P.J. van Koppen (2013) Spoiled judges and the fluidity of evidence in criminal cases, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law, Rome, June 10-14, 2013* (pp. xxii-xxiv). New York: Association for Computing Machinery.

Fourteenth International Conference on

Artificial Intelligence and Law

Rome, June 10 - 14, 2013
Consiglio Nazionale delle Ricerche
(National Research Council of Italy)

Proceedings of the Conference

Sponsored by:
International Association for Artificial Intelligence and Law
ITTIG Consiglio Nazionale delle Ricerche
Artificial Intelligence, Elsevier
e-Codex
OASIS
Springer

In cooperation with:
Association for the Advancement of Artificial Intelligence (AAAI)
ACM SIGART

Under the auspices of:
the Senate of the Italian Republic
CNR - Dipartimento Scienze Umane e Sociali - Patrimonio Culturale

Spoiled Judges and the Fluidity of Evidence in Criminal Cases

Peter J. van Koppen
Maastricht University and Free University Amsterdam

The introduction of DNA changed much in the perception of evidence in criminal cases. Judges, attorneys, prosecutors and police officers were used to having witnesses deliver most of the evidence. Okay, sometimes there was technical forensic evidence, such as fingerprint evidence. The latter was considered infallible; which it, in fact, is not. Apart from fingerprints, evidence was always discussed in terms of 'strong evidence' or 'weak evidence' or 'circumstantial evidence'.

DNA-evidence has changed all this. Now, experts dare to give numerical values to their evidence. They speak of random match probabilities: the chance that by sheer accident an individual has the same DNA-profile as the sample from the crime scene. They come up with enticingly small values like 1 in a billion, in cases where a full DNA-profile is available from the crime scene that matches the profile of the suspect.

The numerical values of DNA-evidence have led to three dramatic changes in the field of criminal law. First, judges and prosecutors and attorneys now expect expert witnesses to produce numbers. I particularly notice such if I am teaching judges and prosecutors together with colleagues from the forensic laboratory. After their class on DNA-evidence, judges start to whine about number for all sorts of evidence. They suddenly expect me to give an exact diagnostic value or likelihood ratio for, for instance, a positive identification in a line-up. I have a hard time explaining to them how silly that is. Of course we could produce some number for a particular line-up as we study them in the psychological laboratory. Line-ups in police practise tend to differ markedly from those in our labs. When the police line-ups differ dramatically from what is desirable, calculations might become of interest. But also: if the police make all kinds of errors, these fine laboratory studies are not very good models for what is happening and are useless for estimating a diagnostic value.

A second dramatic recent change is that judges have become spoiled by these small numbers like 1 in a billion, meaning that the evidence is extremely strong. Since they are lawyers who are not used to calculations, they misperceive numbers. The following fine case can illustrate this. In a place near Rotterdam two men robbed an Aldi Supermarket. Something rare happened: shots were fired and the supermarket owner died. The shooter fled the scene of the crime and his accomplice was apprehended shortly after. He told the police his robbery buddy was Barry and that Barry was the shooter. Barry was arrested but denied all allegations. Apart from the witness statement of the other robber, proof against him was a shoe the shooting robber had left at the scene of the crime. The shoe was Barry's size. But, amazingly, the laboratory did not succeed in securing a DNA-profile from the shoe. There was something else: the

supermarket owner had started a fight such that blood of the robber was found on his nails. The laboratory found a partial DNA-profile in the blood that matched the DNA of Barry. The DNA-expert estimated the rarity of the DNA-profile with a random match probability of 1 in 100,000. Now, assume that evidence with this property would be the only evidence against the defendant in 100,000 different cases. And assume that a judge would use such a single piece of evidence to convict each of these 100,000 robbers. That would mean that the judge would commit a miscarriage of justice in roughly 1 out of 100,000 cases. To me, that seems quite an acceptable error rate.

Yes, I know some of you would accuse me that I now commit the so-called prosecutor's fallacy. And indeed, I do, but that only is a problem if you are a Bayesian. Bayesianism, however, is a religion that, if applied to evidence in criminal cases may cause more problems than it solves. That is another discussion. But, even Bayesians should admit that although DNA evidence with a random match probability of 1 to 100,000 is not as strong as DNA evidence can be, it is still quite strong, nevertheless. But in the Aldi case it was not strong enough for the judges: the defendant was wrongly acquitted both by the trial court in Rotterdam and the appellate court in The Hague. It is fair to say that both courts misunderstood the strength of the evidence, mainly because they are so spoiled by DNA experts presenting them with silly small random match probabilities. They are so used to evidence that is presented as very strong that, if such evidence is not present, they become unsure and tend to acquit, rather than taking the chance of making an error, a miscarriage of justice.

In reality all these numbers are nonsense. At best a number given to the strength of DNA evidence is a shot in the dark, but most of the time just rubbish. There are so many assumptions that must hold before such a number may be true, that these numbers are almost always meaningless. One must assume that the database that is used to estimate the strength is in all respects representative for the specific DNA-comparison at hand. One must assume that no mistakes have been made in the whole chain from the crime scene to the final report by the DNA-expert. One must also assume that the DNA-comparison is simple, between a full profile from the defendant to a full profile found in a crime scene sample. In practice, however, the samples found at the crime scene almost always are mixtures from two or more individuals. With these mixtures calculations soon become quite difficult, often too difficult to do anyway.

A third consequence of the apparent calculability of evidence is that some people think that it is possible to model evidence in a manner that would lead to a meaningful analysis of evidence that in turn can be used to decide the case. I am sorry to say that is a misconception. I should admit that I secretly know that such models in the form of evidence charts are much older that DNA profiles. But admitting that would spoil my argument.

My argument begins with the following. The evidence structure of criminal cases is very simple. In most cases that come to court it is evident that the defendant committed the crime. There is abundant evidence and the defendant confessed to the crime or both. Whatever system one uses to evaluate the evidence and to decide the case, it does not matter. The outcome is always the same: guilty. There may be all kinds of legal problems and the punishment may be a point of discussion, but the question of

guilt does not pose a problem. One could even leave the decision to laymen like a lay jury.

Only a relatively small proportion of the criminal cases are difficult from an evidentiary point of view. For these cases it may matter what method is used to evaluate the evidence. Would then modelling like evidence charts be the right option? Let us follow for a moment the work of the New Evidence Scholarship, represented by people like Terry Anderson, David Schum and William Twining, work that is rooted in the work by Wigmore, and that has recently inspired researchers in the field of AI & Law. They propose a plan by which it can be decided whether the ultimate probandum has been proven. They start out by making a key list, a list of all propositions in a case, all elements that are relevant. Apart from the simplest case where, we just ascertained, no special decision method is necessary, such a key list is almost always very extensive, too extensive to be viable. At that very point, a lot of choices have to be made to limit the extent of the discussion on the case. That limiting is done at every stage of the development of the case. Suppose, for instance, that in the room where you are now reading this piece, a murdered person is found and suppose you are the technical detective investigating the crime scene. Where would you take samples? Would you, for instance, dust all the pages of all books in the bookcase for fingerprints? Would you take DNA-samples from all surfaces in the room? And what would you do with the other rooms in the house? From the very beginning of the case choices are made and the choices are based on scenarios the people involved have, implicitly of explicitly. If a key list is made, the key list is guided and limited by what we think might be relevant; and what we think that might be relevant again depends on what we expect to be relevant scenarios.

After the key list, charts are made. A simple problem will produce a simple chart. A more complicated problem will not just produce a more complicated chart, but very soon a chart that is prohibitively complicated. Especially cases where an analysis using charts might be useful will produce charts that are so complicated that they are not useful any more. And, again, a chart represents a particular scenario of a limited set of scenarios for the case. In reality there are often all types of scenarios that may be more or less relevant for an analysis of the case. Also, charts are not suitable for complicated pieces of evidence, as for instance pieces of evidence that can mean different things, depending on the context in which they are found.

In short, charts are useful for an analysis of local problems in evidence evaluation. But they will never succeed in modelling the evidence in a complete case in such a way that it is essential for making a good decision.