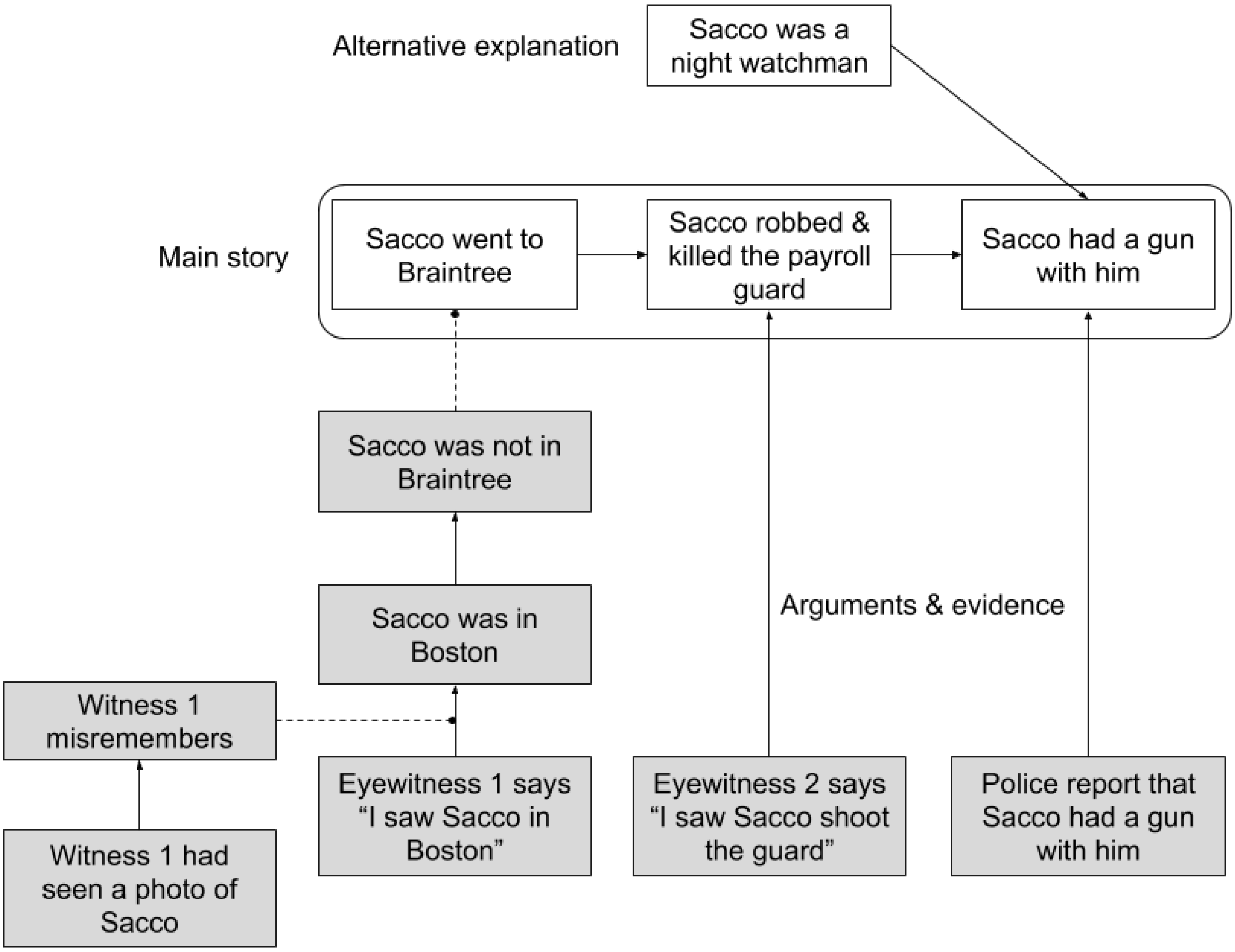

Figure 4. A Story Anchored in Evidence

Several criteria have been formulated that can be applied using the hybrid theory. Broadly speaking, these criteria fall into two categories:

(1) whether the explanation conforms to the specific evidence in the case at hand, that is, how does the available evidence support and contradict the explanation, and which elements of the explanation are not supported by the evidence; and
(2) whether the explanation conforms to general, commonsense knowledge of the world, that is, how plausible are the assumptions and generalizations the explanation makes and how complete, detailed and consistent is the explanation.

These criteria can be defined in more detail using the machinery in the hybrid theory. We can look at the number of pieces of evidence that support or contradict an explanation through arguments. Note that only arguments which are themselves not attacked support or contradict an explanation. Arguments can also be used to reason about the plausibility of an explanation, as the plausibility of elements of an explanation or the explanation as a whole can become the subject of an argumentation process. Arguments about the plausibility of explanations are based on assumptions, as reasoning about plausibility is done using commonsense knowledge about how the world generally works. For plausibility, we do not just consider the individual elements of an explanation (events, causal relations), but also the story as a whole by seeing if the story fits a plausible story scheme.

After Floris' PhD (Bex, 2011), he and Bart kept working together on story schemes in AI (Bex & Verheij, 2010a), the link between legal reasoning and evidential reasoning in AI (Bex & Verheij, 2011, 2013) and also an informal, less schematic version of the hybrid theory for practitioners (Bex & Verheij, 2010b). In this article, of which an extended English version was published later (Bex & Verheij, 2012), we formulated critical questions that point to typical sources of doubt in a hybrid case. Some of these questions concern the arguments in the case – these are the questions associated with the argumentation schemes. Other critical questions in the hybrid theory concern the causal coherence of the explanations. For example, we can ask what kind of general scheme the explanation adheres to – for example, if an explanation speaks of a restaurant visit which lasted only two minutes this is not clear. Furthermore, we can ask whether the explanation is complete – for example, why does a restaurant story not mention any of the actors eating food? Finally, there are general questions that concern the case and comparison of explanations on a higher level. For example, we can ask whether the search for alternative explanations was thorough enough, how decisively some explanation surpasses the alternatives, or how good an explanation is independently of the alternatives. The latter two concern the quality of explanations, and we can use the criteria for explanations given above, taking into account the various standards of proof given by the law.

Floris' PhD project was very successful, being awarded a *cum laude*. At the graduation party, Peter gave a speech, emphasizing how he loved the meetings with the team, just chatting and arguing about evidential reasoning, arguments and stories. The project also included Peter in a publication in an AI journal (Bex *et al*., 2010). Peter would later mention Floris' work in his 2011 book. He lauded the critical thinking involved with 'scheme-building', the use of argumentation and story schemes, but questioned whether the formal schemes and diagrams such as the ones by Wigmore and Bex *et al*. (Figures 2-4) could contain all the necessary knowledge needed in a complex case. Such diagrams and schemes, he argues, necessarily abstract from the complexities of the world around us, which can only be captured by stories. In later work (Bex & Verheij, 2012; Bex, 2019), the focus of the hybrid theory would be less on the diagrams and more on stories, natural language arguments and critical questions.

On Strong and Weak Stories (and the Role of Numbers)

A third part of the journey took the third primitive for the rational handling of evidential reasoning on board: probabilities. In contrast with the qualitative primitives of arguments and scenarios that formed the core of ANT and the hybrid theory, probabilities add a quantitative, numeric perspective. And indeed, both from theory and from practice, there was, and still is, a strong incentive to combine qualitative and quantitative elements. Theoretically, for instance, because the combination of logic (analyzing the logical combinations of propositions) and probability theory (studying probabilistic variables and their relations) has a long and intricate history, connecting to deep debates in epistemology and the philosophy of science (see *e.g.* Dewey *et al*., 2019). Practically, in particular, because of the rise of forensic DNA evidence, that revolution in crime investigation with a strong scientific and statistical underpinning.

Consider, for instance, the issue of whether the suspect is the source of a tissue sample found on the crime scene or someone else. When the found sample has a rare DNA profile, as measured by a database with reference DNA, a good match between the sample's and the suspect's DNA has a high evidential value. The match then comes with a high likelihood ratio distinguishing between the suspect or someone else being the source of the DNA. Using Bayesian calculations, the prior odds that the suspect is the source instead of someone else (before considering the DNA evidence) can be ‘updated' to find the posterior odds (after the evidence) by multiplying the prior odds with the likelihood ratio. With a high likelihood ratio, even a low prior odds can then lead to a high posterior odds.

The success and value of the statistics about DNA sample matching was sometimes used to support that all evidential analysis should be cast in statistical terms, suggesting likelihood ratio as a measure of the strength of evidential support. In practice, it is often not straightforward how to perform a numerical, statistical analysis of the evidence, especially when modelling all the evidence and all the relevant interpretations. For instance, it is not obvious how to combine stories with their emphasis on temporal and causal links between events, with probabilities and the Bayesian calculus.

One attempt to combine scenarios with the Bayesian calculus was performed by Charlotte Vlek in her PhD research (Vlek, 2016).[5] Charlotte’s project was part of the project ‘Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios’ in the NWO Forensic Science research programme.[6] In her PhD research, supervised by Bart, Henry Prakken and Silja Renooij, Charlotte used the AI tool of Bayesian networks as a means to combine qualitative and quantitative elements. Bayesian networks combine a graphical structure (formally a directed acyclic graph) with conditional probability tables. Each node in the graph represents a probabilistic variable and has an associated table of the probabilities of the node’s values conditioned on the values of the node’s parents in the graph. Charlotte developed a method to model the evidence and possible scenarios about a criminal case in a Bayesian network. In the method (Vlek *et al*., 2014), scenarios are modelled as clusters of possible events, and then connected to the evidence. Conditional probability tables make explicit how the events and the evidence are probabilistically related. Charlotte also developed a method to extract an explanatory text from a Bayesian network designed using her method. In the text, the scenarios in the network are made explicit, scenario quality is analyzed, and for each scenario it is shown how it is connected to the evidence (Vlek *et al*., 2016). Charlotte tested her theory with case studies, such as the Anjum case (Crombag & Isräels, 2008). The method was successful in the sense that it combined scenarios with probabilistic degrees of uncertainty. In other words, it provided a way to systematically analyze weakly and strongly supported stories in terms of probabilities. Also, it showed a way to model the coherence of scenarios probabilistically. At the same time, it proved hard to sensibly provide all probabilistic information needed for a Bayesian network (a standard issue for Bayesian networks since they represent full probability distributions).

A second attempt to connect scenarios to probabilities uses the method of case models that provide a formal analysis of argument validity (Verheij, 2014, 2017). Case models consist of cases that (in the setting of evidential reasoning) model the conjunctive combination of hypothetical events and the evidence. The cases come with a preference ordering, that can

---

[5] Other recent attempts include Di Bello (2013) and Urbaniak (2018). See also Di Bello & Verheij (2018).
[6] Further information about the project is available at https://www.ai.rug.nl/~verheij/nwofs/.

represent the plausibility of a case. Since a preference ordering is qualitative in nature, but can be given a numeric representation, the approach can be said to be simultaneously ‘with and without numbers’. Figure 5 shows an example of the use of a case model to analyze a crime case involving the murder of a student in her apartment (Van Leeuwen & Verheij, 2019). The main suspect was the landlord’s son (another tenant of the apartment, P in the figure), who was indeed convicted. The figure illustrates graphically how a case model is gradually constructed by considering the evidence one piece at a time. The construction starts at the top when the victim’s body is found. The case is immediately considered a murder case because of the signs of violence. A friend reported that she had had a phone conversation with the victim that ended abruptly after she heard someone say ‘good morning’, and then screaming and loud noise. All tenants other than the landlord’s son were found to be at work at the time, so two possibilities are considered: P is guilty, or he is not. Gradually, the guilty scenario becomes more concrete after P has fled in the victim’s car and her bloodstains are found in the car. The suspect has a motive, as follows from witness testimony about a conflict, and when the suspect provides a contradictory testimony about a medical examination and confesses to his parents (in a phone call), only one hypothetical scenario is considered possible given the totality of the evidence: the suspect killed the victim with the gun found because of a conflict about a washing machine.

In a 2019 honour’s project at the University of Groningen, Ludi van Leeuwen compared the Bayesian network and case model methods, and concluded that both can model scenarios as clusters of events, and connect them to the evidence. Bayesian networks allow for a more fine-grained numerical representation of the weakness or strength of a scenario’s support, but require more numbers than available. Case models allow for a more crude, qualitative modelling of the level of support of a scenario, but have the advantage that no commitment to explicit probability estimates is required. Both allow for an analysis of the effect of evidence as it is gradually added. Further details can be found in Van Leeuwen & Verheij, 2019.

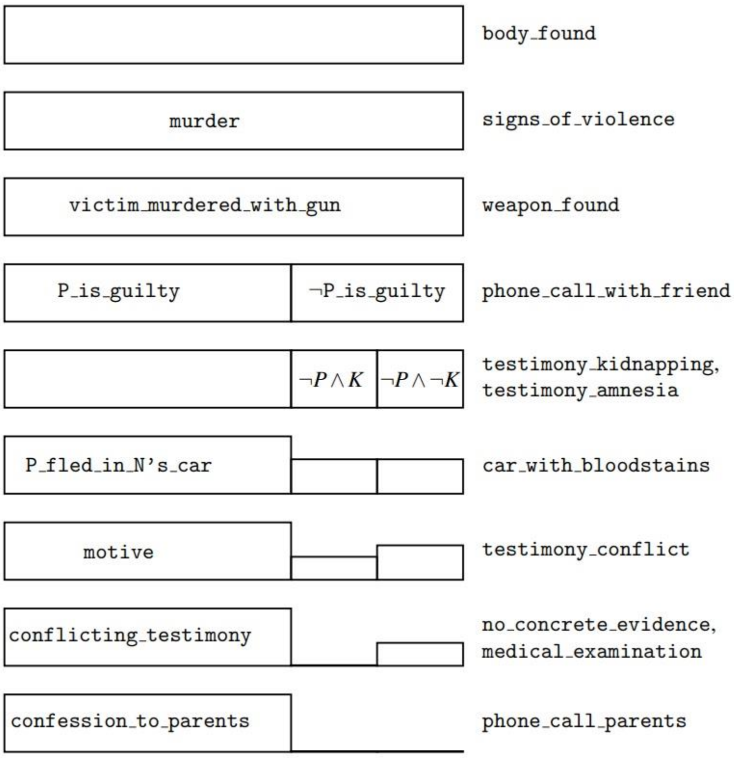

Figure 5. Constructing a Case Model (Van Leeuwen and Verheij, 2019)

Conclusion

We have described our journey with Peter and his stories about evidence in three parts. In the first part, we described the development of our understanding of the anchored narratives approach to the rational handling of the evidence in crime investigation as developed by Peter and his colleagues Hans Crombag and Willem Wagenaar. In the second part, we discussed the combined use of stories and arguments for evidential analysis in the hybrid theory developed by Floris in his PhD research, supervised by Peter, Bart and Henry Prakken. In the third part, we addressed the role of numbers for evidential reasoning, and in particular of probabilities, by summarizing Charlotte Vlek's approach to the embedding of stories in Bayesian networks and Bart's case models 'with and without numbers'.

The journey led us to the belief that stories indeed have an important role to play in the rational handling of the evidence in crime cases: guarding the local and global coherence of the events in light of the evidence. This role is characteristic for the use of stories, and not equally well or straightforwardly addressed by the other two important rationality tools of arguments and stories (see Di Bello & Verheij, 2018 for a systematic discussion of the roles of stories, arguments and probabilities).[7]

Peter's emphasis on case studies has played a significant role in the journey, as again recently in a series of symposia on evidential reasoning and AI.[8] During these symposia, the

[7] And the story continues. See the special issue on interdisciplinary approaches to evidence & decision making in the law in the *Artificial Intelligence and Law* journal (Di Bello and Verheij, 2020).
[8] In 2015 a three-day workshop at the Zentrum für Interdisziplinäre Forschung (Centre for Interdisciplinary Research) in Bielefeld, Germany was held, with legal scholars and practitioners,

idea arose to model a real legal case using the various approaches, in order to compare their strong and weak points. Peter proposed the Simonshaven case, which concerns the death of a woman who was violently killed near the village of Simonshaven, close to Rotterdam, while walking there with her husband, Ed Lourens. A recent special issue on Models of Rational Proof in Criminal Law (Prakken, Bex & Mackor, 2020) collects the different analyses, including an analysis of the case in the hybrid theory, an analysis using the 'scenario-approach' (an amended version of ANT) and numerous Bayesian or probabilistic approaches. Interestingly, the hybrid analysis as well as the scenario-based analysis, which Peter developed together with Anne Ruth Mackor (Mackor & Van Koppen, 2019), are presented as tables of stories, sub-stories, arguments and evidence, showing that conceptually, the hybrid theory and ANT have relevant connections.

Peter has always been simultaneously interested and skeptical about AI approaches to evidential reasoning. He sees a role for 'scheme building' (whether Wigmore or in another way) but has consistently emphasized that human dossier analysis cannot quickly be replaced by computers (see, for instance, recently his comments on computer-supported cold case research[9], a 'crazy plan'; see also Schraagen *et al*., 2018, which proposed an automated intake system for the police based on the hybrid theory).

Indeed, the journey has not ended. New parts will follow and many stories on evidence and its rational use in crime investigation (with or without computer support) will follow. We hope to often find Peter as a companion on this journey and look forward to the stories that he will contribute.

References

Bex, F. J. (2011). *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Dordrecht: Springer.

Bex. F. (2019). The Hybrid Theory of Stories and Arguments applied to the Simonshaven Case. *Topics in Cognitive Science – special issue on Models of Rational Proof in Criminal Law*. Wiley.

Bex, F. J., Van Koppen, P. J., Prakken, H., & Verheij, B. (2010). A Hybrid Formal Theory of Arguments, Stories and Criminal Evidence. *Artificial Intelligence and Law 18*(2), 123–152.

Bex, F.J., Prakken, H., Reed, C., & Walton, D. N. (2003). Towards a formal account of reasoning about evidence: argumentation schemes and generalisations. *Artificial Intelligence and Law 11*, 125–165.

computer scientists, philosophers and cognitive scientists. Followup workshops were organized at the Isaac Newton Institute for Mathematics in Cambridge, England in 2016 and at Queen Mary University London, England in 2017.

[9] 'Hoe de politie oude zaken nieuw leven inblaast met slimme computers', *De Volkskrant*, 31 August 2018. https://www.volkskrant.nl/wetenschap/hoe-de-politie-oude-zaken-nieuw-leven-inblaast-met-slimme-computers~bd5aeaec/.

Bex, F. J., & Verheij, B. (2010a). Story schemes for argumentation about the facts of a crime. *Proceedings of the 2010 AAAI Fall Symposium on Computational Narratives. AAAI Technical Report FS-10-04*, Menlo Park CA: AAAI Press.

Bex, F.J, & Verheij, B. (2010b). Het onderbouwen van een feitelijk oordeel in een strafzaak: methode, casus, aanbevelingen. In P.J. van Koppen, H. Merkelbach, M. Jelicic & J.W. de Keijser (Eds.) *Reizen met mijn Rechter – de Psychologie van het Recht* (pp. 935–952). Deventer: Kluwer.

Bex, F.J., & Verheij, B. (2011). Legal Shifts in the Process of Proof. *Proceedings of the 13th International Conference on Artificial Intelligence and Law*. New York: ACM Press.

Bex, F.J., & Verheij, B. (2012). Solving a Murder Case by Asking Critical Questions: An Approach to Fact-Finding in Terms of Argumentation and Story Schemes. *Argumentation*, *26*(3), 325–353.

Bex, F.J., & Verheij, B. (2013). Legal Stories and the Process of Proof. *Artificial Intelligence and Law 21 (3)*, 253-278.

Buchanan, B. G., & Shortliffe, E. H. (Eds.) (1984). *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison Wesley: Reading, MA, 1984.

Cornish, D. (1994). The procedural analysis of offending and its relevance for situational prevention. *Crime prevention studies*, *3*, 151–196.

Crombag, H. F. M., & Isräels, H. (2008). *Moord in Anjum. Te veel niet gestelde vragen*. Den Haag: Boom Juridische uitgevers.

Crombag, H. F. M., Van Koppen, P. J., & Wagenaar, W. A. (1994). *Dubieuze zaken: De psychologie van strafrechtelijk bewijs*. Amsterdam: Uitgeverij Contact, (2ndedn. (in Dutch).

Demey, L., Kooi, B., & Sack, J. (2019). Logic and Probability. In E.N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2019 Edition). https://plato.stanford.edu/archives/sum2019/entries/logic-probability/

Di Bello, M. (2013). *Statistics and Probability in Criminal Trials: The Good, the Bad and the Ugly.* Dissertation. Stanford University, Stanford.

Di Bello, M., & Verheij, B. (2018). Evidential Reasoning. In Bongiovanni, G., Postema, G., Rotolo, A., Sartor, G., Valentini, C., & Walton, D. (Eds.), *Handbook of Legal Reasoning and Argumentation.* (pp. 447–493). Dordrecht: Springer.

Di Bello, M. & Verheij, B. (Eds.) (2020). Special issue `Evidence & decision making in the law: theoretical, computational and empirical approaches'. *Artificial Intelligence and Law*.

Dung, P.M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n–person games. *Artificial Intelligence 77*(2), 321–357.

Kadane, J. B. & Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. John Wiley and Sons Inc.

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science 11*(4), 481–518.

Prakken, H. (1997.) *Logical Tools for Modelling Legal Argument*, Dordrecht: Kluwer Academic Publishers.

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, *1*(2), 93–124.

Prakken, H., Bex, F.J. & Mackor, A.R. (Eds.) (2020) *Special issue on Models of Rational Proof in Criminal Law, Topics in Cognitive Science*, Wiley.

Reed, C., Walton, D. & Macagno, F. (2007). Argument diagramming in logic, law and artificial intelligence. *Knowledge Engineering Review*, *22*(1), 87–109.

Schank, R.C., & Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: an Inquiry into Human Knowledge Structures*. Hillsdale, New Jersey: Lawrence Erlbaum.

Schraagen, M., Testerink, B., Odekerken, D., & Bex, F. (2018) Argumentation-driven information extraction for online crime reports. CKIM 2018 International Workshop on Legal Data Analysis and Mining (LeDAM 2018), CEUR Workshop Proceedings.

Twining, W. L. (1995). Anchored Narratives, A Comment. *European Journal of Crime, Criminal Law and Criminal Justice 3*, 106–114.

Twining, W. L. (2006). *Rethinking Evidence: Exploratory Essays*. Cambridge: Cambridge University Press.

Urbaniak, R. (2018). Narration in judiciary fact-finding: a probabilistic explication. *Artificial Intelligence and Law* 26 (4), 345–376.

Van Koppen, P. J., & Roos, N. H. M. (Eds.) (2000). *Rationality, Information and Progress in Law and Psychology. Liber Amicorum Hans F. Crombag*. Maastricht: Metajuridica Publications.

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

Van Leeuwen, L., & Verheij, B. (2019). A Comparison of Two Hybrid Methods for Analyzing Evidential Reasoning. In M. Araszkiewicz & V. Rodríguez-Doncel (Eds.), *Legal Knowledge and Information Systems. JURIX 2019: The Thirty-second Annual Conference* (pp. 53–62). Amsterdam: IOS Press.

Verheij, B. (2000). Dialectical Argumentation as a Heuristic for Courtroom Decision Making. In P.J. van Koppen and N. Roos (Eds.), *Rationality, Information and Progress in Law and Psychology. Liber Amicorum Hans F. Crombag* (pp. 203–226). Maastricht: Metajuridica Publications.

Verheij, B. (2005). *Virtual arguments: on the design of argument assistants for lawyers and other arguers*. Den Haag: T.M.C. Asser Press.

Verheij, B. (2014). To Catch a Thief With and Without Numbers: Arguments, Scenarios and Probabilities in Evidential Reasoning. *Law, Probability & Risk 13*, 307–325.

Verheij, B. (2017). Proof With and Without Probabilities. Correct Evidential Reasoning with Presumptive Arguments, Coherent Hypotheses and Degrees of Uncertainty. *Artificial Intelligence and Law 25*(1), 127–154.

Verheij, B., and Bex, F. J. (2009). Accepting the Truth of a Story about the Facts of a Criminal Case. In H. Kaptein, H. Prakken & B. Verheij (Eds.), *Legal Evidence and Proof: Statistics, Stories, Logic (Applied Legal Philosophy Series)* (pp. 161–193). Farnham: Ashgate.

Vlek, C. S. (2016). *When Stories and Numbers Meet in Court Constructing and Explaining Bayesian Networks for Criminal Cases with Scenarios*. Dissertation University of Groningen.

Vlek, C., Prakken, H., Renooij, S., & Verheij, B. (2014). Building Bayesian Networks for Legal Evidence with Narratives: a Case Study Evaluation. *Artificial Intelligence and Law 22*(4), 375–421.

Vlek, C., Prakken, H., Renooij, S., & Verheij, B. (2016). A Method for Explaining Bayesian Networks for Legal Evidence with Scenarios. *Artificial Intelligence and Law 24*(3), 285–324.

Wagenaar, W. A., van Koppen, P. J., & Crombag, H. F. M. (1993). *Anchored Narratives. The Psychology of Criminal Evidence*. London: Harvester Wheatsheaf.

Walton, D. N., Reed, C., & Macagno, F. (2008). *Argumentation Schemes*. Cambridge: Cambridge University Press.

Wigmore, J. H. (1913). *The Principles of Judicial Proof or the Process of Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials* (2nd edn. 1931). Boston (Massachusetts): Little, Brown and Company.

Prof. dr. Floris Bex is in 2009 gepromoveerd aan de Rijksuniversiteit Groningen met een proefschrift getiteld "Arguments, stories and criminal evidence: A formal hybrid theory". Peter van Koppen was, naast Henry Prakken, één van zijn promotores. Sindsdien heeft Floris op meerdere manieren samengewerkt met Peter aan theorieën van verhaal, argument en bewijs, waar Floris de wiskundige aanpak uit de Artificiële Intelligentie (AI) probeert te rijmen met de meer natuurlijke verhaalsaanpak van Peter. Thans is Floris hoogleraar bij het Tilburg Institute for Law, Society & Technology (TILT) van de University of Tilburg, wetenschappelijk directeur van het Nationaal Politielab AI bij het Innovation Centre for AI (ICAI) en universitair docent AI bij het departement Informatica, Universiteit Utrecht. Floris' onderzoek concentreert zich op de verantwoorde inzet van AI voor de politie en de rechtspraak.

Prof. dr. Bart Verheij bekleedt de leerstoel kunstmatige intelligentie en argumentatie aan de Rijksuniversiteit Groningen, waar hij hoofd is van de afdeling kunstmatige intelligentie van het Bernoulli Institute of Mathematics, Computer Science and Artificial Intelligence. Bart en Peter ontmoetten elkaar toen ze beiden werkten bij de capaciteitsgroep Metajuridica van de rechtenfaculteit van de Universiteit Maastricht (2003-2004). Het gesprek over verstandig omgaan met bewijs werd voortgezet bij de gezamenlijke begeleiding van Floris Bex' proefschrift (met Henry Prakken en Peter als promotores en Bart als copromotor; 2005-2009). Peters lucide skepsis en nadruk op het belang van realistische case studies bleven invloedrijk toen Bart een project leidde in het NWO Forensic Science programma (met Henry en Floris) over de combinatie van argumentatie, verhalen en kanstheorie (2012-2017). Bart werkt aan de verbanden tussen kennis, data en redeneren, als bijdrage aan goede kunstmatige intelligentie (zoals uitgelegd in zijn oratie `Arguments for Good Artificial Intelligence'). Hij is co-coordinator van de onderzoekslijn `Responsible Hybrid Intelligence' van het NWO Zwaartekrachtproject Hybrid Intelligence (net gestart, 10-jarig).