

Bewijskracht 10, volle vaart recht
vooruit

Bewijskracht 10, volle vaart recht vooruit

Rede

in verkorte vorm uitgesproken bij de aanvaarding van het ambt van
bijzonder hoogleraar Forensische Statistiek
aan de Universiteit van Amsterdam
op vrijdag 4 november 2011

door

Marjan Sjerps

Dit is oratie 417, verschenen in de oratiereeks van de Universiteit van Amsterdam.

Opmaak: JAPES, Amsterdam

© Universiteit van Amsterdam, 2012

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voorzover het maken van kopieën uit deze uitgave is toegestaan op grond van artikel 16B Auteurswet 1912 j° het Besluit van 20 juni 1974, St.b. 351, zoals gewijzigd bij het Besluit van 23 augustus 1985, St.b. 471 en artikel 17 Auteurswet 1912, dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 882, 1180 AW Amstelveen). Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16 Auteurswet 1912) dient men zich tot de uitgever te wenden.

*Mevrouw de rector magnificus,
Mijnheer de decaan,
Geachte leden van het curatorium,
Geachte leden van het bestuur van de Stichting Leerstoel Criminalistiek,
Dames en heren,*

De forensische statistiek

In de New Scientist van 5 september dit jaar stond een artikel met de alarmerende titel (vrij vertaald): ‘DNA super-netwerk vergroot risico van warboel’ (Geddes 2011). In het artikel wordt beschreven hoe de totaal onschuldige Engelse barman Peter Hamkin in 2003 gearresteerd werd en twintig dagen ten onrechte verdacht werd van moord op een Italiaanse vrouw. Deze verdenking ontstond na een zoekactie in de Engelse DNA databank, waarbij het DNA profiel van Hamkin bleek te matchen met het DNA-profiel van een spoor dat gevonden was op de plaats van het misdrijf. Het artikel beschrijft vervolgens hoe onder het verdrag van Prüm alle Europese DNA databestanden met elkaar zullen worden gekoppeld, waardoor dagelijks miljoenen vergelijkingen worden gemaakt tussen DNA-profielen van sporen en mensen uit de verschillende landen. Het is niet te voorkomen dat dit proces, net als in de Hamkin zaak, een aantal “toevals” matches oplevert, met name wanneer niet het volledige DNA profiel kan worden vergeleken, waardoor onschuldige personen ten onrechte kunnen worden beschuldigd van een ernstig misdrijf. Het artikel besluit met de wijze woorden (vrij vertaald): “Het hele gebied van waarborgen rond het uitleveringsproces en het gebruik van deze informatie in de rechtszaal is cruciaal.”

Het artikel slaat met deze woorden de spijker op zijn kop. De DNA-databestanden zijn goud waard in de opsporing, en hebben dat in de dagelijkse praktijk al talloze malen bewezen, waaronder in Nederland geruchtmakende zaken als de Puttense moordzaak of de moord op Andrea Luten. Ook de internationale zoekacties hebben hun nut nu al ruimschoots bewezen, zoals u kunt lezen in het jaarverslag van de Nederlandse DNA-databank (van der Beek 2011). Maar de Hamkin zaak maakt duidelijk dat de rechter toch niet helemaal blind kan varen op een DNA-match. Hoe kan hij dan verstandig gebruik maken van

dit soort informatie, zoals in het artikel wordt bepleit? Moet een DNA-match anders gewaardeerd worden als we hem vinden via een databank-zoekactie? De intuïtie schiet tekort als we dit soort vragen willen beantwoorden. Gelukkig is het antwoord te vinden in de literatuur van een vakgebied dat ‘forensische statistiek’ heet.

Nee, dat vakgebied heeft niets te maken met optimalisatie van reizigersstromen, zoals het woord ‘forens’ ten onrechte doet vermoeden en zoals men mij regelmatig vraagt. Het gebied houdt zich bezig met de interpretatie van forensisch bewijsmateriaal. Dit bewijsmateriaal is afkomstig van sporen die worden verzameld in relatie met een delict, en dat gebruikt wordt in de opsporing en in de rechtszaal. DNA-profielen, vingerafdrukken en schoensporen zijn bekende voorbeelden. Forensic science is bij het grote publiek bekend uit populaire televisieseries als ‘CSI (crime scene investigation)’. Wat mij altijd opvalt in deze series is dat het interessantste deel van het verhaal altijd wordt overgeslagen. Wanneer bijvoorbeeld de vezels van de zwarte wollen trui van de verdachte treffende overeenkomsten vertonen met zwarte wollen vezels die op het rode katoenen t-shirt van een slachtoffer gevonden worden, dan roepen de CSI acteurs triomfantelijk uit: ‘we have a match!’ Daarmee is de kous dan af, het bewijs is geleverd en de verdachte is erbij. Maar voor mij begint het interessante deel van het verhaal juist hier. Wat kunnen die zwarte vezels eigenlijk bewijzen, met andere woorden, tussen welke hypothesen kunnen zij onderscheid maken? Wat is de bewijskracht dan? Hoe kunnen we die bewijskracht op een objectieve manier bepalen? Wat zeggen de aantallen vezels en de plaats waarop zij gevonden zijn over de aard van een eventueel contact? Zou je eigenlijk geen rode katoenen vezels verwachten te vinden op de zwarte trui? Hoe kunnen we de bewijskracht van de zwarte wol combineren met de bewijskracht van de afwezigheid van het rode katoen? En hoe schrijven we onze conclusies op een begrijpelijke maar wetenschappelijk correcte en volledige manier op in het rapport?

Hierover worden geen tv-series gemaakt helaas. Maar om als bewijs te worden gebruikt in de rechtszaal hebben we in feite helemaal niets aan het enkele feit dat er een match is vastgesteld tussen vezels, DNA-profielen, vuilniszakpatronen of wat voor sporen dan ook. Zo’n match wordt namelijk pas interessant als we weten hoe zeldzaam hij is. Hoe groot is de kans dat een andere vezel, persoon, of vuilniszak ook een match oplevert? Pas als we die kans kunnen inschatten is de match mogelijk bruikbaar als bewijsmiddel. Dit geldt ook in bredere zin. Wat is een bekentenis waard als we niet weten hoe groot de kans is dat de verdachte bekent als hij onschuldig is? Wat is de ‘herkenning’ van de dader door een getuige waard wanneer hem een foto van de verdachte wordt getoond, als we niet weten hoe groot de kans is dat hij de verdachte ‘herkent’

als die onschuldig is? Helemaal niets, is het antwoord, dat iedereen na enig logisch nadenken zal geven. Waarnemingen aan objecten of personen vormen slechts informatie. Deze informatie kan pas bewijs vormen door toepassing van statistiek en kansrekening. Met geavanceerde technologie kan men tot op tien cijfers achter de komma nauwkeurig meten, en getuigen en verdachten kunnen van alles verklaren, maar als bewijs is dit alles nutteloos als de statistiek ontbreekt. Kortom: *zonder statistiek geen bewijs*.

De forensische statistiek bestudeert de stap van waarneming naar bewijs dat relevant kan zijn in een rechtszaak (Finkelstein and Fairley 1970, Robertson and Vignaux 1995, Evett 1998, Aitken and Taroni 2004, Sjerps en Coster van Voorhout 2005, Curran 2009, Berger en Aben 2010). Daarbij komen een aantal fundamentele vragen op zoals: wat is eigenlijk relevant bewijs? Volgens regel 401 van de 'Federal Rules of Evidence' (Federal Evidence Review 2011) is dat: 'evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence'. Ofwel, informatie die de waarschijnlijkheid beïnvloedt van een gebeurtenis die van belang is voor het bepalen van de actie. Door beïnvloeding van waarschijnlijkheid als criterium te nemen belanden we op het terrein van de kansrekening en statistiek, de forensische statistiek in dit geval.

Een voorbeeld kan dit wellicht verduidelijken. Stel dat een moeder en een meisje zich bij de politie melden met het verhaal dat het meisje is aangevallen door een gemaskerde man die plotseling uit de bosjes sprong. De man heeft het meisje met een scherp voorwerp over haar buik gekrast, en ging er daarna meteen vandoor. Het meisje is zo geschrokken dat zij geen enkele beschrijving van de dader of het voorwerp kan geven. De dader heeft volgens haar niets gezegd en er zijn ook geen getuigen. De politie moet in zo'n geval nagaan of er inderdaad sprake is van een misdrijf. Het komt namelijk ook voor dat mensen zichzelf verwonden. Misschien heeft het meisje dat ook gedaan maar wilde zij dit niet toegeven aan haar moeder, en heeft zij als smoes voor de verwonding het hele verhaal uit haar duim gezogen. De verwondingen zouden misschien zelfs kunnen zijn ontstaan door een ongelukje met een rits of iets dergelijks. Stel dat de verwondingen van dien aard zijn dat een ongelukje kan worden uitgesloten. Een gebeurtenis waarvan de waarschijnlijkheid in dit geval van belang is, is dat de krassen op de buik van het meisje door een andere persoon zijn gemaakt. Een forensisch geneeskundige kan in dit geval de verwondingen onderzoeken met in het achterhoofd twee mogelijkheden: het meisje heeft zichzelf verwond, of iemand anders heeft dat gedaan. Bij zijn onderzoek doet hij een aantal waarnemingen zoals de aard van de verwondingen en de positie op het lichaam. Hij kan bijvoorbeeld waarnemen dat de letsels op anatomisch

bereikbare locaties zitten, dat het gegroepeerde, oppervlakkige krasletsels zijn in een parallelle rangschikking, die samen een uniform letselbeeld vormen. Er is één litteken van een vergelijkbaar letsel.

Om tot een conclusie te komen kan hij nagaan hoe goed zijn waarnemingen passen bij de twee mogelijkheden. Hoe groot is de kans om dit soort verwondingen te zien als het meisje zichzelf verwondde, en hoe groot is deze kans als iemand anders dat deed? Wanneer er hierover betrouwbare gegevens voorhanden zijn, kan hij beide kansen schatten. Bijvoorbeeld, stel dat bij zelfverwonding de kans op zijn waarnemingen één op tien is, en bij verwonding door iemand anders één op tienduizend, dan is de kans op de waarnemingen dus duizend keer groter bij zelfverwonding dan bij verwonding door iemand anders. Zijn conclusie zou kunnen luiden: 'Het is veel waarschijnlijker om deze verwondingen waar te nemen als het meisje zichzelf heeft verwond, dan als iemand anders haar heeft verwond'. Het zou ook kunnen zijn dat hij waarnemingen doet waarbij hij inschat dat de kans op deze waarnemingen bij zelfverwonding honderdduizend keer groter is dan bij verwonding door iemand anders, of maar drie keer groter. Dit getal, het quotiënt van twee kansen dus, noemen we het aannemelijkheidsquotiënt, in het Engels de *likelihood ratio*, vaak afgekort als *LR*.

Maar de oorspronkelijke vraag was hoe de informatie over de aard en positie van de verwondingen de waarschijnlijkheid beïnvloedt dat het meisje verwond is door iemand anders. Deze waarschijnlijkheid hangt ook nog af van andere informatie in de zaak. De andere informatie kan bijvoorbeeld zijn dat haar beste vriendin vertelt dat zij het meisje betrapt heeft terwijl zij zichzelf aan het verwonden was. Maar de andere informatie kan ook zijn dat een week ervoor een heel ander meisje aangifte deed met exact hetzelfde verhaal als dit meisje: een gemaskerde man die haar plotseling overvalt en haar buik bekrast met een scherp voorwerp. In het eerste geval is zelfverwonding a priori veel waarschijnlijker dan in het tweede geval. Om de kans op zelfverwonding te bepalen moet deze a priori waarschijnlijkheid gecombineerd worden met de conclusie van de deskundige.

Op dit punt komt de forensische statistiek in beeld. Er is namelijk een eenvoudige wiskundige formule die weergeeft hoe die twee logisch gecombineerd kunnen worden. Deze formule is gebaseerd op de regel van Bayes, en is vernoemd naar de Engelse dominee Thomas Bayes, die leefde in de 18^e eeuw en die kansrekening als hobby had. De formule heeft drie ingrediënten: hypothesen, bewijsmateriaal en achtergrond informatie. Zij geeft weer hoe de kansverhouding van twee hypothesen verandert door nieuw bewijsmateriaal te beschouwen, dit alles gegeven de achtergrondinformatie. In het voorbeeld zijn de twee hypothesen die we beschouwen zelfverwonding (hypothese 1) en verwon-

ding door iemand anders (hypothese 2). Het bewijsmateriaal zijn de aard en positie van de verwondingen en het litteken. Vóórdat we het bewijsmateriaal beschouwen hebben de hypothesen een bepaalde waarschijnlijkheid. De verhouding van deze waarschijnlijkheden noemen we de *a priori kansverhouding*. We zagen dat deze afhangt van de achtergrondinformatie: in de situatie dat een vriendin het meisje betrapte is zelfverwonding a priori veel waarschijnlijker dan in de situatie van eenzelfde tweede aangifte door een ander meisje. Wanneer de deskundige concludeert dat de aard en positie van de verwondingen beter past bij hypothese 1 dan bij hypothese 2, verandert de kansverhouding van de hypothesen. We noemen hem nu de *a posteriori kansverhouding*, de verhouding ná het beschouwen van het bewijsmateriaal. De *regel van Bayes* geeft de relatie weer tussen de kansverhouding van de hypothesen vóór en ná het beschouwen van het bewijsmateriaal:

$$a \text{ priori kansverhouding} \times LR = a \text{ posteriori kansverhouding}$$

De vermenigvuldigingsfactor LR is heel belangrijk. Deze factor is ook een kansverhouding:

$$LR = \frac{\text{kans op waarnemen van bewijsmateriaal als hypothese 1 waar is}}{\text{kans op waarnemen van bewijsmateriaal als hypothese 2 waar is}}$$

De LR geeft weer hoe goed het bewijsmateriaal past bij de ene hypothese ten opzichte van de andere. Als het bewijsmateriaal even goed past bij hypothese 1 als bij hypothese 2, dan is de kans in de teller even groot als de kans in de noemer. De LR is dan gelijk aan één. Als het bewijsmateriaal beter past bij hypothese 1 dan bij hypothese 2, dan is de LR groter dan één: hoe beter het bewijsmateriaal past bij hypothese 1 vergeleken met hypothese 2, hoe groter de LR. Andersom geldt, als het bewijsmateriaal beter past bij hypothese 2 dan bij hypothese 1, dan is de LR kleiner dan één: hoe beter het bewijsmateriaal past bij hypothese 2 vergeleken met hypothese 1, hoe kleiner de LR.

Aan de regel van Bayes zien we dat bewijsmateriaal met een LR van 1 niets verandert aan de kansverhouding van de hypothesen: het is neutraal bewijsmateriaal dat niets toevoegt. Bewijsmateriaal met een LR van 1000 verandert wel iets: hierdoor wordt de kansverhouding tussen hypothese 1 en 2 een factor 1000 keer zo groot. Dit materiaal vergroot dus de waarschijnlijkheid van hypothese 1 ten opzichte van die van hypothese 2 en vormt daarom bewijs ten gunste van hypothese 1. We zien ook dat hoe groter de LR, hoe meer de waarschijnlijkheid van hypothese 1 vergroot wordt ten opzichte van die van hypo-

these 2, ofwel hoe sterker het bewijs ten gunste van hypothese 1. Bewijsmateriaal met een LR van 1/1000 verandert ook iets: hierdoor wordt de kansverhouding tussen hypothese 1 en 2 een factor 1000 keer zo klein. Dit materiaal vormt bewijs ten gunste van hypothese 2: hoe kleiner de LR, hoe sterker het bewijs ten gunste van hypothese 2.

Uit de regel van Bayes volgt dus een belangrijke interpretatie van de LR: *de LR is een maat voor bewijskracht!* Net zoals we kunnen meten hoe hard de wind waait (bijvoorbeeld 10 op de schaal van Beaufort) kunnen we meten hoe sterk het bewijs is (bijvoorbeeld 10 op de schaal van LR). Vaak ontbreken de gegevens om een goede numerieke schatting te maken van de LR. In die gevallen kan de deskundige meestal wel op grond van zijn kennis en ervaring een LR in verbale vorm rapporteren, zoals in het voorbeeld van de forensisch geneeskundige. We zien nu ook meteen wat er bedoeld wordt met de definitie van relevant bewijs in regel 401 van de 'Federal Rules of Evidence' ('evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence.'). Relevant bewijs is bewijs met een LR groter of kleiner dan 1. Deze vermenging van wiskunde en strafrecht is ongelooflijk mooi, omdat het op een simpele manier de fundamentele blootlegt van strafrechtelijk bewijs. Dat levert een krachtig instrument op voor de evaluatie van bewijs, zoals ik zal proberen te laten zien in het vervolg van mijn rede.

Uit de regel van Bayes volgt een tweede belangrijke conclusie: als je wil weten hoe waarschijnlijk twee hypothesen zijn op basis van bewijsmateriaal, de a posteriori kansverhouding dus, dan moet je zowel de LR van het bewijsmateriaal als de a priori kansverhouding van de hypothesen weten. Als het gaat om deskundigenbewijs dan is de deskundige degene die de LR kan inschatten. Zo zagen we in ons voorbeeld dat de forensisch geneeskundige kan bepalen hoe goed de aard en positie van de verwondingen en het litteken passen bij de hypothesen zelfverwonding en verwonding door een ander. De deskundige heeft echter vaak geen weet van de a priori kansverhouding. Sterker nog, hij weet hier bij voorkeur niets van, omdat dit zijn oordeel over de LR oneigenlijk kan beïnvloeden. Bovendien ligt de beoordeling van de a priori kansen op de hypothesen per definitie buiten zijn deskundigheidsgebied. Hier ligt een schone taak voor de politie of de jurist (ik zal het voortaan kortweg over de jurist hebben). In ons voorbeeld is het niet aan de forensisch geneeskundige maar aan hem om de betrouwbaarheid van de verklaring van de vriendin of van de tweede aangifte te bepalen. Een goede deskundige wil dit soort informatie niet eens hebben omdat hij hierdoor onbewust beïnvloed kan raken. Mijn teamleider Reinoud Stoel zorgt er zelfs voor dat die ongewenste informatie de deskundige niet eens kan bereiken.

De deskundige kan dus wel de LR schatten, maar niet de a priori kansverhouding. Dit betekent logischerwijze dat hij geen conclusie kan trekken over de a posteriori kansverhouding van de hypothesen. In ons voorbeeld betekent dit dat de deskundige geen uitspraak kan doen over de kans dat het meisje zichzelf heeft verwond. Alleen de politie of de jurist kan zich hierover een oordeel vormen door de a priori verhouding van de kansen op zelfverwonding en verwonding door een ander te vermenigvuldigen met de LR van de forensisch geneeskundige. Dat vinden veel mensen een verrassende en moeilijk te accepteren conclusie. Zij verwachten van de forensisch geneeskundige dat hij zelfstandig een oordeel geeft over de kans dat het meisje zichzelf verwondde. De regel van Bayes toont echter aan dat hij dat alleen kan doen door aannamen te doen over het overige bewijs en omstandigheden, zaken dus die niet op zijn deskundigheidsterrein liggen. Schoenmaker blijf bij je leest, is het juridische devies voor deskundigen, en de regel van Bayes laat precies zien wat die leest dan is: de LR (in numerieke of verbale vorm).

De regel van Bayes is een basisregel uit de kansrekening. Elementaire wiskunde dus, logisch tot op het bot. De toepassing daarvan in het forensische domein leidt zoals boven beschreven tot de conclusie dat je bewijskracht kunt meten met de LR, dat het de taak is van de deskundige om die te rapporteren en de taak van de jurist om daarmee zijn geschatte a priori kansverhouding bij te stellen. Deze conclusie heeft verstrekkende gevolgen gehad, waaronder de geboorte van mijn vakgebied, de forensische statistiek. Het Bayesiaanse paradigma blijkt namelijk heel breed toepasbaar, op gebieden variërend van DNA-analyse tot schoensporenonderzoek. De wiskundige aanpak wekt bij sommigen de indruk dat het doel is om een complexe rechtszaak te reduceren tot een rekensommetje. Dat is uiteraard niet het geval. *Het Bayesiaanse raamwerk biedt een logische en gestructureerde denkwijze om bewijs te evalueren en te combineren met het overige bewijs.* Het wordt daarom wel gezien als het verbindende element tussen de sterk uiteenlopende forensische deskundigheidsgebieden, datgene wat een losse verzameling toegepaste technieken tot een ‘forensische wetenschap’ smeedt. De kernvragen zijn daarbij telkens:

- Wat zijn de twee hypothesen die we willen beschouwen?
- Hoe waarschijnlijk zijn zij a priori? (het quotiënt van de kansen is de a priori kansverhouding)
- Welke waarnemingen vormen het bewijs?
- Hoe waarschijnlijk is het bewijs onder de twee hypothesen? (het quotiënt van de kansen is de LR)
- Hoe waarschijnlijk zijn de hypothesen a posteriori, ná beschouwing van het bewijs? Dit volgt uit de antwoorden op voorgaande vragen en de regel van Bayes.

Waarom is de forensische statistiek zo mooi?

1. Bayesiaanse raamwerk is leidraad om gestructureerd bewijskracht te bepalen

Het Bayesiaanse raamwerk biedt dus een manier om de kracht van het bewijs ten aanzien van twee hypothesen uit te drukken, in een getal als er voldoende gegevens zijn en anders verbaal. Forensisch DNA-onderzoek is een gebied waar voldoende gegevens voorhanden zijn om numerieke LR's te rapporteren. Laten we eens kijken hoe dat werkt in de praktijk van het forensisch DNA-onderzoek (Kaye 2010). We beginnen met een simpel geval, een DNA-match tussen een DNA-spoor en een verdachte, Jan. Beiden hebben DNA-profiel X. We nemen aan dat er geen fouten worden gemaakt in de keten en dat de effecten van verwantschap te verwaarlozen zijn. Laten we als hypothesen kiezen:

- H1: het DNA-spoor is van Jan
- H2: het DNA-spoor is van een onbekende persoon

Het bewijs definiëren we als:

- E: het DNA-spoor en Jan hebben beide DNA-profiel X.

Nu is eenvoudig af te leiden dat de LR gelijk is aan $1/\text{frequentie van profiel X}$ in de populatie personen die mogelijk het spoor hebben gemaakt. Dit simpele geval komt overeen met de intuïtie: hoe zeldzamer het DNA-profiel, hoe groter de LR, dus hoe sterker het bewijs. Voor 'gewone' DNA-sporen kun je op deze manier aantonen dat de bewijskracht groter is dan 1 miljard. Ofwel, de a priori kansverhouding van de hypothesen wordt door de DNA-match een miljard keer vergroot. Was het vóór de DNA-match bijvoorbeeld even waarschijnlijk dat het DNA van Jan is als dat het van een onbekende persoon is, dan is het na de DNA match een miljard keer waarschijnlijker dat het van Jan is dan van een onbekende.

Op dezelfde wijze kun je in veel ingewikkelder situaties de bewijskracht numeriek weergeven (Buckleton et al. 2005, Balding 2005, Evett and Weir 1998). Zo zijn er modellen ontwikkeld waarbij we wel rekening houden met verwantschap, met de mogelijkheid van een fout in de keten, mutaties, en artefacten van de analyse. Ook zijn LR-formules afgeleid voor 'moeilijke' DNA-sporen, zoals mengsels van DNA van verschillende personen of DNA dat half opgegeten is door bacteriën. En is het spoor eigenlijk wel van de dader? Welk databestand moet je gebruiken als de verdachte een kind is van een Touareg vader en een Girbi moeder, zoals eens ter verdediging werd aangevoerd? In dit soort

ingewikkelde situaties is de intuïtie niet meer toereikend om de bewijskracht te bepalen. Sterker nog, de intuïtie kan ons helemaal om de tuin leiden. Die situaties hebben mij altijd geïnteresseerd. Een voorbeeld daarvan is de situatie waarin de verdachte op een speciale manier is geselecteerd.

2. Bayesiaanse raamwerk is leidraad voor het omgaan met selectie-effecten

Stel dat het DNA-spoor een speekselspoor is in een bivakmuts die in de gang wordt gevonden bij een gewelddadige overval op een bejaard echtpaar. In de bivakmuts worden meerdere speekselsporen aangetroffen en een paar haren. Dat levert telkens hetzelfde profiel op, ook bij een contra-analyse in een ander laboratorium. Het lijkt redelijk om hier aan te nemen dat het om het DNA-profiel van de dader gaat dat zonder fouten is bepaald.

Hiervoor heb ik gesproken over een verdachte Jan. Maar ik heb daarbij niet gezegd waarom Jan eigenlijk onze verdachte is. Het kan zijn dat er ander bewijs is, bijvoorbeeld Jan is net na de overval door zijn vriend Piet gezien terwijl hij hard wegliep van de woning van het echtpaar. In deze situatie is a priori de kans dat het DNA-spoor van Jan is al relatief groot. Nemen we vervolgens de DNA-match in beschouwing, dan kunnen we er vrijwel zeker van zijn dat het DNA-spoor inderdaad van Jan is. In de literatuur wordt dit de ‘probable cause’ situatie genoemd (Balding and Donnelly 1996, Dawid and Mortera 1996). Maar tegenwoordig kan het ook zijn dat Jan op een hele andere manier verdachte werd in de zaak. Met grote regelmaat worden verdachten namelijk ‘gevonden’ via een zoekactie in de DNA-databank. Deze situatie wordt de ‘database search’ situatie genoemd. Stel dat Jan verdachte werd in de zaak toen het profiel van het speekselspoor vergeleken werd met de DNA-databank, en zijn profiel het enige was dat matchte. Zijn profiel zat bijvoorbeeld in de databank wegens pedofilie. De match wordt doorgegeven en de politie gaat op zoek naar aanvullend bewijs voor of tegen Jan. Er kunnen nu twee situaties ontstaan:

- A. Er wordt ander bewijs gevonden tegen Jan, bijvoorbeeld, vriend Piet meldt zich en zeg dat hij Jan net na de overval heeft gezien terwijl hij hard wegliep van de woning van het echtpaar;
- B. Er wordt geen ander bewijs gevonden tegen Jan, of juist bewijs dat in zijn voordeel pleit. Bijvoorbeeld, vriend Kees meldt zich en zegt dat Jan bij hem was ten tijde van de overval.

In het geval A hebben we dezelfde bewijslast als in de ‘probable cause’ situatie aan het begin. Het enige verschil is de volgorde waarin we het bewijs vinden, eerst de DNA-match en daarna de herkenning door Piet, of andersom. Dat

maakt voor de bewijskracht natuurlijk niet uit. Door Piet's herkenning is voordat we de DNA match beschouwen de kans al relatief groot dat het speeksel-spoor van Jan is. Na het beschouwen van de match zijn we er daarom vrij zeker van dat het speeksel-spoor van Jan is. In wiskundige termen: een grote prior keer een grote LR levert een grote posterior.

In het geval B hebben we een veel lagere bewijslast. Door het alibi van vriend Kees is voordat we de DNA-match beschouwen de kans relatief klein dat het speeksel-spoor van Jan is. Na het beschouwen van de match zijn we er daarom misschien niet zo heel zeker van dat het speeksel-spoor van Jan is. Het spoor zou ook door puur toeval kunnen matchen. Dat hangt af van de zeldzaamheid van het profiel X en van de betrouwbaarheid van Kees. In wiskundige termen: een kleine prior keer een grote LR levert een grote of kleine posterior, het hangt er maar net van af hoe klein de prior is en hoe groot de LR. Voor een 'gewone' match is de LR een miljard, enorm groot dus. Kees moet dan wel zéér betrouwbaar zijn om dat via de a priori kans te kunnen compenseren. Maar als het een 'moeilijk' DNA-spoor is, bijvoorbeeld als het half opgegeten is door bacteriën, dan is de LR niet zo groot en is een match door puur toeval een reële mogelijkheid. Deze situatie, waarin we behalve de DNA-match verder geen bewijs tegen Jan hebben of alleen bewijs dat in zijn voordeel pleit, noem ik de 'glad ijs' situatie. Deze kan alleen optreden bij een zoekactie, zoals een databank zoekactie of een grootschalig DNA-bevolkingsonderzoek. In de 'probable cause' situatie is er op dit punt geen reden tot zorg, er is dan immers altijd ander bewijs tegen Jan, zijn DNA-profiel wordt niet voor niets vergeleken. In de 'glad ijs' situatie ligt echter een gerechtelijke dwaling op de loer, en daar maak ik mij zorgen over. Ik ben daarin niet de enige, zie bijvoorbeeld een interview met de oervader van het forensisch DNA, Sir Alec Jeffreys (Lowe 2010). Peter Hamkin verkeerde in deze situatie, de Engelse barman die ten onrechte werd beschuldigd van moord op een Italiaanse vrouw. (Hij werd trouwens ook weer vrijgepleit door aanvullend DNA-onderzoek.)

Hoe voorkomen we nu deze situatie in de praktijk? Nationale en internationale zoekacties in de DNA-databanken zijn van grote waarde voor de opsporing. Er worden regelmatig misdrijven mee opgelost die anders niet of minder snel zouden worden opgelost. Tegelijkertijd genereren deze zoekacties heel af en toe een match met een persoon die het spoor niet maakte, maar toevallig wel hetzelfde profiel heeft als het spoor. Het zal dan meestal gaan om een zoekactie met een 'moeilijk' profiel, niet om een volledig profiel van de huidige generatie DNA profielen die extreem zeldzaam zijn. Dit is helaas niet te voorkomen als we ook met die 'moeilijke' profielen willen zoeken. Ook al komt een profiel bijvoorbeeld slechts bij 1 op de miljoen mensen voor, dan verwachten we puur op grond van toeval een match te vinden als we via een databank

zoekactie profielen van miljoenen mensen vergelijken met dit spoor. In die gevallen verwachten we echter ook dat er geen ander bewijs tegen de matchende persoon gevonden kan worden, of hij moet wel een enorme pechvogel zijn.

Het is dus zaak dat politie en justitie bij databankmatches goed naar de overige informatie in de zaak kijken. Hier ligt een ware uitdaging. Immers, de gemiddelde politiemann of jurist denkt bij een DNA-match eerder aan een rondevaak dan aan de vraag of hij misschien moet corrigeren voor een mogelijk kleine a priori kans. Ronald Meester, Ate Kloosterman, Kees van der Beek, Klaas Slooten en ik hebben het rapporteren van DNA-zoekacties bestudeerd. Wij trekken nu al jaren ten strijde met de boodschap dat de DNA-onderzoeker een waarschuwing in zijn rapport moet opnemen om politie en juristen te attenderen op de ‘glad ijs’ situatie (Meester and Sjerps 2003, 2004, Sjerps, Kloosterman en van der Beek 2010, Slooten en Meester 2011). Het Nederlands Forensisch Instituut (NFI) en de beheerder van de Nederlandse databank nemen die waarschuwing inmiddels op in hun rapporten en ook in een Europese richtlijn (ENFSI DNA working group 2010). Bij buitenlandse rapporten ontbreekt een dergelijke waarschuwing echter vaak.

De bewijskracht van een databank match heeft voor veel discussie gezorgd in de literatuur. Hierbij zien we een aantal drogredenen steeds weer opkomen in verschillende variaties. Die drogredenen komen er meestal op neer dat een match die gevonden wordt via een DNA databank zoekactie niet zoveel waard is als een match die gevonden wordt in de ‘probable cause’ situatie (NRC 1996, Stockmarr 1999, Schneider et al. 2010, van Koppen 2011). Maar zoals we zagen is het niet de waarde van de match die verschilt, maar de mogelijkheid dat het overige bewijs tegen de matchende persoon ontbreekt (Balding and Donnelly 1996). De conclusie uit dit inzicht moet vervolgens niet zijn dat we een databank match hetzelfde kunnen rapporteren als een ‘probable cause’ match (Buckleton et al 2005, Dawid 2001), maar juist dat een waarschuwing in het rapport moet worden opgenomen voor de ‘glad ijs’ situatie.

Het interessante is dat het zoeken in DNA-databanken een voorbeeld is van een algemene situatie die veel vaker voorkomt (Sjerps and Meester 2009). Ik beschrijf die situatie als het selecteren van een item uit enorme hoeveelheden persoonskenmerken, objecten of gegevens volgens een bepaald criterium, waarbij dat item de focus wordt van een vervolgonderzoek, en waarbij hetzelfde selectiecriteria als bewijs wordt gebruikt. Binnen het forensische domein wordt bijvoorbeeld naar matchende vingers gezocht in enorme vingerafdruk-databanken of naar matchende kogelhulzen in hulsbodembdatabanken. Politie en veiligheidsdiensten maken in toenemende mate gebruik van het zoeken naar matchende gezichten in grote hoeveelheden camerabeelden van bijvoorbeeld uitgaanspubliek. Bij huiszoekingen wordt bijvoorbeeld gezocht naar mat-

chende textielvezels, waarbij grote aantallen textielen voorwerpen worden vergeleken, en waarbij vervolgens de matchende vezel als bewijs in de rechtszaak wordt gebruikt. Managers zoeken naar medewerkers die opvallend veel fouten maken, en terroristen worden gezocht door het koppelen van allerlei bestanden en het zoeken naar opvallende patronen. Kortom, dames en heren, wij leven in het tijdperk waarin dit soort grootschalige zoekacties continue worden uitgevoerd, en waarbij we met enige regelmaat met de ‘glad ijs’ situatie te maken zullen krijgen. In mijn ogen zijn de verpleegkundige Lucia de B, die opvallend vaak aanwezig leek te zijn bij een medisch incident, en de Amerikaan Brandon Mayfield, wiens vingerafdruk opvallend leek op die van de bommenlegger in Madrid, voorbeelden van personen die door dit gladde ijs ten onrechte in zware juridische moeilijkheden kwamen. Wij hebben er dus allemaal belang bij om de resultaten van de zoekacties goed op waarde te kunnen schatten, en behalve aan een ronde zaak ook onmiddellijk te denken aan glad ijs.

3. Bewijskracht meten: objectivering van bewijs

De kracht van het Bayesiaanse raamwerk is dat het algemeen toepasbaar is. Er worden momenteel behalve voor het DNA-onderzoek numerieke LR-modellen ontwikkeld voor chemisch georiënteerd forensisch onderzoek zoals verdovende middelenonderzoek, benzinevergelijking, schotrestenonderzoek en vergelijkend glasonderzoek. Maar ook de LR-modellen voor andere gebieden zoals spraak en vingerafdrukken vorderen in rap tempo. Mijn collega's, met name Ivo Alberink en Annabel Bolck schreven mooie artikelen over de berekening (bijvoorbeeld Bolck et al. 2009, Bolck en Alberink 2011) en Didier Meuwly over de methodologie (Meuwly 2006). Het aantal collega's binnen en buiten het NFI dat hiernaar onderzoek doet wordt snel groter. Hierbij komen een groot aantal smulvragen op voor statistici. Sommige daarvan zijn van fundamentele aard, zoals: bestaat er zoiets als een ‘ware’ LR? Heeft het zin om te spreken over een betrouwbaarheidsinterval voor de LR? Hoe schatten we een extreem lage kansdichtheid in een gebied waar geen data meer zijn? De ontwikkelde LR-modellen zijn leidend voor de verzameling van gegevens, zoals frequenties van kenmerken in bepaalde populaties, het onderzoeken van afhankelijkheden etc. Dit geeft ook een impuls aan het ontwikkelen van objectieve meetmethoden, bijvoorbeeld het meten van kleur in vezelonderzoek. Het uiteindelijke doel voor het forensisch onderzoek is om op basis van deze harde gegevens en LR-modellen de bewijskracht uit te drukken in een getal, met een bekende onzekerheid. In de nabije toekomst rapporteert de chemicus dus bijvoorbeeld: ‘Ik heb de elementsamenstelling vergeleken van de referentiemonsters van de gebroken vitrineruit en de glasfragmentjes die zijn aangetroffen in

de kleding van de verdachte. De waargenomen elementsamenstellingen zijn minstens honderdduizend keer waarschijnlijker als de glasfragmentjes afkomstig zijn van de gebroken ruit dan als zij afkomstig zijn van een onbekende andere glasbron.’

Het streven naar het rapporteren van een goed onderbouwd getal als maat voor bewijskracht noemt men ‘objectivering van bewijs’. Dit vereist objectieve waarnemingsmethoden, gegevensbestanden en experimenten, LR-kansmodellen, en validatie van het geheel. Uiteindelijk levert dit een belangrijke verbetering ten opzichte van de huidige praktijk in veel forensische gebieden waarin de deskundige op basis van zijn ervaring een subjectief oordeel rapporteert. Een Amerikaanse wetenschappelijke onderzoekscommissie luidde onlangs met recht de alarmbel over deze werkwijze in het forensisch onderzoek in de Verenigde Staten (NRC 2009). Vanuit wetenschappelijk oogpunt, maar zeker ook vanuit een juridisch oogpunt, is objectivering daarom een belangrijk doel. De rapportages worden hierdoor minder afhankelijk van de deskundige die het oordeel geeft. Het wordt ook duidelijker waar de deskundige zijn oordeel op baseert, en hoe hij is omgegaan met verschillende bronnen van onzekerheid. Ofwel, objectiever, transparanter, en beter onderbouwd. Dit klinkt als een reclameboodschap, maar het is precies waar objectivering om draait, en waarom het terecht bovenaan de onderzoeksagenda staat van onder andere het NFI.

4. Bayesiaanse raamwerk is leidraad voor het combineren van bewijs

4a. Conditioneel onafhankelijk bewijs, zelfde hypothesen

Stel dat de vraag is of een stuk zilverkleurige tape van 5 cm breed waarmee een slachtoffer werd geboeid afkomstig is van de rol tape uit de keukenla van de verdachte. Voor de vergelijking van het stukje tape kunnen we een aantal onderzoeken doen, zoals chemisch onderzoek aan de lijmlaag, de dragerlaag en de eventuele tussenlaag, vormonderzoek om te kijken of de rand van het stukje tape qua vorm past bij het uiteinde van de rol tape en om de opbouw van de laag textiel te vergelijken. Deze drie onderzoeken zouden dan hetzelfde hypothesepaar kunnen beschouwen, bijvoorbeeld:

H1: het stuk tape is afkomstig van de taperol

H2: het stuk tape is afkomstig van een andere zilverkleurige taperol van 5 cm breed.

Stel nu dat de resultaten van het chemisch onderzoek en het vormonderzoek aan de uiteinden en aan de textielopbouw een LR hebben van respectievelijk circa 100, 1000, en 10. Wat is dan de bewijswaarde van de combinatie?

Als we kijken hoe hier in de praktijk mee om wordt gegaan, dan zien we verschillende redematies. Eén daarvan wordt soms wel eens de ‘wegstreep’ redematie genoemd wordt. Hierbij wordt elk onderzoek afzonderlijk gewaardeerd, en daarbij telkens ‘weggestreept’ omdat de bewijskracht van elk onderzoek onvoldoende is, of omdat zekerheid ontbreekt. Aan het einde van dit proces is meestal alles op deze wijze weggestreept, en worden de onderzoeken niet meegewogen in de beslissing. Een variant hiervan zou je de ‘zoeken naar een anker’ methode kunnen noemen. Hierbij zoekt men naar tenminste één onderzoek dat sterk genoeg is om een hypothese te kunnen ‘bewijzen’. Als een dergelijk onderzoek ertussen zit, dan kunnen de overige onderzoeken dienen als ‘steunbewijs’. Als een dergelijk onderzoek er niet tussen zit, worden de onderzoeken niet of nauwelijks meegewogen in de beslissing. Deze redematietypen waarbij naar zekerheid wordt gezocht vormen een scherp contrast met een ander redematietype waarbij waarschijnlijkheid centraal staat. Voor de focus op waarschijnlijkheid in plaats van op zekerheid is al door vele anderen en in eerdere oraties gepleit (Broeders 2005). Hierbij gaat men ervan uit dat de onderzoeken elkaar versterken of verzwakken en probeert men de bewijskracht van de combinatie te vinden door de onderzoeken ‘in onderlinge samenhang’ te beschouwen.

Het Bayesiaanse raamwerk laat zien dat de laatste methode het meest logisch is, en maakt ook expliciet wat dan de gecombineerde bewijskracht is. Je kunt de regel van Bayes namelijk heel eenvoudig uitbreiden naar twee of meer bewijsmiddelen¹ (Robertson and Vignaux 1995). Hieruit volgt dat de bewijskracht van een combinatie van bewijsmiddelen die onafhankelijk van elkaar zijn onder de twee hypothesen simpelweg het product is van de bewijskracht van de afzonderlijke bewijsmiddelen. Voorwaarde is natuurlijk wel dat hetzelfde hypothesepaar wordt beschouwd voor alle bewijsmiddelen. Voor bovenstaand voorbeeld van het vergelijkend tape-onderzoek kun je op die manier afleiden dat de bewijskracht van de combinatie gelijk is aan $100 \times 1000 \times 10 = 1$ miljoen. Dit betekent dat, als de afzonderlijke onderzoeksresultaten niet heel

1. Voor twee bewijsmiddelen E_1 en E_2 en hypothesen H_1 versus H_2 geldt:

$$LR_{E_1 \text{ en } E_2} = \frac{P(E_1, E_2 | H_1)}{P(E_1, E_2 | H_2)} = \frac{P(E_1 | H_1)}{P(E_1 | H_2)} \cdot \frac{P(E_2 | E_1, H_1)}{P(E_2 | E_1, H_2)}$$

sterk zijn, de combinatie wel heel sterk kan zijn. Dit staat lijnrecht tegenover de “wegstreep” redenatie en de ‘zoeken naar een anker’ methode.

4b. Conditioneel afhankelijk bewijs en/of andere hypothesen

Als de onderzoeken niet onafhankelijk zijn onder de hypothesen moeten we de kans weten op de resultaten van het ene onderzoek, gegeven de resultaten van het andere onderzoek, om de bewijskracht van de combinatie te kunnen afleiden. Meestal is het nog ingewikkelder omdat de onderzoeken betrekking hebben op verschillende hypothesen. Bij een worsteling tussen dader en slachtoffer heeft bijvoorbeeld het DNA-onderzoek betrekking op de vraag van wie het DNA is en het vezelonderzoek van welk textiel object de vezels afkomstig zijn. De onderzoeken kunnen gecombineerd worden door in plaats van hypothesen te beschouwen die betrekking hebben op de mogelijke bronnen van de sporen hypothesen te beschouwen die betrekking hebben op de mogelijke activiteiten die de sporen veroorzaakten. Bijvoorbeeld:

H₁: het slachtoffer heeft met de verdachte geworsteld

H₂: het slachtoffer heeft met een onbekende persoon geworsteld

Deze hypothesen brengen echter met zich mee dat een groot aantal factoren beschouwd moeten worden die te maken hebben met de kans dat het DNA en de vezels worden overgedragen, persisteren, en worden gevonden. Ook verandert wat we als bewijs willen definiëren: naast het DNA-profiel en het vezeltype ook de hoeveelheid DNA en het aantal vezels dat wordt aangetroffen, de plaatsen waarop zij worden gevonden, en welke DNA-profielen en vezels er juist niet wordt gevonden maar wel zouden worden verwacht.

Om in deze situatie de bewijskracht van de combinatie af te kunnen leiden wordt het formulewerk te ondoorzichtig. Gelukkig hebben wij tegenwoordig de beschikking over software die de kansrekening voor ons kan doen. Deze software staat bekend onder vele namen, waaronder grafische modellen of Bayesiaanse netwerken (Taroni et al. 2006, Huygen 2004, Sjerps en Kloosterman 2010). Voor de kansrekening zal deze software denk ik even revolutionair blijken als de statistische software pakketten zoals bijvoorbeeld SPSS voor de statistiek. Bayesiaanse netwerken vinden nu al belangrijke toepassingen in bijvoorbeeld de medische, financiële, en wetenschappelijke wereld. Ik denk dat zij perfect aansluiten bij veel vragen uit de forensische statistiek. Om u een indruk te geven van de mogelijkheden en tegelijkertijd te demonstreren dat de forensische statistiek ook uitermate interessant is voor juristen, bekijken we een voorbeeld uit de strafrechtpraktijk.

Stel dat twee personen die met elkaar op straat staan te praten getuige zijn van een bankoverval op klaarlichte dag aan de overkant van de straat. Zij zien beiden een minuut lang haarscherp de overvaller, die zijn bivakmuts weggooit terwijl hij de bank uitrent en in een vluchtauto weggrijdt. Beide getuigen melden zich een minuut later bij de politie die hen onmiddellijk onafhankelijk van elkaar een fotoboek voorlegt van zes hen bekende bankovervallers. Vijf daarvan zitten nog in de gevangenis, maar Jan is zojuist op vrije voeten gesteld. De getuigen zijn welwillend om de politie te helpen en zullen alleen iemand aanwijzen die zij herkennen als de overvaller. Beide getuigen wijzen daarbij Jan aan. De foto in het boek is een recente en goede foto van Jan. De vraag is: wat voegt de tweede getuige toe aan de eerste? Ik heb deze vraag bij voordrachten voor verschillende groepen juristen wel eens gesteld, en wat me opviel was dat de antwoorden sterk uiteenliepen. Sommigen meenden dat door de tweede getuige de kans dat Jan de overvaller is sterk toeneemt, omdat de twee getuigen onafhankelijk van elkaar Jan aanwezen. Anderen meenden juist dat de tweede getuige vrijwel niets toevoegde aan de eerste getuige, immers Jan was al door de eerste getuige herkend en de getuigen hadden exact hetzelfde gezien. Anderen nuanceerden deze mening iets door te zeggen dat de tweede getuige niettemin een bevestiging vormt van de verklaring van de eerste getuige, maar over de vraag of hierdoor de kans dat Jan de overvaller is dan toeneemt en zo ja, hoeveel, verschilden de meningen weer.

Het Bayesiaanse raamwerk levert een middenweg op tussen deze redematies. We definiëren:

H₁: Jan is de overvaller

H₂: een onbekende is de overvaller

De verklaringen van de twee getuigen korten we af met E₁ en E₂.

Door de regel van Bayes voor twee bewijsmiddelen toe te passen, zien we dat we vier kansen moeten bepalen, die afhangen van de zes bankovervallers:

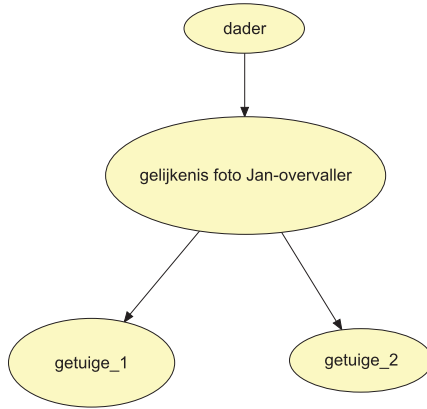
1. De kans dat de eerste getuige Jan aanwijst, als Jan de overvaller is
2. De kans dat de eerste getuige Jan aanwijst, als een onbekende de overvaller is
3. De kans dat de tweede getuige Jan aanwijst, als Jan de overvaller is en de eerste getuige hem heeft aangewezen
4. De kans dat de tweede getuige Jan aanwijst, als een onbekende de overvaller is en de eerste getuige Jan heeft aangewezen.

We zien hierdoor dat er een belangrijke variabele een rol speelt, namelijk de mate van gelijkenis tussen de foto van Jan en de onbekende in geval een onbe-

kende de overvaller is. Als de eerste getuige Jan aanwijst, dan vergroot dit de kans dat de foto van Jan op de overvaller lijkt, zowel in het scenario dat Jan de overvaller is als in het scenario dat een onbekende de overvaller is. Hierdoor wordt de kans dat de tweede getuige Jan aanwijst groter. De twee getuigenverklaringen zijn dus niet onafhankelijk van elkaar onder de twee scenario's. Niettemin voegt de tweede getuige wel degelijk bewijskracht toe. Namelijk, het quotiënt van de kansen onder 3 en 4 hierboven. Dit model kan uitgebreid worden naar andere situaties.

Het is een nogal technische redenatie. Je kunt dit inzichtelijker modelleren met een Bayesiaans netwerk (figuur 1). De knopen in dit model zijn variabelen die met een bepaalde kans in een bepaalde toestand zijn. Een pijl van A naar B geeft aan dat de kansen op de toestanden van B afhangen van de toestand van A. Wanneer je deze kansen specificiert kun je vervolgens het grafische model als rekenmachine gebruiken, maar dan met kansen. Je kunt dan bijvoorbeeld zien hoe de informatie dat de eerste getuige Jan aanwijst de kans vergroot dat Jan de overvaller is, en vervolgens hoe de tweede getuige die kans vergroot. Je ziet dan dat de tweede getuige inderdaad wel iets toevoegt, maar minder dan de eerste. In dit voorbeeld gebruiken we het Bayesiaanse netwerk om de logica te demonstreren van deze conclusie. Je kan zo ook de logica van 'schakelbewijs' demonstreren (een eerste stap is Evett et al. 2006). Het gaat hierbij uitdrukkelijk niet om de numerieke waarden die het model genereert, maar meer om de logica van de redenering transparant en expliciet te maken. Dit is zowel voor forensisch wetenschappers als politie en justitie relevant. Het maken van een LR model op zich, of het nu met een Bayesiaans netwerk is, met formules, of met een verbale argumentatie, dwingt de onderzoeker tot het expliciet maken van zijn redenatie, en leidt alleen hierdoor al tot een beter inzicht (Aitken and Gammerman 1989).

Een andere belangrijke toepassing in de forensische wetenschap is het gebruik van Bayesiaanse netwerken als expertsysteem (bijvoorbeeld Slooten et al. 2011, Cowell et al. 2011). Hier gaat het juist wel om de getallen. Bayesiaanse netwerk software biedt verder mogelijkheden om het kansmodel te analyseren. Je kunt bijvoorbeeld nagaan welke variabelen het grootste effect hebben op de gecombineerde bewijskracht, of hoe gevoelig deze bewijskracht is voor kleine afwijkingen in de gespecificeerde kansen.



Figuur 1. Bayesiaans netwerkmodel voor de twee getuigen

De variabele ‘dader’ heeft twee toestanden: ‘Jan’ en ‘onbekende’. De variabele ‘gelijkenis foto Jan-overvaller’ heeft drie toestanden: ‘sprekende gelijkenis’, ‘lijkt een beetje’ en ‘lijkt niet’. De variabelen ‘getuige_1’ en ‘getuige_2’ hebben drie toestanden: ‘wijst Jan aan’, ‘wijst niemand aan’ en ‘wijst iemand anders aan’.

5. Bayesiaanse raamwerk maakt denkfouten zichtbaar

We hadden het eerder over glad ijs, maar ik kan u meteen waarschuwen voor een heel ander soort van drogredentatie, de prosecutor’s fallacy (Thompson and Schuman 1987). Deze drogreden kent veel varianten, komt vaak voor, en lijkt onuitroeibaar ondanks alle aandacht die eraan besteed wordt. In essentie is deze drogredentatie het omkeren van een voorwaardelijke kans, dus de kans op A als je B weet verwarren met de kans op B als je A weet. In een forensische setting komt dit neer op het verwarren van de kans op het vinden van onderzoeksresultaten als een bepaalde hypothese waar is met de kans dat die hypothese waar is als je de onderzoeksresultaten vindt. Voor de wiskundigen onder ons: $P[E|H]$ wordt verward met $P[H|E]$. Zo rapporteert de DNA-deskundige de kans op een match als het spoor van een onbekende is, en dat is iets heel anders dan de kans dat het spoor van een onbekende is als er een match is. Ook bij het rapporteren van de LR gebeurt een soortgelijke omkering. De forensisch geneeskundige uit het voorbeeld van het verwonde meisje rapporteert bijvoorbeeld ‘Het is veel waarschijnlijker om deze verwondingen waar te nemen als het meisje zichzelf heeft verwond, dan als iemand anders haar heeft verwond’. Uit de praktijk, de literatuur en uit onderzoek (de Keijser et al. 2009) weten wij dat dit heel vaak wordt verward met ‘gezien deze verwondingen is het veel

waarschijnlijker dat het meisje zichzelf heeft verwond dan dat iemand anders haar heeft verwond'. Inmiddels ziet u natuurlijk onmiddellijk wat hier aan de hand is: de LR wordt verward met de posterior odds. Voor de wiskundigen onder ons: $P[E|H_1]/P[E|H_2]$ wordt verward met $P[H_1|E]/P[H_2|E]$.

Een relevante vraag is natuurlijk wat de consequenties zijn in de praktijk van deze prosecutor's fallacy. Is het slechts een woordenspelletje of leidt dit tot gerechtelijke dwalingen? Het antwoord schuilt wederom in de regel van Bayes. We zien dat de LR gelijk is aan de a posteriori kansverhouding van de hypothesen als de a priori kansverhouding gelijk is aan één. Dus als op grond van het overige bewijs in de zaak beide hypothesen a priori even waarschijnlijk zijn, dan heeft de fout geen enkele consequentie. We zien op dezelfde manier dat de fout zowel in het nadeel als in het voordeel van de verdachte kan zijn, al naar gelang de a priori kansverhouding kleiner is dan één of groter. Voor een groot deel van de zakenstroom van het forensisch onderzoek geldt dat de a priori kansverhouding vrij groot is. De items worden niet voor niets voor onderzoek aangeboden. Als de onderzoeksresultaten dan een redelijk grote bewijskracht hebben, dan is zowel de LR als de a posteriori kansverhouding groot en zijn de consequenties van de prosecutor's fallacy zeer beperkt. Maar als op grond van het overige bewijs de a priori kansverhouding heel klein is, dan leidt de prosecutor's fallacy tot grove overschatting van de bewijswaarde, wat kan leiden tot een veroordeling of arrestatie van een onschuldige persoon. Herkent u deze situatie al? Zo'n kleine a priori kansverhouding vinden we vooral in de 'glad ijs' situatie van databank zoekacties en vergelijkbare selectiemethoden. De combinatie met de prosecutor's fallacy levert dan 'spiegelglad ijs' op die kan leiden tot de vervolging en veroordeling van onschuldigen.

Een gerelateerde fout is het totaal negeren van het overige bewijs in de zaak, in wiskundige termen, van de a priori kansverhouding. Een voorbeeld hiervan zien we steeds vaker in zaken met 'moeilijke' DNA profielen, die een lagere bewijskracht hebben dan de standaard berekende frequentie 'minder dan 1 op 1 miljard' die normaliter wordt gerapporteerd. Wanneer het op grond van het overige bewijs al vrij waarschijnlijk is dat het spoor van de verdachte is, dan is er geen LR van een miljard nodig om te kunnen concluderen dat het spoor vrijwel zeker van de verdachte afkomstig is. Is er bijvoorbeeld op grond van het overige bewijs 50% kans dat het spoor van de verdachte is, en zijn er geen fouten gemaakt, dan vergroot een DNA match met de verdachte met een berekende frequentie van het profiel van één op honderdduizend deze kans al naar honderdduizend tegen één (99,999%). Wanneer de rechter in zo'n situatie het DNA-bewijs terzijde schuift omdat het niet zo sterk is als normaal dan ontloopt een dader dus mogelijk ten onrechte zijn straf en kan verder gaan met het plegen van misdrijven. Mijn collega Ate Kloosterman noemt dit 'de

erosie van het DNA-bewijs'. Waar men vroeger verguld was met een berekende frequentie van 1 op duizend, neemt men tegenwoordig het DNA-bewijs niet meer mee als de 1 op 1 miljard niet wordt gehaald. Maar het kan ook gebeuren dat bijvoorbeeld het wapenonderzoek in de prullenbak belandt omdat niet de hoogste tree van de waarschijnlijkheidsschaal werd geconcludeerd, terwijl de officier van justitie betoogt dat er op grond van het overige bewijs een hoge a priori kans is dat de kogel uit een bepaald wapen kwam.

Er is de laatste tijd veel aandacht voor de denkfouten die in rechtszaken worden gemaakt (o.a. Crombag et al. 2011, Derksen 2008, van Koppen 2011). Hieraan mogen wij niet de conclusie verbinden dat alle betrokken personen domme mensen zijn. Niet voor niets wordt er in de vakliteratuur zoveel aandacht aan besteedt. Het gaat om subtiele en geniepige denkfouten die ook door slimme mensen worden gemaakt. Het Bayesiaanse raamwerk maakt deze fouten inzichtelijk en helpt daarmee om ze te voorkomen.

6. *LR is optimale maat om bewijskracht te meten*

In het forensisch onderzoek worden er ook andere methoden gebruikt om de bewijskracht van de observaties weer te geven. Die leiden echter meestal tot een aantoonbaar misleidende weergave van de bewijskracht. Een vaderschapstest is een mooi voorbeeld om dit punt te illustreren. Wanneer op basis van de DNA-profielen van vermeende vader en kind blijkt dat de vermeende vader inderdaad de biologische vader van het kind zou kunnen zijn, dan drukken sommige DNA-laboratoria de bewijskracht hiervan uit door aan te geven wat de kans is dat een willekeurige man niet zou worden uitgesloten als de vader van het kind (in het Engels wordt dit aangegeven met de 'random man not excluded' (RMNE) benadering). Intuïtief lijkt dit een goede benadering, maar de LR-methode toont aan dat dit zowel kan leiden tot overwaardering als onderwaardering van het DNA-bewijs. Stel bijvoorbeeld dat het kind van zijn twee ouders de DNA-kenmerken 12 en 16 geërfd heeft, waarvan kenmerk 12 extreem zeldzaam is en kenmerk 16 vaak voorkomt. Als de vermeende vader ook het zeldzame kenmerk 12 bezit dan is dat natuurlijk sterk bewijs dat hij inderdaad de biologische vader is. In de RMNE komt die bewijskracht echter niet goed tot uitdrukking, omdat ook alle mannen die het andere DNA-kenmerk 16 bezitten niet kunnen worden uitgesloten als vader. De LR geeft in deze situatie de grote bewijskracht wel goed weer. Nog erger wordt het wanneer vervolgens zonder toelichting een kans wordt gerapporteerd dat de vermeende vader inderdaad de vader van het kind is. Deze kans is namelijk meestal gebaseerd op de aanname dat er a priori een kans van 50% is dat de vermeende vader de echte vader is van het kind. Maar waarop baseert de DNA-onderzoek-

ker deze aanname? Niet op zijn eigen expertise, dat is wel duidelijk. Toepassing van de LR methode geeft een neutrale manier om de bewijskracht weer te geven in dit soort situaties en wordt daarom al sinds 1938 geadviseerd als de beste methode (Essen-Möller 1938). Dus, geachte mannelijke toehoorder, mocht u ooit op deze wijze beticht worden van vaderschap, onterecht natuurlijk, dan weet u nu alvast een ijzersterke verdedigingsstrategie.

De RMNE is een voorbeeld van het denken in termen van match-of-geen-match. De verdachte matcht als biologische vader bij het kind, en de kans dat een willekeurige man zou matchen is 1 op de zoveel. Bij dit denken wordt eerst een criterium gedefinieerd voor het al dan niet matchen, en vervolgens wordt de kans gegeven dat een willekeurig subject zou matchen. Deze populaire methode geeft de bewijskracht echter niet optimaal weer. Bij continue metingen leidt zij tot het zogenaamde 'fall-of-the-cliff-effect': wanneer de meting nog net aan het matchcriterium voldoet wordt een match gerapporteerd met een bepaalde matchkans, terwijl een meting die net niet aan het matchcriterium voldoet tot een exclusie leidt. Kijken we bijvoorbeeld naar brekingsindex metingen aan glas dan kan een meting van 1.51994 leiden tot het rapporteren van een match en een meting van 1.51993 tot het rapporteren van een exclusie. Dat een miniem verschil in de metingen tot een zo groot effect op de conclusie leidt voelt niet goed en dat is het ook niet. De LR methode past veel beter bij het continue karakter van de metingen en heeft geen last van dit 'fall-of-the-cliff-effect'.

Gebruikelijke statistische methoden zoals hypothese toetsen zijn ongeschikt om te gebruiken als maat voor bewijskracht, zoals wiskundig kan worden aangetoond. De LR heeft bovendien de beste eigenschappen als maat voor bewijskracht, zoals ondermeer aangetoond door Royall (1997) en Blume (2002).

7. Bayesiaanse denkraam maakt waarschijnlijkheidsredenatie expliciet

De LR-methode wordt dus gebruikt om de bewijskracht numeriek of verbaal te kunnen uitdrukken. Het grote voordeel hiervan is dat de waarschijnlijkheidsredenatie expliciet wordt gemaakt. Het Bayesiaanse raamwerk dient daarbij als leidraad voor de onderzoeker, die daarmee gedetailleerd duidelijk kan maken hoe hij tot zijn oordeel over de bewijskracht komt. Hij moet namelijk expliciet antwoord geven op de volgende vragen: Welke hypothesen heeft hij beschouwd? Welke waarnemingen heeft hij als bewijs beschouwd? Welke aannamen deed hij bij het afleiden van de LR? Welke gegevens heeft hij daarbij gebruikt, en hoe is hij omgegaan met onzekerheden zoals steekproefonzekerheid en meetonzekerheid?

De LR-methode maakt op gestructureerde wijze duidelijk dat er bij al deze vragen keuzen moeten worden gemaakt. Je zou deze moeilijke vragen en vereiste keuzen liever vermijden. Sommige onderzoekers bekritisieren het Bayesiaanse model op dit punt (Derksen 2010, van Koppen 2011). Ik schreef samen met mijn collega Charles Berger een reactie op deze kritiek (Sjerps en Berger 2011). Het Bayesiaanse model maakt inderdaad de moeilijkheden en keuzen zichtbaar die onlosmakelijk verbonden zijn aan de interpretatie van bewijs. In onze ogen is dat een voordeel en geen nadeel. Wie net doet alsof hij de moeilijkheden kan vermijden door op een andere manier te redeneren houdt zichzelf voor de gek. Andere manieren van evaluatie van bewijs hebben ook deze moeilijkheden en keuzen, het verschil is alleen dat zij ze minder zichtbaar maken. De zaken simpeler voorstellen dan ze zijn leidt misschien tot een schijnbaar beter begrip van de rapportage, maar uiteindelijk niet tot een betere rechtspraak. Eenvoudiger ‘alternatieve’ methodes beantwoorden simpelweg een geheel andere vraag. Zo beantwoordt van Koppen (2011) de vraag hoe mensen intuïtief redeneren, namelijk door gebruik te maken van ‘verankering’. In essentie komt dit neer op het afronden van waarschijnlijkheid naar zekerheid. Dat is natuurlijk geen antwoord op de vraag hoe mensen logisch gezien zouden moeten redeneren, de kernvraag van het Bayesiaanse model. Opvallend is dat de basisprincipes van het Bayesiaanse raamwerk ook door de critici ervan worden onderschreven.

Bewijskracht is een relatief begrip

Met name de definitie van de hypothesen, achtergrondinformatie en het bewijs zijn keuzen die de grootte van de LR bepalen. Bijvoorbeeld, we zagen eerder dat de LR van een volledige DNA-match tussen spoor en verdachte ten aanzien van de hypothesen ‘het DNA is van de verdachte’ versus ‘het DNA is van een onbekende persoon’ groter is dan een miljard. Beschouwen we echter als alternatieve hypothese ‘het DNA is van een onbekende persoon met hetzelfde DNA-profiel als de verdachte’, dan is de LR gelijk aan één. Dat is logisch, de DNA-match kan immers geen onderscheid maken tussen de twee hypothesen. Waar is de factor één miljard dan opeens gebleven? Die zien we terug in de a priori kansen: a priori is de kans op de eerste alternatieve hypothese een miljard keer groter dan de kans op de tweede alternatieve hypothese. Voor de a posteriori kansen maakt het niet uit welke alternatieve hypothese we kiezen: als de LR een factor een miljard keer kleiner wordt en de a priori kansverhouding een miljard keer groter dan heft dat elkaar op in het product. We zeggen

dan dat de twee alternatieve hypothesen logisch equivalent zijn gegeven het bewijs, of in jargon, dat zij conditioneel equivalent zijn (Dawid 2001).

Voor de wiskundige is hiermee de kous af, er is geen voorkeur voor een bepaalde keuze. Voor de toepassing op de forensische praktijk ligt hier echter een belangrijke keuze gezien de rolverdeling tussen deskundige en jurist: de deskundige schat de LR, de jurist de a priori kansen. Wanneer de DNA-deskundige bijvoorbeeld zou kiezen voor het tweede alternatief, 'het DNA is van een onbekende persoon met hetzelfde DNA-profiel als de verdachte', dan is hij gauw klaar: de LR is één. Hij laat het dan aan de jurist om de zeldzaamheid van het profiel in te schatten. Dat kan natuurlijk niet de bedoeling zijn, want juist op dit punt is kennis van de deskundige nodig. De deskundige moet de definities van hypothesen, achtergrondinformatie en bewijs daarom zo kiezen dat de LR al zijn kennis (en alleen zijn kennis) weerspiegelt. Er is nog een belangrijk verschil tussen de wiskundige en de forensische wereld, namelijk dat in praktijk voor veel juristen de regel van Bayes geen gesneden koek is. De deskundige mag er daarom niet zonder meer van uit gaan dat zijn keuze tussen twee conditioneel equivalente hypothesen niet uitmaakt omdat de jurist zijn keuze toch wel compenseert via de a priori kans. Een mooi voorbeeld hiervan is de discussie over de rapportage van een DNA-match die gevonden wordt via de DNA-databank.

De keuze van de alternatieve hypothese levert vaak grijze gebieden en dus discussie op. Derksen (2010) noemt dit het probleem van de juiste referentieklassen. Zo discussiëren schoenspoordeskundigen in internationaal verband over de keuze van de alternatieve hypothese. Stel bijvoorbeeld dat zowel het schoenspoor als de schoen van de verdachte maat 42 is met een noppenprofiel. De keuze is dan tussen 'het spoor is veroorzaakt door een andere schoen' en 'het schoenspoor is veroorzaakt door een andere schoen met maat 42 en een noppenprofiel'. De zeldzaamheid van de combinatie maat 42 / noppenprofiel maakt bij de eerste keuze onderdeel uit van de LR en moet dus door de deskundige worden geschat. Bij de tweede keuze maakt deze zeldzaamheid onderdeel uit van de a priori kans, en komt daarmee voor rekening van de jurist.

Principes uit de forensische statistiek zijn algemeen toepasbaar

De inzichten uit de forensische statistiek zijn veel algemener toepasbaar. In deze rede gaf ik al aan dat evaluatie van bewijs ook essentieel is voor politie en justitie. In principe, zo is mijn stelling, is het Bayesiaanse raamwerk voor evaluatie van bewijs van belang voor ieder die zich met evaluatie van bewijs bezig-

houdt, dus ook wetenschappers in algemene zin. Laat ik u een voorbeeld geven uit de medische wereld. In de Volkskrant van 20 augustus 2011 wordt de lezer blij verrast met het bericht ‘Hond ruikt longkanker in vroegtijdig stadium’ (van Raaij 2011). In het artikel legt een hoogleraar pathofysiologie uit dat honden in de adem van patiënten vluchtige stoffen kunnen detecteren die afkomstig zijn van kankercellen, en noemt dit een veelbelovende methode. ‘De honden doen dit met een gevoeligheid van 70 procent en een specificiteit van 90 procent. Ze pikken dus niet alle kanker op, maar sluiten wel vrij trefzeker kanker uit.’, aldus de hoogleraar. We mogen hopen dat de man meer verstand heeft van geneeskunde. Wanneer we namelijk als hypothesen beschouwen:

H₁: de patiënt heeft longkanker

H₂: de patiënt heeft geen longkanker,

en dan de LR uitrekenen van de waarneming ‘hond geeft kanker aan’, dan komen we slechts op 7, en van ‘hond geeft geen kanker aan’ op 1/3. Deskundigen van het NFI zouden dit kwalificeren als de zwakste term in de conclusiereeks en de LR van 7 formuleren als: ‘de bevindingen van het onderzoek zijn iets waarschijnlijker als de patiënt longkanker heeft dan als hij dat niet heeft’ (zie ook AFSP 2009). Als we willen weten hoe groot de kans is dat de patiënt in kwestie longkanker heeft, dan voegt deze hondentest heel weinig toe aan de overige informatie. Gigerenzer (2002) laat zien dat dit voorbeeld symptomatisch is voor de medische wereld. Blume (2002) laat zien dat likelihood methoden zeer geschikt zijn voor medische data analyse.

In volle vaart vooruit

In de afgelopen decennia hebben enkele forensisch statistische pioniers, met name dr. Ian Evett van de voormalige Forensic Science Service in de UK, erkenning weten te krijgen voor het belang van hun vak en de lijnen uitgezet voor het daadwerkelijk berekenen van LR's in de zakenstroom. Zo heeft de vereniging van Europese forensische laboratoria twee belangrijke doelstellingen geformuleerd die worden gedragen door de forensische statistiek (ENFSI 2011). Ook het Nederlandse NWO forensic science programma heeft statistiek als één van de speerpunten. Wij zijn ‘hot’!

Het wetenschappelijk onderzoek komt nu in een stroomversnelling. Het programma voor de forensische statistiek, en voor mij, voor het komende decennium zou ik als volgt willen formuleren:

1. *Objectivering*: het kunnen afleiden van de bewijskracht (LR) op zo objectief mogelijke wijze
2. *Validatie* van de gerapporteerde bewijskracht (LR)
3. *Combinatie*: wat is de bewijskracht van een combinatie van waarnemingen? Dit is nodig voor een eindconclusie bij multidisciplinair onderzoek en ook voor conclusies die betrekking hebben op mogelijke activiteiten in plaats van mogelijke bronnen.
4. *Expansie* van toepassingen binnen en buiten de forensische wetenschap
5. *Communicatie* met politie en justitie

De samenwerking met universiteiten en onderzoeksinstituten om dit programma te verwezenlijken is daarbij essentieel. De kritische blik en innovatieve ideeën die hieruit voortkomen zijn een enorme stimulans voor het onderzoek. Ik ben ervan overtuigd dat de liefde wederzijds zal blijken. De forensische statistiek heeft interessante en verrassende vragen van zowel fundamentele als toegepaste aard. Nederland kan daarbij een voortrekkersrol spelen. De combinatie van diverse universitaire statistiek en kansrekening groepen van hoog niveau met een interesse in forensische statistiek, een academisch forensisch master programma, een groot forensisch laboratorium met een groep forensisch statistici die de theorie in praktijk brengt en die toegang tot data en deskundigen heeft, een geïnteresseerde juridische beroepsgroep, en beleidsmakers die het belang van de forensische statistiek inzien is een gouden combinatie die naar mijn weten uniek is in de wereld. Alle seinen staan op groen. Als wij goed weten samen te werken kan in Nederland forensische statistiek van topniveau ontwikkeld worden. Ik ben blij dat ik daar mijn steentje aan mag bijdragen.

Het Recht vooruit

Evaluatie van bewijs is een vak. Een moeilijk en fascinerend vak, dat onlosmakelijk verbonden is met kansrekening en statistiek. Het Bayesiaanse raamwerk is een leidraad om de vereiste waarschijnlijkheidsredenatie op een logische en gestructureerde wijze op te bouwen. Dit raamwerk behelst de essentie van het forensische en juridische vak: het bepalen van de bewijskracht van de bewijsmiddelen, van de combinatie van bewijsmiddelen, en uiteindelijk het bepalen van de waarschijnlijkheid van de verschillende scenario's. Dit raamwerk is breed toepasbaar, ook buiten de forensische wetenschap en met name op juridische vraagstukken. Het is de onderbouwing van de overtuiging die nodig is voor een veroordeling. De gebruikers van dit raamwerk zijn momenteel voor-

namelijk te vinden onder statistici, forensisch wetenschappers, strafrechtpsychologen en wetenschapsfilosofen. Slechts een enkele jurist weet zich succesvol in dit onderwerp te verdiepen en het toe te passen (zie bijvoorbeeld de conclusie van Aben in LJN BK2094). Dat is eigenlijk een vreemde situatie. Zouden de juristen niet eigenlijk de specialisten moeten zijn op het gebied van evaluatie van bewijs? Ik heb de indruk dat veel juristen zich omkeren zodra het woord 'kans' valt, onder het valse argument dat kansrekening niet hun vak is.

Mijn stelling is dat evaluatie van bewijs een kernexpertise zou moeten zijn van elke strafrechtjurist. Ik moedig hen daarom van harte aan dit terrein op de bèta en gamma wetenschappers te heroveren. Het is niet noodzakelijk dat iedere jurist foutloos kansformules produceert of een correlatie kan uitrekenen. Het is wel noodzakelijk dat zij bekend zijn met de basisprincipes van evaluatie van bewijs. Ik ben blij te zien dat de eerste stappen in die richting worden genomen, onder andere door het studiecentrum rechtspleging SSR en sommige universiteiten. Maar een versnelling is hard nodig. Daarnaast zou er een groep moeten zijn, verbonden aan een universiteit of juridisch instituut, dat zich toelegt op onderzoek op dit terrein. De consequenties voor de praktijk die uit dit onderzoek volgen zouden als praktische richtlijnen kunnen worden verspreid onder de beroepsgroep. Voor het recht betekent dit een inhoudelijke verdieping op een kernvraag en zou daarmee een grote stap vooruit zijn.

Conclusie

Zonder statistiek geen bewijs. Daarmee begon ik dit betoog. Waarschijnlijkheidsredenering is de basis om de bewijskracht van (forensische) bewijsmiddelen te bepalen, en is de grondslag voor de overtuiging van de jurist. Het Bayesiaanse raamwerk is hierbij een belangrijke leidraad om die bewijskracht te meten, en om de taakverdeling te verduidelijken: de deskundige levert de LR, en de jurist stelt daarmee de waarschijnlijkheid van de hypothesen bij. De forensische statistiek, die zich richt op het bepalen van deze LR, heeft zich in de afgelopen decennia voornamelijk bezig gehouden met het aantonen van het belang en de mogelijkheden hiervan. Nu deze worden erkend, is het tijd voor een stroomversnelling. Op de agenda staat objectivering, validatie, combinatie, expansie en communicatie. Het gebied borrelt over van de interessante vragen voor de wetenschapper en de jurist. Wat nu nodig is, is samenwerking en meer mankracht. De kritische en innoverende rol van de universiteiten is hierbij cruciaal. Nederland verkeert internationaal gezien in een uitstekende positie om toonaangevend te worden in de forensische statistiek. Nu nog even doen.

Dankwoord

Graag sluit ik af met een woord van dank aan diegenen zonder wie ik hier nu niet had gestaan. Allereerst dank aan het College van Bestuur van de Universiteit, het bestuur van de Faculteit Natuurwetenschappen, Wiskunde en Informatica, en aan de leden van het bestuur van de Stichting Leerstoel Criminalistiek, voor het in mij gestelde vertrouwen. Jan Wiegerinck en Chris Klaassen van het Korteweg-de Vries Instituut voor wiskunde, voor het herkennen van de wetenschappelijke potentie van de forensische statistiek.

Wim Neuteboom en de directie van het Nederlands Forensisch Instituut, die zich via de Stichting Leerstoel Criminalistiek persoonlijk hebben ingezet voor de instelling van mijn leerstoel. Verder natuurlijk de leden van het NFI statistiek team, Annabel Bolck, Ivo Alberink, Peter Vergeer en Reinoud Stoel, en de andere statistici op het NFI, voor de talrijke discussies over zin en onzin in de forensische statistiek, en voor jullie vriendschap en humor. Mijn andere collega's wil ik bedanken voor de interessante problemen die jullie telkens weer aan ons voorleggen. Met name mijn *partner in crime* Ate Kloosterman is hierbij een onuitputtelijke inspiratiebron. Samen met academische kanjers als Ronald Meester, Richard Gil, Eric Cator, Gabriel Vivo-Truyols, Bert van Es en vele anderen vormen wij een klasse Bayesiaans netwerk. De a priori verwachtingen zijn hoog.

Mijn beide promotoren Hans Metz en Ed van der Meijden, mijn co-promotor Patsy Haccou, en mijn begeleider Evert Meelis: jullie legden de basis waar ik vandaag op sta. Bedankt voor de boeiende tijd in Leiden, ik heb veel van jullie geleerd.

Mijn vader Jan Sjerps, lieve Dad, dankzij mama en u had ik een warm nest en kon ik studeren. Ik heb pas achteraf beseft dat dat helemaal niet zo vanzelfsprekend was als ik toen dacht. Mijn familie en schoonfamilie, bedankt voor jullie steun en komst vandaag.

Roel, Ivo, Sonja en Luc. Ik houd van de forensische statistiek, maar ik kan met zekerheid (!) stellen dat ik het meest van jullie houd.

Ik heb gezegd.

Referenties

- Aitken C.G.G. and Gammerman A.J. (1989) 'Probabilistic reasoning in evidential assessment', *Journal of the Forensic Science Society* 29, 303-316
- Aitken C.G.G. and Taroni F. (2004) 'Statistics and the evaluation of evidence for forensic scientists', 2nd ed., Chichester: Wiley
- AFSP (Association of forensic science providers) (2009) 'Standards for the formulation of evaluative forensic science expert opinion', *Science and Justice* 49, 161-164
- Balding D.J. (2005) 'Weight-of-evidence for forensic DNA profiles', Chichester UK: Wiley
- Balding D.J. and Donnelly P. (1996) 'Evaluating DNA profile evidence when the suspect is identified through a database search', *Journal of Forensic Sciences* 41, 603-607
- Beek C.P. van der (2011) 'Nederlandse DNA databank voor strafzaken, jaarverslag 2010', uitgave van het Nederlands Forensisch Instituut, via www.forensischinstituut.nl
- Berger C.E.H. en Aben D. (2010) 'Bewijs en overtuiging: Rationeel redeneren sinds Aristoteles', *Expertise en Recht* 2, 52-56.
- Blume J.D. (2002) 'Likelihood methods for measuring statistical evidence', *Statistics in medicine* 21, 2563-2599
- Bolck A., Weyermann C., Dujourdy L., Esseiva P., van den Berg J. (2009) 'Different likelihood ratio approaches to evaluate the strength of evidence of MDMA tablets comparison', *Forensic Science International* 191, 45-51
- Bolck A. and Alberink I. (2011) 'Variation in Likelihood Ratios for forensic evidence evaluation of XTC tablets comparison', *Journal of Chemometrics*, 25, 41-49
- Broeders A.P.A. (2005) *Ontwikkelingen in de criminalistiek*, Den Haag: Boom juridische uitgevers
- Buckleton B., Triggs C.M. and Walsh S.J. (eds) (2005) 'Forensic DNA evidence interpretation', Boca Raton: CRC Press
- Cowell R.G., Lauritzen S.L., and Mortera J. (2011) 'Probabilistic expert systems for handling artifacts in complex DNA mixtures', *Forensic Science International Genetics* 5, 202-209
- Crombag H.F.M., van Koppen P.J. en Wagenaar W.A. (2011) 'Dubieuze zaken. De psychologie van strafrechtelijk bewijs'. Amsterdam: Olympus (eerste druk in 1992)
- Curran J.M. (2009) 'Statistics in forensic science', *WIREs Computational Statistics* 1, 141-156
- Dawid A.P. and Mortera J. (1996) 'Coherent analysis of forensic identification evidence'. *Journal of the Royal Statistical Society, Series B*, 58, 425-443
- Dawid A.P. (2001) 'Comment on Stockmarr's "Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search"', *Biometrics* 57, 976-980
- Derksen A.A. (2008) 'Het OM in de fout' Diemen: Veen Magazines

- Derksen A.A. (2010) 'De ware toedracht: een praktische wetenschapsfilosofie voor waarheidszoekers', Rotterdam: Veen Magazines
- ENFSI (European network of forensic science institutes) (approved by the members on October 20 2008) 'ENFSI strategic plan 2008 – 2011', www.enfsi.eu
- ENFSI DNA working group (2010), 'DNA-database management review and recommendations', <http://www.enfsi.eu/page.php?uid=98>
- Essen-Möller E. (1938), 'Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis. Theoretische Grundlagen', *Mitteilungen der Anthropologischen Gesellschaft* 68, 9-53
- Evvett I.W. and Weir B.S. (1998) 'Interpreting DNA evidence', Sunderland MA: Sinauer
- Evvett I.W. (1998). 'Toward a uniform framework for reporting opinions in forensic science case work'. *Science and Justice* 38, 198-202
- Evvett I.W., Jackson G., Lindley D.V. and Meuwly D. (2006) 'Logical evaluation of evidence when a person is suspected of committing two separate offences', *Science & Justice* 46, 25-31
- Federal Evidence Review 2011, download van: www.FederalEvidence.com
- Finkelstein W.O. en Fairley W.B. (1970) 'A Bayesian approach to identification evidence', *Harvard Law Review* 83, 873-887.
- Geddes L. (2011) 'DNA super-network increases risk of mix-ups', *New Scientist* 2828, 24
- Gigerenzer G. (2002) 'Reckoning with risk: Learning to live with uncertainty'. London: Penguin Books
- Huygen P.E.M. (2004) "'Bayesian Belief Networks" voor redeneren over juridische bewijsvoering', in: W.H. van Boom en M.J. Borgers (red.) 'De rekenende rechter', Boom Juridische Uitgevers, Den Haag
- Kaye D.H. (2010) 'The Double Helix and the Law of Evidence', Harvard University Press, Cambridge.
- Keijser J.W. de, Elffers H., Kok R.M. en Sjerps M.J. (2009), 'Bijkans begrepen?', Den Haag: Boom Juridische uitgevers
- Koppen P.J. van (2011) 'Overtuigend bewijs', Amsterdam: Nieuw Amsterdam
- Lowe (2010) 'False DNA matches could be likely', www.Public.service.co.uk, Feb 3
- Meester R. en Sjerps M. (2003) 'The evidential value in the DNA database search controversy and the two-stain problem', *Biometrics* 59, 727-732
- Meester R. en Sjerps M. (2004) 'Why the effect of prior odds should accompany the likelihood ratio when reporting DNA evidence', *Law, probability and Risk* 3, 51-62
- Meuwly D. (2006) 'Forensic individualisation from biometric data', *Science & Justice* 46, 205-213
- NRC National Research Council (1996). 'The Evaluation of Forensic DNA Evidence', Washington, D.C.: National Academy Press
- NRC National Research Council (2009). 'Strengthening Forensic Science in the United States: A Path Forward', Washington D.C.: National Academy Press
- Robertson B. and Vignaux G.A. (1995) 'Interpreting Evidence: Evaluating Forensic Science in the Courtroom'. Chichester: Wiley.

- Royall R.M. (1997) 'Statistical Evidence-a likelihood paradigm', London: Chapman and Hall
- Schneider P.M., Schneider H., Fimmers R., Keil W., Molsberger G., Pflug W., Rothämel T., Eckert M., Pfeiffer H., Brinkmann B. (2010) 'Allgemeine Empfehlungen der Spurenkommision zur statistischen Bewertung von DNA-Datenbank-Treffern' *Rechtsmedizin* 20, 111–115
- Sjerps M. en Berger C. (2011) 'Het Bayesiaanse model biedt een helder zicht op een complexe werkelijkheid', www.forensischinstituut.nl
- Sjerps M.J. en Coster van Voorhout J.A. (red.) (2005) 'Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht', Deventer: Kluwer
- Sjerps, M., Kloosterman A. en van der Beek, K. (2010) 'De interpretatie van een DNA-databank match', *Delikt en Delinkwent* 40, 138-155
- Sjerps, M.J. en Kloosterman A.D. (2010) 'Het gebruik van Bayesiaanse netwerken in de forensische (DNA) statistiek', *Ars Aequi*, 502
- Sjerps M.J. en Meester R. (2009) 'Selection effects and database screening in forensic science', *Forensic Science International* 192, 56-61
- Slooten K. (2011) 'Validation of DNA-based identification software by computation of pedigree likelihood ratios', *Forensic Science International: Genetics* 5, 308-315
- Slooten K. and Meester, R. (2011) 'Forensic identification: the island problem and its generalisations', *Statistica Neerlandica* 65, 202–237
- Stockmarr A. (1999) 'Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search', *Biometrics* 55, 671-677
- Taroni F., Aitken C.G.G., Garbolino P., Biedermann A. (2006) 'Bayesian networks and probabilistic inference in forensic science', Chichester: Wiley
- Thompson W.C. Schuman E.L. (1987) 'Interpretation of statistical evidence in criminal trials-the prosecutor's fallacy and the defense attorney's fallacy', *Law and Human Behavior* 11, 167-187