7

Thinking about Evidence[1]

DAVID LAGNADO

Introduction

LEONARD VOLE is charged with murdering a rich elderly lady, Miss French. He had befriended her, and visited her regularly at her home, including the night of her death. Miss French had recently changed her will, leaving Vole all her money. She died from a blow to the back of the head. There were various pieces of incriminating evidence: Vole was poor and looking for work; he had visited a travel agent to enquire about luxury cruises soon after Miss French had changed her will; the maid claimed that Vole was with Miss French shortly before she was killed; the murderer did not force entry into the house; Vole had blood stains on his cuffs that matched Miss French's blood type.

As befits a good crime story, there were also several pieces of exonerating evidence: the maid admitted that she disliked Vole; the maid was previously the sole benefactor in Miss French's will; Vole's blood type was the same as Miss French's, and thus also matched the blood found on his cuffs; Vole claimed that he had cut his wrist slicing ham; Vole had a scar on his wrist to back this claim. There was one other critical piece of defence evidence: Vole's wife, Romaine, was to testify that Vole had returned home at 9.30 p.m. This would place him far away from the crime scene at the time of Miss French's death. However, during the trial Romaine was called as a witness for the prosecution. Dramatically, she changed her story and testified that Vole had returned home at 10.10 p.m., with blood on his cuffs, and had proclaimed: 'I've killed her.' Just as the case looked hopeless for Vole, a mystery woman supplied the defence lawyer with a bundle of letters. Allegedly these were

[1] This chapter has benefited greatly from the Leverhulme/ESRC Evidence project, and in particular the wise words of David Schum, Philip Dawid, William Twining, Nigel Harvey, Amanda Hepler, and Gianluca Baio. I also thank Cheryl Thomas, Tobias Gerstenberg, Norman Fenton, Tracy Ray, Adam Harris, and two anonymous reviewers for insightful feedback on earlier drafts of the chapter.

Proceedings of the British Academy, **171**, 183–223. © The British Academy 2011.

written by Romaine to her overseas lover (who was a communist!). In one letter she planned to fabricate her testimony in order to incriminate Vole, and rejoin her lover. This new evidence had a powerful impact on the judge and jury. The key witness for the prosecution was discredited, and Vole was acquitted.

After the court case, Romaine revealed to the defence lawyer that she had forged the letters herself. There was no lover overseas. She reasoned that the jury would have dismissed a simple alibi from a devoted wife; instead, they could be swung by the striking discredit of the prosecution's key witness.

This crime story is a work of fiction. It is drawn from Agatha Christie's play *Witness for the Prosecution*. The story contains twists and turns that are not representative of a typical crime case; however, it serves to illustrate the patterns of inference that recur in real-world legal contexts. The task of the 'fact-finder' (e.g. investigator, judge or juror) is to pull together all the diverse threads of evidence and reach a singular judgment of innocence or guilt. One thing that makes this task so difficult is that the different pieces of evidence are often interrelated. You cannot simply sum up the positive evidence on the one hand, and the negative on the other. The evidence interacts in complex ways. For example, in the above story, Vole's enquiry about a luxury cruise is not relevant on its own; it becomes relevant because he had recently been written into the old lady's will. Moreover, it strongly suggests that he knew that she had changed her will. Not only does this give him a motive for the murder, but it also shows that he was lying when he claimed not to know that he stood to benefit from her death.[2]

This is what makes crime stories so fascinating. They cannot be solved simply by adding or subtracting beliefs; rather, one must negotiate the intricacies of how the different parts of the puzzle fit together. Further, the pressure to reach a decisive verdict—in a criminal case beyond reasonable doubt[3]—means that a simple leaning towards one side or the other is no good. One needs to mentally bolster the case for or against the suspect, so that it clearly dominates the alternatives—crowding out other possible construals of the

[2] In the legal context it is common to distinguish between direct and circumstantial evidence. The former is supposed to prove a fact without need of additional inference (e.g. a witness testifying that he saw the defendant kill the victim) while the latter must be supported by other inferences (e.g. a witness testifying that he saw the defendant leave the crime scene; DNA evidence). This terminology is potentially misleading, because direct evidence is still open to doubt (e.g. the witness might be mistaken or lying in their testimony). Nevertheless, the fact that most of the evidence in a case is circumstantial supports our contention that items of evidence are often interrelated. In the 'Witness for the prosecution' story all of the evidence is circumstantial.

[3] In England jurors 'must be sure that the defendant is guilty'.

case. This compels one to construct a story that is both one-sided[4] and comprehensive, and thus likely to fill in many gaps left unsupported by the evidence at hand. There is seldom the leisure to tinker away slowly, as in science, accumulating support for each step; instead, one must sketch a picture all in one go, and hope that it captures the essential truths of the case.

Despite the enormity of the task, untrained jurors are expected to reach verdicts that can have life-changing consequences. For the large part they achieve this (although mistakes are made!). How do they do this? How should they do it? These are the questions that this chapter will address.

Before we start, it is important to clarify the intended domain of the chapter. It is not about the reasoning of legal experts, such as judges, barristers, or investigators. It is about the reasoning of lay people when confronted with complex bodies of evidence. This might involve an individual juror on a criminal case, but it could also be a member of the general public following the unfolding of evidence through reading snippets in the media. Moreover, it is the individual juror that is the focus of attention here, not the group of jurors that participate in the jury deliberation process. This is not to say that what we discover about the psychology of the individual juror, or non-expert, does not have implications for the jury as a whole, or for the expert judge (they are human after all). But extrapolation to these more complex cases would require a lot more argument and evidence.

Networks of relations

Evidence is typically sorted as positive or negative with respect to a particular hypothesis. For example, evidence can either exonerate or incriminate a suspect. However, there are various different ways in which the evidence can exert its influence on a hypothesis, and these different routes are masked by the simple dichotomy of positive versus negative. To illustrate, consider the distinction between *affirmative* and *rebuttal* evidence (cf. Binder and Bergman, 1984). Affirmative evidence directly supports the case made by either prosecution or defence. For example, the maid's testimony that she heard Vole talking to Miss French shortly before the murder is affirmative evidence for Vole's guilt. Rebuttal evidence is less direct; it serves to undermine a claim made by the opposing side. For instance, the maid's testimony is rebutted by the evidence

[4] This is most applicable to the adversarial system in criminal trials in England and the USA, but less so to the inquisitorial systems typical in Europe.

that she disliked Vole, and sought to incriminate him. Affirmative and rebuttal evidence can be presented by both sides to the dispute, and thus crosscuts the distinction between incriminating and exonerating evidence. Indeed some evidence is both affirmative and rebuttal. For example, the wife's statement that Vole returned home much later than 9.30 p.m. rebuts his alibi, and also supports the claim that he was with Miss French shortly before the murder.

More generally, a crucial difference between affirmative and rebuttal evidence is that the latter is only relevant to a hypothesis (e.g. guilt or innocence) because it targets a piece of evidence presented by the opposing side.[5] Without an item of evidence to oppose, rebuttal evidence exerts no influence on the target hypothesis. Thus evidence that the maid disliked Vole is only relevant to his guilt given her testimony that he was with Miss French shortly before the time of her murder.

To capture the distinction between affirmative and rebuttal evidence, and various other structural subtleties, evidence and hypotheses need to be represented in a *network.* It seldom suffices to gather positive items on one side, negative items on the other, and compute a weighted sum (as Charles Darwin famously did when deciding whether or not to marry). Instead, account must be taken of how these items of evidence might interrelate.

Formal models of evidential reasoning

How can complex interrelations between evidence and hypotheses be represented? Before looking at how people do this in practice, it is instructive to consider how it can be done in principle. There have been substantial advances in normative models of evidential reasoning over the past decade, and a variety of network models have been developed, including Wigmore charts (Wigmore, 1913), cognitive maps (Axelrod, 1976), and Bayesian networks (Pearl, 1988). We will focus on Bayesian networks. They have well-established foundations in probability theory, and are currently applied in many practical contexts, including legal and forensic reasoning (Dawid and Evett, 1997; Dawid, Mortera and Vicard, 2007; Fenton and Neil, 2008; Hepler, Dawid and Leucari, 2007; Taroni, Aitken, Garbolino and Biedermann, 2006).

[5] Schum (2001) makes the distinction between 'evidence about events' and 'evidence about evidence'. Here affirmative evidence is a subset of the former, and rebuttal evidence is a subset of the latter.

Bayesian networks

Bayesian networks (BNs) consist of two parts: a graph structure and a set of conditional probability tables. The graph structure is made up of a set of nodes corresponding to the variables of interest, and a set of directed links between these variables corresponding to causal or evidential relations. In legal contexts the variables will include hypotheses about the details of the crime, the culprit, the suspect and numerous evidence statements. This yields a directed graph that compactly represents the probabilistic relations between variables, in particular the conditional and unconditional dependencies. Thus the graph can be used to read off which items of evidence are relevant to each other, or to particular hypotheses. For example, the graph in Figure 7.1 represents a small portion of the evidence in the *Witness for the prosecution* story outlined above.

This graph has four binary variables, each taking values of either true or false. The variable 'Vole Guilty' represents whether or not Vole murdered Miss French. The variable 'Vole Present' represents whether or not Vole was with Miss French at 9.30 p.m. on the night of the murder. The variable 'Maid Testimony' represents whether or not the maid testified that Vole was present at that time. The variable 'Maid dislikes Vole' represents whether or not the maid disliked Vole.

The link from 'Vole Present' to 'Vole Guilty' indicates that Vole's committing the murder that night depends on his being with her at 9.30 p.m.. Obviously this link is probabilistic—Vole might have been present at that time but not guilty of the subsequent murder. Moreover, Vole's presence at the crime scene is clearly not a sufficient cause of his murdering Miss French. Various additional causal factors, such as motive and intent are necessary for Vole to have murdered her. Some of these variables are represented in the fuller Bayesian Network in Figure 7.3. Nevertheless Vole's presence at the crime scene is evidence of opportunity, and thus raises to some degree the probability that

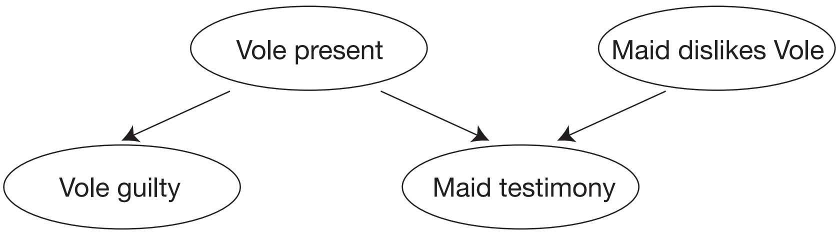

Figure 7.1. A simple Bayesian Network capturing a few variables in the *Witness for the Prosecution* story.

he did murder Miss French. This explains the link from 'Vole Present' to 'Vole Guilty'.

The link from 'Vole Present' to 'Maid Testimony' indicates that the Maid's claim—that she heard Vole speaking to Miss French around 9.30 p.m.—depends on whether or not Vole was actually there at that time. This link is also probabilistic. Perhaps the maid misidentified Vole's voice. Indeed the defence lawyer suggested that she might have heard voices from the radio. Another alternative cause of the maid's testimony is that she fabricated it because she disliked Vole. After all, in the trial she expressed a strong dislike for him. This possibility is explicitly represented in the network by the link from 'Maid dislikes Vole' to 'Maid Testimony'.

In addition to the graph, a Bayesian network also requires a conditional probability distribution table for each variable. This dictates the probability of the variable in question conditional on the possible values of its parents (the nodes with direct links into that variable). When a node has no parents, this table simply contains the base-rate value for that variable. This base-rate corresponds to the prior probability of the variable before any of the case evidence is taken into account. In some cases the exact values of these conditional probabilities are not too important, so long as they obey the qualitative relations encapsulated by the links in the graph. For example, the presence of a link from 'Vole present' to 'Maid Testimony' requires that the probability of the testimony given that Vole is present is greater than (or less than) the probability of the testimony given that Vole is not present. However, there are various aspects that are not given by the graph structure alone, but are furnished by the probabilities themselves: for example, whether the link is positive or negative; how the values of different parent nodes combine to dictate the value of the child node. In addition, most of the algorithms that allow a BN representation to be used for inference require exact numbers.

Representation and inference

A significant bonus of Bayesian networks is that once the representation is constructed, it can be used for inference. This sets it apart from most other forms of networks (e.g. Wigmore charts and Cognitive maps), which serve mainly as descriptive tools. Indeed representation is intertwined with inference in a BN. The arrangement of nodes and links, plus the conditional probability tables for each node, dictate what inferences are licensed (via the laws of probability). One way to draw novel inferences is to set a subset of variables to particular values (instantiate the variables), and then see what effect this has on the other variables of interest (e.g. the crime hypothesis). This corresponds to standard cases of inferential reasoning in legal cases. Moreover, it enables several kinds of inference: inference based on evidence that is known for sure (e.g. the Maid's testimony), evidence that is believed with some probability (e.g. that Vole was poor), and evidence that is presumed for sake of argument (e.g. if we suppose that Vole was present at 9.30 p.m., what else follows?). The latter can be very useful at the investigative stage of enquiry, when new pieces of evidence are sought. For example, a detective might suppose that Vole was indeed present at 9.30 p.m., and then infer the likely consequences of this, such as Vole being seen and heard by the maid, or leaving some trace evidence.

Patterns of inference

The network structure captures several patterns of inference critical to evidential reasoning.

Screening-off (conditional independence)

A basic feature of BNs is the screening-off relation. This holds when two variables that are probabilistically dependent are rendered independent by the knowledge of the state of a third variable. For example, the maid's testimony depends on Vole's guilt, but if we know for certain that Vole was with Miss French at 9.30 p.m., her testimony becomes independent of his guilt. In other words, if we already know that Vole was with Miss French at 9.30 p.m., the maid's testimony does not add anything to our knowledge of whether or not Vole is guilty. Of course such certainty is often hard to come by. Even CCTV footage confirming Vole's presence at 9.30 p.m. is open to doubt. Was it really Vole? Could the footage have been tampered with? Is the timing correct? Nevertheless there are many situations where a proposition is assumed or accepted as true (or false).

The BN representation rests upon the assumption that the parent nodes of a variable screen it off from all other variables in the network (except for variables that themselves depend on that variable).[6] This is a powerful condition: it

[6] It is important to be clear about the provisional nature of BN representations and the inferences they sanction. The 'screening-off' condition is just an assumption, and in some cases its applicability is debatable, see Cartwright (2007) and Williamson (2005). This reinforces the fact that any inference requires certain assumptions, and there is no guarantee that these assumptions hold true of the domain to be modelled. This problem strikes at the status of BNs as 'true' models of the world, but it need not undermine their status as 'useful' models. Any practical model is

can greatly simplify the computations needed to draw inferences, allowing various items of evidence to be ignored all together. For example, if it is established that Vole was with Miss French at 9.30 p.m., then no other witness testimony about this event can influence the probability that Vole is guilty. This screening-off assumption can also aid future investigation and information gathering. Thus, if an event E is established for certain, there is no additional inferential benefit to be gained from further witness testimonies that attest to E. Of course in many cases the truth of a key event will remain in dispute; hence the adding of extra witnesses to the same event will be a reasonable policy.

In legal investigations, as in everyday life, it is crucial to distinguish between what people claim and what actually happened. The network structure is ideally suited to this, and readily distinguishes witness reports from the events or situations that these reports are about. Thus the maid's testimony that Vole was with Miss French at 9.30 p.m. is represented separately from the event that he was in fact with her at that time. One advantage in representing the report E* and the reported events E separately is that the probative force of the events (if true) is kept distinct from the credibility of the witness source.[7] This is important because the factors relevant to the probative force of E on the target hypothesis H are quite different from those relevant to the relation between the report E* and E. For example, there are various reasons why Vole's presence at 9.30 p.m. does not guarantee that he killed Miss French; perhaps someone else was there too, or broke in shortly afterwards. But a different set of factors potentially undermine the reliability of a witness report, and thus the inference from E* to E. This is a place where rebuttal evidence can exert its force. Perhaps the maid misidentified the voice, or heard the radio, or simply lied. The distinction between E* and E also greatly facilitates inference in situations where there are several witness testimonies to the same event, and clarifies the differences between corroborating, conflicting, or contradictory testimony.[8]

---

bound to involve simplifications and assumptions. The real test is how well they serve the inferential goals of the user. And human users, with their bounded computational abilities, might be well-served by principled simplifications such as the screening-off assumption. Indeed we will argue later that the BN framework needs to be simplified further, if it is to provide a reasonable tool for human cognition. It is also important to distinguish questions as to whether a BN is appropriate for a specific case (e.g. is there a link from H to A? Does E screen-off H from A?), and the justifiability of the 'screening-off' assumption in general.

[7] This distinction is explicitly introduced in Schum (2001), and his notation is used here.

[8] See Schum (2001) for extensive discussion of these issues.

Explaining away

The screening-off relation holds when three variables are in a chain (A→B→C)[9] or a divergent structure (A←B→C). In both cases, A and C are dependent, but become independent conditional on B. The converse situation occurs with a convergent structure (A→B←C). In this case, A and C are independent, but become dependent conditional on B (or conditional on a variable that itself depends on B). This encapsulates a distinctive pattern of inference termed 'explaining away'.[10]

Explaining away typically occurs in situations where there are multiple independent hypotheses (explanations) for an observed item of evidence. The observed evidence leads to some increase in probability for all these hypotheses; however, if one of these is found to be true, the others then decrease in probability. The evidence that previously supported them is explained away. A notable feature of this situation is that the possible hypotheses are independent when the status of the evidence is unknown, but become conditionally dependent given knowledge of its status. This is distinct from cases where hypotheses are mutually exclusive.

To illustrate, consider again the network in Figure 7.1. There is no direct link between 'Vole present' and 'Maid dislikes Vole'. This indicates that these two variables are unconditionally independent. Whether or not Vole was present at that time is irrelevant to whether or not the maid disliked Vole, and vice versa. However, once the maid gives her testimony, these two variables become conditionally dependent—they compete as alternative explanations of her testimony. Suppose further that we are sure that the maid does have a strong dislike for Vole, and wishes to incriminate him. This hypothesis will 'explain away' the maid's testimony.

Consider another example from the *Witness for the Prosecution*. Vole's cuffs were found to have traces of blood on them (type O). This was advanced as evidence that Vole had murdered Miss French (who had blood type O).[11]

[9] An example of an A→B→C chain, drawn from our crime story, is as follows: Vole's being a legatee in Miss French's will (A) is a potential cause of him murdering Miss French (B) (it is definitely a motive!), and his murder of Miss French is a potential cause of the blood on his cuffs (C). This chain is embedded in Figure 7.3.

[10] See Pearl (1988).

[11] Christie's story comes from the days before DNA evidence. Nowadays DNA testing would be the standard method to determine whether the bloodstain was from Vole or Miss French. DNA tests have much greater power to discriminate, but still yield conclusions with a degree of uncertainty. In such forensic contexts, Bayesian networks provide the normative standard for quantifying the impact of DNA evidence (Taroni *et al*., 2006).

However, the defence sought to explain away this evidence by claiming that the blood belonged to Vole (also type O), who had cut himself when slicing ham. This claim in turn was backed-up by a recent scar on Vole's wrist. Clearly there are two competing explanations for the presence of blood on Vole's cuffs. Moreover, these explanations were independent prior to the discovery of blood on the cuffs.

Legal scenarios, and evidential contexts in general, are replete with 'explaining away' inferences, and it is a substantial advantage of the BN framework that it models this inference so naturally. This also explicates the distinction between affirmative and rebuttal evidence mentioned earlier. Rebuttal evidence serves to explain away an opposing piece of affirmative evidence. For instance, the blood on Vole's cuffs is affirmative evidence that he murdered Miss French, whereas the claim that he cut himself slicing ham (and the scar on his wrist) is rebuttal evidence. Whether or not Vole cut himself is only relevant to the question of his guilt because it rebuts (explains away) the evidence provided by the blood on his cuffs. BNs provide a natural format for representing this kind of evidential subtlety.

Alibi testimony and failure of screening-off

In the simplest case of witness testimony, the event testified to (E) will screen-off the witness report E* from the target hypothesis H (see Fig. 7.2, model ~~2~~A). This screening-off relation can hold even if there are serious reasons to

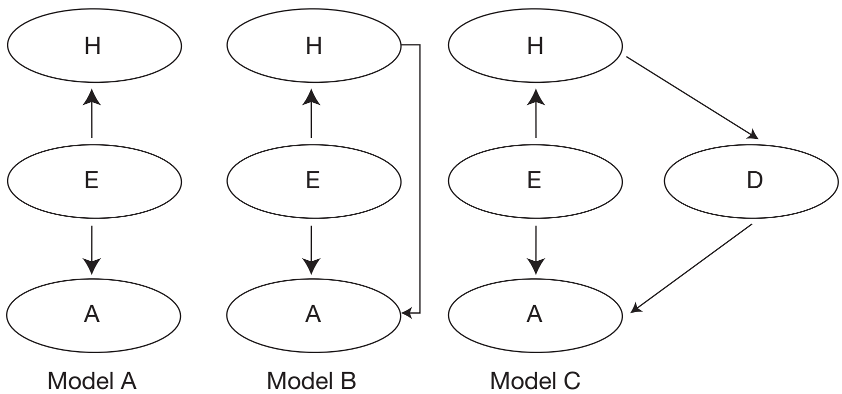

Figure 7.2. Witness vs Alibi models. Model ~~2~~A is an impartial witness testimony in which E screens-off H from E*. Model ~~2~~B is a partial alibi testimony where E does not screen-off H from A, and Model ~~2~~C represents the same situation as model ~~2~~B but with the deception variable D explicitly represented.

doubt the reliability of the maid's testimony, or the probativeness of the event E. However, this will not always be the case. Sometimes the witness report E* will exert an independent influence on H. This is most clearly illustrated by considering alibi testimony.

An alibi involves the claim that the defendant was somewhere else at the time the crime was committed. Assuming that nobody can be in two places at the same time[12] an alibi is potentially very strong evidence in favour of the defendant. This is because the probability that the defendant committed the crime, on the supposition that he *was not* at the crime scene, is very low, and much lower than the probability that he committed the crime, given that he *was* at the crime scene. However, alibi evidence is often considered weak evidence, especially if it is only the defendant's word, and there is no corroborating evidence (Gooderson, 1977).

The alibi context is depicted by model B in Figure 7.2. The variable H corresponds to the hypothesis that the defendant is guilty; variable E to the defendant's presence at the crime scene; variable A to the defendant's claim that he was somewhere else. The link from E to H indicates that his committing the crime depends on his presence at the crime-scene; the link from E to A indicates that his alibi statement depends on whether or not he was at the crime scene. In the case of alibi testimony the inference from H to E is usually taken to be much stronger than the inference from A to E.

There is also a direct link from H to A. This represents our claim that the event E does not screen-off H from A. Why is this the case? Recall that the screening-off relation states that once you know the value of the intermediate variable E, knowledge of A tells you nothing more about H (and vice versa). But consider the situation in which the defendant gives his alibi, but you have independent evidence (e.g. CCTV footage) that he was in fact at the crime scene. Does the fact that he said he was not at the crime scene tell you anything more about whether or not he is guilty? Well now you know that he lied.[13] And this information seems incriminating, over-and-above the fact that you know that he was at the crime scene. Of course he might be lying for other reasons: perhaps he was having an affair; or committing a different crime. But it seems reasonable to assume that the probability that he will lie in his alibi is greater when he is guilty than when he is innocent.

[12] And ignoring situations where something or someone is remotely controlled—e.g. detonating a bomb; hiring an assassin.

[13] It is possible that he was mistaken, for example if he suffers from severe memory loss; but in most cases this will be unlikely.

On this reading of the situation, the event E no longer screens-off H from A. The probability that the defendant is guilty (H) given that he was at the crime scene (E) is lower than the probability that he is guilty (H) given that he was at the crime scene (E) *and* said he was not at the crime scene (A). In numbers, $P(H|E\&A) > P(H|E)$. For screening-off to hold, these two probabilities would have to be equal.[14]

In other words, finding out that the defendant has lied in his alibi tells you more about his guilt than simply knowing that he was at the crime scene. The possibility that the alibi provider is lying can itself be represented in the graph (see Figure 7.2 model C) with an additional node D representing whether or not the defendant is motivated to lie. The link from H to D corresponds to the assumption that the motivation to lie depends on whether or not the defendant is guilty (i.e. he is more likely to lie when guilty than when innocent). The link from D to A indicates that whether the defendant says he was present at the crime scene depends on whether he is motivated to lie (i.e. he is more likely to say he was not at the crime scene if he is motivated to lie).

This 'alibi network' is readily applied to Vole's alibi. Recall that Vole claimed that he returned to his home at 9.30 p.m., and therefore was not with Miss French at that time. However, if Vole is indeed guilty he would have strong motivation to lie about this. This furnishes an alternative explanation for his alibi. Therefore it is not too surprising that a defendant's alibi is often treated with scorn. The alternative explanation in terms of deception explains away the alibi testimony.

The situation changes if someone else corroborates the alibi. An impartial witness, with nothing to gain, might bolster the alibi considerably (although not always as much as expected[15]). A partial witness, such as a friend or relative, is less convincing. After all, they too have a motive to lie. Romaine knew this, and realised that the jury would not be overly impressed by a supporting alibi from Vole's beloved wife.

An intriguing consequence of the alibi network is that the failure of screening-off only seems to apply when the alibi-provider *knows* whether or not the defendant committed the crime. This is because the link from H to D is only present if the guilt of the defendant influences the alibi-provider's motivation to lie. But if the alibi-provider does not know whether or not the defendant is guilty, there is no such link. This is not to say that the alibi-provider is not motivated to lie in his favour, but just that this motivation is not dependent on the defendant actually being guilty. For example, one might

[14] For details see Hepler, Lagnado and Baio (forthcoming).
[15] See empirical study by Culhane and Hosch (2004).

expect a wife to lie for her beloved husband even if she does not know whether he is guilty or innocent. But in this case the event E will screen-off H from A. If we know that the defendant was at the crime scene (E), finding out that the alibi-provider lied to protect the defendant does not add anything extra to our assessment of guilt. Their motivation to lie was not affected by whether or not the defendant was indeed guilty.

This twist demonstrates the considerable subtleties that can arise in reasoning about witness or alibi evidence, and reiterates the need for a network representation. In a later section we will present an empirical study showing that people are sensitive to this subtlety in their inferences about alibi evidence.

Thus far it appears that BNs provide a promising framework to represent the complex interrelations between bodies of evidence and hypotheses. In particular, BNs capture important patterns of reasoning such as screening-off and explaining away, and elucidate some of the subtleties involved in alibi and witness testimony.[16] Moreover, it is the graph structure rather than the exact conditional probabilities that play the key role. The network representations and the qualitative 'relevance' relations between variables do much of the inferential work.

Holistic vs atomistic approaches

Bayesian approaches to legal reasoning are often criticised for their 'atomistic' evaluation of evidence (e.g. Pardo and Allen, 2008). In contrast, proponents of a holistic approach argue that evidence should be assessed as a whole—not piecemeal. This claim also draws on psychological research that jurors compose coherent stories to make sense of the evidence presented at trial (Pennington and Hastie, 1986, 1992). However, this criticism is based on a restricted notion of the Bayesian approach that does not take into account the holistic relations implicit in Bayesian networks. The network model defended in this chapter maintains that people organise evidence and hypotheses in coherent networks, and that this is often the right thing to do from a rational viewpoint. Thus, an item of evidence can only be evaluated with respect to its relation to other items of evidence and relevant hypotheses. For example, the scar on Vole's wrist is only relevant to the hypothesis of his guilt given his claim that he cut himself slicing ham, and the evidence of blood on

[16] The Bayesian network framework also has the potential to formalise and explicate ~~and formalise~~ other patterns of legal inference, for example the distinction between cable and chain inferences (see Hamer, 1997).

his cuffs. This is not to say that smaller subsets of evidence cannot be analysed in isolation from other subsets. Presumably the complexities of the blood evidence are largely independent of the issues that surround the maid's testimony about Vole's presence on the night of the crime—these subsets are only linked via the superordinate proposition that Vole committed the crime. Indeed it is the possibility of isolating small subsets of evidence that makes a network approach tractable to the human mind.

A related issue is whether people's ultimate focus is on the probability of the issue in question, for example, the probability that the defendant is guilty, or on something more holistic, such as the probability (or plausibility) of the prosecution's account as a whole (as compared to the defendant's account). The Bayesian account is usually portrayed as concentrating on the former—the probability of the crime hypothesis given all the evidence. And a standard objection is that in contrast fact-finders are concerned with holistic judgements, such as how believable the prosecution story is as compared to the defence story. However, the latter kind of judgement is readily accommodated within the broader Bayesian framework. Thus, Pearl discusses inference mechanisms that revise the probability of composite sets of beliefs rather than updating individual probabilities (Pearl, 1988). In short, the Bayesian network framework is not restricted to probability judgements about singular propositions, but can extend to judgements about sets of propositions. It supplies the tools for assessing the probability of a connected set of beliefs, as well as individual beliefs.

Bayesian networks as models of human reasoning?

Bayesian networks have considerable appeal as normative models of evidential reasoning, especially in domains with quantitative data (Aitken, Taroni and Garbolino, 2003; Dawid, Mortera and Vicard, 2007; Hepler and Weir, 2008). However, fully fledged BNs are unlikely to provide a comprehensive model for human reasoning. The specification of precise conditional probabilities, and the complex computations required to draw inferences, seem beyond the capabilities of human reasoners, especially when large numbers of variables are involved. For example, even a partial BN for the *Witness for the Prosecution*, including only a subset of the available evidence, would present an intimidating picture for the uninitiated juror (see Fig. 7.3). Indeed many psychological studies suggest that people are poor at estimating and calculating with probabilities (Gilovich, Griffin and Kahneman, 2002; Kahneman, Slovic and Tversky, 1982).

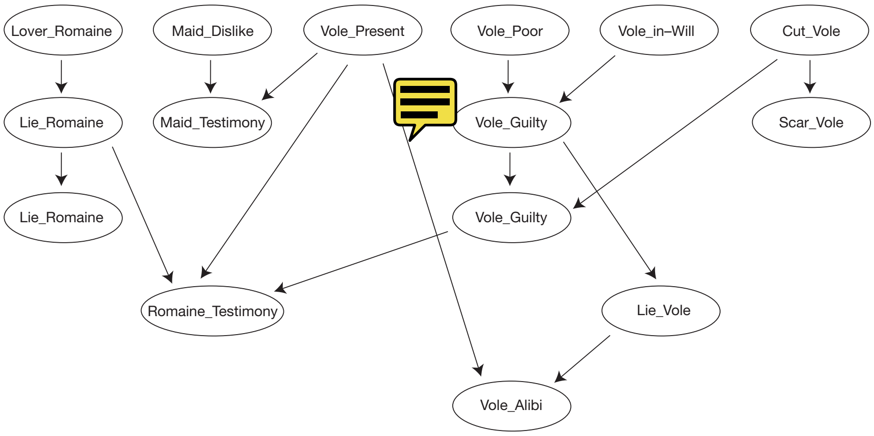

Figure 7.3. Partial Bayesian Network covering some of the major pieces of evidence in *Witness for the Prosecution*. Note that not all the items of evidence have been included in this network.

It thus appears that BNs are a non-starter as a descriptive model. However, this conclusion is premature. It overlooks the fact that key aspects of the BN formalism are qualitative rather than quantitative, and that BNs can be hierarchically structured to overcome human processing limitations.

Qualitative network models

The network structure of a BN is purely qualitative, representing the presence or absence of dependencies between variables. Thus a link from A to B tells us that certain values of A will change the probability of certain values of B, without needing to specify exactly how much. And although the standard BN framework requires a precise set of conditional probabilities, many of the important characteristics of the network are retained without a full and exact set of probabilities (Biedermann and Taroni, 2006; Wellman and Henrion, 1993).

This means that someone can construct the graphical part of a BN without having access to any precise probabilities. Moreover, in many cases they will still be able to draw inferences based solely on this qualitative structure, albeit less precise ones than with a quantitative network. Indeed most of our examples and discussions so far have relied on this qualitative sense of probabilistic relations.

Therefore the fact that people cannot estimate or calculate with exact probabilities does not undermine the possibility that they use networks to represent relations between evidence and hypotheses. Furthermore, such qualitative representations can apply equally well in domains where no precise figures are available. This is particularly significant in the legal context, where much of the evidence does not admit of quantification. For example, it might not be possible to quantify the exact probative force of a witness testimony that places the defendant at the crime scene; but most people would agree that it raises the probability of guilt, however slightly. Moreover, people will often be able to make comparative probability judgements; for instance, judging that a certain piece of forensic evidence (e.g. traces of the victim's blood found on the defendant's coat) raises the probability of guilt more than the testimony of a partial witness.

The important point is that even if people lack a fully specified probabilistic model, they can still express their uncertain knowledge in terms of a qualitative network. In a similar vein, even if people are unable to perform exact Bayesian computations over this network, they can still draw approximate inferences (perhaps using heuristic methods). And this might not be such a bad thing, especially in contexts where exact figures are unavailable, or where there are large numbers of interacting variables, so that exact inference is intractable. The restriction to qualitative networks might actually increase the applicability and feasibility of these networks to the problems faced in legal contexts.

The idea that people utilise qualitative networks is not new. And in computer science there are various qualitative reasoning systems designed to mesh with the natural propensities of human users (Druzdzel, 1996; Keppens, 2007; Parsons, 2001). These provide an impressive array of representational formats and inference algorithms. There are also several considerations from psychology that speak in favour of qualitative approaches.

First, psychophysical studies show that for a range of sensory phenomena people are poor at making absolute judgements, and instead make ordinal comparisons (e.g. Stewart, Brown and Chater, 2005). Estimates of likelihood or strength of evidence seem no exception to this. Thus, although someone might not be able to judge the precise probative force of evidence E1 on hypothesis H, they might be able to judge that E1 makes H more probable, and perhaps that E1 is stronger evidence for H than E2 is.

Second, analyses of a wide range of predictive tasks (e.g. clinical and medical diagnosis) suggest that statistical models that use unit weights often outperform more complex models (Dawes, 1979). The key requirement for these simpler models is that the sign of each variable in the model is correct; the exact weights placed on these variables is not critical. While the generalisability of these findings is open to question, the success of unit weight models in these specific environments remains impressive. One lacuna is that the models typically assume a simplistic structure, where the relevant predictor variables are independent. This assumption might work well in certain environments, but is unlikely to succeed in more complex contexts, such as the legal domain, where the interrelations between evidence and hypotheses are crucial. The proposed qualitative networks thus go beyond simple linear models, but share the intuition that precision in the weights is not a necessary condition for successful inference (and in fact might impair performance, by adding a spurious degree of precision).[17]

Third, psychological research on causal reasoning (Griffiths and Tenenbaum, 2005; Lagnado, Waldmann, Hagmayer and Sloman, 2007; Sloman, 2005) suggests that people are initially concerned with judgements about causal structure (is there a link between A and B?) rather than causal strength (how strong is the link between A and B?). People focus on qualitative causal relations, and use these causal models to guide their inference and actions. There is also growing empirical evidence that people reason in accordance with the qualitative prescripts of causal BNs (Krynski and Tenenbaum, 2007; Sloman and Lagnado, 2005). Clearly, causal beliefs are critical in the construction of our models of the world. If these are predominantly qualitative, then so too will be the network structures that they underpin. Indeed it has been argued that people's network representations of the world are causal through-and-through. Pearl (2000) advances an argument for the primacy of qualitative causal beliefs that also serves as a strong empirical hypothesis about human cognition. He argues that the best way for people to organise their knowledge of the world is in terms of invariant (stable) qualitative causal relations. These relations, once known or assumed, will not change according to the particularities of the information we have, whereas probabilistic relations can. For example, consider a chain A→B→C. There is a probabilistic dependency between A and C, but this disappears when we know B (screening-off). Conversely, consider a common effect model A→B←C. On this model A and C are probabilistically independent, but become dependent conditional on our knowing B (explaining away). What remains constant across all these changes in the probabilistic relations are the underlying causal relations in the models. Of course we might have the structure wrong, but this is a separate issue. In essence, the argument is that people should prefer to organise their knowledge on the basis of invariant rather than unstable aspects of the world. This is not to say that causal relations do not change. They often will, especially

[17] Note that in the judge's directions to the jury on how they should reach their verdict, the terminology used is 'the law does not attempt to provide a scale of answers to juries'.

when we interact with the world. But unlike with probabilistic relations, these changes will reflect changes in the world rather than changes in our knowledge about it. This is a powerful conjecture about human cognition, and psychologists are exploring its implications. However, it is not an essential part of the current argument, which emphasises the *qualitative* nature of our 'mental networks'.

The proposal that the fundamental building blocks of human reasoning are qualitative fits with Gilbert Harman's claim that people think and reason primarily in terms of beliefs (all-or-none) rather than degrees of belief (Harman, 1986). To speculate, one reason for the central role of categorical beliefs might derive from the close relation between thought and action. Events in the world, including actions and outcomes, are typically all-or-none; so our representations of these events are likely to follow suit. For instance, the suspect was either at the crime scene or not—he cannot have been 67 per cent present. Consequently, when reasoning about this possibility we either suppose that he was present, or suppose that he was not. It seems less plausible that we can suppose and reason with some mixed state in which he is partially there, and partially elsewhere. This is not to say that more complex situations cannot admit of gradations, or that we cannot assign probabilities or degrees of belief. The claim is that our cognitive systems have primarily developed to reason with situations involving categorical events, actions and outcomes.

Small-scale networks

Not only do people reason qualitatively, they also seem to reason with just a few variables at a time. This is not to argue that they cannot build up large-scale knowledge sets containing many variables; but the active reasoning process is likely to involve only a few variables at a time. This seems to be a natural consequence of our limited working memory capacity (Cowan, 2001; Halford, Cowan and Andrews, 2007; Miller, 1956). This restriction suggests that active evidential reasoning takes place with network fragments (e.g. the model in Fig. 7.1) rather than the full-scale networks (e.g. the larger model in Fig. 7.3). These small-scale networks are sufficient to carry out key inference patterns such as explaining-away or screening-off, but may require that some variables be 'chunked' together or ignored entirely. Reasoning with network fragments can lead to modifications in the implicitly stored knowledge base, so that in the long term people effectively represent larger structures. In other words, people construct and reason with network fragments, and these are stitched together to yield a large-scale picture of the world. On this view, working memory acts as a bottleneck between the world and our large-scale representation of this world.

Hierarchical representations

The notion of chunking is widely recognised to be a key feature of human memory, but it has not been connected with inference and reasoning in a legal context.[18] In the context of evidential reasoning, it seems that people use richly structured representations that can be unpacked at various levels of grain. In this way people can negotiate a complex problem domain with multiple interacting variables, while respecting the limitations of working memory and active reasoning.[19] For example, in many crime cases the initial level of representation is in terms of individual people (e.g. victim, perpetrator, suspect, accomplices, witnesses), objects (e.g. weapons, blood traces, fibres), and the spatiotemporal relations and interactions between them. At a finer level of grain, each individual possesses various attributes (age, race, gender, personality traits, dispositions etc.), along with beliefs, desires and intentions that serve to explain and predict their behaviour. When reasoning about a crime case the fact-finder can switch between these levels of abstraction, at one moment reasoning about the interactions between several individuals (e.g. the locations of Vole, Miss French and the maid on the night of the crime), at another moment reasoning about the motives, beliefs and intentions of a specific individual (e.g. the maid). The key point is that by using rich hierarchically structured representations human reasoners can overcome the limitations imposed by their limited-capacity working memory.

How does this ability to chunk information fit with the idea that human reasoners use Bayesian networks? At first sight it would seem that the hierarchically structured representations used in human reasoning are far-removed from Bayesian networks, which represent all variables at the same level. However, recent work in computer science (Koller and Pfeffer, 1997; Hepler *et al.*, 2007) has developed probabilistic systems that introduce hierarchical structures ('objects') to deal with more complex real-world domains. These systems are still based upon Bayesian networks, and exploit the notion of conditional independence, but allow for richer representations and more efficient inference procedures. This highlights an intriguing parallel between

[18] See work on chunking in expert chess players, Chase and Simon (1973), and in medical diagnosis, Schmidt and Boshuizen (1993).

[19] For a related argument about the use of stereotypes in decision making in legal contexts, see Davies and Patel (2005).

the human reasoning system and artificial systems developed in computer science. In order to cope with the computational demands of large and complex domains, both systems make use of hierarchical structured representations, and simplify inference by using network structures that exploit conditional independence relations. Thus future research into human evidential reasoning can profitably draw on advances in probabilistic graphical frameworks.

Our guiding hypothesis is that people reason about legal evidence using small-scale qualitative networks. These networks are generated on the basis of background assumptions, generic causal knowledge, and case-specific information, and utilise hierarchically structured representations to overcome computational limitations and support efficient inference. Such networks often require only comparative judgements of relevance and probability, rather than precise probability estimates. In addition, it is likely that people flexibly adopt the format that best suits the data available to them. For example, combining quantitative evidence such as that provided by a DNA match with qualitative evidence such as that furnished by a witness testimony. Reasoning and inference might also be conducted in an approximate fashion, rather than through full-scale Bayesian computation. There is a range of possible inference mechanisms, including sign propagation, belief propagation, spreading activation and constraint satisfaction. Here again the choice of mechanism might depend on the available data and the format of the network. For example, if all network links have signs (positive or negative),[20] but no strength value, then sign propagation is the most appropriate inference mechanism. If the strength of links can be ranked or quantified, then other algorithms might be more suitable.

Another important feature to note is that the qualitative network models that people construct are subjective. These models depend on an individual's background knowledge and assumptions, the evidence available to them, their interpretation of this evidence, and myriad other factors. And of course these can differ substantially from individual to individual. Presumably the prosecution's model will be very different from that of the defence. However, this does not mean that they are unconstrained, or that reasonable people cannot end up agreeing on a shared model. The requirements of consistency and coherence (and fit with real-world causal knowledge), and the need to accommodate the undisputed elements of the case, will place constraints on the viable models and inferences that can be legitimately drawn from the available

[20] The standard way of determining the sign of a link is via the following comparative probabilistic relations: $A \rightarrow B$ is +ve if $P(B|A) > P(B|{\sim}A)$; –ve if $P(B|A) < P(B|{\sim}A)$; 0 if $P(B|A) = P(B|{\sim}A)$. For details see Wellman and Henrion (1993).

evidence. The trial is a social structure that will ideally converge on a model that is an appropriate reflection of what actually happened in the case.

Sources of reasoning errors

The central claim is that qualitative networks lie at the heart of people's evidential reasoning. This is not to say that people will conform perfectly to the dictates of any specific qualitative reasoning system. Rather, the claim is that the fundamental vehicle for representation and inference is qualitative, and often derived from causal understanding. Indeed, there will be many ways in which evidential reasoning can fall short of normative theory. With regard to representation, people can err due to inadequate models, failure to include crucial variables or links, inappropriate collapsing or grouping of variables. With regard to computation, people can err because of short-cut heuristic procedures that can lead to suboptimal inference. In both cases errors will typically arise from capacity and processing limitations, and the reasoner's attempts to overcome these by simplifying representation and inference. As with many cognitive biases, these errors are a necessary price to pay in order to maintain and use workable models of the world.

Dynamic networks

A key feature of human cognition is that it adapts to a changing environment. These changes might involve novel evidence, hypotheses and goals. Thus network construction in the face of evidence is dynamic. People adapt their network online as information is received and hypotheses are developed or thought up. This is not a strictly Bayesian process (in which a complete set of hypotheses are continually updated). Rather, people seem to introduce and eliminate hypotheses in a more all-or-nothing manner. This saves greatly on storage and computation, but can lead to biases and errors. For example, an early piece of evidence might be ignored because the right hypothesis had not yet been entertained. And the interpretation of ambiguous evidence will depend on what hypotheses are entertained at that time. This can lead to substantial order effects—the final evaluation of a body of evidence being heavily dependent on the order in which that evidence is processed (Hogarth and Einhorn, 1992).

A parallel can be drawn with action and practical reasoning. At an early stage one might not have the requisite knowledge to take advantage of an opportunity; later on one acquires the knowledge, but the opportunity has passed. This would not happen in the ideal world of flawless memory and unrestricted reasoning abilities, but will be commonplace in the bounded and imperfect world that we inhabit (especially as untrained jurors). For instance, consider a case in which the defendant is charged with assault, and pleads self-defence. According to the law it is crucial to establish whether the defendant used force that was 'reasonable' in the circumstances. But jurors are often not given a definition of self-defence until the end of trial. This can be problematic, because critical evidence about the suspect acting in self-defence might have been presented before the juror has an appropriate understanding of the legal notion.

Despite these potential shortcomings, the online generation and adaptation of a small range of hypotheses will often be a good adaptive solution to the problems that we face. People do not use full-scale (static) BN representations—they are more likely to adopt small-scale networks that undergo discrete changes as new evidence is encountered and hypotheses are refined, abandoned or augmented. The small-scale nature of these networks can facilitate rapid and flexible adjustment. Moreover, the qualitative nature of the networks spares the numerous probability estimations and computations that a full BN would require.

Current models of evidential reasoning

We have speculated about the form that a human model of evidential reasoning might take. Thus far only general principles have been articulated; desiderata that we would expect any plausible theory of human reasoning to satisfy, without specifying the cognitive processes that implement these principles. How does this network model fit with other psychological theories of evidential reasoning? There are three dominant models, each with pros and cons.

Belief-adjustment model

The belief-adjustment model is a general purpose model for how people update their beliefs in the face of new evidence (Hogarth and Einhorn, 1992). Applied to the legal context, the model assumes that people start with an initial degree of belief in the guilt of a suspect, based on background information. This prior information can include both specific details about the case (e.g. nature of the charge; gender and race of the defendant etc.) and general assumptions or knowledge (e.g. the presumption of innocence).[21]

[21] See Hastie (1993) for more details.

When a new item of evidence is received, it is encoded as positive or negative in relation to the guilt hypothesis, weighted according to its judged strength, and then additively combined with the prior belief. This produces an adjusted degree of belief in the suspect's guilt, which then serves as the new prior when the next piece of evidence is received. This process is iterated until all items of evidence are integrated, and a final degree of belief is reached.

The belief-adjustment model has some attractive features. It specifies a simple processing model that avoids heavy computational or storage demands. At any one stage, the decision maker only needs to consider their prior opinion and the impact of the new item of evidence. It has been applied to a wide variety of cognitive tasks, and can account for a rich pattern of empirical results. In particular, it can explain both primacy effects—where people over-weight items of evidence that are presented early in a sequence, and recency effects—where people over-weight items of evidence that are presented later in the sequence. In the case of legal judgments, where the evidence is encoded as positive or negative relative to a target hypothesis, it predicts that later items of evidence are over-weighted relative to earlier items (Kerstholt and Jackson, 1999).

The Achilles heel of the belief-adjustment model, when applied to legal contexts, is that it treats all pieces of evidence as independent. It ignores the interrelations between items that make legal cases so compelling. This is a substantial shortcoming, even in relatively simple situations. For example, it cannot capture instances of explaining away, and thus cannot distinguish rebuttal from affirmative evidence. Consider again the *Witness for the Prosecution* story. The police had evidence that there was blood on Vole's cuffs, and this matched the victim's blood. The defence rebutted this piece of evidence by showing that Vole himself had the same blood type as the victim, and claiming that he had cut himself slicing ham. This rebuttal was in turn supported by the fact that Vole had a fresh scar on his wrist. According to the belief-adjustment model, the blood on the cuffs constitutes positive evidence of guilt, and the claim that Vole cut his hand, and the scar on his wrist, constitute negative evidence. But the fact that this negative evidence impacts on the guilt hypothesis only by undermining the positive evidence cannot be represented.

The inability to represent interrelations between items of evidence can rapidly lead to counter-intuitive consequences. For example, suppose that further forensic tests reveal that the blood on Vole's cuffs does not match the victim's. This finding rules out this piece of incriminating evidence against Vole, but it also renders the claim that he cut his hand slicing ham, and the scar on his wrist, irrelevant to whether or not he is guilty. These items of rebuttal evidence are no longer needed (the blood does not match); they no longer count as positive evidence. The belief-adjustment model cannot capture these changes. It has no mechanism, or representational structure, that allows it to re-evaluate earlier items of evidence.

It might be argued that while the neglect of interrelations undermines the model as a normative account, this feature might fit with actual human reasoning. However, there is a wealth of empirical data (see below), plus every crime writer's intuitions, suggesting that people can accommodate these patterns.

Story model

In direct contrast to the belief-adjustment model, the story model (Pennington and Hastie, 1986, 1992) places strong emphasis on the interrelations between items of evidence. This model maintains that people construct stories in order to make sense of the evidence presented in court, and these narratives are key determinants of the final verdicts reached. According to Pennington and Hastie these stories involve complex networks of causal relations between both physical events and psychological states (e.g. intentions, desires and motives). These causal networks draw on evidence presented in the case, as well as prior assumptions and common-sense knowledge. They are used to construct a plausible narrative for the unfolding of the crime.

The story model has garnered broad empirical support, but remains vaguely specified with respect to the underlying cognitive processes and mechanisms. For example, no precise account is given for how people construct or update their causal models, or how they draw inferences from them. Moreover, Pennington and Hastie argue that the story model only applies to global judgements (those made once all the evidence is processed). In the case of online judgements, people are supposed to revert to a simpler belief-adjustment model.

The qualitative network perspective argued for in this paper shares many of the insights of the story model. In particular, that evidential reasoning involves the construction of causal networks that organise and make sense of the available evidence. In contrast to the story model, however, our perspective emphasises the role of probabilistic links between evidence and hypotheses. The story model explicitly rules out any role for probabilistic evaluations.[22]

[22] Note that Pennington and Hastie's main argument against the Bayesian approach rests on participants' failure to conform to Bayes' rule for exact computations (people tend to be more conservative). However, this does not count against people's use of qualitative notions of probability.

Our account uses a formal Bayesian framework to represent these relations, and capture important patterns of inference. The story model is silent on these details. In addition, our account applies equally to online judgements, as the evidence comes in, and to global judgements, once all the evidence is presented.

Coherence-based models

Coherence-based models (Simon, Snow and Read, 2004; Simon and Holyoak, 2002; Thagard, 2000) also accentuate the complex interrelations between items of evidence. These models derive from earlier psychological theories of cognitive consistency (Heider, 1946), and are based on the idea that the mind strives for coherent representations of the world. Evidence and hypotheses are represented as nodes in a large-scale connectionist network. The relations between nodes are represented as bidirectional links that are either excitatory (where the activation of one node increases the activation of the linked node) or inhibitory (where the activation of one node decreases the activation of the linked node). Some of these nodes are instantiated as true statements (e.g. observational reports), while others are set at random or prespecified values.

Judgments (e.g. is the defendant guilty on the basis of the given evidence?) are supposed to emerge from an interactive process that maximises the overall coherence of the network via constraint satisfaction (Thagard, 2000). This process can lead to substantial re-evaluation of hypotheses and evidence. In particular, even the evaluation of assumed facts can shift to achieve greater coherence with the emerging verdict (Simon *et al.*, 2004).

Coherence-based models, like BNs, represent multiple interrelations between items in a probabilistic fashion, and representation is closely tied to inference. The nature of the representation is somewhat different. Whereas BNs aim to represent events and processes in the world, coherence-based models represent the flow of inference in the mind itself. Another important difference is that BNs use directed links (often corresponding to causal direction) whereas coherence-based models have bidirectional links. This is a critical difference, and means that the latter models are unable to represent basic forms of inference such as 'explaining-away'. Returning to the example of the blood on Vole's cuffs. There are two main explanations for this piece of evidence: the blood came from the victim, and thus raises the probability of guilt, or it came from Vole himself, and thus lowers the probability of guilt. However, to capture this inference, a coherence-based model must assume that the two explanations are exclusive (or at least negatively correlated). But this is an inappropriate representation of their true relation in the world.

Whether or not Vole cut himself slicing ham is unrelated to (independent of) whether or not he is guilty of murder. The two explanations only become dependent given the evidence of blood on the cuffs that they both seek to explain. As noted earlier, this pattern of inference is naturally captured in a BN representation. An important question for future empirical research is the extent to which human reasoners engage in explaining away inferences, without assuming that alternative explanations are always exclusive or negatively correlated.

Recent studies on the role of network models in evidential reasoning

Most of the empirical studies conducted so far provide support for either the story or coherence models. However, the key findings, that people are sensitive to both causality and coherence, can also be explained on the network model proposed in this chapter. One major factor that discriminates between the network model and the other two models is the role of probability. Both the story model and the coherence model explicitly reject the idea that people think about evidence in a probabilistic fashion. This conclusion is largely based on the fact that people do not seem to follow the prescripts of Bayes' rule when evaluating evidence (Pennington and Hastie, 1986, 1992). In particular, people seem to be more conservative than Bayesian updating requires. We believe that although such findings militate against the notion that people carry out precise probability computations, they do not undermine the possibility that people are sensitive to the qualitative structure of probabilistic reasoning. For example, judging that one piece of evidence is relevant (or irrelevant) to a specific hypothesis does not require precise numerical estimates. We can all agree that the presence of the suspect at the crime scene increases the likelihood of guilt, especially if the suspect has no other reason to be there, even if we cannot assign exact probabilities to these events. Comparative judgements like these will often suffice to build up a qualitative network that links relevant pieces of evidence. This suggests a slightly different line of research to assess the extent to which people can engage in probabilistic reasoning: how closely does their reasoning approximate the qualitative prescripts of Bayesian networks? In the final section of this chapter we will review some recent studies that address this question.

Models of Alibi evidence

Alibi evidence is often critical in criminal cases. It has great potential to exonerate a suspect—if the suspect was not at the crime scene at the time of the crime, it is very unlikely that they are guilty. However, alibi evidence is often treated with suspicion (Gooderson, 1977), especially if it is proffered by a friend or relative of the accused. This is because the alibi testimony is readily explained away by the possibility that the alibi provider is lying to protect the suspect.

Despite the ubiquity of alibis in court cases, alibi evidence has not been subjected to much formal or empirical study. The few psychological studies on alibis (Culhane and Hosch, 2004; Olson and Wells, 2004) have reached sensible conclusions; for example, that people are more convinced by an alibi when it is provided by an impartial stranger rather than by a partial friend or relative; and more convinced by physical corroboration (e.g. a time-stamped receipt) than by personal corroboration (e.g. the verbal statement of a cashier). However, this leaves open a whole raft of questions about how people evaluate and assess alibis.

Of particular interest for the current chapter is how alibi evidence is assessed in the light of other pieces of evidence. As discussed above, one subtlety of alibi evidence is that it can provide information over-and-above the issue of whether the suspect was actually at the crime scene. We developed a basic alibi model[23] that represents the network of relations in typical alibi situations. The model implies that when a suspect's alibi is refuted by another piece of evidence (e.g. CCTV footage showing that the suspect was at the crime scene) this can incriminate the suspect along two separate routes: (a) evidence that the suspect was at the crime scene raises the probability of guilt; (b) knowing that the suspect lied in their statement also raises the probability of guilt. Furthermore, we distinguished two situations: (1) when the alibi provider does not know whether or not the suspect is guilty (e.g. when the alibi provider is an impartial stranger); (2) when the alibi provider does know whether or not the suspect is guilty (e.g. when the suspect is the alibi provider).We argue that the discredit of an alibi statement incriminates the suspect along both routes in case 2 but only along route (a) in case 1.

This is captured by the two models in Figure 7.4. Model 1 represents the case in which the alibi provider does not know whether the suspect is guilty,

[23] See earlier section on alibis, as well as Hepler, Lagnado and Baio (forthcoming), and Lagnado (forthcoming).

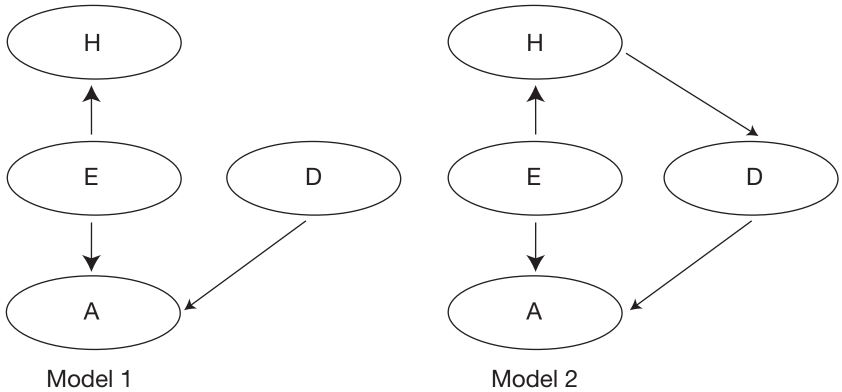

Figure 7.4. Two alibi models: (1) when alibi provider does not know whether suspect committed the crime (e.g. stranger provides alibi); (2) when alibi provider does know whether suspect committed the crime (e.g. suspect provides alibi).

Model 2 represents the case where the alibi provider does know. The key difference is the link from guilt to deception in Model 2 but not Model 1.

How well do these two alibi models capture people's actual reasoning when confronted with alibi evidence? We conducted an experimental study to explore this question (see Lagnado, 2010). All participants[24] in the study were given the background story to an assault case (based on the details of a real case). The gist of the case was that a young man was accused of assaulting a woman shortly after she left a nightclub. The suspect matched the description given by the victim, and he admitted to being at the nightclub on the night in question. Participants gave initial estimates for the probability that the suspect was guilty, which served as baselines for their subsequent judgements. They were then presented with alibi evidence in one of three versions. In condition (i) the suspect provided the alibi statement. He claimed that he caught a night bus from the club and was at home at the time of the crime. In condition (ii) the suspect's mother provided an equivalent alibi statement, and in condition (iii) the night bus driver provided it. The content of the alibi statement was kept as constant as possible across the three conditions, so that the essential difference between conditions was the nature of the alibi provider (suspect, mother or stranger). The conditions were constructed so that they varied as to whether the alibi provider was motivated to lie to protect the

[24] There were eighty jury-eligible participants in the study, all from the UCL student population. This experiment has since been replicated with 100 participants.

suspect (conditions i–ii) and whether the alibi provider knew if the suspect was guilty (condition i). See Table 7.1 for a summary.

Once participants received the alibi information they again estimated the probability that the suspect was guilty. Next, participants in all conditions were presented with a piece of evidence that undermined the alibi. This consisted in a statement to the effect that CCTV footage and face recognition analysis revealed that the suspect had followed the victim very near to the crime scene at about the time of the crime. After reading this statement, participants made their final estimates as to the guilt of the suspect. In addition to the three alibi conditions, a control condition was run on another group of participants. This control group was presented with the background information and then the CCTV evidence, but with no alibi evidence in between. This provided a critical comparison, because it served as a measure of how much the CCTV evidence alone was judged to raise the probability of the suspect's guilt.

The results for this experiment are shown in Figure 7.5. The bars labeled 'alibi' represent the differences between the baseline probability ratings (based on the background information) and the ratings after the alibis were presented. Effectively these bars show the impacts of the three alibis. It is notable that only the stranger's alibi lowers the judged probability of guilt. In all the other conditions the alibi has little effect. This fits with our expectations that alibis from partial witnesses are treated with great suspicion—there is an obvious alternative explanation for their statement (e.g. they might be lying to protect the suspect). What happened when the alibis were subsequently undermined by the CCTV evidence? This is shown by the bars labeled 'CCTV', which correspond to the difference between baseline ratings and those given after the CCTV evidence is presented. In all three cases the judged probability of guilt rises, reflecting the incriminating effect of the CCTV evidence. However, the level for the suspect is significantly higher than for the other alibi providers. And, most importantly, it is higher than the judged probability of guilt based on CCTV evidence in *the absence of an alibi*. (In numbers, $P(H|E\&A) > P(H|E)$.)

This pattern was predicted by our alibi model. The suspect is incriminated by the CCTV evidence along two routes: it shows (a) that he was at the crime scene, and (b) that he was lying in his alibi statement. This pattern is not

Table 7.1. Three conditions in the alibi study.

Alibi provider	Motivated to lie?	Knows that suspect is guilty?	Alibi Model	Model prediction
Suspect	Yes	Yes	2	$P(H\|E\&A) > P(H\|E)$
Mother	Yes	No	1	$P(H\|E\&A) = P(H\|E)$
Stranger	No	No	1	$P(H\|E\&A) = P(H\|E)$

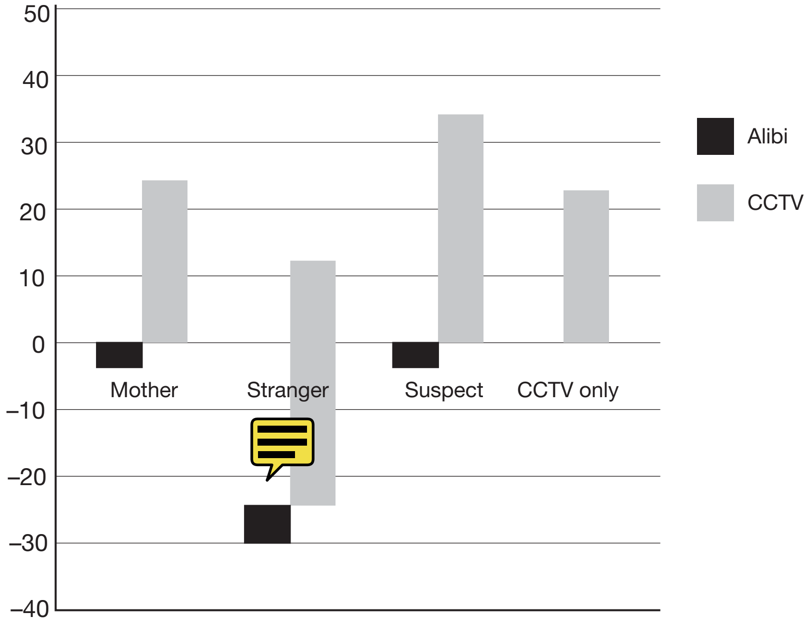

Figure 7.5. Baseline subtracted ratings for three alibi conditions and control (± standard error).

predicted by an additive model of evidence integration. The CCTV evidence by itself is less incriminating than the combination of CCTV evidence and alibi statement, despite the fact that the alibi statement by itself slightly lowers the judged probability of guilt. This suggests that Model 2 best captures people's reasoning about the suspect's alibi. In contrast, Model 1 best captures people's reasoning about the mother and stranger. In the case of the bus driver, we suppose that the preferred explanation for his false alibi was error rather than deception. This would account for why the CCTV evidence does not raise the level of guilt as high as with the other alibi providers. This possibility is supported by some other studies on alibi evidence (Olson and Wells, 2004), which show that when a stranger provides an alibi for the suspect, the question of correct identification is raised (which is much less likely for a friend or relative of the suspect).

This experiment supplies initial support for the alibi models proposed earlier. More generally, it confirms the claim that people's reasoning can be sensitive to the interrelations between hypotheses and items of evidence as predicted by the qualitative aspects of Bayesian network models.

The impact of discredited evidence

Recall one of the major twists in the plot of the *Witness for the Prosecution*. Vole's wife, Romaine, fabricated the letters that undermined her own testimony against Vole. She reasoned that the jury would not believe her if she simply provided an alibi for her husband (and our studies suggest that she was right about this). Instead, Romaine thought that a better way to persuade the jury of Vole's innocence was to undercut a substantial pillar of the prosecution's case, namely, her own testimony against Vole. A more general psychological maxim can be extracted from this line of reasoning—that once a story is mounted in favour of one side (in this case the prosecution), the discrediting of one element of that story can serve to collapse the whole story, even if there still exists incriminating evidence unaccounted for.

Exactly this pattern of reasoning has been explored in a recent set of studies (Lagnado and Harvey, 2008). We were interested in how the discrediting of one piece of evidence might affect other items of evidence. According to a purely normative account, the discredit should only affect related items, i.e. those that bear some causal or evidential relation to the discredited item. For example, consider a situation in which a suspect is accused of house burglary, and a witness testifies that the suspect was loitering in the area a few days before the crime (statement A). This same witness also testifies that she saw the suspect near the crime scene on the night of the crime (statement B). What happens if it is subsequently discovered that the witness has fabricated statement B. For example, perhaps there is strong evidence that she was out of town on the night in question, but fabricated her statement because she dislikes the suspect? Clearly one should disregard statement B—it no longer provides valid evidence against the suspect. But what about statement A? Should this also be disregarded?

In the situation where the items are related (e.g. both are produced by the same individual), it seems appropriate to extend the discredit of statement A to statement B. After all, if the witness is lying on one occasion, shouldn't we doubt her other statements too? Especially now that we know that she has reason to fabricate evidence against the suspect. At the opposite extreme, it also seems clear that we should not extend this discredit to pieces of evidence that are entirely unrelated, for example, forensic evidence such as a match between the suspect's shoes and footprints found at the crime scene. There is, however, a large grey area in between, where the extent to which a discredit should be generalised is unclear, and will depend heavily on the precise details of the case. For example, should we also call into question the testimony of other neighbours? What if they are friends of the discredited witness?

We conducted a set of studies to investigate such questions. Participants[25] were presented with various crime scenarios, and judged the probability that a suspect was guilty on the basis of several pieces of evidence. In the first study people always received two items of exonerating evidence in a row, followed by information that discredited the second item. For example, at stage one they were told about a footprint match, at stage two they were told about the witness testimony, and at the final stage they were informed that the witness had fabricated their statement. Participants gave probability of guilt judgements at each stage. The key question was whether the judged probability at the final stage, after the discrediting of the second item of evidence, simply returned to the estimate given at stage one (after the first item of evidence). If it did return, this would suggest that the discrediting information had only affected the second item of evidence. If it did not, this would suggest that the discredit was being extended in some way.

There were two main experimental conditions: the two items of evidence were either related (e.g. two statements from the same witness) or unrelated (e.g. a footprint match and a witness statement). A reasonable prediction—and indeed that sanctioned by a normative theory based on BNs—would be that the discredit should only be extended to related items of evidence. For example, the discredit of a witness testimony should affect other statements from the same witness, but should not affect footprint match evidence. This can be illustrated by constructing two BN models—one for the case of unrelated evidence, the other for related evidence (see Fig. 7.6).

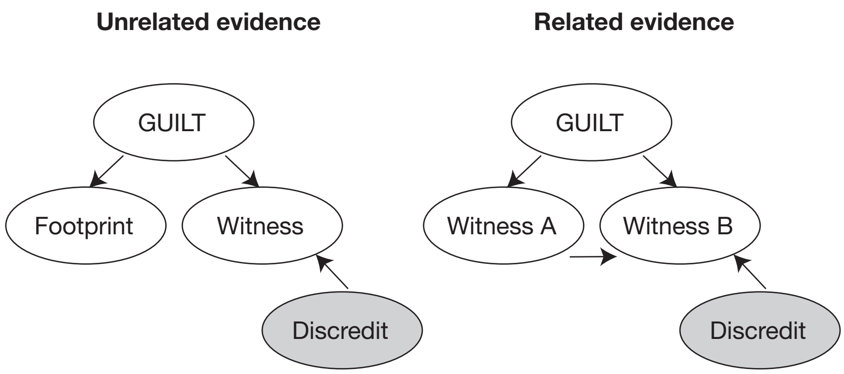

Figure 7.6. Two main conditions in the discredited evidence study.

[25] Participants were jury-eligible students from the UCL population. In total there were ninety-eight participants across the three studies. For details see Lagnado and Harvey (2008).

The results for this first study did not completely fit with the normative predictions (reinforcing the importance of empirical studies!). In those cases where the two items were related (e.g. two statements from the same witness) the final judgement was significantly lower than that given at stage one. Thus the discrediting of the second item was extended to the first (related) item. This was as expected by a normative model. If the witness has lied in one statement, there is increased reason to suppose that he had lied in another statement. However, this extension of the discrediting information also took place when the two items of evidence were *unrelated*. For example, the discrediting of a witness testimony was extended to an unrelated item of footprint evidence. This is clearly not sanctioned by the normative theory.

At first blush these findings can be most simply explained by the belief adjustment model. On this model later items of evidence are over-weighted relative to earlier items. It can therefore explain the results by assuming that the final discrediting information is over-weighted (hence the final stage judgement is lower than the first stage judgement). This belief adjustment model also assumes insensitivity to the interrelations between items of evidence, and thus explains why this effect occurs irrespective of whether the items are causally related.

To test out this simple explanation, a second study varied the order in which the items of evidence were presented. In particular, participants received information in one of two orders: (i) *late discredit*, as in study 1, in which the discrediting information was presented at the final stage, and (ii) *early discredit*, in which the discrediting information was presented after the first stage, and then another item of evidence was presented afterwards. For example, a witness statement was presented first, and at the second stage it was discredited. At the final stage another piece of incriminating evidence was presented (either related or unrelated to the initial item).

The results are displayed in Figure 7.7. The late discredit condition replicates the finding from Study 1—the discrediting is extended irrespective of the relations between the items of evidence (and against the predictions of a normative model). In contrast, the early discredit condition falls in line with the normative predictions. The discredit of one item is only extended to the other item when they are related, not when they are unrelated.

This is a puzzling pattern of results—why should people make normatively appropriate judgements when they receive information in one order, but not when they receive it in a different order? We advanced an explanation for this pattern that draws on coherence-based models of decision making. Coherence models (e.g. Simon *et al.*, 2004) presume that the mind strives for the most coherent representation of the evidence and hypotheses in the decision

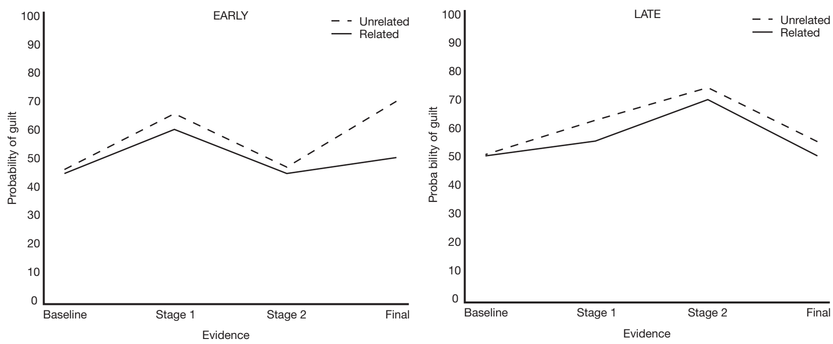

Figure 7.7. Mean probability of guilt ratings (± standard error) for Study 2 in late condition (left panel) and early condition (right panel).

problem. In particular, elements that cohere with each other will tend to be accepted or rejected together. Applied to our experiments, we assume that people tend to group items of evidence together according to whether they incriminate or exonerate the suspect. This basic division is strongly encouraged by the legal context, and the distinction between prosecution and defence evidence. During the decision-making process these two groupings (evidence for or against the accused) will compete with each other to be accepted, whereas the within-group elements will mutually cohere, irrespective of the exact causal relations between them.

On this view of the decision process, the over-extension of the discrediting information in Study 1 arises because items with a common direction (e.g. two pieces of incriminating evidence) are grouped together, and a subsequent discredit of one of these items hurts the whole group, bringing people's final estimates of guilt below their first stage estimates. It can also explain the difference between the late and early discrediting conditions in Study 2. In the late condition participants receive items A+ and B+,[26] and group these as positive evidence against the suspect (see Fig. 7.8). The two items cohere because they are both incriminating. Then participants receive information C, which discredits B. This discredit is extended to item A because of the prior grouping. The mutual coherence between A and B, and C's subsequent discredit of B, means that A and B are both undermined. In contrast, in the early condition participants receive item B+ first; this item is then discredited by C. When they receive A+ this is not grouped with B+ because B+ has already

[26] Note: '+' = incriminating evidence; '−' = exonerating evidence.

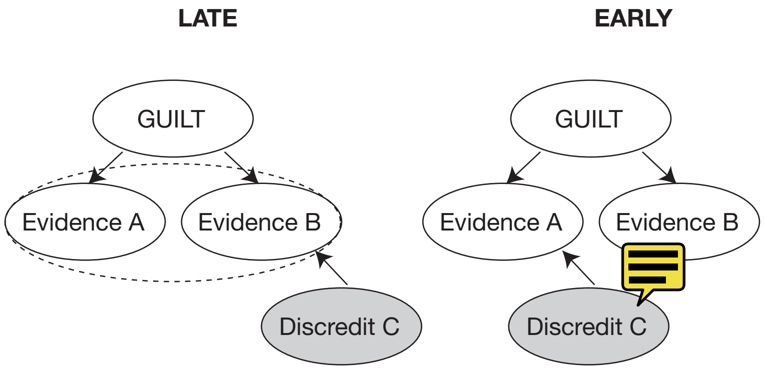

Figure 7.8. The effect of grouping on discrediting. In the late condition grouping of A and B leads to over-extension of the discredit C from B to A; in the early condition no grouping occurs, so discredit C is restricted to causally related evidence.

been discredited. Thus A+ is only discredited if C applies to it directly (i.e. if there is an appropriate causal relation).

This grouping hypothesis has some testable predictions. In particular, it predicts that coherent groupings will only emerge when evidential elements share the same direction (e.g. both incriminating or both exonerating). This implies that when the two items of evidence are mixed (e.g. one item incriminating and the other exonerating) the discredit of one item will not be extended to the other. This prediction was tested in Study 3. There were four conditions:

(1) Two incriminating items (A+, B+), followed by the discredit of B+
(2) Two exonerating items (A–, B–), followed by the discredit of B–
(3) One incriminating and one exonerating item (A+, B–), followed by discredit of B–
(4) One exonerating and one incriminating item (A–, B+), followed by discredit of B+

In short, there were two *non-mixed* conditions (1 and 2) and two *mixed* conditions (3 and 4). As in the previous studies, the causal relations between the two items of evidence were also varied within each condition. For example, in condition (3) the two items were either related (e.g. item A+ and B– were both witness statements, A+ stating that the suspect was seen at the crime scene, B– stating that the suspect was seen elsewhere) or unrelated (e.g. item A+ was footprint evidence placing the suspect at the crime scene, B– a witness statement stating that the suspect was seen elsewhere at that time).

The results for Study 3 are shown in Figure 7.9. The predictions of the grouping model were supported in all conditions. In the non-mixed conditions (both items positive or both items negative) the discredit was extended irrespective of whether the two items of evidence were related or unrelated. In contrast, in the mixed conditions (e.g. one positive item and one negative item) there was no extension of the discredit, again irrespective of whether the two items of evidence were related or unrelated. This suggests that the flow of inference in people's evidential networks was dictated by the positive vs negative grouping of evidence, rather than the precise causal relations between the elements. This finding does not show that people do not use causal reasoning; they clearly used 'explaining-away' inferences when evidence was discredited. However, their reasoning does not appear to be neatly captured by a Bayesian model of inference.

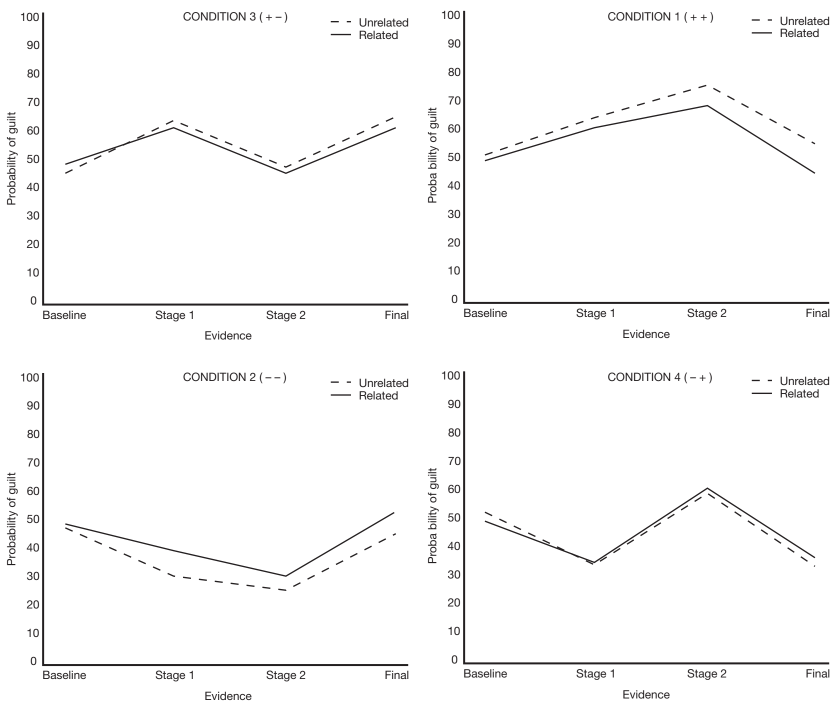

Figure 7.9. Mean probability of guilt judgements (± standard error) for the four evidence conditions in Study 3. Note: '+' = incriminating evidence; '−' = exonerating evidence.

Summary of discrediting studies

The experiments on discrediting evidence have thrown up a number of findings: that people use causal models to explain-away discredited evidence; that this explaining-away can also lead to over-extension of the discrediting information; that people's inferences are dependent on the order in which information is received. To explain these findings, we drew on coherence-based models of juror reasoning and introduced the idea that people group evidence according to its direction. Items of evidence with a shared direction are supposed to mutually cohere, irrespective of their exact causal relatedness, and mutually coherent groupings will fall together in the face of information that discredits one item of that grouping. This grouping hypothesis can explain why discrediting information is extended to causally unrelated items, so long as they share a common direction. It also explains the difference between early and late conditions: items only form a grouping in the late condition; in the early condition the discredit undermines the prior item before it can be grouped.

Returning to the *Witness for the Prosecution* story, we now see that Agatha Christie, and her character Romaine, had the right intuitions about the effect of discrediting evidence. By having Romaine testify for the prosecution, and then undermining her testimony, the case against her husband Vole collapsed. And this occurred despite the existence of independent incriminating evidence (such as the blood on Vole's cuffs, the maid's testimony, Vole being left in Miss French's will, Vole's enquiry about cruises, etc.). Our studies confirm this pattern of inference, and also suggest that had Romaine's discrediting occurred earlier in the trial, then its affect might have been less catastrophic!

These experimental findings need to be explored with a wider range of materials and conditions. If robust, they will reinforce previous studies showing that people do not integrate evidence in a fully Bayesian way (cf. Pennington and Hastie, 1992; Schum and Martin, 1982; Simon *et al.*, 2004). However, this 'over-extension' effect is a sign of a sensible cognitive mechanism at work. By grouping information people can overcome memory and processing limitations.[27] Without such organising and simplifying strategies it is unlikely that people could draw firm conclusions from complex and interrelated bodies of evidence.

One way of reconciling the current findings with previous research is to maintain that people use causal networks to represent the interrelations between evidence, but allow that they use coherence-based mechanisms for

[27] Cf. 'chunking' in memory research, see above, p. ~~000~~, for references.

inference and decision. This separation of representation and inference-mechanisms is not uncommon in cognitive science, and might prove a fruitful path to explore in future work.

Conclusions

This chapter has argued that people integrate and evaluate uncertain evidence by constructing network models. These network models are typically qualitative and incomplete, and based on people's causal beliefs about the specifics of the case as well as the workings of the physical and social world in general. These models can differ widely from person to person according to the knowledge and presuppositions that each individual brings to the situation. Despite these differences, individuals share the common aim of producing a consistent and coherent picture of the evidence: one that supports further inference, and the reaching of a decisive verdict.

Although we have concentrated on legal contexts, we believe that the network framework can be extended to other areas where people must integrate and reason about bodies of uncertain evidence (e.g. clinical and medical contexts, business and management, social interactions etc.). In all of these areas there are large bodies of interrelated evidence, and the decision maker must somehow combine these to reach singular judgements.

This chapter has only covered a subset of the issues arising in evidential reasoning. There are many other aspects that require careful treatment, such as the process of hypothesis generation, the search for evidence, and the social context in which evidence is proffered and judged. Indeed the role of evidence as an exchange of information between one person (e.g. defendant, witness) and another person or group (e.g. judge, investigator, jury) needs extensive study, especially in situations where this exchange involves strategic interactions between parties. Our optimistic perspective is that all of these issues will benefit from the central idea that people construct and reason with qualitative network models.

References

Aitken, C., Taroni, F. and Garbolino, P. (2003), 'A graphical model for the evaluation of cross-transfer evidence in DNA profiles', *Theoretical Population Biology*, 63: 179–90.

Axelrod (1976) ???details???.
Biedermann, A. and Taroni, F. (2006), 'Bayesian networks and probabilistic reasoning about scientific evidence when there is a lack of data', *Forensic Science International*, 157: 163–7.
Binder, D. A. and Bergman, P. (1984), *Fact investigation: from Hypothesis to Proof* (American Case book series) (St Paul, MN: *West* Publishing Company).
Cartwright, N. (2007), *Hunting Causes and Using Them* (Cambridge, Cambridge University Press).
Chase, W. G. and Simon, H. A. (1973), 'Perception in chess', *Cognitive Psychology*, 4: 55–81.
Christie, A. (1953/2002), *Witness for the Prosecution* (London, HarperCollins).
Cowan, N. (2001), 'The magical number 4 in short-term memory: a reconsideration of mental storage capacity', *Behavorial and Brain Sciences*, 24: 87–185.
~~Cowan and Andrews (2007), ???details???.~~
Culhane, S. E. and Hosch, H. M. (2004), 'An alibi witness's influence on jurors' verdicts', *Journal of Applied Social Psychology*, 34: 1604–16.
Davies, G. and Patel, D. (2005), 'The influence of car and driver stereotypes on attributions of vehicle speed, position on the road and culpability in a road accident scenario', *Legal and Criminological Psychology*, 10: 45–62.
Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. American Psychologist, 34, 571–82.
Dawid, A. P. and Evett, I. W. (1997), 'Using a graphical method to assist the evaluation of complicated patterns of evidence', *Journal of Forensic Science*, 42: 226–31.
Dawid, A. P., Mortera, J. and Vicard, P. (2007), 'Object-oriented Bayesian networks for complex forensic DNA profiling problems', *Forensic Science International*, 169: 195–205.
Druzdzel, M. (1996), 'Qualitative Verbal Explanations in Bayesian Belief Networks', *Artificial Intelligence and Simulation of Behaviour Quarterly*, 94: 43–54.
Fenton, N.E. and Neil, M. (2008), 'Avoiding Legal Fallacies in Practice Using Bayesian Networks', Seventh International Conference on Forensic Inference and Statistics. Lausanne, Switzerland.
Gilovich, T., Griffin, D. and Kahneman, D. (2002), *Heuristics and biases: the psychology of intuitive judgment* (Cambridge, Cambridge University Press).
Griffiths, T. L. and Tenenbaum, J. B. (2005), 'Structure and strength in causal induction', *Cognitive Psychology*, 51: 354–84.
Gooderson, R. N. (1977), 'Alibi' (London, Heinemann Educational).
Hamer, D. (1997), 'The Continuing Saga of the Chamberlain Direction: untangling the cables and chains of criminal proof', *Monash University Law Review*, 23: 43–76.
Halford, G. S., Cowan, N. and Andrews, G. (2007), 'Separating cognitive capacity from knowledge: a new hypothesis', *Trends in Cognitive Science*, 11: 236–42.
Harman, G. (1986), *Change in View: principles of reasoning* (Cambridge, MA, MIT Press/Bradford Books).
Hastie, R. (1993), *Inside the Juror* (Cambridge, Cambridge University Press).

Heider, F. (1946), 'Attitudes and cognitive organization', *Journal of Psychology*, 21: 107–12.

Hepler, A. B., Dawid, A. P. and Leucari, V. (2007), 'Object-oriented graphical representations of complex patterns of evidence', *Law, Probability and Risk*, 6: 275–93.

Hepler, A. and Weir, B. (2008), 'Object-oriented Bayesian networks for paternity cases with allelic dependencies', *Forensic Science International: Genetics*, 2(3): 166–75.

Hepler, A., Lagnado, D, Baio, G (forthcoming), 'Subtleties of alibi evidence'. (details???)

Hogarth, R. M. and Einhorn, H. J. (1992), 'Order effects in belief updating: the belief-adjustment model', *Cognitive Psychology*, 24: 1–55.

Kahneman, D., Slovic, P. and Tversky, A. (1982), *Judgment Under Uncertainty: Heuristics and Biases* (Cambridge, *Cambridge* University Press).

Keppens, J. (2007), 'Towards Qualitative Approaches to Bayesian Evidential Reasoning'. *Proceedings of the 11th International Conference on Artificial Intelligence and Law*, 17–25.

Kerstholt, J. H and Jackson, J. L. (1999), 'Judicial decision making: order of evidence presentation and availability of background information', *Applied Cognitive Psychology*, 12: 445–54.

Koller, D. and Pfeffer, A. (1997), 'Object-Oriented Bayesian Networks', *Proceedings of the 13th Annual Conference on Uncertainty in AI (UAI)*, 302–13.

Krynski, T. R. and Tenenbaum, J. B. (2007), 'The role of causality in judgment under uncertainty', *Journal of Experimental Psychology: General*, 136: 430–50.

Lagnado, D. A. (2010), 'Adverse inferences about alibi evidence' (Forthcoming, details???).

Lagnado, D. A. and Harvey, N. (2008), 'The impact of discredited evidence', *Psychonomic Bulletin and Review*, 15: 1166–73.

Lagnado, D. A., Waldmann, M. A., Hagmayer, Y. and Sloman, S. A. (2007), 'Beyond covariation: cues to causal structure', in A. Gopnik and L. E. Schultz. (eds.), *Causal Learning: psychology, philosophy, and computation* (Oxford, Oxford University Press), pp. 154–72.

Miller, G.A. (1956), 'The magical number seven, plus or minus two: some limits on our capacity for processing information', *Psychological Review*, 63: 81–97.

Olson, E. A. and Wells, G. L. (2004), 'What makes a good alibi? a proposed taxonomy', *Law and Human Behavior*, 28: 157–76.

Pardo, M. S. and Allen, R. J. (2008), 'Juridical Proof and the Best Explanation', *Law and Philosophy*, 27: 223–68.

Parsons, S. (2001), *Qualitative Methods for Reasoning under Uncertainty* (Cambridge, MA, MIT Press).

Pearl, J. (1998), ???details???.

Pearl, J. (2000), *Causality: Models, Reasoning and Inference* (Cambridge, Cambridge University Press).

Pennington, N. and Hastie, R. (1986), *Evidence evaluation in complex decision making*, *Journal of Personality and Social Psychology*, 51: 242–58.

Pennington, N. and Hastie, R. (1992), 'Explaining the evidence: test of the story model for juror decision making', *Journal of Personality and Social Psychology*, 62: 189–206.

Schmidt, H. G. and Boshuizen, H. (1993), 'On Acquiring Expertise in Medicine', *Educational Psychological Review*, 5(3), 205–21.

Schum, D. A. (2001), *The Evidential Foundations of Probabilistic Reasoning* (Evanston, IL, Northwestern University Press).

Schum, D. A. and Martin, A. W. (1982), 'Formal and empirical research on cascaded inference in jurisprudence', *Law and Society Review*, 17: 105–51.

Simon, D., Snow, C. and Read, S. J. (2004), 'The redux of cognitive consistency theories: evidence judgments by constraint satisfaction', *Journal of Personality and Social Psychology*, 86: 814–37.

Simon, D. and Holyoak, K. J. (2002), 'Structural dynamics of cognition: from consistency theories to constraint satisfaction', *Personality and Social Psychology Review*, 6: 283–94.

Sloman, S. A. (2005), *Causal Models: how people think about the world and its alternatives* (Oxford, Oxford University Press).

Sloman, S. A. and Lagnado, D. A. (2005), 'Do we "do"?', *Cognitive Science*, 29: 5–39.

Stewart, N., Brown, G. and Chater, N. (2005), 'Absolute identification by relative judgment'. *Psychological Review*, 112: 881–911.

Taroni, F., Aitken, C., Garbolino, P. and Biedermann, A. (2006), *Bayesian Networks and Probabilistic Inference in Forensic Science* (Chichester, John Wiley and Sons).

Thagard, P. (2000), *Coherence in Thought and Action* (Cambridge, MA, MIT Press).

Wellman, M. P. and Henrion, M. (1993), 'Explaining "explaining away"', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15: 287–91.

Wigmore (1913), ???Details???

Williamson, J. (2005), *Bayesian Nets and Causality* (Oxford, Oxford University Press).