

P.J. van Koppen (1995) Judges' decision-making. In:
R.H.C. Bull & D. Carson (red.), *Handbook of psychology
in legal contexts* (pp. 581-610). Chichester: Wiley

Chapter 6.7

Judges' Decision-making

Peter J. van Koppen,

Netherlands Institute for the Study of Criminality and Law Enforcement

INTRODUCTION

In each trial two main decisions are taken. In civil cases the decision concerns the liability of the defendant and, if liable, the compensation (or special order) to be awarded. In criminal cases the decision concerns the guilt or innocence of the accused and the sentence if guilt is sufficiently proved. This chapter deals with the decision on guilt or innocence in criminal cases.

In many countries criminal procedure is modelled on a trial in which a jury decides on guilt and a judge decides on the sentence. A jury trial, however, is comparatively rare and usually both decisions are left to the judge or magistrate, as occurs in most other countries. This chapter assumes that the decision is always made by a judge or a panel of judges. But it should be kept in mind that these models, *mutatis mutandis*, apply to jury decision-making.

DESCRIPTIVE AND PRESCRIPTIVE

Psychological models of decision-making in criminal cases share a characteristic that is peculiar in the field of psychology: all are both descriptive and prescriptive in nature. The mixed nature of these models derives from the mixed nature of law itself. Law is a 'social system created with a view of regulating the conduct of members of a community' (Blackman, Müller and Chapman, 1984, p. 3). Thus law is a behavioural technology and both the law and the legal system can be judged by the extent to which they serve that purpose successfully (Crombag and Van Koppen, 1991). At the same time, however, the law is an expression of a social philosophy in which, depending on place, time and circumstances, an ideal state of affairs is described which

society at large should strive for (Crombag, 1982; Van Koppen and Hessing, 1988). As a consequence, the study of decision-making in criminal cases has always been a mixture of prescriptive and descriptive theories. This has been most prominent in research on the decision of guilt or innocence of the suspect—the subject of this chapter—but also in research on sentencing. In sentencing studies the assumption always has been that disparity between sentences in comparable cases is evil and should be removed as much as possible (see for instance Homel and Lawrence, 1992; Berghuis, 1992).

This mixture of prescription and description stems from the mixed nature of the law and thus cannot be avoided. In each legal system the decision on guilt or innocence, for instance, is governed by a set of legal rules which prescribes how the fact finder—judge or jury—is to handle the evidence. These rules appear to emerge from the common understanding that without them too many innocent citizens might be convicted or too many guilty suspects might be acquitted. That, in itself, is an understanding which is descriptive in nature. But, the social philosophy character of law is also reflected in rules of evidence. Some of these rules, for instance, may serve to control police behaviour, by declaring evidence inadmissible if it is generated by police practices which are unwanted for whatever reason. Other pieces of evidence are inadmissible because they are considered biased against the defendant, even though they might be highly relevant for the decision on guilt or innocence; for example, in some legal systems, the prior criminal record of the defendant.

The standard of proof is another area where the mixture of description and prescription is profound. At various stages in criminal procedure, the evidence available must amount to a particular level of certainty to warrant a decision. 'Probable cause', necessary to arrest and for searches and seizure, requires 40–50 per cent of certainty (Melton et al., 1987, p. 27). In *Nugent v. Superior Court for San Mateo County* (254 C.A.2d 420, 62 Cal.Rptr. 217, 221), however, probable cause justifying an arrest without warrant is defined as a situation where the arresting officer has more evidence favouring a suspicion than against it, constituting a percentage of more than 50. For a conviction 'beyond a reasonable doubt' 90–95 per cent certainty seems to be required (Melton et al., 1987, p. 125). American jurors apparently consider around 90 per cent enough to convict (Hastie, Penrod and Pennington, 1983, p. 11).

Where do these percentages come from? They are not given by nature, because there were times when a suspicion was enough to convict (see Langbein, 1977). Indeed, in too many countries such a low standard of proof is still applied. Juries seem to use a standard as 'reasonable doubt' because they are instructed to do so (Kagehiro, 1990); judges probably do so because they have been taught that way.

What do such percentages mean? One manner in which this percentage can be interpreted is that American jurors consider it acceptable that of the convicted

Table 6.7.1 Estimated performance of juries

Reality	Jury verdict		Total
	Guilty	Not guilty	
Guilty	60.3	29.7	90.0
Not guilty	6.7	3.3	10.0
Total	67.0	33.0	100.0

defendants, 90 per cent are guilty and 10 per cent innocent. Given the base-rate of guilty and innocent defendants who appear at trial—which is highly skewed towards guilt because of the filtering done by police and prosecution—that is an amazing percentage. The standard that jurors say they apply would amount to a huge number of miscarriages of justice. A fair estimate might be the following. Kalven and Zeisel (1966) found that the jury convicts in 67 per cent of the criminal cases that come up for jury trial in the USA. Then, jurors apparently deem it acceptable that in 6.7 per cent of all cases an innocent defendant is convicted. If we assume that only 10 per cent of the defendants coming up to trial are innocent, still this juror attitude causes 3.3 miscarriages of justice in every 100 cases, while innocent defendants have only a 33 per cent chance of coming out acquitted (see Table 6.7.1). Is this the reason why so many miscarriages of justice are reported from the USA and the UK (see Borchard, 1932; Gardner, 1952; Frank and Frank, 1957; Hale, 1961; Hill, Radin, 1964; Zimmermann, 1964; Brandon and Davies, 1973; Young and Hill, 1983; Young and Sergeant, 1985; Bedau and Radelet, 1987, 1988; Gross, 1987; Woffinden, 1987; Rattner, 1988; Waller, 1989; Fletcher, 1992)?

The above analysis, however, is based on the assumption that the strength of the evidence is 'normally' distributed among guilty and innocent defendants respectively. That is not a realistic assumption, at least at the start of the trial. Under the (admittedly unrealistic) assumption that the 'evidence' which the police use for search and seizure is unrelated to the evidence presented at trial by the prosecution, it can be assumed that the strength of the evidence is distributed normally among guilty and innocent defendants respectively at the time of their arrest (see Figure 6.7.1).

Not all of these cases reach trial. The cases with too little evidence—in the judgement of police or prosecution—are dropped somewhere along the line. Under a system of plea-bargaining, the cases in which the strength of the evidence is overwhelming may not be decided by juries either. Figure 6.7.1 probably is not completely realistic, because the selection of cases which are left for a full trial depends on the policy of police and prosecution and estimates made by the defendants during plea-bargaining. The effect of the selections made prior to trial, however, is that the cases which go to trial resemble each other much more in a system with plea-bargaining than in systems without.

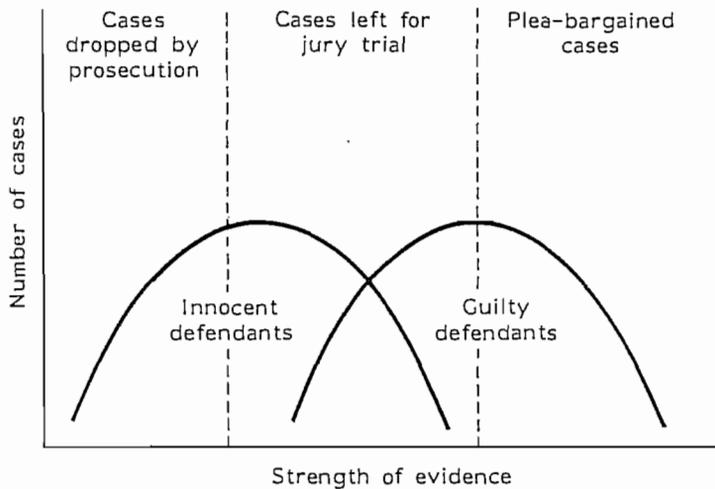


Figure 6.7.1 Hypothetical distribution of strength of evidence among guilty and innocent defendants

This example might be developed much further, but is used here to show that both descriptive and prescriptive elements play a role. Such an analysis might be used in a descriptive sense to compare inquisitorial to accusatorial systems by studying the effects of, for instance, plea-bargaining or a jury-system; but such an analysis is meaningless if the end results and the percentage of resulting miscarriages of justice are not taken into account.

The above discussion was necessary to argue that a valid theory of decision-making in criminal cases needs to account both for the empirical reality and for the normative elements involved. This chapter discusses the two kinds of models which are most widely used in the psychological study of decision-making in criminal cases: the hypothesis testing model and the so-called story model. The discussion is mainly based on the work undertaken by Crombag, Wagenaar Van Koppen (Crombag, Van Koppen and Wagenaar, 1992; Wagenaar, Van Koppen and Crombag, 1993) and will favour a story model. Before discussing the models, I will try to describe the decision-making problem faced by the fact finder in a criminal trial.

THE PROBLEM OF GUILT OR INNOCENCE

The finder of fact in a criminal trial is faced with the problem of distinguishing guilty from innocent defendants. In more general terms, that is a problem handled by signal detection theory (Green and Swets, 1966) which deals with distinguishing signals against a background of 'noise'. Distinguishing guilty defendants against the background of innocent ones is such a decision problem.



Figure 6.7.2 Confidence dimension with decision criterion for criminal cases, according to signal detection theory

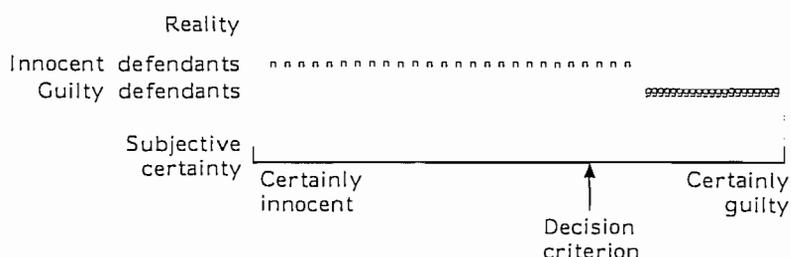


Figure 6.7.3 Decision criterion for criminal cases with perfect discrimination between innocent and guilty defendants

An essential element in the theory is the level of confidence of the decision-maker. After evaluating all the evidence, the decision-maker has a level of confidence in the guilt of the defendant. The levels of confidence can be depicted on a continuous dimension, going from 'certainly innocent' to 'certainly guilty'. On the same dimension a decision criterion can be depicted. If the level of confidence in the guilt of the defendants surpasses the decision criterion the defendant is found guilty (see Figure 6.7.2). Where a particular defendant falls on this dimension, depends on the subjective evaluation by the decision maker.

The 'true' state of affairs can be depicted on the same dimension, as is done in Figure 6.7.3: guilty defendants with a 'g', not guilty defendants with an 'n'. Truth is a relative subject in criminal cases, but for the sake of the present discussion let us assume that defendants fall into one of two classes (the guilty ones and the innocent ones) and that the evidence presented against them has some relation to their guilt. In Figure 6.7.3 a few innocent defendants are convicted, because they fall above the decision criterion. This problem, however, can be amended easily if the decision-maker only slightly adjusts his or her decision criterion. If a higher standard of proof is required, and thus the decision criterion is shifted slightly to the right in Figure 6.7.3, all innocent defendants are acquitted, while still all guilty ones are convicted.

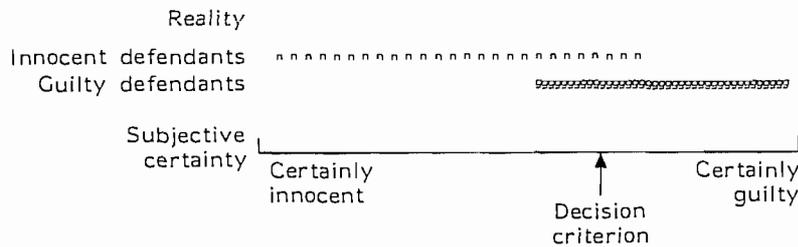


Figure 6.7.4 Decision criterion for criminal cases with imperfect discrimination between innocent and guilty defendants

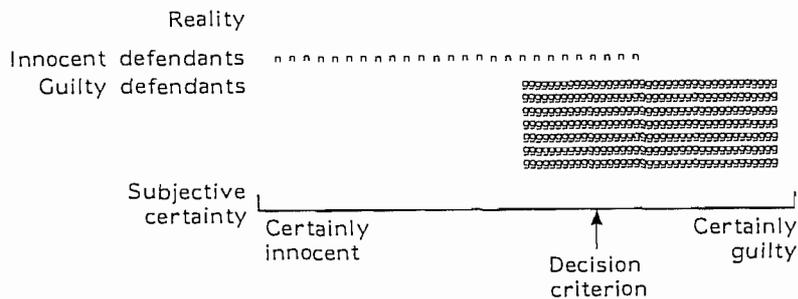


Figure 6.7.5 Decision criterion for criminal cases with imperfect discrimination between innocent and guilty defendants under a 95 per cent base-rate of guilty defendants

In reality, however, there are innocent defendants who *appear* to be guilty and there are guilty defendants who *appear* to be innocent (because too little evidence has been found against them). Figure 6.7.4 is more realistic in that sense, and shows that no shift of the decision criterion to either the left or the right can discriminate between guilty and innocent defendants without making errors.

Figure 6.7.5 is even more realistic for decision-making in criminal trials, introducing base-rate information into the decision problem. In most criminal trials the defendant is guilty of the crime charged. Crombag, Van Koppen and Wagenaar (1992) estimated that about 95 per cent of Dutch defendants are guilty (there is no reason to expect other percentages in other countries). If the judge simply convicted every defendant without further investigation then only in 5 per cent of the cases an error would be made; a rate that is considerably less than in many other areas of decision-making. Of course, there are many good reasons not to proceed in that manner, one of them being that it would leave the behaviour of the police and the prosecution completely unchecked. But it should be realised that judicial decision making must be of considerably high quality to beat a base-rate of 95 per cent guilty defendants.

There is no way in which the decision criterion can be set such that the evidence perfectly discriminates between guilty and innocent defendants, but Figure 6.7.5 also shows that a shift of the decision criterion has unequal effects on conviction rates of guilty and innocent defendants. A shift to the left—to a more lenient position—causes many more guilty defendants to be convicted, but also leads to the conviction of some innocent defendants. Likewise, a shift to higher standards of proof—when the decision criterion moves to the right—will save innocent defendants from being convicted, but causes an acquittal for many more guilty defendants.

Signal detection theory predicts that the precise placing of the decision criterion on the dimension—that is the level of confidence required for a conviction—is influenced by the utility of wrong and right decisions. And, indeed, the relation between the number of wrongful convictions and unjust acquittals has long been recognised in the legal community. Wigmore (1937), for instance, argued that avoiding one wrongful conviction is worth 20 unjust acquittals. Others have argued for other levels (see Williams, 1963, chapter 7). Apparently the costs of the suffering inflicted upon the convicted innocent is considered more grave than the hazards for society if a guilty person is turned loose.

The risks and costs involved in decision-making in criminal cases are not invariant across types of crimes. For minor crimes, the risks for society of acquitting a guilty defendant are not very large, while a wrongful conviction may cause much suffering for the defendant. For a serious crime like rape, however, the risks involved with an unjust acquittal may outweigh the costs for the innocently convicted defendant. Signal detection theory, then, predicts something which Crombag, Van Koppen and Wagenaar (1992, ch. 5) called the 'conviction paradox': for serious crimes where a true perpetrator is likely to repeat the crime, the decision criterion moves to the left—a lower standard of proof—while for minor crimes the decision criterion is moved to the right. In countries with a jury system the seriousness of the offence charged may have a counter effect: the more serious the offence, the more likely it is that the defendant is tried before a jury, which in turn makes it more likely that the defendant will be acquitted (Kalven and Zeisel, 1966).

THE LOGIC OF EVIDENCE

It would be most appealing if the evaluation of evidence in a criminal case could lead to a decision through a strictly logical process of consecutive steps. Such a process would require that the facts presented as evidence to the court or jury can be established in some objective manner. Then some inferential process almost automatically and inescapably would lead to the conclusion on guilt or innocence of the defendant.

The possibility of such a process is debatable, both on philosophical and practical grounds. It assumes that there are facts 'out there' which can be known with certainty if we just look hard enough. It has long been established that each of these individual steps in such a logical process needs some form of interpretation (in law, for instance, by Cuesta-Rua, 1981, pp. 113 ff.; Scholten, 1974, pp. 11 ff.). At the same time, lawyers behave as if such a process is possible. Wigmore (1937), for instance, described this process as a regression by which the *probandum* is specified into a large number of *facta probanda*, which are at some point matched with *facta probantia*. He appears to assume that the *probandum* can be specified as much as we want, which implies that legal proof can attain any required degree of precision. This conclusion does not follow. The regression to conditions of conditions postpones the problem without solving it. In the end the evidence must be matched with conditions and it is far from obvious that any degree of specification will make this matching unproblematic. Only a solid match would allow a perfectly safe conclusion that the condition is met. Such a conclusion can only be safe when the evidence allows just one interpretation. In reality there are always more interpretations.

Not only are the individual steps to be taken in such a logical process impossible; also the structure of the process as a whole poses problems. Any process of logical inference is a bottom-up process: one starts with the facts—the evidence—and infers conclusions from these facts. But how do we know which facts are relevant for the decision and which not? We only know which facts are relevant if we anticipate the decision. But, at the same time, we only know which decision is anticipated, if we know the facts of the case. This circular problem can only be solved by taking a decision first and working back to the facts. Such a decision need not be final but can take the shape of a working hypothesis. This working hypothesis is splendidly provided for by the prosecution.

Together, these problems shape the manner in which the decision is made into a top-down process, in which the hypothesis (the allegation of theft, rape or murder) comes first, and the fact-finding is derived from it. Intent is not inferred from facts that happened to be presented, but the other way around: certain facts are sought and presented because they may serve to prove intent. This way of describing the process comes closer to what appears actually to be happening during the trial. The trial starts with a presentation of the indictment, i.e. before facts are presented. The charge is not inferred by the judge or jury from the evidence, but the other way around. Basically, this is a process in which the evidence is used to verify the indictment; not one in which the innocence of the defendant is falsified. This problem will be looked at further below.

HYPOTHESIS TESTING

If decision-making in criminal cases is not a logical bottom-up process, one might turn to Hart's (1963, p. vii) conclusion that courtroom decision-making comes closer to rhetoric than to reason. In psychology two alternative paths have been chosen to model judicial decision-making. One draws upon the analogy to typical scientific top-down reasoning: a hypothesis is tested against evidence, as is done in most empirical sciences. The Reverend Thomas Bayes gave a mathematical formulation for this process which has become quite popular as a means of modelling decision-making in criminal cases (Kaplan, 1968; Finkelstein and Fairly, 1970; Tribe, 1971; Schum, 1979; Goldsmith, 1980; Saks and Kidd, 1980; Edwards, 1988). The principle is simple. It is assumed that the decision-maker has a certain prior belief in the truth of the hypothesis. This degree of belief is mathematically expressed by odds, i.e. a number between zero and infinity. These odds are obtained by dividing the probability that the hypothesis is true, p_{true} , by the probability that it is false, p_{false} . The prior belief, or the prior odds, then, are expressed by:

$$\text{prior odds} = \frac{P_{\text{true}}}{P_{\text{false}}} \quad (6.7.1)$$

For instance, when p_{true} is .80, and p_{false} is .20 (the hypothesis is either true or false), then the prior odds are $.80/.20 = 4.0$. In a criminal trial the two mutually exclusive hypotheses are 'guilty' or 'innocent'. New evidence offers the opportunity to revise the prior belief, and turn it into a *posterior* belief. This is achieved by multiplying the prior odds by the diagnostic value (D) of the evidence:

$$\text{posterior odds} = \text{prior odds} \times D_i \quad (6.7.2)$$

The diagnostic value of evidence is obtained from a performance table. An example is given in Table 6.7.2, showing the performance of the so-called Dolls test which is used in sexual abuse cases to establish the truthfulness of

Table 6.7.2 Performance of the Dolls test ($n = 20$)

The child was:	Results of the Dolls-test	
	Positive	Negative
Sexually abused	Hit .9	Miss .1
Not sexually abused	False alarm .2	Correct rejection .8

Source: Based on Jampole and Weber, 1987

the child's account. (Whether the Dolls-test has a diagnostic value which is high enough to make it admissible is not without debate. The Dutch Supreme Court (HR 28 February, 1989, *Nederlandse Jurisprudentie* 1989, no. 748) considered the Dolls-test so unreliable, that it decided that a court cannot ignore counter-arguments by the defence. By this Supreme Court decision the test was, in practice, abolished in The Netherlands. In the USA the Dolls-test is the most commonly used tool in the validation of sexual abuse allegations (Skinner and Berry, 1993). However some US courts have ruled that evidence based on Anatomically Correct Dolls is inadmissible because the test does not meet basic psychometric criteria (*In re Amber B. and Tella B.*, 191 Cal. App.3rd 682, 1987; *in re Christine C. and Michael C.*, 191 Cal. App.3rd 676, 1987).)

The diagnostic value (D) of the Dolls test is the ratio of hits and false alarms; in the table: $0.9/0.2 = 4.5$, meaning that a positive test makes it 4.5 times more likely that the child was sexually abused. After this piece of evidence the prior odds—whatever they were—are adjusted to become posterior odds which may serve as prior odds for the next piece of evidence. With N pieces of evidence, the initial prior odds are turned into final posterior odds as follows:

$$\text{final posterior odds} = \text{prior odds} \times D_1 \times D_2 \dots \times D_1 \times \dots \times D_N \quad (6.7.3)$$

If the final posterior odds surpass a pre-set level of confidence, the defendant can be convicted. This Bayesian approach thus seems an elegant model of decision-making. This chapter argues, however, that modelling decision-making in criminal cases as hypothesis testing is the right solution to the wrong problem. There are four defects of this approach: (1) setting the initial probability; (2) determining the diagnostic value of evidence; (3) revision of the probability; and (4) taking the final decision (the arguments advanced here are more fully discussed in Wagenaar, Van Koppen and Crombag, 1993).

Prior Probability

The presumption of innocence assumes that the defendant is innocent until proven guilty. This would require the decision maker to give a prior probability of zero to the hypothesis that the defendant is guilty. Then, of course, evidence of whatever quality cannot change the odds to anything higher than zero—the result of a multiplication with zero is, of course, always zero—and the Bayesian model would be useless. A solution might be to keep the initial probability 'very, very low' (proposed by Wagenaar, 1988, p. 149), say at 0.0001, or give it the value of the base-rate of guilty defendants, say 0.95. For both levels—and many more levels—arguments can be given. This problem is avoided in a variation of the model, proposed by Goldsmith (1980). He suggests that it is possible to leave the prior odds of guilt before the presentation of evidence undetermined. But his 'evidentiary value theory' still suffers from the drawbacks listed below.

On a theoretical level the initial probability of guilt is not of overwhelming importance, since it can be shown that during the process of adjustment of odds the influence of the initial value becomes less and less. In practice, however, the initial probability of guilt can be of decisive importance to the final decision as Schünemann and his co-workers demonstrated (Schünemann and Bandilla, 1989; Schünemann, 1983). In one of their experiments professional judges who had read the file of the case before the trial all convicted the defendant, while only 27 per cent of the judges who had not read the file beforehand convicted. Reading the file before the trial—as is common in most inquisitorial systems—apparently introduces a prior belief in guilt of the defendant which subsequently requires less evidence to come to a conviction at trial. Schünemann's results are confirmed in research by Koehler (1991), who concluded that '[a]ny task that prompts a person to temporarily accept the truth of a hypothesis will increase his or her confidence in that hypothesis' (p. 502).

Diagnostic Value of Evidence

In the Bayesian model—or in any quantitative approach to evidence in criminal cases for that matter—it is a *conditio sine qua non* that the value of evidence can be quantified in some way. Sometimes this can be done. An example is the Dolls test, given in Table 6.7.2. If a Dolls test in a case proved positive, the diagnostic value of such evidence is known. That is, it is known if certain conditions are met. We at least must have empirical facts or research, before we can give a fair estimate of the diagnostic value of some specific piece of evidence, provided that the research is of decent quality and the facts are close to the evidence at hand. But, usually research is scarce and the facts are difficult to assess. Research on identification line-ups, for instance, has shown that the diagnostic value of a splendidly performed line-up is about 15 (Wagenaar and Veefkind, 1992), but what if not all the requirements are met to call it a very good line-up? And even then, some argue, the present tradition of laboratory research does not apply to forensic practice (Bekerian, 1993; Egeth, 1993; Wells, 1993; Yuille, 1993). Another example is DNA-profiling, where it remains unclear which population must be taken as the reference group to determine the probability of a match (see Easteal, McLeod and Reed, 1991; Roberts, 1991) and thus some argue that the probability of a match is at best unknown (see for the softness of other kinds of 'hard' forensic evidence Saks, 1989; Saks and Koehler, 1991). So, even with well researched kinds of evidence we usually know much about what problems and pitfalls can be encountered, but the effect on the diagnostic value of such evidence remains an enigma. With other kinds of evidence, we are much more in the dark. How to assess, for instance, the diagnostic value of the testimony of a witness who might have an interest in the outcome of the trial?

In the absence of robust figures on the diagnostic value of pieces of evidence, judges and juries cannot but step in and guess. It would be most appealing if they could be helped in decision-making with at least some indication of the

lower and upper limit of the diagnostic value of evidence. The theory of anchored narratives, discussed below, provides for such a decision aid.

Revision of Probability

The process of revision, as described by the Bayesian model, assumes a number of properties that are, in fact, quite unrealistic. One of these properties is compensation: one very diagnostic piece of incriminating evidence can be cancelled out by a number of facts that point in the opposite direction through the multiplication operation. Then, a very strong piece of evidence favouring the defendant—for instance, he was in prison at the time of the crime—can be compensated by a large number of fairly weak pieces of evidence against him. Or the other way around: the fingerprints of the defendant found at the scene of the crime—often decisive evidence for a conviction—can be cancelled out by a good explanation which places the defendant at the scene at some other time. A sufficient number of ‘facts’ of low diagnostic value may, in the end, compensate for a fact that might be considered as decisive evidence.

Another problem is the independence of the supposed diagnostic values. All sorts of interactions cannot be represented in the Bayesian model. The same is true for contradictions. The Bayesian model, however, has no provision for such dependencies among pieces of evidence and would lump their diagnostic values together (see Cohen, 1977; Wagenaar, 1991).

The Decision Criterion

Finally there is the problem of the decision criterion. When are the odds in favour of guilt high enough to convict? The model does not explain how such a criterion is chosen, nor even whether the criterion is constant or variable. For instance, should the criterion be the same in the case of a traffic violation for which only a fine is given, as in the case of a capital crime? One could argue that the judge should take fewer risks in the latter case. Should a judge in a case of multiple rape put the criterion on a lower more risky level because after a wrongful acquittal the criminal may rape again? Does consideration of consequences justify variations in the decision criteria? The Bayesian model does not answer such questions.

Conclusion

Taken together, the problems posed by the Bayesian model of legal decision-making are overwhelming. The most important argument against the use of models of hypothesis testing for decision-making in criminal cases, however, is that judges and juries do not argue and decide in that way. In fact, hypothesis testing is so far off what actually happens in court that it is not only unsuccessful as a descriptive model, but also too alien to the legal tradition to be of use as a prescriptive model either. That does not mean that models of

hypothesis testing might not be useful to discuss certain pieces of evidence or might not apply to some—but atypical—cases. The Collins case (*People v. Collins*, 68 Cal.2d 319, 1968) is such a case and has been analysed many times (for instance Edwards, 1991). But, the hypothesis testing models will not hold in most cases—and indeed in most cases with problems of evidence—for the reasons stated above. Thus, these models are the right solution to the wrong problem.

ANCHORED NARRATIVES

Alternatively, I will argue that a narrative model comes closer to a description of what judges and juries in fact do and at the same time gives a base for grounding prescriptions for evaluating evidence. A narrative is, as Bruner (1984) puts it, 'concerned with the explication of human intentions in the context of action'.

Story Context and Facts

The theory of anchored narratives (Crombag, Van Koppen and Wagenaar, 1992; Wagenaar, Van Koppen and Crombag, 1993), discussed here, is not entirely original. Others have proposed similar ideas and prepared the way. Bennett and Feldman (1981) begin their book *Reconstructing Reality in the Courtroom* by stating that 'the criminal trial is organised around story telling'. The idea is that the work of the judge or jury consists of determining the *plausibility* of the stories presented by the prosecution and the defence. Narrative theories, which have a long history in cognitive psychology (cf. Bartlett, 1950; Rumelhart, 1975), hold that evidence derives its meaning from a story context. Detached from a story, facts do not prove anything; a court cannot decide on mere facts, only on a story. Two aspects of a story may determine its believability: the 'goodness' of the story in itself (to be defined later), and the degree to which it is supported by facts (not *the* facts, which means something vastly different).

The derivation of meaning from story contexts is a well-known and easily illustrated effect. Consider the following story.

I brought my daughter Vere to Jan's birthday party. Upon arrival I got coffee and Vere lemonade. We sang for Jan, after which he blew the candles. I left, to pick Vere up at the end of the afternoon.

This story strongly suggests a number of things that remain implicit, and that may not even be true. For instance, it is suggested that more children were at the party; and most of the readers will have imagined a cake. People will automatically fill in gaps in stories, and in doing so, give meanings to statements which maybe were never intended. On the other hand, one may

re-word stories carefully, with the intention of suggesting inferences that cannot be proved by the facts.

The Quality of Stories

Researchers in many disciplines, such as literature, anthropology, and artificial intelligence, have tried to answer the question of what constitutes a good and plausible story. Rumelhart (1975), Robinson (1981), and Van Dijk (1980) have all designed story grammars: sets of rules to which a well-formed story must conform. Among the story grammars which have been proposed, Bennett and Feldman's (1981) was specially designed for judicial contexts (cf. Jackson, 1988). The manner in which they obtained their grammar is quite interesting. They asked 58 students to tell a story; half of them were asked to tell a *true* story, the other half to *invent* a story. Every time a story had been told, the others were asked to guess whether this was a true or an invented story. The guesses were not better than chance. But stories that were *accepted* as true shared some properties that the rejected stories were lacking. These properties were: a readily identifiable central action, and a context (setting) that provides an easy and natural explanation of why the actors behaved in the way they did.

In a good story all elements are connected to the central action. The context provides a full and compelling account of why the central action should have developed in this particular manner. If the context does not achieve that effect, then the story is said to contain ambiguities. There are two types of ambiguities: missing elements and contradictory elements.

The analysis of what makes stories plausible was extended by Pennington and Hastie in a series of subsequent publications (1986, 1988, 1991). Without going into too much detail, it can be said that, according to them, in good stories all actions are explained by factors of three kinds: physical conditions, psychological conditions, and goals. Each of these can influence the other two, and is also determined by the general setting. Given the goals of the actors and the situation they found themselves in, a good story makes it 'logical' that a psychological condition developed that made the crime a 'logical' outcome.

Pennington and Hastie's story grammar, and their thesis about the importance of stories, is supported by empirical research. In one study (Pennington and Hastie, 1986) they presented over 200 prospective jury members with a filmed documentary of a criminal trial. The possible verdicts were: first-degree murder, second-degree murder, manslaughter, and self-defence. On the basis of the evidence presented in the film all options were chosen by a reasonably large number of subjects. Next they were asked to explain what they thought had happened. All story structures conformed to the postulated story grammar; but different judgements were based on different selections of facts. In the authors' words: 'variability in the story construction stage is systematically related to verdict choice' (p. 253). But the variety of stories was not based upon

an extremely rich and multi-interpretable set of evidence. The differences were mostly brought about by inferences about what people did, thought, felt, wanted. Inferred facts which were not present in the film constituted 45 per cent of all story components. If people are allowed to 'invent' 45 percent of the evidence, it is not surprising that they can construct different stories about what happened and reach different conclusions. (See also Hogarth (1971) who found that disparity among judicial decision of magistrates vanishes if they agree on the facts of the case.)

In another study, Pennington and Hastie (1986) showed that the order in which evidence is presented has a major influence on the judgement (which is in conflict with predictions of Bayesian theory). Both the prosecution and the defence could present their evidence in a random order, or in story order. The combination of these two variables results in four groups. The dependent variable was whether the subjects thought that the defendant was guilty of first-degree murder. The results of the study indicate that the party who presents the evidence in story order was believed more readily, even though the evidence itself was exactly the same in both conditions. The effect can be as large as changing a 31 per cent chance of conviction (when the defence presents its evidence in story order and the prosecution in random order) into a 78 per cent chance (when the prosecution presents evidence in story order and the defence in random order). Clever presentation of the story is half of the work! What is the other half?

Anchoring

Stories told in a criminal court must not only be good, they must be considered true before a conviction can follow. The truth of a story is established by means of evidence. In and of itself, however, evidence does not prove anything at all. Any piece of evidence only proves something if we are willing to believe in a general rule which we hold to be true most of the time. For instance, the testimony by two eye-witnesses will only prove something, i.e. support the story, if it is assumed that, generally, eye-witnesses do not lie or make mistakes; the post mortem report can only prove something as long as the pathologist does a good job.

Such general rules, however, are seldom true without exception. Witnesses sometimes err or lie, and experts occasionally do make mistakes. The rules that make evidence prove something should more accurately be phrased: witnesses speak the truth *most of the time*, and pathologists *almost* never make mistakes. The possibility of exceptions to rules means that, on a particular occasion, we must show that a possible exception does not apply. This argument calls on the general rule that it is very improbable that two lying eye-witnesses come up with the same lie. But, of course, this rule allows exceptions: if witnesses have had the opportunity to confer before testifying, they can easily lie and be mutually consistent at the same time. Hence the

prosecution must prove that this exception does not apply, for instance by showing that the two witnesses had no opportunity to confer, or if they had, did not do so.

In a similar manner every piece of evidence needs further support, until it can be safely anchored in a general rule that cannot be sensibly contested because all parties acknowledge it to be true in the given case. These general rules are usually common-sense facts of life. We often accept an argument because we unwittingly believe the underlying rule which gives it an anchor, even though an explicit formulation of the rule would cause us to protest or even reject it. Cohen (1977) calls these generally accepted rules 'common-sense presumptions, which state what is normally to be expected but are rebuttable in their application to a particular situation if it can be shown to be abnormal in some relevant respect' (p. 247). I shall qualify the role of general rules later but for the time being define them as 'common knowledge of the world in the form of rules which are usually valid'. A pictorial representation of the anchoring heuristic is presented in Figure 6.7.6.

At the top of Figure 6.7.6 there is the story of the original indictment, of which the soundness has already been judged satisfactorily. Next comes an ordering of the evidence in such a way that it forms anchors between the story and a 'ground' of generally accepted common-sense rules. For three details evidence is offered, but each piece of evidence forms a sub-story in itself, which needs an anchor in the form of further evidence, which in turn forms a sub-sub-story in need of an anchor. Whether a sub-story is safely anchored depends on our willingness to accept as true the common-sense rule of which the sub-story is an instance.

For the first detail of the original story the anchoring is quite complicated, constituting a long anchor chain. Two pieces of evidence are offered, which apparently cannot be safely anchored as such in safe common-sense rules. Hence more evidence is sought, constituting sub-sub-stories. The first of these is anchored on to the ground through a sub-sub-sub-story; the second one is not anchored at all. Hence the anchor chain as a whole is ineffective. The second point in the story is directly anchored on to the ground of some common-sense rule. That could have been a statement such as: 'A man who runs into a bank with a mask on and a gun in his hand, has the intention to rob the bank.' There is no point in doubting this rule to be generally true and hence no reason to probe deeper for a safer anchor. The third point is anchored through one intermediate story; a police officer's sworn testimony is an example of that sort of anchoring.

Evaluating the story of the indictment and the evidence might seem to be a deductive process: it appears as if the individual case is brought under the general common-sense rule and the reasoning departs from the general rule. It is not. The point of departure is the individual case from which a

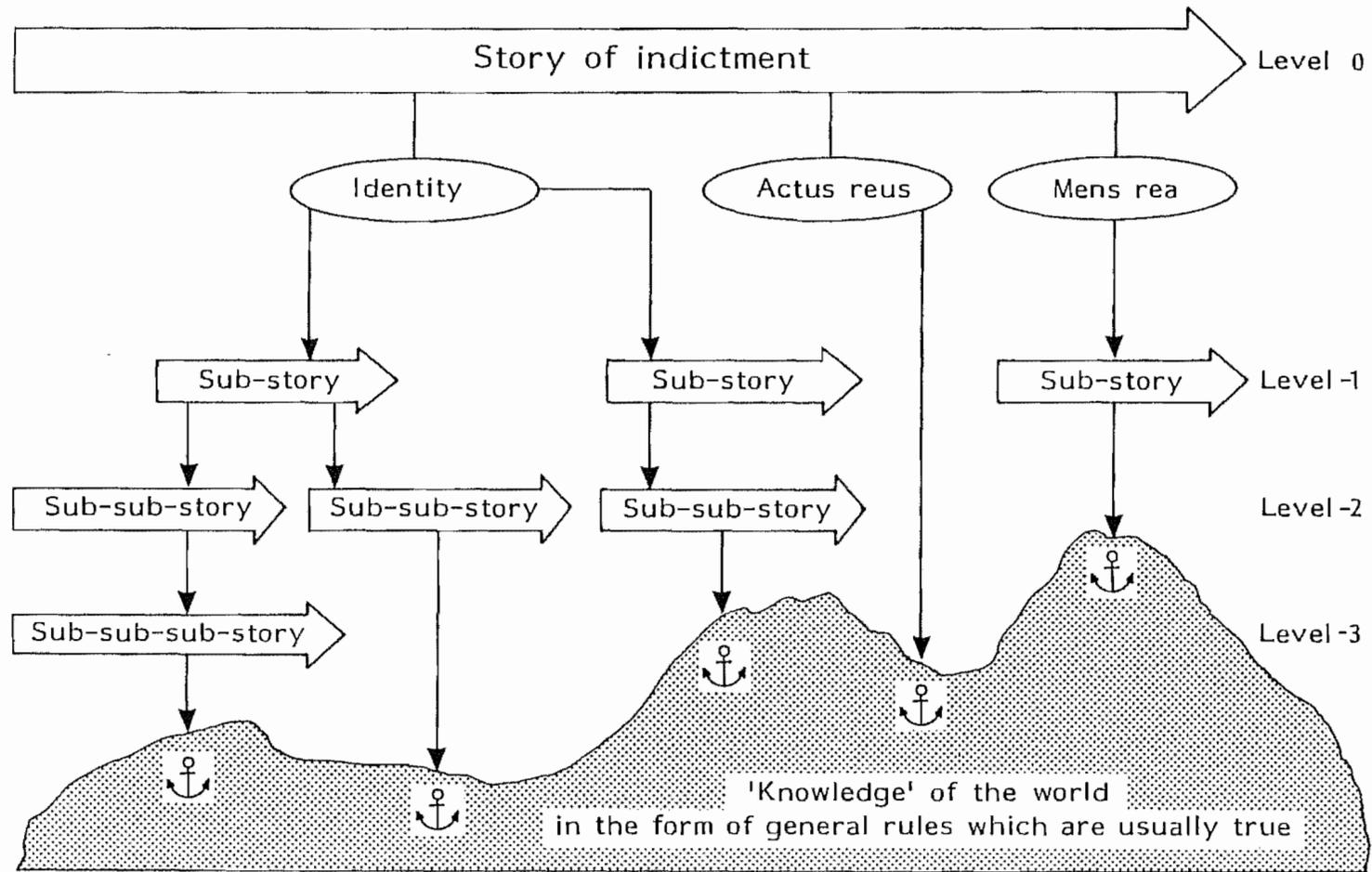


Figure 6.7.3 The theory of anchored narratives

generalisation is sought to explain the individual story (see Cohen, 1977, p. 247 ff.). It is a basic human need to explain the world around us and we are proficient in explaining odd or unexpected occurrences by generating general rules which might explain them (see Semin and Manstead, 1983).

ANCHORED NARRATIVE MODEL OF DECISION-MAKING

The model of decision-making in criminal cases, encompassed in the theory of anchored narratives, provides a means of describing these decisions, but also can serve as a vehicle to show on what points decisions based on evidence went wrong or are about to go wrong. A detailed analysis of such errors in Dutch cases was presented in Crombag, Van Koppen and Wagenaar (1992) and Wagenaar, Van Koppen and Crombag (1993). These points are discussed in more general terms below.

A Description of Courtroom Behaviour

The first task of the prosecutor is to tell a good story; a story with a central action and a context that makes the central action—the crime—so plausible, that judge and jury will react with ‘Aha, of course’. At this point the defence has already half lost the case. And indeed, our study of Dutch cases showed that with a very good ‘story’ the prosecution can get defendants convicted, even when solid evidence is virtually absent. It is a good ‘story’ because it readily makes sense and fits expectations. Being a good story is independent of the quality of the evidence.

In most cases, however, the prosecutor’s narrative needs anchoring in common-sense rules through chains of embedded sub-stories, i.e. pieces of evidence. Every sub-story is itself a piece of evidence that needs further anchoring, either in another sub-story or in generally accepted common-sense rules. If a court or jury accepts a piece of evidence without further sub-stories, it can only do so if it accepts one or more common-sense rules as valid and applicable to the specific case. Although such acceptance almost always remains implicit, the common-sense rules can usually be reconstructed with some knowledge of the case. If a defendant is convicted on, among others, the eye-witness testimony of his neighbour without further evidence, the following common-sense rules may be implied: ‘witnesses under oath usually do not lie; neighbours have little interest, most of the time, in the conviction of each other; neighbours usually know each other well enough to recognise each other under bad lighting conditions.

The role of the defence is threefold. First, the defence can challenge the story itself. The defence can either try to introduce information that makes the

prosecution's narrative less plausible, or it can try to come with a better, more plausible story. A solid alibi falls in the latter class. Showing that the defendant, accused of armed robbery, has a weapon-phobia falls in the first class.

The second role for the defence—anticipating that attacks on the narrative (the prosecution story) itself almost always fail—is to attack the evidence, the sub-stories. The defence must try to show that acceptance of a specific sub-story proposed by the prosecution would involve the acceptance of a common-sense rule which is silly, not commonly accepted or simply not true. In the latter, expert witnesses sometimes play a role, to demonstrate that commonly held beliefs are wrong. The defence can also try to argue that, although the (implicitly) accepted common-sense rule is true, this case falls under the exceptions. Each common-sense rule knows exceptions, but some have more than others. Identification evidence, for instance, is relatively often not valid, while we know that as a general rule the Dutch National Forensic Laboratory makes few mistakes.

The third tactic of the defence is to come up with sub-stories that falsify the prosecutor's narrative. As noted above, the way evidence is evaluated in court is most consistent with a system of verification: evidence is used to verify the narrative told by the prosecution. This is at odds with 'beyond a reasonable doubt' standard, as Allen (1991) noted. If the 'beyond a reasonable doubt' standard is taken literally, the doubt should be the focus of attention at trial and the State should suffer the burden of demonstrating that there is no plausible account consistent with innocence (see also Zuckerman, 1989, pp. 122 ff.). This would mean the prosecution should demonstrate in court that it undertook serious attempt to falsify its own narrative, and that all attempts have failed. The practice in court is, however, that the prosecution only brings forward verifying evidence and the defence is given an opportunity to bring forward falsifying information. If the defence fails to do so, it is generally seen as evidence that supports the prosecutor's narrative, probably under the general common-sense rule that it is usually easy for innocent defendants to produce exculpatory evidence.

How Decisions Go Wrong

The model we propose, and daily practice in court, does not guarantee that the decisions of the court or the jury are logically sound. At the same time it must be recognised that the rules of criminal evidence in most civilised countries *usually* produce sound decisions. Judicial errors, wrongful convictions and miscarriages of justice are more interesting to study, because these failures of the system give much more insight into how decision-making operates. A good model of decision-making must therefore provide a framework which can be used to point out where and why decisions in the more 'difficult' cases go wrong. I will discuss some rules that can be derived from the theory of anchored narratives are discussed below.

Table 6.7.3 Ten universal rules of evidence

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

The prosecutor's narrative can take quite complicated shapes, especially when the defendant is charged with a prolonged or complicated crime. It would be practically and presumably also in principle impossible for the prosecution to offer evidence for each and every element in its narrative. Legally, the prosecution is not obliged to do so. Under Anglo-Saxon law there are, generally, no minimum requirements for evidence; a single piece of evidence is enough to convict. Dutch law specifies, with some exceptions, that at least two pieces or sources of evidence are necessary to convict. This so-called *unis testis*-rule—one witness is not enough—however, does not say anything about the quality of these two pieces of evidence. Somewhere between the legal requirements and a full burden of proof for the prosecution, the level of evidence can be found which is necessary for sound decision-making.

One requirement which can be imposed on the court or jury is that each piece of evidence which it uses to convict, should be anchored in common-sense rules which are held to be true by most people. To establish whether or not the evidence used is indeed anchored adequately requires that decision-makers cite the evidence used and the anchorings chosen. Juries are required to do neither. Under Dutch law an essential element of criminal procedure is that the court must justify its sentence. In practice, courts mention the evidence which convinced them, but they never discuss the anchors used and rarely explain *why* evidence listed in the opinion supports the decision.

A further requirement that can be imposed on decision-making in criminal cases is that at least the who-what-why in the prosecution's narrative—depicted in Figure 6.7.2 as identity, *actus reus* (including causation and absence of defence) and *mens rea*—are supported with evidence. Often it is rather easy to prove two of the three, but problems surround proving the third. The prosecution, for instance, will have no problem—using the testimony of the bank tellers—in proving that a bank was robbed and that this constitutes an offence. *Mens rea*, then, can be proved with the common-sense rule that people seldom, by accident, run into a bank with a gun in their hand demanding the contents of the safe. In these kinds of cases proof that the defendant was the perpetrator usually poses the problem.

Using this example we can take the analysis one step further to set conditions on a sound decision. The bank robbery example shows that often the same evidence is used to prove more than one component of the who-what-why. In this instance, the story told by the witnesses is used both for proving *actus reus* and *mens rea*. In some cases, their story can also be used to prove identity; for instance when they described the tattooed butterfly on the robber's cheek. It can hardly be expected of the prosecutor in each case to offer separate evidence for each of the three (or five) central elements. What can be asked, however, is that the prosecutor offers separate anchorings for identity, *actus reus* and *mens rea* which must pass the 'beyond a reasonable doubt' test for each of the three separately. The quality of anchors is looked at again below.

Rules of criminal procedure have been made to facilitate decisions without errors. Most rules, however, do not touch upon the content of the evidence and how evidence should be evaluated. From the theory of anchored narratives these rules can be derived and, because they should hold for every judicial system, we called them universal rules of evidence. They are more fully explained in Wagenaar, Van Koppen and Crombag (1993, ch. 12). Here, we will look at only one procedural requirement.

Many of the rules of criminal procedure are concerned with the selection of evidence. In countries with a jury system this is accomplished with extensive rules for the admissibility of evidence. Some of these rules are not aimed at facilitating sound decision-making but are, for instance, aimed at checking police behaviour. Most, however, are designed to withhold evidence from the jury that may unduly bias them against the defendant or evidence that is too unreliable to be of any use. After the admissible and inadmissible evidence is separated, it is left to the judge's instructions and the wisdom of the jury to give a further unchecked decision.

In continental Europe another path is chosen. Since the courts decide both on the admissibility of evidence and give the verdict, a refined tradition of rules on the admissibility of evidence never emerged. Rather, all evidence is presented to the court and only afterwards does the court decide which

evidence it will or will not use in support of the decision. Of course that removes the check on the evidence to another part of the trial. Dutch judges, therefore, are obliged to render a reasoned opinion in which the evidence used is at least listed. For the same reason, both the facts of the case and the legal arguments are fully reconsidered on appeal by the Court of Appeal.

The workload of Dutch courts has recently changed this tradition dramatically and the change is soon going to be legalised by a Bill which is pending in the Dutch Parliament. Starting at the Amsterdam Trial Court the habit has grown of rendering so-called head-tail opinions. In these opinions only the formal steps taken in the procedure are listed (the head) and the sentence is given and defended (the tail). The middle part, where the evidence should be discussed, is only written if, and after, one of the parties appeals, often many months later. Anyone who has ever tried to write down an argument that seemed perfectly sound at the time knows that confiding it to paper provides a thorough check on the soundness of the argument—as is asking colleagues for comments. If such checks are missing, the carefully balanced criminal procedure may fall apart because a decision, which already has been announced publicly, will of course not be withdrawn by the court many months later if the evidence turns out to be weaker than was believed at the time of the decision. Some of the recent opinions of trial courts I have seen provide evidence which seemed rather meagre to support the decision.

Thus writing down the evidence and specifying why it supports the decision is essential for sound decision-making under an inquisitorial system. The interesting question then is whether the finely tuned rules on the admissibility of evidence under a jury system might compensate for the lack of such an essential part of criminal procedure. I doubt it.

The above does not imply that no selection of evidence is necessary. In almost every case there is at least some conflicting evidence, so courts and juries would be unable to render a verdict without at least some selection. But, they should always explain why evidence which is not used to support a decision has been ignored.

THE QUALITY OF ANCHORS

All evidence must be anchored on to common-sense rules that are safe enough to serve as anchors. In fact, the strength—that is, the diagnostic value—of evidence wholly depends on the rules chosen as anchors. As discussed above, sometimes a fair estimate of the diagnostic value of a piece of evidence can be given on the basis of empirical research. For most types of evidence that is not possible, and an impression of the diagnostic value must be obtained by scrutinising the anchors used. As noted earlier, both courts and juries leave the common-sense rules used as anchors implicit, but with some knowledge of the

case, the rules can usually be reconstructed. Such a reconstruction might produce the following rule: if evidence is anchored on to rules which are silly or simply not true the diagnostic value of the evidence can be estimated at zero. If a court, for instance, was willing to accept the testimony of a witness who says he recognised the defendant as the perpetrator, while he saw the perpetrator at night, at a distance of about 100 metres, then the court has implicitly accepted as true the general rule that recognition under these circumstances is possible. However, we know that a valid recognition in this situation is virtually impossible and thus the general rule the court accepted is silly.

In less extreme cases, the exceptions possible under the common-sense rule can give a rough—but hardly quantifiable—estimate of the diagnostic value of the evidence. It is, for instance, less safe to use evidence which is anchored onto the rule that *most* allegations of sexual abuse are true, than evidence which is anchored onto the rule that 'the Dutch National Forensic Laboratory makes *few* mistakes'.

A discussion in court of anchors rather than evidence has the advantage that it points the court or jury to the most important issue. When I appear as an expert witness in court on, for instance, line-up recognitions, the court always wants to know whether it is possible that the witness made a valid recognition. The answer is yes; but that is not the issue. The discussion should focus on the question how much risk the court takes if it accepts this recognition as evidence and that can only be derived from an explicit discussion of the general rule, which might be something like 'after witnesses saw a perpetrator under the conditions that were present in this case, identifications of the defendant by witnesses are often wrong'. A court must only use evidence which is anchored onto rules that are safe enough to pass the criterion of beyond a reasonable doubt.

Discussing anchors will also expose general rules which have been shown to be not true or which leave too much room for doubt. One such rule, for instance, is 'Sound stories are more likely to be true than weak stories' (Bennett and Feldman, 1981). Another one is 'witnesses under oath rarely lie'. That may not be true. One can consider the oath in two ways. First the oath is the title under which a witness can be prosecuted for perjury if it turns out he lied. Second, this threat can be used to deter a witness from lying. But a witness may have all sorts of reasons to lie anyway. Possibly, the likelihood of lying by a witness is *reduced* by the oath and the threat of prosecution upon perjury, but without considering all other circumstances and interests of the witness, his or her testimony should not be assumed to be true under such a rule. Courts, however, do assume this.

A third example is the rule 'confessions are usually true'. That rule may also be invalid; Gudjonsson and Sigurdsson (1993) found that an astounding 12 per cent of inmates in Iceland made a false confession. Nevertheless, English

defendants are commonly convicted on the basis of their confession alone, even when it is disputed at trial (Gudjonsson, 1992).

ALL EVIDENCE IS WITNESS EVIDENCE

There are many technical resources available to the police in investigations. However, in the vast majority of cases technical evidence plays no role whatsoever and most convictions are based solely on witness testimony. With few exceptions, we know little of the validity of witness testimony. How often do witnesses err; under what circumstances do they lie? In most cases we can only find out by thoroughly questioning witnesses and by comparing their statements with statements made by others or with other kinds of evidence. That tradition is highly developed in the USA and England and Wales, but little developed in The Netherlands. If witnesses appear in Dutch courts—and they seldom do, as most cases are decided on the statements written down by the police—a typical question is: ‘You made such a statement to the police. Do you stick to that?’ Most Dutch judges and attorneys seem to believe that a deep look into the witness’ blue eyes is enough to judge whether he or she is lying or not (see for instance Mols, 1989); or they believe that the confidence of the witness is a good indicator of whether the witness is erring (it is not; see Deffenbacher, 1980).

Only on rare occasions do Dutch courts and attorneys call on expert witnesses to provide a sub-story to support or refute the witness’ narrative. Ideally, an expert should be called in when the court or jury is unsure of the rules which underlie certain pieces of evidence, or are unsure of the exceptions the general rule allows (see Crombag, 1992). It should be realised that expert testimony is always anchored on to the reputation of the expert. This is returned to below.

It may be argued that all evidence becomes witness evidence if one probes deep enough into the anchoring structure. An example of sturdy evidence is a match of the fingerprints of the defendant with prints found on the murder weapon (there is doubt about the sturdiness of fingerprint evidence too; see Saks and Koehler, 1991). The relevant witnesses here are, for instance, a police officer who examined the scene of the crime, a bullet expert and a dactylographer (i.e. the fingerprint expert). They come into play as soon as a sub-story for the fingerprint evidence is sought and the question is asked of how we know the prints form a match. Although this question is almost never asked and fingerprint evidence is taken for granted, the method used to match the prints leaves room for some doubt (Macalister, 1989), if it is no more than that the expert making the match might have had a bad day. Whether or not the method is reliable enough can be established only by questioning an expert; whether or not the police officer who made the match knows his or her job can only be established by questioning that police officer. If one of the participants in court is not satisfied with the officer’s testimony, one or more steps can be

taken into the potentially endless regression of sub-sub-...-sub-stories. One might seek evidence to establish the quality of the police officer's training; one might examine the teachers or even the quality of the teachers, and so on for ever. At some point the court or jury has to accept somebody's word and reputation as good enough or not to use the fingerprint evidence.

Continuing the example, one might ask how we can be sure that the fingerprints indeed come from the weapon and how we can be so sure it was the *murder* weapon. We can only find out by questioning the police officer who was at the scene of the crime. Assuming that police officers have a reputation for truthfulness, we may accept the officer's testimony that bullets shot from the gun with the fingerprints matched the bullet recovered from the victim's body. This testimony will, if the court probes deeper into the structure of sub-stories, lead to the bullet expert who compared the bullets and the pathologist who cut open the corpse. Pressing on, we again arrive at a consideration of their or other people's reputations.

COMMON-SENSE RULES AND REPUTATION

The arguments advanced above support the thesis that all evidence is witness evidence; that is, as long as one probes deep enough into the structure of embedded sub-stories. The analysis also seems to imply that all evidence—again if probed deeply enough—is grounded on somebody's reputation. Before this conclusion can be drawn, a further discussion of the process of probing and anchoring is needed.

A sub-story is as much a message as any other and thus has two aspects: the message and the messenger. The diagnostic value of a sub-story can be established by anchoring the message, or by anchoring the messenger (again, either to a sub-story one level deeper or to a general rule). In general, however, the message, the testimony by a witness, cannot be evaluated without taking the credibility of the witness into account. That is not a practical constraint, but one of principle: why else question the witness? In practice, the testimony of many witnesses is accepted as valid without questioning the credibility of the witness, most of the time because the story told by the witness fits in very nicely with what the court knows about the crime from other sources. If five independent witnesses all tell more or less the same story, there is no reason to question the credibility of each individual witness. Evaluating evidence in this manner nevertheless bears a small risk. Gross (1987) reports a miscarriage of justice which occurred after 17 independent witnesses had identified the innocent defendant. Apart from this kind of exceptional cases, there seems no practical reason to probe any further if a number of witnesses confirm most of each other's stories. Then the stories are anchored onto the general rule that if independent witnesses give testimony that support one another, they almost always speak the truth. If the defence accepts anchoring to this rule and does

not challenge the independence of the witnesses, the decision is going to be a sound one. If, however, the defence wants to probe deeper, the credibility of the witnesses comes into play and the trial goes into a probing of sub-stories which always ends at somebody's reputation.

The common-sense general rules which serve as anchoring grounds in the theory on anchored narratives seem to serve practical purposes. Without them, trials would go on almost for ever and probably would raise questions about many well kept reputations.

COMPARING MODELS

Why, then, does the theory of anchored narratives form a better model of decision-making than the Bayesian model? Anchored narratives are a better vehicle for prescription for courtroom behaviour and decision-making because they are less alien to the tradition in law and to what actually happens in court than the Bayesian model. In a descriptive sense this theory is better because it predicts instances of courtroom behaviour and decision making which cannot be predicted by the Bayesian model. I refer to the extensive analysis of Dutch cases in Crombag, Van Koppen and Wagenaar, 1992; Wagenaar, Van Koppen and Crombag, 1993). Here are just a few examples.

Sometimes the defence introduces a piece of evidence which at once destroys what seemed until then an iron-clad case for the prosecution. Although the fictional Perry Mason comes up with such a defeating piece of evidence in each and every case, this is a rather rare occasion in reality. In a Bayesian model a very strong piece of evidence favouring the innocence of the defendant would not destroy the prosecution's case completely; it would just adjust the odds of guilt downward. Under a story model it can easily be demonstrated how a single piece of evidence destroys a case.

With the theory of anchored narratives it can also be demonstrated how evidence that is hardly quantifiable contributes to the decision of judge or jury. In the Bayesian model that is impossible.

Most important is that the theory of anchored narratives provides for a more complete model of decision-making and what goes on at trial. It can be used to show why prosecution and defence fail or succeed in making their case; it can show how the selection of evidence operates; it can also show why some evidence is taken for granted, and other evidence is scrutinised and requires a deep descent into anchoring structure of embedded sub-stories.

ACKNOWLEDGEMENT

Steven Penrod and David Carson gave most useful comments on an earlier draft of this chapter.

REFERENCES

- Allen, R.J. (1991). The nature of juridical proof. *Cardozo Law Review*, 13, 373–422.
- Bartlett, F.C. (1950). *Remembering: A Study in Experimental and Social Psychology* (2nd edn). Cambridge: Cambridge University Press. (1st edn, 1932.)
- Bedau, H.A. and Radelet, M.L. (1987). Miscarriages of justice in potential capital cases. *Stanford Law Review*, 40, 21–179.
- Bedau, H.A. and Radelet, M.L. (1988). The myth of infallibility: a reply to Markman and Cassell. *Stanford Law Review*, 41, 161–70.
- Bekerian, D.A. (1993). In search of the typical eye-witness. *American Psychologist*, 48, 574–6.
- Bennett, W.L. and Feldman, M.S. (1981). *Reconstructing Reality in the Courtroom*. London: Tavistock.
- Berghuis, A.C. (1992). De harde en de zachte hand: Een statistische analyse van verschillen in sanctiebeleid. *Trema*, 15, 84–93.
- Blackman, D.E., Müller, D.J. and Chapman, A.J. (1984). Perspectives in psychology and law. In D.J. Müller, D.E. Blackman and A.J. Chapman (Eds), *Psychology and Law*. New York: John Wiley.
- Borchard, E.M. (1932). *Convicting the Innocent: Sixty-Five Actual Errors of Criminal Justice*. Garden City, NJ: Doubleday.
- Brandon, R. and Davies, C. (1973). *Wrongful Imprisonment: Mistaken Convictions and their Consequences*. London: Allen & Unwin.
- Bruner, J. (1984). Narrative and paradigmatic modes of thought. Invited address, Division 1 of the American Psychological Association, Toronto, August 25, 1984.
- Cohen, L.J. (1977). *The Probable and the Provable*. Oxford: Clarendon.
- Crombag, H.F.M. (1982). Wat is rechtspsychologie? *Ars Aequi*, 32, 237–40.
- Crombag, H.F.M. (1992). Expert witnesses as vicarious anchors. *Expert Evidence*, 1, 127–31.
- Crombag, H.F.M. and Van Koppen, P.J. (1991). Praktische bezwaren: Psychologie voor juristen. In P.J. Van Koppen and H.F.M. Crombag (Eds), *De Menselijke Factor: Psychologie voor Juristen*. Arnhem, The Netherlands: Gouda Quint.
- Crombag, H.F.M., Van Koppen, P.J. and Wagenaar, W.A. (1992). *Dubieuze Zaken: De Psychologie van Strafrechtelijk Bewijs*. Amsterdam: Contact.
- Cuenta-Rua, J.C. (1981). *Judicial Methods of Interpretation in Law*. Baton Rouge: Louisiana State University, Paul M. Herbert Law Center.
- Deffenbacher, K.A. (1980). Eye-witness accuracy and confidence: can we infer anything about their relationship? *Law and Human Behavior*, 4, 243–60.
- Easteal, S., McLeod, N. and Reed, K. (1991). *DNA Profiling: Principles, Pitfalls and Potential*. Chur: Harwood.

- Edwards, W. (1988). Summing up: the Society of Bayesian Trial Lawyers. In P. Tillers and E.D. Green (Eds), *Probability and Inference in the Law of Evidence*. Dordrecht, The Netherlands: Kluwer Academic.
- Edwards, W. (1991). Influence diagrams, Bayesian imperialism, and the Collins case: an appeal to reason. *Cardozo Law Review*, **13**, 1025–79.
- Egeth, H.E. (1993). What do we not know about eyewitness identification? *American Psychologist*, **48**, 577–80.
- Finkelstein, M.O. and Fairly, W.B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review*, **83**, 489–517.
- Fletcher, H. (1992). Supplementary evidence from the National Association of Probation Officers. Paper for the Royal Commission on Criminal Justice.
- Frank, J.N. and Frank, B. (1957). *Not Guilty*. New York: Doubleday.
- Gardner, E.S. (1952). *Court of Last Resort*. New York: Sloane.
- Goldsmith, R.W. (1980). Studies of a model for evaluating judicial evidence. *Acta Psychologica*, **45**, 211–21.
- Green, D.M. and Swets, J.A. (1966). *Signal Detection Theory and Psychophysics*. New York: John Wiley.
- Gross, S.R. (1987). Loss of innocence: eyewitness identification and proof of guilt. *Journal of Legal Studies*, **16**, 395–453.
- Gudjonsson, G.H. (1992). *The Psychology of Interrogations, Confessions and Testimony*. London: John Wiley.
- Gudjonsson, G.H. and Sigurdsson, J.F. (1993). How frequently do false confessions occur? An empirical study among prison inmates. *Psychology, Crime, and Law*, **1**, 1, 21–26.
- Hale, L. (1961). *Hanged in Error*. Harmondsworth: Penguin.
- Hart, H.L.A. (1963). Introduction In C. Perelman (Ed.), *The Idea of Justice and the Problem of Argument*. London: Routledge.
- Hastie, R., Penrod, S. and Pennington, N. (1983). *Inside the Jury*. Cambridge, MA: Harvard University Press.
- Hill, P., Young, M. and Sergeant, T. (1985). *More Rough Justice*. Harmondsworth: Penguin.
- Hogarth, J. (1971). *Sentencing as a Human Process*. Toronto: University of Toronto Press.
- Homel, R.J. and Lawrence, J.A. (1992). Sentencer orientation and case details. *Law and Human Behavior*, **16**, 509–37.
- Jackson, B.S. (1988). *Law, Fact and Narrative Coherence*. Liverpool: Deborah Charles.
- Kagehiro, D.K. (1990). Defining the standard of proof in jury instructions. *Psychological Science*, **1**, 194–200.
- Kalven Jr., H. and Zeisel, H. (1966). *The American Jury*. Boston, MA: Little, Brown.
- Kaplan, J. (1968). Decision theory and the fact finding process. *Stanford Law Review*, **20**, 1065.
- Koehler, D.J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin*, **110**, 499–519.
- Langbein, J.H. (1977). *Torture and the Law of Proof: Europe and England in the Ancien Régime*. Chicago: University of Chicago Press.
- Macalister, P. (1989). From fingerprints to genetic codes: indisputable evidence? *Law Society Journal*, **27**, 43.
- Melton, G.B., Petrila, J., Poytress, Jr, N.G. and Slobogin, C. (1987). *Psychological*

- Evaluation for the Courts: A Handbook for Mental Health Professionals and Lawyers*. London: Guilford.
- Mols, G.P.M.F. (1989). *Staaende de zittng: Een beschouwing over het onmiddelijkheidsbeginsel*. Arnhem: Gouda Quint (oratie RU Limburg).
- 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. (1988). Explanantion-based decision making: effects of memory and structure on judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 521–33.
- Pennington, N. and Hastie, R. (1991). A theory of explanation-based decision-making. In G. Klein and J. Orasanu (Eds), *Decision-making in Complex Worlds*. Hillsdale, NJ: Ablex.
- Radin, E.D. (1964). *The Innocents*. New York: Dell.
- Rattner, A. (1988). Convicted but innocent: wrongful conviction and the criminal justice system. *Law and Human Behavior*, **12**, 283–93.
- Roberts, L. (1991). Fight erupts over DNA fingerprinting. *Science*, **254**, 1721–3.
- Robinson, J.A. (1981). Personal narratives reconsidered. *Journal of American Folklore*, **94**, 58–85.
- Rumelhart, D.E. (1975). Notes on a schema for stories. In D.G. Bobrow and A. Collins (Eds), *Representation and Understanding: Studies in Cognitive Science*. New York: Academic.
- Saks, M.J. (1989). Prevalence and impact of ethical problems in forensic science. *Journal of Forensic Science*, **34**, 772–93.
- Saks, M.J. and Kidd, R.F. (1980). Human information processing and adjudication: trial by heuristics. *Law and Society Review*, **15**, 123–160.
- Saks, M.J. and Koehler, J.J. (1991). What DNA 'fingerprinting' can teach the law about the rest of forensic science. *Cardozo Law Review*, **13**, 363–72.
- Scholten, P. (1974). *Mr C. Asser's Handleiding tot de Beoefening van het Nederlands Burgerlijk Recht* (3rd edn, revised by G.J. Scholten). Algemeen Deel. Zwolle, The Netherlands: Tjeenk Willink.
- Schum, D.A. (1979). A review of the case against Blaise Pascal and his heirs. *Michigan Law Review*, **77**, 446.
- Schünemann, B. (1983). Experimentelle Untersuchungen zur Reform der Hauptverhandlung in Strafsachen. In H.J. Kerner, H. Kurry and K. Sessar (hrgeb.), *Deutsche Forschungen zur Kriminalitätsentscheiden und Kriminalitätskontrolle*. Köln: Heymanns.
- Schünemann, B. and Bandilla, W. (1989). Perseverance in courtroom decisions. In H. Wegener, F. Lösel and J. Haisch (Eds), *Criminal Behavior and the Justice System: Psychological Perspectives*. New York: Springer.
- Semin, G.R. and Manstead, A.S.R. (1983). *The Accountability of Conduct: A Social Psychological Analysis*. London: Academic.
- Skinner, L.J. and Berry, K.K. (1993). Anatomically detailed dolls and the evaluation of child sexual abuse allegations: psychometric considerations. *Law and Human Behavior*, **17**, 399–421.
- Tribe, L.H. (1971). Trial by mathematics: precision and ritual in the legal process. *Harvard Law Review*, **84**, 1329–93.
- Van Dijk, T.A. (1980). *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition*. Hillsdale, NJ: Erlbaum.
- Van Koppen, P.J. and Hessing, D.J. (1988). Legal psychology or law and psychology.

- In P.J. Van Koppen and D.J. Hessing (Eds), *Lawyers on Psychology and Psychologists on Law*. Amsterdam: Swets and Zeitlinger.
- Wagenaar, W.A. (1988). *Identifying Ivan: A Case Study in Legal Psychology*. New York: Harvester Wheatsheaf.
- Wagenaar, W.A. (1991). Waar logica faalt en verhalen overtuigen: Een beschouwing over het strafrechtelijk bewijs. *Onze Alma Mater*, 45, 256-78.
- Wagenaar, W.A. and Veeffkind, N. (1992). Comparison of one-person and many-person lineups: a warning against unsafe practices. In F. Lösel, D. Bender and T. Bliesener (Eds), *Psychology and Law: International Perspectives*. Berlin: de Gruyter.
- Wagenaar, W.A., Van Koppen, P.J. and Crombag, H.F.M. (1993). *Anchored Narratives: The Psychology of Criminal Evidence*. New York: Harvester Wheatsheaf.
- Waller, Sir George (Chair) (1989). *Miscarriages of Justice*. London: Justice.
- Wells, G.L. (1993). What do we know about eyewitness identification? *American Psychologist*, 48, 553-71.
- Wigmore, J.H. (1937). *The Science of Judicial Proof as Given by Logic, Psychology and General Experience* (3rd edn). Boston: Little Brown.
- Williams, G. (1963). *The Proof of Guilt: A Study of the English Criminal Trial* (3rd edn). London: Stevens.
- Woffinden, B. (1987). *Miscarriages of Justice*. London: Hodder & Stoughton.
- Young, M. and Hill, P. (1983). *Rough Justice*. London: British Broadcasting Corporation.
- Yuille, J.C. (1993). We must study forensic eye-witnesses to know about them. *American Psychologist*, 48, 572-3.
- Zimmermann, I. (1964). *Punishment Without Crime*. New York.
- Zuckerman, A.A.S. (1989). *The Principles of Criminal Evidence*. Oxford: Clarendon.