8. The Story of Criminal Proceedings: From Fact-finding to Police Decision-making

PETER J. VAN KOPPEN

Introduction

My first study in the field of law and psychology was on decision-making by judges in civil cases, done together with Jan ten Kate (ten Kate and van Koppen, 1984; van Koppen and ten Kate, 1984). We asked a large number of Dutch judges to decide on the same nine civil disputes and had them fill out a number of personality questionnaires. They key question was how personality characteristics of the judges might influence their decisions. To reach maximum probability that we would find any relation in our study, we carefully—with the aid of a group of judges and attorneys—designed the nine cases such that the arguments in favour of the position of both the plaintiff and the defendant were equally strong. So, we could expect half of the judges to decide in favour of the plaintiff and the other half in favour of the defendant. In the study two nasty things happened. First, the decisions by the judges were not equally divided; in all nine cases more than 80 per cent chose one party and a meagre number, less than 20 per cent, the other. Second, we found no relationship between our personality characteristics and decision-making by the judges. Our conclusion was that apparently judges are socialized to such an extent that they are able to minimize differences in decision-making and thus individual differences do not play a role in the run-of-the-mill cases.

Our study on individual differences in judicial decision-making fitted nicely into the tradition that had grown up by then, especially in political science (for a recent overview, see Baum, 1997). And this tradition fitted nicely into the dramatic changes that took place in the 20th century in the

1. I wish to thank Jan de Keijser for valuable comments on an earlier draft.
views on legal decision-making. In the 19th century, Netherlands legal decision-making was modelled as a rather straightforward process. Statutes were supposed to be clear and the only task of the judge was to render the one and only right application of the law. This 19th century ideology was drawn from Montesquieu (1748-1834) who, within the division of powers, envisaged the role of the judge as nothing but being the boucle de la loi. The judge’s decision had the form of a simple syllogism. Starting from a sharp division of general rules and the facts of a case to which the rule was to be applied, the judge’s decision was supposed to be the only right conclusion from a major (rule) and a minor (facts). In the days of old, rules were rules, and facts were facts, and all were out there just to be found. Interpretation of facts or of rules was a very limited process indeed.

Now, views of legal decision-making have changed dramatically. I have always found these changes fascinating, not just because they lead to a changing role for judges and the judiciary as a whole, but also because they affect the behaviour of attorneys, prosecutors and the police and principally because they emphasize the peculiar dual nature of decisions in a legal context. Unlike many other decisions, legal decisions are both descriptive and prescriptive. Judges, for instance, try to establish what happened but at the same time they append values to these facts. These are intertwined processes, as I shall discuss below. The study of legal decision-making thus goes beyond much of the research in the field of psychology of law, where the main focus – with the exception of jury research – is on contributions to evidence, especially as far as witnesses are concerned (see for overviews Bull and Carson, 1995; Roesch, 1999; Wójcikiewicz, 2000; Traverso, 2000; Milne, 1999; Horowitz, 1998; Boros, 1998; Ainsworth, 1998; Kapardis, 1997).

I have been working in the field of criminal law for the last 15 years. In this chapter I shall discuss the changes in legal doctrine, the present state of affairs in the theories of legal decision-making, and some possible future directions. I limit the discussion to a consideration of facts in criminal cases. Not because interpretation of rules and especially the interactions of rules and facts are not of interest, but because, limiting the discussion to facts brings in the authors who form the central focus of this chapter: Reid Hastie and Nancy Pennington. Their work has been and still is a major input to the work I have done, for a large part in collaboration with Hans Crombag and Willem Albert Wagenaar.
Changes in decision-making

Nineteenth and early 20th century life seemed simple for Dutch judges. They followed a four step procedure: find out what the facts are, take the law out of the bookcase and find the relevant rule, apply the rule to the facts found, and decide what consequences this should have for the future. Now we know that these steps cannot be distinguished from each other. For instance, the relevant rule of law can only be found after interpretation of the facts and facts can only be interpreted in light of the relevant rule.

The attack on a mechanical view of legal decision-making started a little earlier in the USA than on the European continent. Holmes (1881; 1897) considered law nothing more than what a judge in fact does: 'The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law' (Holmes, 1897, p. 43). Roscoe Pound (see the discussion by Grossman, 1935) - together with Cardozo (1921) exponent of the so-called sociological jurisprudence - then gave the judge a central role in the law. The judge is, according to Pound, the really creative element in the law, the one who takes care that the law can keep the pace of modern times. This line of reasoning was continued by proponents of the legal realists as Karl Llewellyn (1960; Llewellyn and Hoebel, 1941) and Jerome Frank (1950). They asked attention for the more pragmatic aspects of legal decision-making (see for German and Dutch lawyers in this area Esser, 1970; Scholten, 1974). This has led to extremes as some lawyers consider the legal decision as a highly individual creative act by the judge (for instance Van Dunne, 1974).

This line of legal thinking fits neatly with an interest in research on social and psychological influences on legal decision-making (see for some early attempts Herzog, 1917; Schröder, 1918; Sturm, 1910) both in the Western (see for instance Hogarth, 1971; ten Kate and Van Koppen, 1984; De Keijser, 2000) and Non-Western world (Fallers, 1969; Gluckman, 1967; Von Benda-Beckmann, 1984).

Two general conclusions can be drawn from all these studies on court behaviour. First, legal decisions have a principled open character; facts and rules can be interpreted in many differing ways. Second, at the same time, interpretations of facts and rules are limited as well. Judges in bench trials are, for instance, limited in their decisions by the opinions of their colleagues and, moreover, their anticipation of what the superior court may do if one of the parties appeals. Juries, to give another example, are limited in their freedom of decision-making because they have to come to a shared decision.
Psychological models of legal decision-making

As said, 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 is derived 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 law and the legal system can be judged by the extent to which they successfully serve that purpose (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; one for which society at large should strive (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 decisions about the guilt or innocence of the suspect – the subject of the present chapter – but also in research on sentencing and in research on distributive and procedural justice. 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 Berghuis, 1992; Homel and Lawrence, 1992). The psychological study of justice considerations by individuals draws on the same mixture of prescription and description (Adams, 1965; Deutsch, 1975; Lerner, 1975, 1980; Tyler, 1994; Walster, Walster and Berscheid, 1978) I will limit myself, however, to the decision on guilt and innocence.

In each legal system the decision on guilt or innocence, for instance, is governed by a set of legal rules that prescribes how the fact-finder – court or jury – is to handle the evidence. To me, 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 that is descriptive in nature. But, the social philosophic 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 that 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 above 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. In this chapter I will discuss the models that are most widely used in the psychological study of decision-making in criminal cases. As I shall discuss below, some of these models are, in my modest opinion, too far away from how reasoning is done psychologically. It should be noted, however, that I found, for instance, the Bayesian model very enlightening in discussing individual pieces of evidence. This framework, however, never appealed to me as one that could in any way produce a model that comes near to anything a judge is, in practice, doing. The ‘story’ model comes much closer to that. It also accords much more with my experience in teaching in the law school in Rotterdam for some 12 years. Dutch law students do learn the law, but they get much more excited when they discuss Supreme Court cases. These stories are, in the Dutch legal mind, used to investigate and define the boundaries of the law and its application. Moreover, these are considered the heart of law, where law gets its shape. Supreme Court cases and the stories from these cases, however, are anomalies. After the selection process in the inferior courts these are cases in which both parties still think they have good arguments for their position. That is far away from normal legal life, where usually one of the parties – both in civil and in criminal cases – just is dead wrong. In this sense, law is the science of anomalies and it teaches its students to be more attentive to the exception than to the average.

For empirical psychologists, who are much more concerned with averages and group differences, this is a weird experience. I later found out that in teaching or lecturing for lawyers, a well chosen example is much more convincing to them than any sound or hard empirical data. Among psychologists carefully chosen examples are considered cheap argumentation tricks; among lawyers it appeals to the heart of their trade. That very mechanism makes the presentation of evidence by the prosecution from the case file, which is not more than well chosen examples of what might have happened, so dangerous. How these dangers can be tackled, is the subject of the rest of my chapter.
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 suspect.

The possibility of such a process is debatable both on philosophical and on practical grounds, because 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 Cueta-Rua, 1981, p. 133 ff.; Scholten, 1974, p. 11 ff.). At the same time, lawyers often 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 less problematic. It seems that 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 are 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 needs not to be final but can take the shape of a working hypothesis. In criminal cases 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 (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 to be actually 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. Again, this is a gruesome way of reasoning for psychologists, but to a large extent it seems to be inevitable in legal proceedings.

Hypothesis testing

If decision-making in criminal cases is not a logical bottom-up process, one might turn to Hart’s conclusion that courtroom decision-making comes closer to rhetoric than to reason (1961, p. vii). 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 us a mathematical formula for this process; one that has become quite popular in modelling decision-making in criminal cases (Edwards, 1988; Finkelstein and Fairly, 1970; Goldsmith, 1980; Saks and Kidd, 1980; Schum, 1994; Tribe, 1971). This model has never appealed to me. It is a fine and elegant model when used to structure scientific decision-making, but is in my opinion too far from how real people reason. It does not accord with the mixed prescriptive and descriptive nature of law, since it assumes the ‘evidence’ used as input is value free. I simply do not like this model, but let me first explain it and then give some more fundamental objections to it.

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, \( P_{\text{true}} \), by the probability that it is false, \( P_{\text{false}} \). The prior belief or the prior odds, then, are expressed by:

\[
\text{Prior odds} = \frac{P_{\text{true}}}{P_{\text{false}}}
\]
For instance, when $P_{\text{true}}$ is .80, and $P_{\text{false}}$ is .20 (the hypothesis is either true or false), then the prior odds are $0.80/0.20 = 4.0$. 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 a piece of evidence:

$$\text{Posterior odds} = \text{prior odds} \times D,$$

After this piece of evidence the prior odds – whatever they were – are adjusted to become posterior odds that 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:

$$\text{Posterior odds} = \text{prior odds} \times D \times D_2 \times \Lambda \times D_3 \times \Lambda \times D_n$$

If the final posterior odds surpass a preset level of confidence, the defendant can be convicted. This Bayesian approach thus seems an elegant model of decision-making. I will argue, however, that modelling decision-making in criminal cases as hypothesis testing is the right solution to the wrong problem. I will discuss 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).

The presumption of innocence assumes that the defendant is innocent until proven guilty.\(^2\) 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 always zero – and the Bayesian model would be useless. A solution might be to keep the initial

---

\(^2\) I am not quite sure if this sentence is correct. I have participated in a lot of criminal trials, not as a suspect, but as an expert witness. My experience is that a court can either start from a high prior odd of guilt or from a low prior odd and that does not make much of a difference for the end result. Having a trial starts from the assumption that the suspect may be guilty as hell. In fact, questioning the suspect in court is only possible if the court assumes guilt, at least during the interrogation. To me it seems that the presumption of innocence is much more relevant for the manner in which a suspect is treated prior to his conviction than as an element of judicial decision-making.
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 the Swedish psychologist 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. However, in real life, as I know from the courts, the initial probability of guilt can be of decisive importance to the final decision. This was demonstrated by Schünemann and his co-workers (Schünemann, 1983; Schünemann and Bandilla, 1989). 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, who concluded that any task ‘that prompts a person to temporarily accept the truth of a hypothesis will increase his or her confidence in that hypothesis’ (Koehler, 1991, p. 502).

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 identification line-up in which a witness recognises the suspect (see Cutler and Penrod, 1995; Wells et al., 1998). If a line-up 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 exceeds 15 (compare the formulas above, see 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). 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.

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 or she was in prison at the time of a crime that was committed elsewhere – can be compensated by a large number of fairly weak pieces of evidence against him or her. 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).

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, or 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 can be given, as in the case of a capital crime? One could argue that the fact-finder should take fewer risks of a wrongful conviction in the latter case. Should a court or jury in a case of multiple rapes put the criterion on a lower more risky level, since 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.

Taken together, the problems faced 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 from 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 as well. That does not mean that models of hypothesis testing might not be useful to discuss certain individual 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. Thus, these models are the right solution to the wrong problem.

The story model

For the reasons discussed above, the hypothesis-testing model of legal decision-making seems a less plausible model. Rather, a story model seems to come closer to what actually happens in the courtroom. ‘The criminal trial is organized around story telling’, Bennett and Feldman (1981) wrote in a book that set the tone for my research in decision-making in criminal cases: Reconstructing Reality in the Courtroom. Story models are based on the notion that stories give meaning to behaviour of individuals, or as Bruner (1984) puts it, are ‘concerned with the explication of human intentions in the context of action.’

Scientists in many disciplines have tried to establish what makes a story believable. Rumelhart (1975), Robinson (1981), and Van Dijk (1980) all designed story grammars; sets of rules to which a well-formed story must obey. The story grammar proposed by Bennett and Feldman (1981) was designed specifically for judicial contexts (comp. 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. In short these properties were: (1) A readily identifiable central action and (2) 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; nothing sticks out on its own. 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. Authors of detective fiction will recognize this, as I have described elsewhere (van Koppen, 1991).

Nancy Pennington and Reid Hastie extended the analysis of what makes stories plausible in a series of subsequent publications (Hastie and Pennington, 1991; Pennington and Hastie, 1986, 1988, 1991, 1992, 1993a, 1993b, 1993c). Indeed, using a story model was not new. Their approach was new, however, because for the first time a rigorous application was given to decision-making in a legal context, which resulted in a remarkable precision of claims and hypotheses. Pennington and Hastie focus on jury decision-making but there is no reason why judges or panels of judges would behave differently, as far as interpretation of the facts of a case is concerned.

Pennington and Hastie propose that the central cognitive process in jury decision-making is based on the construction of a story. They distinguish three component processes: (1) evidence evaluation through story construction, (2) representation of the decision alternatives by learning verdict category attributes, and (3) reaching a decision through the classification of the story into the best fitting verdict category (an overview of their theory and experiments is given in Pennington and Hastie, 1993b). Their main thesis is that the story constructed by a juror determines the juror’s decision. Their story model is based on the hypothesis that jurors impose a narrative story organisation on trial information. In the story three types of knowledge are used: (1) case-specific information acquired during the trial, (2) knowledge about events similar in content to those that are the topic of dispute, and (3) generic expectations about what makes a complete story. This three-step activity results in an interpretation of the evidence in narrative form. As a consequence, different jurors will construct different stories of the same criminal case. These differences arise from differences in world knowledge among jurors.

In the whole process, the structure of the story is of paramount importance. Without going into too much detail, it can be said that, according to Pennington and Hastie, all actions are explained in good stories by factors
Decision-making of three kinds: physical conditions, psychological conditions, and goals. Gaps in the story are filled with inferences made by the jurors themselves. In one study, for instance, they found that of the references in protocols made by their subject-jurors, 55 per cent consisted of information that was actually included in the trial presented to them, while an amazing 45 per cent referred to inferred events, consisting of actions, mental states and goals that filled in the story (Pennington and Hastie, 1981).

Of course, Pennington and Hastie acknowledge that jurors can entertain several competing stories before a decision is reached. The processes by which the story that in the end is considered the 'best' story, is called certainty principles. They propose three of these principles: coverage, coherence and uniqueness. Coverage refers to the extent to which a story accounts for the evidence presented at trial. Coherence is considered to be based on three principles: consistency, plausibility and completeness of the story. Finally, uniqueness, or better lack of uniqueness, refers to the extent to which different competing stories are judged to be coherent.

The second component of jury decision-making consists of representation of the decision alternatives by learning verdict category attributes. Pennington and Hastie propose that this is a quite difficult one-time learning task, usually centred around the instruction to the jury. Since this part is typical for jury systems, I will not discuss it here. But, please note that in the discussion of anchored narratives below, this component returns in a different fashion, namely as the anchoring of evidence. The third component is reaching a decision through the classification of the story into the best fitting verdict category. Pennington and Hastie give, I think, a rather optimistic view of how juries handle this component. Pennington and Hastie, for instance, argue that jurors will only accept a story if its quality surpasses the 'beyond a reasonable doubt' criterion, assuming that juries entertain such a rigorous decision criterion. I am quite sure they are wrong here. I must admit I am biased because I do not like the jury system for many reasons, one being the list of described miscarriages of justice in countries with a jury system (see for instance Bedau and Radelet, 1987; Belloni and Hodgson, 2000; Borchard, 1932, 1970; Brandon and Davies, 1973; Callaghan, 1994; Carrington, Dever, Hogg, Bargen, and Lohrey, 1991; Chamberlain, 1990; Crispin, 1987; Dennis, 1993; du Cann, 1960; Engelmayer and Wagman, 1985; Floriot, 1972; Folsom, 1994; Frank and Frank, 1957, 1971; Frasca, 1968; Gardner, 1952; Greer, 1994; Gross, 1987, 1996, 1998; Hale, 1961; Hill, Young and Sergeant, 1985; Huff and Rattner, 1988; Huff, Rattner and Sagarin, 1986; Huff, Rattner and Sagarin, 1996; Karp and Rosner, 1991; Kee, 1986; Mullin, 1989; Nobles and
The most fundamental claim in the theory of Pennington and Hastie is that the story construction, with all its falters and flaws causes the decision by the jury. Pennington and Hastie's story grammar, and their thesis about the importance of stories, was strongly supported by their research. I will not, here, go over all their empirical studies which cover more than a decade, but shall only present what I consider the most telling support for their theory. In an earlier study, Pennington demonstrated that story structures differed systematically for jurors who chose different verdicts (Pennington, 1981). That left open the question whether these story constructions were post hoc adaptations to the decisions reached, or that indeed story structure caused the decisions. This question was settled in a study in which Pennington and Hastie (1986) showed that the order in which evidence is presented has a major influence on the judgement. In their study, both the prosecution and the defence could present their evidence in witness order (i.e. a rather random order) or in story order. The combination of these two variables results in four groups. The dependent variable was the answer to the question whether the subjects thought that the defendant was guilty of first-degree murder. Their results are presented in Table 8.1.

Table 8.1 Effects of presentation of evidence at trial: Percentages of subjects choosing a verdict of guilty of murder by prosecution and defence order condition (source Table 4 in Pennington and Hastie, 1988, p. 529)

<table>
<thead>
<tr>
<th>Prosecution evidence</th>
<th>Defence evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Story order</td>
</tr>
<tr>
<td>Story order</td>
<td>59</td>
</tr>
<tr>
<td>Witness order</td>
<td>31</td>
</tr>
</tbody>
</table>
The data indicate that the party which presents the evidence in story-order was believed more readily, even though the evidence itself was exactly the same in all four conditions. The effect can be as large as changing a 31 per cent chance of conviction into a 78 per cent chance. Clever presentation of the story is half of the work. What is the other half? We tried to answer that question (Crombag, Van Koppen and Wagenaar, 1992; Wagenaar et al., 1993).

**Anchored narratives**

The theory of *Anchored Narratives* (Crombag et al., 1992; Wagenaar et al., 1993), discussed here, is strongly influenced by the work done by Pennington and Hastie. Hans Crombag, Willem Albert Wagenaar and I did not start our work on anchored narratives with the intention to contribute to the theory of judicial decision-making. We had started with the intention to write a draft for a Dutch Code of Criminal Evidence. Dutch rules on evidence are very scarce and the reason is simple: in the Netherlands the judge is both gatekeeper for the admissibility of evidence and decides on all issues in a trial. Thus, having rules on evidence serves no purpose. The few Dutch rules therefore are all on how the court can use evidence in its decisions. We felt that the draft code should at least reckon with the problematic cases, where the evidence can be of the essence. Being empiricists, we wanted to start by reading a number of difficult cases. So, we asked a number of attorneys for files of cases in which, in their opinion, their client had been convicted on flawed or absent evidence.

We were shocked by the cases we received. The most notorious example is a case in which the suspect was convicted of sexual abuse on the basis of a police report of an interrogation by a mother who related that her daughter had told her that she had heard from her eight-month-old sister that she also had been abused. For those readers who are not parents: eight-month-old children do not talk at a level that they can discuss sexual abuse. These cases changed our project dramatically to answering the question: how is it possible that a court would convict with such flawed or little evidence? We started with the idea that the work of the fact-finder—judge or jury—consists of determining the plausibility of the stories presented by the prosecution and by the defence. Part of this plausibility is found in the quality of a story as described by Pennington and Hastie.

The most important extension we gave to their model is that the story, as presented by the prosecution, needs evidence. How evidence is treated is the
anchoring part of anchored narratives. Supporting the story is done with general rules. If a fact-finder accepts a story or part of a story, it always involves the acceptance of a general rule, which is part of the fact-finder’s knowledge of the world. If, for instance, a witness statement is accepted as evidence, this can only be done if the rule is accepted that witnesses tend to speak the truth.

Such general rules, however, are seldom true without exception. Witnesses sometimes err or lie. The rules that make evidence prove something should more accurately be phrased: witnesses speak the truth most of the time; or pathologists almost never make mistakes, to give another example. The possibility of exceptions to rules means that, on a particular occasion, we must show that a possible exception does not apply. This argument, for instance, may call on the general rule that it is very improbable that two lying eyewitnesses 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. These general rules are usually common-sense facts of life. We often accept an argument because we unwittingly believe the underlying rule that gives it an anchor, even though an explicit formulation of the rule would cause us to protest or even reject it. Cohen (1977, p. 247) 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.’ 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 8.1.

At the top of Figure 8.1 there is the story of the original indictment, of which the soundness, in terms of Pennington and Hastie, has already been judged satisfactory. 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. At three, detailed evidence is offered, but each piece of evidence forms a sub-story in itself, which needs an anchor in the form of further evidence, and which, in turn, forms a sub-sub-story in need of an anchor in the form of further evidence, and which, in turn, forms a sub-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.

**Figure 8.1 Model of Theory of Anchored Narratives (after Crombag et al., 1992)**

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. 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. 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 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).

The Theory of Anchored Narratives is both descriptive and prescriptive, and that is the element I like about it most. This model of decision-making in criminal cases 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).

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 only will react with ‘Aha, of course’. At this point the defence has already lost half the case. And indeed, our study of Dutch cases showed that with a very good story the prosecution could get defendants convicted, even when solid evidence is virtually absent. It is a good story because it readily makes sense and fits expectations. We argue that 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). 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 the basis of, among other things, fingerprints at the scene of the crime, the following common-sense rules may implied:

- If someone’s fingerprints are found on the scene of the crime, he almost always has been there;
- Technical detectives make small mistakes in securing and comparing fingerprints; and,
- The technique of comparing fingerprints produces small mistakes (in fact this rules is debatable, see Evett and Williams, 1996; Stoney, 1997).

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...
The second role for the defence – anticipating that attacks on the narrative (the prosecution’s 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 given by witnesses, for instance, is fairly 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. The way evidence is evaluated in court, however, 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, p. 122 ff). 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.

The model we proposed, 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 recognized that the rules of criminal evidence in most civilized 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 that can be used to point out where and why decisions in the more ‘difficult’ cases go wrong.
Moulding criminal stories

The work by Pennington and Hastie and our work on anchored narratives take a story as presented by the prosecution as a starting point. Until now I have said little on how this story came about, and indeed the theories have little to say on where the prosecution got its story from or how the police moulds all the data they gathered during an investigation into a story presented at trial.

Some important differences between the police investigation and evidence at trial are obvious. When a criminal offence comes to the notice of the police, their first objective is to find a suspect. In most cases, however, it is not the police, but members of the public who report that a crime has been committed and who point out a suspect to the police (Black, 1970; Bottomley and Coleman, 1976; Erickson, 1981; Greenwood, Chaiken and Petersilia, 1977; Steer, 1980). Professional investigators, who must abide by the rules governing these investigations, perform the gathering of information following such notifications.

Another distinction between investigation and proof is concerned with the level of certainty required for taking decisions. At various stages in criminal procedure, the evidence available must amount to a certain level of certainty to warrant a decision. For instance, 'probable cause', which is necessary before arrest and for searches and seizure are permitted, may require 40 to 50 per cent certainty according to some authorities (Melton, Petrila, Poythress and Slobogin, 1997, p. 27).

Wagenaar, Van Koppen and Crombag (1993) have demonstrated that the roots of many dubious convictions can be found in the police investigation, even in the early stages of investigation (see also, for instance, Belloni and Hodgson, 2000; Gross, 1996; Nobles and Schiff, 1995; Radelet et al., 1992). I shall now briefly discuss some of the mechanisms we identified, namely offence-driven and suspect-driven searches, the problem of verification and falsification, and the so-called trawler method.

One important distinction is between offence-driven and suspect-driven searches. This is related to the starting point of the investigation. In an offence-driven search the starting point is the crime and the facts related to the crime. The identity of the culprit, then, is inferred from these facts. In a suspect-driven search someone becomes a suspect for no clear reason, or at least no reason that is explained by the known facts of the crime. Only then is an attempt made at finding evidence which links this particular suspect to the crime. Such a search is limited right from the start. An example of the latter is showing photographs of known criminals to a witness (Wells et al., 1998). The relevance of the
distinction between offence-driven and suspect-driven investigations lies in the diagnostic value of the resulting evidence. In an offence-driven search the narrative is the product of an inferential process, based on information. In a suspect-driven search the narrative is the starting point, and the information is its product. In offence-driven search one collects so much information that the search logically excludes all possible alternative suspects. In suspect-driven search one needs only enough information to make the suspect look bad. It can even be argued that one may take any citizen, investigate him thoroughly, and connect him to one of the many unsolved crimes in the police files. This may be done by way of recognition tests, an accusation by another suspect, attributing a motive, identifying some piece of intimate knowledge, forensic analysis of traces, a report by a psychiatrist, the absence of an alibi, or even a confession obtained under prolonged interrogation. Especially the latter in the form of prolonged police interrogations seems prone to produce bad evidence (Gudjonsson, 1992; Van Koppen, 1998).

In many of the cases discussed by Wagenaar, Van Koppen and Crombag (1993) the suspects became suspects because of their criminal records. Gross (1987) reported that in 60 per cent of 92 miscarriages of justice in which a suspect was incorrectly identified by eyewitnesses, the first suspicion was based on outer appearance, while nothing else related the suspect to the crime. This 60 per cent is an extremely high figure, compared to Steer’s (1980, Table 4:2 at p. 97) finding that in of all crimes 21 per cent of the suspects are connected to crimes through suspect-driven searches. Suspect-driven searches appear to promote unsafe convictions.

The second mechanism is related to the distinction between verification and falsification. Logically, hypotheses are tested by two complementary processes; verification and falsification. An attempt of verification means looking for facts that are predicted by the hypothesis; falsification means looking for facts that are excluded by the hypothesis. To test a hypothesis both processes are necessary. Falsification is not some sort of luxury, in which one only engages when there is an excess of means. As long as alternative hypotheses are not excluded, they may be more likely than the verified hypothesis.

An uncommon example may clarify the distinction between verification and falsification (described by Rossen, 1992, p. 9 ff; Rossen and Schuijer, 1992, p. 437). Some years ago, a twelve-year-old boy called Patrick living in the Dutch town Enschede was accused of sexually abusing about 200 children. The case had started with an accusation made by a single child, but had grown rapidly after the police solicited for more cases in the neighbourhood where the
boy lived. The boy confessed to all cases. One may wonder how a twelve-year-old boy can find the time to abuse 200 children in a time-span of less than one-and-a-half years. Patrick was interviewed as follows: the interviewing policeman showed him a photograph of a child in the neighbourhood and asked whether or not the boy also abused this child. To almost all the pictures presented, Patrick said ‘yes’; a verification of the accusation. The police never tried a simple falsification by, for example, showing the boy pictures of children from other towns. The police afterwards said it was not necessary because the boy had a photographic memory.

An attempt at falsification rarely occurs in criminal investigations. As a rule the prosecution limits its activities to verification attempts. In the perspective of research logic this practice is absurd, but it is predicted by the theory of anchored narratives. Anchors are only verifications of the hypotheses included in the indictment’s narrative. Falsifications, unless they are definite, do not destroy such anchors, the anchoring structure simply has no place for them. If falsification attempts occur at all, they are most often the initiative of the defence. But even falsifications offered by the defence are rare, probably because the defence realises that falsifications have no place in the anchoring structure, and may therefore not affect the court’s decision process; they may easily be considered irrelevant.

Particularly deceptive results may be obtained by means of the third method by which the police investigation may generate miscarriages of justice, which is called the trawling method. This method falls into the class of suspect-driven search methods. The term is used for police investigations that start from a generalised and little specified suspicion against a person or, more often, a group of persons, in which a large police force is engaged, investigating every conceivable detail of his or its behaviour until some sort of mischief is discovered. It is like trawling a very wide net in a place where there may not be many fish, but if the net is wide enough one always may be expected to catch some fish.

A less conspicuous but equally misleading set of trawling methods is related to identification procedures. One of these methods is to show a suspect’s picture in a nationwide broadcast (such as Opsporing Verzocht in The Netherlands, and Crime Watch in Great Britain), with the question ‘Who saw this person?’ Usually this elicits many responses. It is quite likely that at least one of these will be incriminating for the person shown. Such trawling methods capitalize on chance and are therefore to be avoided if at all possible. While trawling for evidence, one can find enough incriminating facts against virtually anybody.
The story of the police

The examples of problems in the police investigation I have discussed above all focus on how the investigation may cause wrongful convictions at the trial stage. The construction of proof during the trial will always be suspect-driven, because there is a defendant present right from the beginning. So the question to be answered by the court is not ‘Who did it?’, but ‘Did he do it?’. In other words, the objective of the trial is not to uncover the truth, but to evaluate the believability of the indictment’s narrative by testing the quality of the available evidence. All parties involved in criminal proceedings, with the possible exception of the defence, have a strong preference for suspect-driven search and verification, both in the investigative stage and during the construction of proof. As a consequence investigation and construction of proof, although essentially different, may become indistinguishable, especially after the investigation did not involve discovery of a suspect, but the construction of sufficient proof against a known suspect. In these cases the investigators become judges, although they were never meant to be so. Judges are doomed to repeat what the investigators have already done, allowing innocent defendants little hope that the outcome of the trial will be anything else than a predictable confirmation of the indictment. Thus, the police have the best opportunities to prevent miscarriages of justice.

My discussion of some characteristics of the police investigation is based on ‘looking back’ from what happens at trial in difficult cases. We now know a little on how problems in the generation of a story during police investigations may influence what happens at trial. This, however, provides for little information on how the police generate a story during the investigation. We may suspect that the decision that turns an ordinary citizen into a suspect of a criminal offence is taken on the basis of a narrative, in the same manner as at a later stage the decision to convict is to a large extent taken on the basis of a narrative and that any of the problems in court decision-making also apply to decisions of the police.

There are, however, important differences between the story of the police and the story at trial. At an early stage of their investigation the police can probably only work from an incomplete narrative, mostly based on what was found at the scene of the crime. Some elements of the narrative are available from the start, found by accident, by reports from citizens, through combining and verifying facts, or mere guesswork. Other parts of the narrative only emerge during the investigation. From this incomplete story the investigators may work their way up; ambiguities or contradictions are further investigated or put aside.
as irrelevant. Gradually a coherent and complete story emerges. In the investigation it is hoped that the narrative emerges from the facts.

All these elements, and probably many more, make police work essentially different from the work of fact-finders. How precisely the police build their story of a crime is a virtually un-researched subject (an exception is Binder and Bergman, 1984). Therefore, I consider a thorough study of police decision-making the most important challenge for future research on how stories form criminal proceedings.

References


Decision-making


