

Het juiste gewicht in de schaal

Charles Berger^{*}

‘Meten is weten’, luidt de gevleugelde kreet, en meten kan inderdaad belangrijke informatie opleveren. Alleen zal voor het trekken van conclusies uit meetresultaten vrijwel altijd meer nodig zijn. Gevolgtrekkingen zijn niet alleen op die meetresultaten gebaseerd maar ook op logica en populatiegegevens. Daarmee kan worden aangegeven in welke mate de resultaten bewijs vormen voor relevante hypothesen in een zaak. Als het gaat om waarheidsvinding is de wetenschappelijke kant van bewijs essentieel om het juiste gewicht in de schaal te leggen.

Juristen - en zij zijn daarin niet de enigen - verwarren regelmatig wetenschap en techniek, en wetenschappers met technici die metingen doen. Dat is jammer, omdat de wetenschap juristen veel meer te bieden heeft dan alleen techniek en metingen. De wetenschap is ook van onschatbare waarde als het gaat om het interpreteren van metingen en waarnemingen en het redeneren met onzekere gegevens. Wanneer een jurist daarover geen kennis heeft en slechts vaart op ervaring en intuïtie, loopt hij de kans essentiële fouten te maken. Dat is een belangrijk probleem, te meer daar is geconstateerd dat: “Judges and lawyers usually react to science with all the enthusiasm of a child about to get a tetanus shot”.¹

De interpretatie van bewijs is een centraal thema in de criminalistiek, maar deze kennis geldt net zo goed voor ‘technisch’ bewijs als voor ‘tactisch’ bewijs, en rechterlijke waarnemingen. Het interpreteren van bewijs, en het redeneren met bewijs vindt nu nog vaak plaats op een te laag niveau.

Zo verwacht een rechter van de deskundige vaak een oordeel op onderdelen. In een oordeel in de zin van een zekerheidsuitspraak kan de rechter dan een

^{*} Dr. ir. C.E.H. Berger is principal scientist bij het Nederlands Forensisch Instituut (NFI), en lid van het College van het Nederlands Register Gerechtelijk Deskundigen (NRGD).

¹ Science and the Law in the Wake of Daubert: A New Search for Scientific Knowledge, B. Black, F.J. Ayala, C. Saffran-Brinks, Texas Law Review 1994, 72:715-802.

uiteindelijke beslissing ‘verankeren’. In de meeste gevallen kan een deskundige zo’n uitspraak echter niet doen. Hij heeft niet de informatie of de rol om beslissingen te nemen, maar om de wetenschappelijke bewijskracht van zijn waarnemingen te beschrijven. Daarmee stelt hij de jurist in staat dat onderdeel in het geheel van het bewijs mee te wegen. Het is de taak van de rechter om op grond van die weging van alle informatie in de zaak tot een eindoordeel, een beslissing, te komen.

Tot het moment van die beslissing moet een rationele beslisser de (beperkte) bewijskracht van alle aangedragen informatie meewegen. Wanneer onzekerheden al eerder tot zekerheid (in bevestigende of ontkennende zin) verheven worden, kan dat leiden tot een verkeerde afweging. Net zoals een tros druiven niets zou kosten wanneer de prijs van een druif eerst naar nul zou worden afgerond. De bewijsstandaard dient dus niet te worden toegepast op de afzonderlijke onderdelen, maar op het uiteindelijke totaal. Dat betekent tevens dat het bewijs eerst gecombineerd moet worden. In een ander artikel in deze uitgave zal op dat combineren worden ingegaan.² In dit stuk beperken we ons tot het bepalen van de bewijskracht op de afzonderlijke onderdelen.

In het denken over bewijskracht van onderzoeksresultaten is men in de wetenschap verder dan in de juridische wereld. Het hangt immers direct samen met de manier waarop het geheel van wetenschappelijke kennis voortdurend wordt uitgebreid. In de criminalistiek zien we dat bij subjectieve vormen van onderzoek zogenoemde conclusiereeksen worden gebruikt: reeksen verbale termen die een maat voor de bewijskracht willen geven. Maar soms gingen de termen, zelfs in één reeks, over verschillende dingen zoals zekerheid, mogelijkheid, waarschijnlijkheid, en aanwijzing.³

Traditioneel geven veel conclusiereeksen verbale termen voor de mate van waarschijnlijkheid van een hypothese. Dit hangt samen met de misconceptie dat forensisch onderzoek aan een spoor zou kunnen bepalen wat de kans is dat een verdachte dat spoor zou hebben achtergelaten. Maar een forensisch onderzoek in isolatie kan deze kans niet bepalen, omdat die kans ook wordt bepaald door vele andere factoren. Daarbij kan gedacht worden aan overige forensische sporen, maar

² Zie de bijdrage van Sjerps en Kloosterman in deze uitgave van *Ars Aequi*: M. Sjerps en A. Kloosterman, ‘Het gebruik van Bayesiaanse netwerken in de forensische (DNA)-statistiek’, *AA* juli/augustus 2010, p. 502.

³ D.A. Rudram, *Science & Justice* 1996; 36: p. 133-138. Op p. 137 haalt hij de conclusiereeks van de documentonderzoekers van de Metropolitan Police Forensic Science Laboratory aan.

ook aan andere informatie. Hoe zat het bijvoorbeeld met de kans dat de verdachte op de plaats delict aanwezig was, had hij een alibi? Naarmate zo'n alibi sterker is zal de kans dat de verdachte het spoor achterliet afnemen. Daarom kan het onderzoek aan het spoor alleen, die kans niet bepalen.

Maar wat is dan de waarde en de bewijskracht van het forensisch onderzoek aan het spoor? Die waarde zit hem in de mate waarin het onderzoek helpt om onderscheid te maken tussen hypothesen, in dit voorbeeld hypothesen over wie de sporen achterliet. De bewijskracht bepaalt niet de kans op de juistheid van een hypothese, maar doet deze kans (of onze gerechtvaardigde mate van overtuiging) wel toe- of afnemen. Bewijskracht is in die zin een relatief begrip.

Al in 1908 werd de gerenommeerde forensische onderzoeker Alphonse Bertillon in het Dreyfus-proces op deze relativiteit gewezen.⁴ De door hem waargenomen overeenkomsten in een vergelijkend onderzoek konden niet de kans op een hypothese (van gelijke herkomst) bepalen, zoals Bertillon had beweerd. Het was de wiskundige Henri Poincaré van de *Academie des Sciences* die er toen al op wees dat de bewijskracht van de waargenomen overeenkomsten slechts de relatieve toe- of afname van de waarschijnlijkheid van een hypothese kon geven, en niet die waarschijnlijkheid zelf.

De bewijskracht doet de verhouding tussen de waarschijnlijkheden van twee concurrerende hypothesen (bijvoorbeeld over de bron van een spoor) veranderen. De vermenigvuldigingsfactor van die verandering volgt uit de wetten van de kansrekening. Hij wordt de *likelihood ratio* (LR) genoemd en geeft een universele maat voor de bewijskracht. De LR (en dus de bewijskracht) is gelijk aan de waarschijnlijkheid van de waarneming wanneer de ene hypothese waar is, gedeeld door de waarschijnlijkheid datzelfde waar te nemen wanneer de concurrerende hypothese waar is.⁵

Poincaré beschikte over deze kennis, maar deze werd niet overgenomen door de forensische wereld, die weinig academische verankering kende en sterk gelieerd was aan de wetshandhaving. Door het uitbreken van de Eerste Wereldoorlog kreeg de forensische wetenschap nog minder aandacht en dit belemmerde de ontwikkeling van

⁴ L'affaire Dreyfus : la revision du procès de Rennes, Henry Mornard (1907), Ligue française pour la défense des droits de l'homme et du citoyen, p. 334.

⁵ Zie voor een laagdrempelige uitleg: Criminalistiek is terugredeneren, C.E.H. Berger, Nederlands Juristenblad 2010, p. 784.

de forensische kennis. De bestaande kennis werd tot een technisch, geprotocolleerd instrument voor de politie. De waarschuwingen van Poincaré werden vergeten.

Sir Harold Jeffreys kwam in 1939 met een reeks verbale termen voor bewijskracht. Hij koppelde de numerieke LR aan verbale termen als ‘*not worth more than a bare mention*’ (LR: 1-3), ‘substantial’ (3-10), ‘strong’ (10-30), en ‘very strong’ (30-100).⁶ In recenter tijden was het Irving John Good⁷ die de kansrekening gebruikte voor het wegen van bewijs, waarbij hij ook forensische toepassingen zag. Maar een serieuze toepassing van de probabilistische benadering in de moderne criminalistiek liet op zich wachten tot de jaren zeventig van de vorige eeuw in het forensische glasonderzoek.⁸ Het was daarna vooral de opkomst van het DNA-onderzoek⁹ vanaf de tachtiger jaren die ervoor zorgde dat deze benadering ook door begon te dringen tot rest van de forensische wereld.¹⁰

Het stelde de criminalistiek in staat om onderzoek en bewijskracht te objectiveren, en maakte duidelijk welke methodologie daarbij toegepast dient te worden. Bij vergelijkend onderzoek van sporen¹¹ is de vraag of een spoor van onbekende (of betwiste) oorsprong dezelfde herkomst heeft als een referentiespoor van bekende oorsprong. Voor het objectiveren van een dergelijk onderzoek gelden de volgende stappen:

- Eerst dient de waarneming zelf geobjectiveerd te worden, door een aantal relevante, objectieve kenmerken van de sporen (in algemene zin) vast te stellen.
- Daarna wordt een objectieve maat voor het verschil (of de overeenkomst) tussen twee sets van kenmerken gekozen: de score.
- Voor de interpretatie van deze score zijn achtergrondgegevens noodzakelijk. Hoe verschillend zijn sporen uit dezelfde bron, en hoe verschillend zijn sporen uit verschillende bronnen? Daarom worden voor vele bronnen uit een verzameling die representatief is voor de populatie, vergelijkingen tussen sporen uit dezelfde bron

⁶ Theory of Probability, Harold Jeffreys (1939), Oxford: The Clarendon Press.

⁷ Probability and Weighting of Evidence, I.J. Good (1950), Charles Griffin & Company Ltd.

⁸ A problem in forensic science, D.V. Lindley, Biometrika 1977, 64:207-213. The interpretation of refractive index measurements, I.W. Evett, Forensic Sci 1977, 9:209-217.

⁹ Interpreting DNA evidence - statistical genetics for forensic scientists, I.W. Evett & B.S. Weir (1998), Sinauer Associates.

¹⁰ Interpreting Evidence: Evaluating Forensic Science in the Courtroom, B. Robertson & G.A. Vignaux (1995), Wiley. Statistics and the Evaluation of Evidence for Forensics Scientists, Statistics in practice, 2nd edition. C.G.G. Aitken & F. Taroni (2004), Wiley.

¹¹ De meeste forensische onderzoeken behoren tot dit type, of het nu om DNA, kogels, vezels, glas, of vingersporen gaat.

gemaakt (kleinere verschillen). Datzelfde gebeurt voor sporen uit vele willekeurige verschillende bronnen (grotere verschillen).

- Met deze achtergrondgegevens kan voor een mate van verschil die is waargenomen in een zaak nu bepaald worden hoe waarschijnlijk die waarneming is onder de hypothesen van dezelfde of verschillende bron. De verhouding van die twee waarschijnlijkheden is gelijk aan de bewijskracht (LR) voor de hypothesen van gelijke of verschillende herkomst van de sporen. Daarmee is ook de interpretatie van de onderzoeksresultaten geobjectiveerd.

Ook wanneer de kenmerken niet gekwantificeerd kunnen worden geldt dezelfde logica. Om de bewijskracht te bepalen dient ook nu de waarschijnlijkheid van de waarnemingen onder beide concurrerende hypothesen geëvalueerd te worden. Zo luidt de verbale conclusiereeks van het NFI bijvoorbeeld:¹²

De onderzoeksresultaten zijn...

- ongeveer even waarschijnlijk...[als]
- iets waarschijnlijker...
- waarschijnlijker...
- veel waarschijnlijker...
- zeer veel waarschijnlijker...

...wanneer hypothese 1 juist is, dan wanneer hypothese 2 juist is.

Door de vooruitgang van de forensische interpretatie en evaluatie van bewijs is meteen ook duidelijk geworden hoe ingewikkeld dat eigenlijk is. Daarbij zijn een aantal problemen die er altijd al waren nu duidelijk aan het licht gekomen: de zogenoemde ‘*fallacies*’ of denkfouten. De twee belangrijkste *fallacies* zijn genoemd naar de aanklager en verdediging: de zogenoemde *prosecutor's fallacy* en *defence attorney's fallacy*.¹³

Hoewel de benamingen suggereren door wie deze denkfouten vooral zouden worden begaan, blijkt in de praktijk de prosecutor's fallacy een denkfout die vaak en door alle partijen gemaakt wordt. Deze denkfout wordt ook wel aangeduid met ‘verwisseling van voorwaarden’. Een veel voorkomend voorbeeld is de verwisseling waarbij een conclusie als “*de kans op een matchend DNA-profiel, gegeven dat de*

¹² De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs, Vakbijlage NFI, 2008.

¹³ Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defence attorney's fallacy. W.C. Thompson & E.L. Schumann, Law and Human Behaviour 1987, 11:167-187.

verdachte niet de donor is van het spoor, is 1 op 1 miljard”, ten onrechte wordt gelezen als “*de kans dat de verdachte niet de donor is van het spoor, gegeven het matchende DNA-profiel, is 1 op 1 miljard*”.

Hoe fout deze verwisseling is zien we duidelijker met een eenvoudiger voorbeeld als “*de kans dat een dier 4 poten heeft, gegeven dat het een koe is*” wat duidelijk niet hetzelfde is als “*de kans dat een dier een koe is, gegeven dat het 4 poten heeft*”. In zijn algemeenheid wordt bij deze denkfout “*de kans op de waarneming gegeven de hypothesen*” verward met “*de kans op de hypothese gegeven de waarneming*”.

Bij de *defence attorney's fallacy* wordt een conclusie als “*de kans op een matchend DNA-profiel, gegeven dat de verdachte niet de donor is van het spoor, is 1 op 1 miljard*” weggewuifd door daaraan toe te voegen dat de verdachte (met de huidige wereldbevolking) dus slechts 1 van de 7 mensen is die de donor van het spoor kunnen zijn. De denkfout is dat hier impliciet, en ten onrechte, is aangenomen dat de gehele wereldbevolking in gelijke mate in aanmerking komt als donor, iets dat in werkelijkheid nooit zal gelden. Een soortgelijke fout is om ook bij dergelijk sterk bewijs te stellen dat verdachte “*hoogstens niet is uitgesloten*”.

Het bovenstaande is geen gemakkelijke materie, maar hopelijk heeft u deze ‘wetenschappelijke injectie’ niet als al te pijnlijk ervaren, en is tenminste duidelijk geworden dat meer kennis hierover van groot belang is voor het wegen van bewijs en daarmee voor het nemen van een juiste beslissing.