

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/272156095>

Over Statistisch Bewijs

Article · January 2008

CITATIONS

0

READS

255

2 authors:



[Richard David Gill](#)

Leiden University

245 PUBLICATIONS 17,771 CITATIONS

[SEE PROFILE](#)



[Aernout Schmidt](#)

Leiden University

35 PUBLICATIONS 62 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



AOVA-2010 [View project](#)



Forensic statistics [View project](#)

Over statistisch bewijs

1. Inleiding en verantwoording

Op 16 februari van dit jaar ontmoetten de auteurs elkaar in een publiek debat over de betekenis van statistisch bewijs en rechtspraak. We spraken af onze argumenten op schrift te stellen voor een bijdrage aan dit tijdschrift. Het stuk bestaat uit drie delen: de visie van Schmidt, jurist (paragraaf 2); de visie van Gill, statisticus (paragraaf 3) en een gezamenlijke analyse (paragraaf 4). Doel is te komen tot concluderende aanbevelingen aan 'de' strafrechter voor de beoordeling van voorgelegde statistische expertise. Daarbij moeten disciplinaire verschillen in aanpak en denkwijze zichtbaar worden gemaakt en behandeld, en, zo leert de ervaring, moeten we er rekening mee houden dat botssende vakjargons en -culturen er nogal eens aan bijdragen dat een multidisciplinair discours ontaardt in wederzijdse stigmatisering die ons niet verder helpt. We kozen de volgende aanpak, die zijn weerslag vindt in hoe dit stuk is vormgegeven.

1. Paragrafen 2 en 3 zijn onafhankelijk opgetekend door de auteurs.
2. De auteurs hebben, onafhankelijk, in elkaars bijdrage die passages aangegeven die zij ofwel onjuist vonden, ofwel als onnodig kwetsend voor hun discipline ervoeren. Aan de hand van die passages is intensief, langdurig en informatief overleg gevoerd ter opheldering van de geïdentificeerde aandachtspunten. De resulterende annotaties, waarin de auteurs elkaar vonden, zijn in een herkenbare typografische stijl weergegeven en dichtbij de passages in kwestie ingevoegd in de oorspronkelijke tekst. In vervolg hierop is de concluderende paragraaf gezamenlijk geformuleerd. (Omdat Gill geen Nederlandstalig *native speaker* is heeft Schmidt de eindredactie op zich genomen).
3. Het resultaat is onderworpen aan een proces van *peer-review* dat tot verbeteringen in de tekst heeft geleid, en voorts aanleiding heeft gegeven tot het opnemen van enkele eindnoten ter opheldering van onduidelijkheden.

2. Rechtspraak: kennis of gokwerk? (Aernout Schmidt)

2.1. Inleiding

Wie, als ik – overigens met Socrates – van oordeel is dat de wetenschap en de strafrechter steeds *moeten* zoeken naar de waarheid, maar nooit *mogen* menen die in pacht te hebben ziet van verre aankomen waar dit betoog naar toe gaat. Bewijs is nooit absoluut.

Anders gezegd: elke op wetenschap gebaseerde conclusie is in zekere mate gokwerk – en zo ook het strafrechtelijk oordeel. Dat betekent dat er knopen moeten worden doorgehakt. In het strafrecht laten we dat over aan rechters. Daarmee verschepen we een hoop vertrouwen in hun richting. Draagvlak voor dat vertrouwen schuilt in de openbare verantwoording voor rechterlijke beslissingen, en, wanneer er door technici of andere deskundigen achteraf wordt geoordeeld dat er fouten zijn gemaakt, in een authentiek debat (en, eventueel, herziening). Voor dit blad, dat over de relatie tussen expertise en recht gaat, is vooral van belang dat er tussen rechters en experts hoe dan ook sprake is van wederzijdse kennisasymmetrie. En dat het, juridisch bezien, van grote betekenis is dat rechters de argumentaties van experts kunnen beoordelen.

Andersom is het belang dat experts de argumentaties van rechters kunnen beoordelen vooral daarin gelegen, dat zij de voorgelegde vraag zo getrouw mogelijk vertalen in de modellen en werkwijzen van hun discipline.

Aanleiding tot dit betoog is de veenbrand in het publiek debat die de behandelingen van de zaak Lucia de B. hebben ontstoken. Op de zaak zelf ga ik niet in. Waar ik naar toe wil is – uiteindelijk – het benoemen van de kennis waarover een rechter behoort te beschikken om tegelijkertijd een zorgvuldige rechter te zijn én om statistisch bewijs te beoordelen dat door deskundigen wordt voorgespiegeld. Daartoe eerst iets over kansrekening, dan iets over statistiek, ten slotte iets over kennisasymmetrie.

2.2. Kansrekening en verdelingen

Kansrekening gaat vooraf aan de statistiek en pleegt ons regelmatig op het verkeerde been te zetten. Velen van ons hebben intuïtief het gevoel dat wie met een dobbelsteen vijf keer achtereenvolgens een 1 heeft gegooit, dat de zesde keer niet zal doen – dat de kans van die zesde keer een 1 minder is dan éénzesde. Quod non. Tenminste, wanneer de dobbelsteen deugt. En een dobbelsteen deugt – als dobbelsteen – wanneer het werpen van de dobbelsteen ertoe leidt dat elk vlak, gemiddeld genomen, even vaak bovenaan eindigt; dat de kans om dat te doen voor elk vlak even groot is. De kansberekening over de uitkomst van dobbelen gaat daarvan uit. Afgezien van de wiskunde (die weliswaar soms intuïtief verrassende resultaten oplevert, maar die moeiteloos kan worden overgelaten aan specialisten) is de berekende kans zo goed en zo slecht als de aannamen over de achterliggende verdelingen. Een

* Prof. dr. R.D. Gill is hoogleraar mathematische statistiek Universiteit Leiden.

** Prof. mr. A.H.J. Schmidt is hoogleraar recht en informatica Universiteit Leiden en redacteur van dit tijdschrift.

rechter die wordt geconfronteerd met een door een deskundige geformuleerde kans op een gebeurtenis moet dus niet alleen nagaan of er goed gerekend is, maar vooral hoe goed de aannamen over de achterliggende verdeling(en) zijn.

Annotatie 1. Dit beeld is onvolledig. In elk geval moet óók zijn/worden beoordeeld hoe goed de informatie waarmee wordt gerekend als datagegevens zijn geselecteerd, geïnterpreteerd, gecodeerd, geschaald, gecombineerd en gemodelleerd. Ook moet heel goed worden nagedacht of de kans in kwestie werkelijk de meest relevante of betekenisvolle kans is. De kans op A, of de kans op A gegeven B, of de kans op B gegeven A? En misschien is het veel betekenisvoller om het quotiënt tussen twee kansen te weten, bijvoorbeeld: hoeveel waarschijnlijker is A onder omstandigheden B dan onder omstandigheden C? Ten slotte valt te overwegen om eenzelfde kansvraag vanuit twee verschillende paradigmata (zie annotatie 10) te doen berekenen en te vergelijken.

En daarvoor is zelden meer nodig dan de algemene ontwikkeling waarover rechters toch al plegen te beschikken.

Annotatie 2. Mede op grond van de vorige annotatie lijkt het verdedigbaar dat voor beoordeling van door statistische deskundigen aangedragen rapportages meer nodig is dan algemene ontwikkeling. In de afsluitende beschouwing komt dit punt terug.

Kansrekening kan in twee situaties worden gebruikt. Ten eerste om de waarschijnlijkheid te voorspellen dat een bepaalde gebeurtenis zich zal voordoen. Hoe groot is de kans dat Nederland, bij ongewijzigd beleid, binnen honderd jaar zal overstromen? Bij dit soort kansberekening gaat het om het inschatten van risico's en wat ertegen te ondernemen. En bij de kwaliteit van de gepresenteerde kans gaat het niet alleen om de vraag of de achterliggende verdelingen goed zijn gekozen, maar ook of alle relevante omstandigheden zijn meegenomen. Toen we dergelijke berekeningen uitvoerden ter voorbereiding van de deltawerken waren we er nog niet aan toe om de invloed van mogelijke klimaatveranderingen mee te nemen. In de Nederlandse strafrechtspraak waren we er in die tijd evenmin aan toe om na te denken over de kans dat onder verantwoording van het OM aangevoerd bewijs kwalitatieve gebreken zou kunnen vertonen. Ook daarin is verandering opgetreden, maar daarover later. In beide gevallen spelen causale verklaringen (bijvoorbeeld: CO₂-uitstoot, kokervisie) een rol bij de beoordeling van de achterliggende verdelingen.

De tweede situatie waarin kansrekening kan worden gebruikt is wanneer we een gebeurtenis die heeft plaatsgevonden beoordelen. Wie er getuige van is dat een dobbelaar niet vijf, maar tien, twintig keer achter-

een een 1 gooit gaat zich afvragen of de dobbelsteen wel deugt. In dat geval is de gebeurtenis op zichzelf niet het *bewijs* dat er iets mis is met de kansrekening of de achterliggende aannamen, maar wel een aanwijzing die zou kunnen aanzetten tot nader onderzoek. De kansrekening wijst op de ongewoonheid van de gebeurtenis. De kansrekening geeft er geen causale verklaring voor. Die verklaring (bijvoorbeeld: waargenomen manipulatie met magneetjes) levert eventueel bewijs. Kansrekening levert op zichzelf *nooit* bewijs.¹

2.3. Statistiek en statistisch bewijs (Bovarysme)

Wat is niet zo met statistisch bewijs. Maar: wat is dat eigenlijk? Statistisch bewijs wordt bijvoorbeeld gebruikt om na te gaan of er tussen variabelen een verband bestaat. Wie van een groot genoeg aantal Nederlanders meet hoe lang ze zijn en hoeveel ze wegen zou de overtuiging kunnen krijgen dat er verband bestaat tussen lichaamslengte en -gewicht. Ook daaraan komt weer rekenwerk te pas. En ook daar spelen – naast de selectie van personen die worden gemeten en gewogen, de steekproef – kansrekening en aannamen een belangrijke rol. De gedachte is hier dat, wanneer geen vooroordeel wordt ingebakken in de opbouw van de steekproef, alle andere redenen die de oorzaak zouden kunnen zijn voor individuele afwijkingen van een eventueel te vinden verband elkaar zullen opheffen. Op die manier kan voor verbanden tussen variabelen statistisch bewijs worden gevonden, waarvan het gezag kan worden uitgedrukt in de kans dat het gevonden verband toevallig is (als gebaseerd op genoemde aanname). Het verband kan worden uitgedrukt in een ratio tussen lengte en gewicht, en aangevuld met een bandbreedte die behoort bij de waarschijnlijkheid dat wanneer de lengte van een willekeurige Nederlander bekend is, diens gewicht erbinnen zal vallen. Zo kan het verband tussen lengte en gewicht van Nederlanders statistisch worden bewezen.

Annotatie 3. Statistisch bewijs is geen causaal bewijs. Wanneer de oorzaak voor de gevonden relatie een andere is dan die tussen de gemeten variabelen wordt gesproken van een *spurious correlation*, een verborgen verband. Wie in 1854 vaststelt dat de bewoners van enkele woonblokken in Soho een uitzonderlijke kans lopen op een cholerabesmetting vindt een relatie tussen woongebied en ziekte. In feite leidde dat ertoe dat de bewoners het gebied ontvluchtten. Toch ging het om een *spurious* statistisch verband. Het 'verborgen' causale verband was, zo werd en wordt aangenomen, tussen het besmette water dat destijds uit de pomp van Broad Street vloeide en het nuttigen ervan.

Voor zover het over gedrag gaat – en dat is in het strafrecht het geval – doet statistisch bewijs uitspraken over *algemeenheden* in sociaal gedrag, niet over individuele gevallen. Statistisch bewijs is doorgaans onbruikbaar in het strafproces en wel om de volgende reden.

De mooiste term die ik ooit gebezigd zag, toepasselijk op het projecteren van statistisch bewijs op individuele gevallen is 'Bovarysme', de met Madame Bovary in verband te brengen behoefte om een model (in haar geval: een modieuze romantische levensambitie) toepasselijk te zien op het individuele geval (in haar geval: haar eigen leven). Bovarysme is de soms tragische noodzaak voor wie rationele keuzen wil maken en te weinig weet.²

Maar Bovarysme in strafzaken op basis van statistisch bewijs moeten we niet doen. En elke redelijke statisticus zal zich beroepsmatig verzetten tegen de manoeuvre waarbij een statistisch verband van toepassing wordt verklaard op een individueel geval. Het is dan ook opmerkelijk dat er zoveel deskundigen bereid zijn gebleken om over de zaak Lucia de B. in termen van kansen te spreken.

Annotatie 4. Aan statistici wordt verzocht om een deskundigenbericht. Doorgaans gebeurt dat in de opsporingsfase en soms ter terechtzitting, maar dan meestal om uitleg te geven over wat eerder werd bericht. En al ware dat anders: een statisticus hoeft niet te zwijgen omdat zijn uitspraken verkeerd kunnen worden gebruikt in een context waar hij geen bevoegdheid heeft.

Om aan te geven hoe belangrijk de daarbij gehanteerde aannamen zijn geef ik een overzicht van vier verschillende, door deskundige aangedragen kansen dat Lucia de B. toevallig zo vaak bij overlijdens was als kennelijk het geval was:³

Statisticus 1: 1 op 342 miljoen
Statisticus 2: 1 op 1 miljoen
Statisticus 3: 1 op 50
Statisticus 4: 1 op 9

Deze verschillen zouden elke rechter (en elke statisticus) te denken moeten geven. Ze ondersteunen mijn aanbeveling om in voorkomende gevallen de gebruikte aannamen te doen expliciteren en toelichten.

Annotatie 5. De 'kansen dat Lucia de B. toevallig zo vaak bij overlijdens was als kennelijk het geval was' zijn andere dan in de weergegeven getallen worden geschat. Dit is namelijk 'de kans dat een onschuldige verpleegster even veel of meer incidenten zou meemaken als Lucia.' Als jurist geeft Schmidt de voorkeur aan een schatting van de eerste (er al dan niet toevallig bij zijn) boven een schatting van de tweede kans (een onschuldige evenveel incidenten zou meemaken). Bij de tweede kansberekening moeten immers variabelen en indicatoren worden bedacht, gecodeerd, gemeten en gemodelleerd voor moeilijk kenbare begrippen als 'onschuldige' en 'incident' (waaronder te verstaan: iets als verdacht of onverwacht overlijden op de betreffende afdeling). En daarmee worden zijns inziens, onnodig veel aan-

namen en foutmogelijkheden in het statistische werk verweven. Dat neemt niet weg dat de voorstelling van zaken als door Schmidt in de hoofdtekst gegeven incorrect is.

Maar er is in mijn ogen weinig mis met de aandacht die erdoor bij het OM is gewekt voor de zaak, ook al is die gebaseerd op ondeugdelijke kansberekening. Wat wél ernstig zou zijn is wanneer die voorstelling van zaken bij het OM tot een vorm van kokervisie zou hebben geleid bij de eigenlijke bewijsgaring, -selectie en -presentatie.

2.4. Kennisasymmetrie (mechanism design)

Als op zichzelf staand strafrechtelijk bewijs zijn kansrekening en statistiek betekenisloos of dienen dat te zijn. Kansrekening en statistiek zijn in de strafrechtketen bruikbaar om regelmatigheden te beschrijven en, daarmee, om uitbijters (extreem afwijkende gevallen) te identificeren voor nader onderzoek. Ook daarbij is het zorgvuldig om na te gaan met welke gegevens is gerekend, hoe is gerekend en op basis van welke vooronderstellingen over verwachte verdelingen. Daarmee zou ik deze bijdrage kunnen afsluiten, ware het niet dat ik van oordeel ben dat de discussie over statistisch bewijs, ook naar aanleiding van het geval Lucia de B., een symptoom is van de manier waarop we in de strafrechtketen omgaan met kennisasymmetrieën. En daarover valt misschien iets te leren vanuit de economie.

De Nobelprijs voor de economie is in 2007 aan een drietal economen toegekend, onder wie Myerson. In zijn aanvaardingsbetoog, dat via internet kan worden gehoord en gezien wordt aandacht besteed aan de betekenis van kennisasymmetrie (zij het met de kanttekening dat Myerson geen hoge ogen zal gooien in een redevoerings-contest). Opmerkelijk is daarbij een aantal van de aannamen die worden gemaakt, en die vooral tot uitdrukking komen in de benoeming van de mogelijke strategische houdingen van wie deelnemen aan de speltheoretisch beschreven situaties. Die situaties gaan allemaal over economische transacties en hoe daarbij informatieverschillen bij de deelnemers zullen uitwerken. Aangenomen wordt dat de (economisch) beste resultaten voor iedereen het gevolg zullen zijn van transacties waarbij partijen hun relevante kennis met elkaar delen. Als probleem voor het bereiken van die toestand wordt gezien dat partijen daartoe gemotiveerd dienen te worden, omdat ze anders veelal op suboptimale resultaten zullen aansturen. Myerson onderscheidt een aantal strategische houdingen die, mogelijk in combinatie, om adequate motivering (hier niet: toelichting) vragen. Van individuen die voor zich opereren:

1. de bereidheid om een uitgebracht bod al dan niet bij te stellen (hier verder het risico van de onwrikbare stellingname) en
2. de bereidheid om al dan niet te liegen (hier verder het onwaarachtigheidsrisico),

en, aanvullend van individuen die een organisatie vertegenwoordigen:

3. de bereidheid om het eigen belang al dan niet voor te laten gaan op het belang van de organisatie (hier verder het loyaliteitsrisico).

Myerson cum suis hebben zo een nieuwe tak van economensport ontwikkeld die ze *mechanism design* noemen. Die nieuwe tak van sport formuleert, onderzoekt en toetst causale relaties die liggen achter maatschappelijke evenwichten die volgen op de manier waarop regelcomplexen de subjecten motiveren. En die wiskundig kunnen worden uitgedrukt. En die – erger nog – daadwerkelijk worden toegepast bij het beoordelen en in elkaar zetten van juridische modellen voor, bijvoorbeeld, het strategisch op juiste wijze motiveren bij milieu- en ontwikkelingsproblemen.

Het is in zekere zin ironisch om op de plaats waar serieus wordt nagedacht over kennisasymmetrieën, opnieuw terecht te komen bij ingewikkelde wiskundige modellen die sterk steunen op aannamen. Ze zijn evenwel van geheel andere aard. Ze richten zich op de rationele benadering van de problemen die aan transacties zijn verbonden waar sprake is van kennisasymmetrieën – met name met het oog op het inrichten van bijbehorende regelstelsels die de deelnemers motiveren om zoveel mogelijk relevante informatie bekend te maken. Het is sterk de vraag of de modellen die daarbij in de economie naar voren treden, en die gericht zijn op de uitwisseling van op geld waardeerbare waarden, van toepassing zullen zijn op door het strafproces gereguleerd gedrag. Toch komt het me voor dat er alle aanleiding is om te onderzoeken of een analoge analyse van procesrechtelijke regulering zinvolle resultaten zou opleveren voor de relaties die in de strafrechtketen door kennisasymmetrieën worden gekenmerkt. Wat beweegt de vier statistici om tot hun uiteenlopende expert-inbreng te komen? Wil de eerste zich onderscheiden door aan het begin van een spraakmakende strafzaak te staan? En vindt de laatstgenoemde het lekker om achteraf de onkunde van de rechtspraak aan de kaak te stellen? We weten het niet, maar moeten met beide mogelijkheden rekening houden. En ervoor zorgen dat dergelijke motieven geen kans krijgen. Waar het gaat om de beoordeling van statistisch bewijs en de resultaten van kansberekeningen zijn de met deze vragen gemoeide problemen minder groot dan het wel eens lijkt.

Overigens is het geenszins uitgesloten dat er naast de kennisasymmetrieën die bestaan tussen de deskundige en de rechter, er ook kennisasymmetrieën zijn tussen het OM en de rechter. En, in dat licht, dat de rechter ertoe moet zijn uitgerust om door eventuele waarheidsvijandige motieven van het OM heen te zien bij de beoordeling van het materiaal dat hem wordt voorgehouden.⁴ Dat zou nog wel eens moeilijker kunnen zijn dan bij deskundigeninbreng. Maar dit is een zijlijn (die mogelijk elders zelfstandig aandacht verdient).

2.5. Afronding

Samenvattend gaat het bij de beoordeling van statistisch bewijs om de volgende kengebieden:

1. Onwetendheid (we weten in het strafproces nooit alles en moeten dat inzicht blijven respecteren, ook wanneer statistische expertise wordt voorgelegd. Bijvoorbeeld wanneer het verleidelijk is minder aandacht te besteden aan informatie waaruit ontlastende c.q. alternatieve verklaringen kunnen worden afgeleid (kokervisie));
2. De aannamen over verdelingen waarop uitspraken over waarschijnlijkheid en statistische verbanden rusten (hierop wordt in de bijdrage van Gill nader ingegaan);
3. De wiskundige correctheid van de gebruikte berekeningen;
4. Waar statistisch 'bewijs' juridisch toelaatbaar is en waar niet (wel ter signalering van uitbijters, niet als *causa* bij verklaringen);
5. Het eventueel *in casu* voorkomen van waarheidsvijandige houdingen die in het algemeen in verband worden gebracht met kennisasymmetrieën: de bereidheid al dan niet (i) vast te houden aan een ingenomen positie, (ii) te liegen en/of (iii) het eigen belang te laten voorgaan op dat van de organisatie die wordt vertegenwoordigd.

Met uitzondering van de wiskundige correctheid liggen geen van deze kengebieden buiten bereik van het beoordelingsvermogen van de rechtspraak.

Annotatie 6. Met wiskundige correctheid wordt hier het rekenwerk bedoeld dat ingezet wordt nadat de aannamen zijn gedaan en de informatie is geselecteerd, geïnterpreteerd, geschaald, gecodeerd en gemodelleerd. Op de vraag of het beoordelen van genoemde activiteiten binnen bereik van niet in het bijzonder daartoe geschoolde rechters ligt wordt, als gezegd, teruggekomen.

En voor het controleren van die wiskundige correctheid is contra-expertise eenvoudig te verwezenlijken (er is geen tekort aan experts op dit gebied).

3. Over verschillen tussen statistici (Richard Gill)

Mijn denken over expertise in rechtszaken is sterk gekleurd door de paar gevallen die ik grondig bestudeerd heb en waar – in mijn opvatting – het vrijwel zeker is dat er is gedwaald, mede door de kwaliteit van expertinbreng. In het bijzonder, de zaken Lucia de B. en Kevin Sweeney. Beide gevallen zijn met tunnelvisie door politie en OM gepresenteerd. Al spoedig meende men te weten wie de moordenaar was en was alles erop gericht om dat aan te tonen. Een opmerkelijk aspect is dat er in deze twee zaken mogelijk niet eens sprake is geweest van op moord of doodslag gerichte opzet van een dader; in beide gevallen ontbrak een reconstructie van wat in overeenstemming met de evidentie feitelijk gebeurd kon zijn. Kortom: ik

vind dat het onderzoek niet deugde, dat het bewijs minimaal was, dat de onderbouwing ontbrak en – wat hier ons eigenlijke onderwerp is – dat de inbreng van experts tegelijkertijd aanvechtbaar was en bij de rechtszaken een belangrijke rol speelde. En dus: dat het belangrijk is de aandacht te richten op de vraag of, en zo ja hoe, deze problematiek beter onder ogen kan worden gezien. In mijn verdere betoog gebruik ik de zaak Lucia de B. als illustratiemateriaal.

Annotatie 7. De *statisticus* Richard Gill heeft vanuit zijn vakkennis in de genoemde rechtszaken gebreken gezien die hem *als burger* hebben doen besluiten in verzet te komen en actie te ondernemen, onder meer door het plaatsen van een advertentie. Statisticus en burger zijn één. Dat moet de lezer weten, omdat daarmee het gevaar van wetenschappelijke kokervisie niet denkbeeldig is. Dat neemt niet weg dat deze bijdrage – voor zover dat in zijn vermogen ligt en Schmidt daarop kan toezien – geen bijdrage is van de burger, maar van de statisticus.

Het is dus inderdaad buitengewoon zinnig om in te gaan op de verschillen tussen de cijfers van de vier statistici als genoemd door Schmidt, en dat doe ik graag meteen.

3.1. Statisticus 1

Statisticus 1 heeft een technische fout gemaakt bij het presenteren van een combinatie van drie p-waarden (overschrijdingskansen). Onder collega's heeft statisticus 1 dat inmiddels volmondig erkend. Er waren namelijk gegevens van drie afdelingen, van twee verschillende ziekenhuizen, die gecombineerd moesten worden. Van de ene afdeling was de coïncidentie van incidenten met diensten van Lucia overrompeling, bij de twee andere was de coïncidentie veel zwakker.

Annotatie 8. Een bekende aanpak van het combineren van kansen is door hen te vermenigvuldigen. Wie de kans wil weten dat er 18 keer achtereens 1 wordt gegooid met een dobbelsteen voert een vermenigvuldiging uit van 18 factoren, alle gelijk aan 1/6. Die aanpak is echter aanvechtbaar wanneer niet één reeks van 18 worpen wordt beschreven, maar drie willekeurige reeksen van zes worpen worden gecombineerd – en zeker wanneer reeksen worpen worden bekeken vanuit de vraag of de dobbelsteen niet deugt en te vaak een één laat zien. Wat een aanwijzing had kunnen zijn is dat de combinatie van groepen gegevens met zeer verschillende overschrijdingskansen niet leidde tot een tussenliggende, maar tot een nog extremer waarde voor het geheel. Wie vandaag met één worp Yahtzee gooit wordt niet in extrema *gesterkt* in het vermoeden dat de stenen niet deugen wanneer hij morgen twee paren, waaronder tweemaal een 1, gooit en de dag daarna nog eens.

Iedere statisticus, ook statisticus 1, is het er mee eens dat, binnen zijn eigen paradigma, de p-waarde van 1 op 342 miljoen nooit en te nimmer genoemd had mogen worden. Verschillende correcte methoden om de cijfers van de drie afdelingen te combineren, binnen het gebruikte paradigma van statisticus 1, leiden tot p-waarden rondom de 1 op 100 duizend. Ik vermoed dat het noemen van zo'n cijfer, noch in de rechtszaal, noch in de media, noch in het openbare debat een vergelijkbare psychologische *impact* zou hebben gehad. Zoals ik later zal suggereren, leefden de kansen voort in de overwegingen van het hof, lang nadat men besloten had een statistische kansberekening te verbannen uit de motivatie van de uiteindelijke veroordeling.

Annotatie 9. Hoe zeer het ook begrijpelijk is om opinies te uiten over de waarachtigheid van rechterlijk handelen wanneer daar twijfels over zijn opgekomen, toch moet benadrukt worden dat het niet om een statistisch gefundeerd deskundigenoordeel gaat maar over een persoonlijke, op *common sense* rustende interpretatie. Te overwegen valt om dergelijke beschouwingen, als *in loco* op het eerste gezicht irrelevant, geheel weg te laten, vooral ook omdat ze kunnen bijdragen aan de indruk dat (ook) in de op wetenschap rustende argumentatie vooroordelen zijn geslopen. Wij hebben dat niet gedaan. Ten eerste omdat de passage laat zien hoe verleidelijk het is (voor statistici én voor juristen) om elkaars terreinen te betreden en ten tweede omdat zij een aanknopingspunt biedt erop te wijzen dat wel degelijk relevante *meer algemene* sociaal-wetenschappelijke evidentie bestaat omtrent de beïnvloeding van de objectieve meningsvorming door vooraf gepresenteerde retoriek.

We moeten op dat laatste wél bedacht blijven. Bijvoorbeeld op de maatschappelijke en individuele doorwerking van *moral panics*, die niet zelden ontstaan door met overtuiging in de media gepresenteerde, al dan niet met wetenschappelijke autoriteit verzwaarde, speculatie over maatschappelijke risico's. (Verwezen zij, bijvoorbeeld, naar: Stanley Cohen, *Folk devils and moral panics*, MacGibbon and Kee, 1972 en, vanuit een andere invalshoek: Mary Douglas, *Risk and Blame: Essays in Cultural Theory*, Routledge, 1992). En ook op de beïnvloeding van individuele oordeelsvorming door de vormgeving van voorafgaande stellige uitspraken waaraan gezag wordt toegekend. (Verwezen zij hier, voor het verschijnsel *framing*, naar Todd Gitlin, *The Whole World is Watching: Mass Media in the Making and Unmaking of the Left*, University of California Press, 1980 en, vanuit een meer institutioneel-psychologische invalshoek naar de bekende experimenten van S.E. Asch, 'Effects of Group Pressure Upon the Modification and Distortion of Judgements', in: H. Guetzkow (ed.) *Groups, Leadership, and Men*, 1951).

Overigens, Aernout Schmidt heeft het over 'de kans dat het gevonden verband toevallig is.' Statistici 1 en 2 hebben hun best gedaan om het hof ervan te verwittigen dat deze woorden beslist niet de juiste interpretatie van de gepresenteerde p-waarden zijn. In onze situatie is de p-waarde de kans dat een onschuldige verpleegster evenveel of meer incidenten zou meemaken als Lucia uitgaande van allerlei veronderstellingen over het normale voorkomen van incidenten in een ziekenhuisafdeling. Dit misverstand over de betekenis van de gepresenteerde p-waarde heeft ook zijn bijdrage gehad in de opinievorming via de media. Wil je Schmidts kans weten, dan moet je een Bayesiaanse aanpak volgen, en moet je onder meer aannamen doen over hoe vaak seriemoordenaars actief zijn in Nederlandse ziekenhuizen. Des te zeldzamer een seriemoordenaar, des te waarschijnlijker dat het gevonden verband juist wel gevolg van toeval is (logisch, toch?). Verschillende onderzoekers hebben aangetoond dat onder de meest voorzichtige aannamen over het voorkomen van seriemoord, Lucia's coincidentie net zo goed aan toeval te wijten kan zijn als aan moord. Het is een aanwijzing dat er misschien iets aan de hand is. Alleen, het zegt heel weinig.

Annotatie 10. Statistische benaderingen verschillen in clusters, die worden gekenmerkt door verschillen in uitgangsposities, in basisaannamen. Die uitgangsposities worden hier paradigmata genoemd. Ze betreffen de interpretatie van het begrip waarschijnlijkheid. Drie paradigmata zijn van bijzonder belang.

Ten eerste het *frequentistische* (waarbij waarschijnlijkheden van willekeurige gebeurtenissen verwijzen naar hun relatieve frequentie als waargenomen in herhaalde experimenten; de frequentistische benadering is afhankelijk van laboratoriumcondities, van de *ceteris paribus clause* en van herhaalbaarheid, en speelt een belangrijke rol bij sociaal-wetenschappelijke experimenten).

Ten tweede het *logische* (waarbij waarschijnlijkheden worden opgevat als de objectieve gradaties van oorzakelijkheid – bijvoorbeeld wanneer wordt gesproken over de kans op het krijgen van cholera voor iemand met een bepaalde bloedgroep na het drinken van in een bepaalde graad besmet water).

Ten derde het *Bayesiaanse* (waarbij waarschijnlijkheden worden opgevat als subjectieve schattingen – overtuigingen dus, of opinies – over de waarschijnlijkheid van indicatoren, die bij de berekening van de waarschijnlijkheid van de daarvan afhankelijk geachte gebeurtenissen een rol spelen; het Bayesiaanse model wordt, bijvoorbeeld, met succes toegepast om de kans te bepalen dat een e-mailbericht SPAM is, op basis van de eigenschappen van een berichtenverzameling die door de ontvanger eerder als SPAM werd geduid).

Om terug te keren tot het dobbelsteenvoorbeeld: de kans dat de volgende worp een 1 zal zijn bedraagt

in alledrie de paradigmata $1/6$. Vanuit frequentistisch gezichtspunt omdat herhaalde experimenten met grote aantallen worpen dat hebben laten zien; vanuit logisch gezichtspunt omdat de fysieke vorm van de dobbelsteen dat, in combinatie met onze kennis over de variabiliteit in luchtweerstand en overgedragen energie, voorspelbaar maakt; vanuit Bayesiaans gezichtspunt omdat we de ervarings-overtuiging hebben dat die kans $1/6$ is.

Alle drie de benaderingen zijn succesvol gebleken in praktische toepassingen. In de ontwikkeling van de statistische wetenschap hebben de drie benaderingen tot discussie en tegenstellingen geleid die gaandeweg minder scherp zijn geworden, mede doordat in de natuurwetenschappen het besef heeft postgevat dat er verschijnselen zijn die niet anders dan in waarschijnlijkheden kunnen worden beschreven. Thans is het zo dat elke statisticus desgevraagd vanuit alle drie de paradigmata een uitspraak zou moeten kunnen doen.

Omdat er bij de verdediging geen statistische tegenexpertise werd ingezet, en omdat de rapporten van de deskundigen lange tijd voor disciplinegegoten geheim waren, is de 'combinatiefout' tijdens het proces in eerste aanleg niet aan het licht gekomen en niet gecorrigeerd.

3.2. Statisticus 2

Statisticus 2 is wellicht statisticus 1 die zich beperkt tot de gegevens van die ene, opvallende periode in die ene afdeling waar de hele affaire gestart is. Lucia is ook aangeklaagd voor moorden op nog twee eerdere werkplekken, naast de twee die gegevens leverden voor statisticus 1. Daar is nooit enige cijferwerk mee gedaan. Men kan zich afvragen wat er zou zijn gebeurd als men nog meer statistische gegevens had verzameld.

Annotatie 11. De verschillen tussen statisticus 1 en 2 berusten op (a) een modelleerfout bij de combinatie van datasets en op (b) een verschil in gegevensselectie. Het komt ons voor dat het niet vanzelf spreekt dat de combinatiefout, ook als de daarbij gehanteerde aannamen worden geëxpliciteerd, door de rechter vanuit zijn eigen expertise zal kunnen worden herkend. Ten aanzien van de gegevensselectie geldt dat, zo menen we, niet.

3.3. Statisticus 3

Statisticus 3 is teruggegaan naar de oorspronkelijke ziekenhuisgegevens en heeft vastgesteld dat de ziekenhuizen die gegevens aan het OM leverden, dat op nogal subjectieve manier deden. Er was geen systematische registratie van 'reanimaties,' noch anderszins van 'incidenten.' Medewerkers moesten uit hun geheugen putten en vervolgens zoeken naar indicaties in incomplete logboeken, patiëntendossiers, die niet voor dat doel waren ingericht. Een helder criterium voor

wat als incident geldt is niet geformaliseerd en in een coderingsvoorschrift vastgelegd. Men heeft gevoelsmatig gekozen voor gebeurtenissen rondom de aanzet of het einde van een medisch incident dat zich over meerdere uren uitstrekte. Wegens overlap tussen de diensten van verpleegsters was evenmin helder of een incident wel of niet in de dienst van een bepaalde verpleegster viel. Een verpleegster die normaal een van de drie diensten per dag doet, is feitelijk aanwezig tijdens gedeelten van alle diensten op een dag. Kortom: doordat niet is vastgehouden aan een helder onderzoeksprotocol zijn de gegevensverzamelingen waarmee is gerekend ruimschoots voorzien van vooroordeel, *bias*.

Statisticus 3 is teruggegaan naar oorspronkelijke ziekenhuisregistraties en heeft gepoogd op een uniforme, objectieve manier incidenten te bepalen en aan diensten toe te wijzen. Men heeft dit tijdrovende en arbeidsintensieve werk niet afgemaakt: een stuk of drie nieuwe incidenten buiten Lucia's diensten om vond men, als aanwijzing al krachtig genoeg. Tegelijkertijd heeft men een aantal incidenten waar Lucia aanwezig was verwijderd, omdat de rechter geoordeeld heeft dat deze incidenten onschuldig waren. De gegevens waarop statisticus 3 zich baseert zijn dus óók onzuiver, en wel aan twee kanten: het is goed mogelijk dat er nog meer incidenten buiten Lucia's diensten zijn; maar statisticus 3 heeft ook incidenten bij Lucia verwijderd op juridische, niet-statistische gronden. Verder heeft statisticus 3 hetzelfde (frequentistische) paradigma gehanteerd als statisticus 1 (alleen met een verantwoorde manier van combineren van de gegevens van 3 afdelingen).

Annotatie 12. Het ligt strikt genomen nogal voor de hand dat statistische uitspraken en andere uitspraken die berusten op kansrekening, en die worden gebaseerd op gegevensverzamelingen, kwalitatief staan of vallen bij de objectiviteit en de nauwkeurigheid van de verzamelde informatie. Het verschil tussen de uitkomsten van statisticus 1 gecorrigeerd (1 op 100 duizend), en statisticus 3 (1 op 50), is dramatisch en laat dat ook zien.

3.4. Statisticus 4

Statisticus 4 heeft een nieuw element toegevoegd, eigenlijk in navolging van het rapport van statisticus 1. Deze had terecht opgemerkt dat een statistisch verband nog geen causaal verband is. De eigenlijke vraag voor de statisticus is een causale vraag: verhoogt de aanwezigheid van Lucia de kans op incidenten?

Annotatie 13. Duidelijk moet hier zijn dat het gaat om statistisch bewijs voor een relatie (in de zin als eerder aangegeven) tussen de variabelen 'wel/niet aanwezigheid van X' en 'wel/geen incidenten' die, als hij al gevonden wordt, nog geenszins hoeft te wijzen op een direct causaal verband. In dit licht is het gebruik van de term 'causale vraag' verwarrend,

maar niet ongewoon, in zekere zin misschien een contaminatie van betekenissen, omdat de term wel toepasselijk zou zijn vanuit het eerdergenoemde logische paradigma (waarvan in dit geval natuurlijk geen sprake kan zijn).

We vragen ons af of dezelfde incidenten zouden zijn gebeurd als Lucia geen dienst had gehad. Deze vraag heeft betrekking op een niet-feitelijke situatie. We kunnen niet experimenteel de klok terugdraaien en een paar jaar ziekenhuisgeschiedenis herhalen, met alles precies hetzelfde als voorheen, maar met een andere verpleegster dan Lucia (wat een zuivere, frequentistische benadering zou zijn geweest). Statisticus 1 moest zijn oordeel baseren op het vergelijken van diensten met Lucia met andere diensten zonder Lucia, en ging er daarbij van uit dat alle diensten in een heel jaar volstrekt vergelijkbaar met elkaar zijn.

Annotatie 14. Wanneer een zuiver frequentistische benadering onmogelijk is, zoals vrijwel steeds bij toepassingen in de strafrechtketen, is bijzondere aandacht vereist voor de beoordeling van de wijzen waarop rekening is gehouden met de (daarmee noodzakelijk geworden) corruptie van de *ceteris paribus clause* die nu immers moet worden benaderd.

Statisticus 4 gaat ervan uit dat systematische verschillen kunnen optreden tussen de diensten van de ene en van de andere verpleegster. Achteraf was inmiddels bekend dat in de jaren voor Lucia's aanwezigheid, en ook in de jaren erna, enkele even grote clusters van sterften zijn voorgekomen in dezelfde afdeling. Dit zou er een aanwijzing voor kunnen zijn dat niet alle diensten vergelijkbaar zijn, dat er periodes met minder, en periodes met meer zwakkere patiënten op een afdeling kunnen voorkomen. Als dat zo is, zijn er (in strijd met de aannamen van statistici 1-3) naast periodes met verhoogde kansen op (natuurlijke) incidenten ook periodes van rust. Het verschijnsel van willekeurige clustering van incidenten leidt tot een grotere variatie van het aantal incidenten dat individuele verpleegsters meemaken.

Annotatie 15. Dit verschijnsel is bekend in de econometrie, de medische statistiek en de epidemiologie: het heet *overdispersion* wegens *unmeasured heterogeneity of confounding factors*. Statistische onderzoekers op dit terrein hebben inmiddels geleerd dat de tijd een van de belangrijkste confounding factors is in empirisch onderzoek, een van de vaakstvoorkomende oorzaken van schijnrelaties.

Hij deed de berekening opnieuw, onder de (Bayesiaanse) aanname dat de ene verpleegster vrij gemakkelijk systematisch ongeveer twee keer zoveel incidenten kan meemaken als een andere. De kans verandert opnieuw aanzienlijk; de coïncidentie is alweer een

stuk minder bijzonder dan eerst leek. Hij deed deze berekening op basis van de mogelijk te optimistische cijfers van statisticus 3.

3.5. Afronding

De enorme verschillen tussen de vier statistici zijn dus inderdaad heel belangrijk. Het is bijzonder leerzaam om de bron van de verschillen op te sporen: (i) de correctie van een steekproefcombinatiefout, (ii) een bijzondere selectie van onderzochte data, (iii) het 'opschonen' van de data en (iv) het rekening houden met *confounding factors*. Hieruit valt maar een conclusie te trekken:

Buiten laboratoriumcondities is het doen van waarschijnlijkheidsuitspraken over een door individueel menselijk gedrag veroorzaken van incidenten (nog) geen eenvoudig toepassen van gevestigde wetenschap. Niet voor niets waren veel statistici bereid om een eigen analyse te doen en met allerlei alternatieve voorstellen te komen. Het type probleem is bekend, berucht, controversieel en mediageniek. Er bestaat nog lang geen consensus over de juiste analyse onder statistici.

3.5.1. Wat niet in de verschillen tot uitdrukking komt

Omdat we ons tot nu toe hebben gericht op de verschillen, hebben we nog geen aandacht besteed aan een bijzonder bezwaar tegen de analyses van *alle vier* de statistici: het *post hoc* probleem. Door allen is uitgegaan van het doen van experimenten om een vooraf opgestelde hypothese te toetsen. Maar de hypothese is gesuggereerd door de gegevens zelf, de gegevens van statisticus 2. Diens resultaat kan mogelijk een aanleiding zijn tot nader medisch en/of justitieel onderzoek, maar diens data zijn *daarmee* onbruikbaar geworden voor verdere statistische conclusies. Men kan een vermoeden niet toetsen door het te herhalen. De enige bekende manier om aan dit bezwaar tegemoet te komen is door een switch te maken naar een empirisch Bayesiaanse aanpak waarbij de bijzonder kleine kans op wat men hier heeft waargenomen, wordt afgezet tegen de vele omstandigheden (vele ziekenhuizen, jaren, verpleegsters) waarin het zich potentieel kon voordoen en men expliciet rekening houdt met de alternatieve hypothese (hoe waarschijnlijk zijn de bevindingen bij moord en de a-priorikans: hoe waarschijnlijk is een moordende verpleegster). De schijnbare hardheid van de cijfers van statistici 1, 2 en 3 wordt dan vervangen door een heel zacht cijfer, met een volstrekt andere interpretatie, gebaseerd op ruwe aannamen. Dit ondersteunt een tweede conclusie:

De zaak Lucia de B. heeft voor statistici dan wel aanleiding gegeven tot wetenschappelijke herbezinning, nieuw onderzoek, publicaties en wetenschappelijke vooruitgang, hij heeft niet bijgedragen aan een betere reputatie van statistische expertise bij de rechtspraak.

3.5.2. Aandachtspunten voor de discussie

Er blijft een aantal aspecten over die mogelijk additionele aandacht verdienen in de discussie:

1. Statisticus 1, die in een zeer vroeg stadium door de politie was ingeschakeld deed een eenvoudige berekening die resulteerde in een simpele, spectaculaire, maar foute uitkomst die hij helder kon uitleggen aan de rechter. In eerste aanleg ging de rechter daar blijkens zijn motivering geheel in mee, mogelijk mede omdat de experts van de verdediging op minder direct aansprekende theoretische gronden – terecht – betoogden dat genoemd getal niet tot bewijs zou mogen dienen. En, daarom, zelf geen enkele 'tegen-uitkomst' wilden noemen. Op het eerste gezicht een niet ongevaarlijke combinatie van kennisasymmetrieën en de uitwerking van verschillen in retorisch talent.
2. Waar was de verdediging? Statisticus 1 heeft aangegeven dat allerlei factoren tot 'onschuldige' systematische verschillen tussen verpleegsters kunnen leiden. Hij noemt er vier bij wijze van voorbeeld. Hij ging ervan uit dat de verdediging zijn lijst zou aanvullen met nog tien voorbeelden, en dat er een serieus onderzoek zou worden ingesteld naar mogelijke alternatieve verklaringen.

Annotatie 16. Het kan niet worden uitgesloten dat het gemengd accusatoire-contradictoire-inquisitoire karakter van het Nederlandse strafproces onduidelijkheden in zich bergt. In hoeverre, bijvoorbeeld, mag van het OM een zozeer magistratelijke opstelling worden verwacht dat het er bij zijn 'eigen' deskundige op aandringt om 'serieus onderzoek te doen naar mogelijke alternatieve verklaringen'?

3. Het is belangrijk om voldoende technische scholing te hebben om de enorme potentiële verschillen te vatten tussen: de kans op toeval, gegeven de data (uitgaande van het frequentistische paradigma); en de kans op de data, gegeven de waarschijnlijkheid (uitgaande van het logische of Bayesiaanse paradigma).

Ik sluit af met een opinie. De lamentabele rol die in de zaak Lucia de B. is gespeeld door wat wel statistisch bewijs is genoemd valt moeilijk toe te schrijven aan een enkele van de institutionele functies die bij het strafproces verantwoordelijkheid droegen – in dit geval het OM, de verdediging, de rechter, de expert. Naast aandacht voor een eventueel tekort aan kennis, dient mijns inziens ook onder ogen te worden gezien in hoeverre het karakter, de inrichting, de ingesloten taakopvattingen, de bewerktuiging en de financiering van het huidige strafproces een rol hebben gespeeld en zullen blijven spelen.

4. Discussie, conclusies en aanbevelingen

De belangrijkste elementen voor discussie tussen de auteurs zijn in het voorafgaande verweven. Er zijn weinig echte verschillen. Drie hoofdpunten vroegen

aandacht. Voor zover statistische expertise een rol speelt in de strafrechtketen: (i) is veel van de voor de beoordeling ervan benodigde kennis inmiddels de revue gepasseerd; (ii) zijn we het er over eens dat – naast kennisasymmetrie – aspecten van organisatorische, culturele en sociaal-wetenschappelijke aard een rol kunnen spelen bij expertise-inbreng en de beoordeling ervan; (iii) is de belangrijkste conclusie die we delen dat kansrekening en statistisch bewijs, zoals de zaken er nu voor staan, een rol kunnen spelen bij de opsporing, maar niet behoren te worden meegewogen in het juridische bewijs. We nemen deze drie punten ter afsluiting nog kort nader onder de loep.

4.1. Benodigde beoordelingskennis

Afgezien van de vraag welke rol statistische expertise in het strafproces ook speelt, is er, wanneer dat van belang wordt geoordeeld, kennis nodig om de betrouwbaarheid van de aangeboden expertise te beoordelen. De meeste van die kennis ligt dicht tegen de algemene ontwikkeling waarover rechters al beschikken. We doelen hierbij op:

- het onderkennen van de verschillende basisaannamen die aan de verschillende paradigma's ten grondslag liggen en die bijdragen aan het beoordelen van de toegepaste subjectieve elementen: de oplossingen voor het '*ceterus paribus*'-corruptieprobleem in het frequentistische paradigma en de schattingen van a-prioriwaarschijnlijkheden in het Bayesiaanse paradigma; het eventueel doen berekenen van een relevant geoordeelde kans vanuit beide paradigma's;
- het onderkennen van de gevolgen die variaties in steekproefselectie en steekproefcombinatie kunnen hebben en het invoeren van contra-expertise wanneer daarover onzekerheid bestaat;
- het onderkennen van de noodzaak om geformaliseerde, objectieve onderzoeksvoorschriften te gebruiken bij de datasetselectie en codering;
- het onderkennen van de verschillen tussen statistisch en causaal bewijs;
- het desgewenst blootstellen van de gebruikte wetenschappelijke modellen en berekeningen aan contra-expertise;
- het onderkennen van de risico's, verbonden aan Bovarysme, en het eventueel verantwoorden van daaromtrent gemaakte keuzen die hebben bijgedragen aan de innerlijke overtuiging waarop het juridische bewijs uiteindelijk rust.

We menen dat aan de hand van deze eenvoudige checklist veel naderheid kan worden voorkomen bij de inhoudelijke beoordeling van expertise. Maar dan steeds wel in het besef dat als je een heel complexe vraag wilt beantwoorden met een enkel getal of ja/nee antwoord (toeval of geen toeval), je geen recht kunt doen aan de wetenschappelijke twijfels, behorende bij de materie. Verschillende paradigma's geven verschillende uitkomsten; het zijn antwoorden op verschillende vragen die gebruik maken van verschillende gegevens.

Vanuit beide perspectieven is er veel onbekend. Voor de frequentist hangen de kansen af van onbekende parameters en mogelijk zelfs van een onbekende structuur. De Bayesiaan weegt al dat onbekende af met behulp van a-prioriverdelingen die soms nauwelijks gegrond zijn in de empirie. In beide gevallen komt er een getal uit dat even zacht is als de aannamen die erin werden gestopt. En: er zijn zachtere antwoorden op meer relevante vragen en hardere antwoorden op minder relevante vragen. Vooral: omdat cijfers een magische uitstraling van exactheid hebben, is deze aanpak ronduit gevaarlijk, wanneer inhoudsdeskundige en statisticus te veel aan elkaar overlaten. Toepassing van de statistiek in empirisch wetenschappelijk onderzoek kenmerkt zich door een tweerichtingswisselwerking tussen de verschillende betrokken disciplines, tussen enerzijds de inhoudsdeskundige en anderzijds de statistisch deskundige. Vergelijking van en betekenis toekennen aan modellen en uitkomsten vanuit het frequentistische en het Bayesiaanse paradigma over een en dezelfde casuspositie vraagt de combinatie van inhouds- en statistische deskundigheid. Zorgvuldige wisselwerking tussen rechter en statisticus zijn noodzakelijk; genoemde checklist kan daarbij behulpzaam zijn. Het gaat mis als de inhoudsdeskundige de statisticus gebruikt als een orakel. Maar we menen ook dat de recente turbulentie over statistisch bewijs en, meer algemeen, over de rol van experts in het strafproces heeft laten zien dat andere dan wetenschappelijk-inhoudelijke factoren meespelen.

4.2. Kennisasymmetrie en aspecten van sociaal-wetenschappelijke aard

Het komt ons aannemelijk voor dat de risico's die door economen met kennisasymmetrieën worden verbonden, met name die welke gericht zijn op de vraag of het betreffende regelstelsel motiveert om relevante informatie uit te wisselen, ook een rol spelen (deels vanzelfsprekend – waar een verdachte niet zal willen meewerken aan zijn eigen veroordeling) bij de kennisasymmetrieën die zich voordoen in het strafproces. Van de drie genoemde risico's lijkt ons dat van de onwrikbare stellingname het meest herkenbaar, ook omdat kokervisie daar een *species* van is. De in dit stuk als rode draad gebruikte casus roept bijvoorbeeld de vraag op wat vroegtijdig publiciteit zoeken kan betekenen. In de zaak Lucia de B. verstreken twee weken tussen de sterfte van Amber en de aangifte van een serie moorden door het ziekenhuis. Binnen die twee weken werden brieven aan ouders gestuurd, een persconferentie gehouden en een TV-optreden van de directeur gearrangeerd. De motieven waren honorabel – de (verpletterende, maar onjuiste) statistische uitspraak had overtuigd dat er een seriemoordenaar actief was die moest worden gestopt. Dus moest – als wel vaker – een risicovolle en mogelijk tragische keuze worden gemaakt. Toch: mét het vroegtijdig verkrijgen van dergelijke publiciteit wordt een eventuele weg

terug, zoals die bij een *verdachte* nog behoort te worden opgehouden, misschien heel wat moeilijker te vinden en uit te leggen dan zonder. En: de rechter wordt geacht rekening te houden met de mate waarin de rechtsorde wordt verstoord, die vaak wordt afgelezen aan commotie in de media. Al met al gaat het om iets veel ingewikkelders dan statistische kennis – het gaat om een normatief klimaat waarin media-aandacht wordt benut ter beïnvloeding van het karakter van ons strafproces. Veel belangen wijzen in dezelfde richting. De pers gedijt erbij, het OM kan er zijn processuele positie mee versterken, de expert kan er zijn status mee vestigen en de rechter kan respect tonen voor wat in de samenleving speelt. Expertise kan zo, al in de allereerste voorfase en in combinatie met de media, worden gebruikt voor het creëren van een *moral panic* die nauwelijks ruimte overlaat voor onafhankelijke, objectieve behandeling. Als deze analyse juist is, staat onze juridische cultuur onder een druk die de rechtspraak uitnodigt om tegenkrachten te mobiliseren.

4.3. Opsporing, bewijs, tragic choices

Nog eens: de belangrijkste conclusie die we delen is dat kansrekening en statistisch bewijs, zoals de zaken er nu voor staan, een rol kunnen spelen bij de opsporing, maar niet behoren te worden meegewogen in het juridische bewijs. Daarmee lijkt het juridische belang van wat aan de orde is gesteld in wezen voor de praktijk ondergeschikt, maar genoemde conclusie is gemakkelijker uitgesproken dan verwezenlijkt. Daarvoor zijn verschillende, opnieuw niet-statistische redenen.

Ten eerste is het moeilijk, zo niet onmogelijk, om eenmaal opgenomen informatie en de daarmee verbonden emoties op bevel uit te schakelen, te vergeten. Men kan wel willen dat het statistische materiaal dat tot opsporing aanzette niet wordt hergebruikt bij de bewijsvoering, maar dat valt in de praktijk niet mee, zoals ook uit de vonnissen te lezen valt waarin afstand is genomen van statistisch bewijs. Het 'telkens aanwezig zijn' blijft hoe dan ook een hardnekkig element dat kennelijk bijdraagt aan de innerlijke overtuiging van de rechter.

Ten tweede lijkt – als gezegd – het gemengde karakter van ons strafproces in beweging, en wel zodanig, dat niet steeds meer duidelijk is wat de verschillende spelers van elkaar mogen verwachten. Hoe magistraatelijk is het OM, hoe belangrijk is contra-expertise door de verdediging? Wanneer de tendens zou doorzetten in de richting van een meer contradictoir karakter van ons strafproces, wordt het moeilijker om statistische en juridische bewijsvoering van elkaar gescheiden te houden.

Ten derde speelt ook nog eens de factor tijd. Die speelt in het licht van wat ons op het gebied van de opsporing nog te wachten staat, in combinatie met de mogelijkheden van ICT-gestuurde gegevensregistratie en -verwerking, gelegitimeerd door nieuwe wetgeving die reageert op terroristische dreigingen van onvoor-

spelbare omvang. Wij verwachten dat juridische schermutselingen over statistisch bewijs bij opsporing en vervolging een enorme vlucht zullen nemen samen met de toegenomen digitale profileringsmogelijkheden en de groeiende tendens naar proactief justitieel optreden. En dat de rechtsbescherming zich daarmee zal moeten kunnen verstaan.

Noten

- 1 De stelling dat bewijs nooit absoluut is, past in een wetenschappelijke houding die de geschriften van Popper serieus neemt (bijvoorbeeld: *Conjectures and Refutations*, Routledge & Keegan Paul 1963) en roept discussie op, met name aan de hand van ingeburgerde overtuigingen, verbonden aan tegenvoorbeelden. Zo wordt een match tussen het DNA-profiel van een spoor en het profiel van een persoon *in de praktijk* doorgaans terecht opgevat als strafrechtelijk bewijs van donorschap, maar niet als absoluut bewijs omdat nu eenmaal niet het volledige genoom wordt bepaald, het bestaan van een eenzige tweelingbroer of -zus of het verwisselen van samples niet kan worden uitgesloten.
- 2 De term 'Bovarysme' is hier ontleend aan M. Februari & M. Drenth, *Een prui van paardenhaar & over het lezen van een boek: Amartya Sen en de onmogelijkheid van de Paretiaanse liberaal*, E.M. Querido's uitgeverij bv 2000. Door haar ontleend aan: Jules de Gaultier, *Bovarysme*, Philosophical Library 1970.
- 3 De weergave van de betekenis van de p-waarden is onjuist, zij het niet ongebruikelijk, zie ook annotatie 5 en paragraaf 3.1. De genoemde waarden zijn p-waarden. Een p-waarde geeft aan hoe extreem de gevonden waarde voor de waargenomen grootheid is, afgezet tegen de aangenomen, achterliggende verdeling (technisch: de verdeling van de nulhypothese). Een p-waarde geeft dus de waarschijnlijkheid dat de aangenomen achterliggende verdeling van toepassing is op de waarnemingen. Hoe kleiner, hoe minder waarschijnlijk, en hoe meer reden de verdeling van de nulhypothese *in casu* te verwerpen. In het geval van onze dubbelsteen geeft de p-waarde aan hoe waarschijnlijk het op basis van de waargenomen waarden is dat hij deugt (de achterliggende verdeling als norm). Hoe kleiner de waarde, hoe minder waarschijnlijk dat is.
- 4 Hier wordt er een lans voor gebroken om in situaties die worden gekenmerkt door kennisasymmetrie, in elk geval bedacht te zijn op de mogelijke werking van waarheidsvijandige motieven die eventueel door het procesrecht c.q. de procespraktijk worden gestimuleerd. Te verwachten valt dat deze aanbeveling door de partners waarover het *in casu* gaat (deskundigen en OM) als merkwaardig en insinuerend wordt ervaren, ook wanneer zij genoemde aanbeveling, wanneer het om verdachten en verdediging gaat, vanzelfsprekend vinden. 'Bedacht zijn op' heeft, zo menen we, te maken met zorgvuldigheid en is niet hetzelfde als 'verwachten dat'.