

De *Blue Bus*-paradox: enkele overwegingen bij het gebruik van statistisch bewijs in het (civiele) recht

35

De Blue Bus-paradox draait om het gebruik van statistisch bewijs in het (civiele) recht. De paradox doet zich voor in een situatie waarin het bewijs voor het (betwiste) feit waarop de aansprakelijkheid van de gedaagde is gebaseerd, louter of hoofdzakelijk statistisch van aard is. Het bewijs voldoet aan de civielrechtelijke bewijsstandaard, waardoor aansprakelijkheid als rechtsgevolg gerechtvaardigd lijkt. In de rechtspraak wordt de aansprakelijkheid in dergelijke zaken echter veelal onaanvaardbaar geacht. Dit artikel bespreekt drie strategieën om de paradox op te lossen: (1) het afwijzen van het statistische bewijs vanwege het referentieklassen-probleem, (2) het verwerpen van een benadering van de bewijsstandaard als een kwantificeerbare, absolute waarschijnlijkheid waaraan het bewijs getoetst moet worden, en (3) het (alsnog) aanvaarden van de – op het eerste gezicht – ongemakkelijke aansprakelijkheid door te beseffen dat het ongemak voortkomt uit een (irrationele) intuïtie.

1. Inleiding

'Iudex non calculat': de rechter rekent niet. Een hedendaagse toepassing van dit aloude adagium kan zijn dat rechters terughoudend zijn bij het interpreteren en gebruiken van statistische gegevens. Tegelijkertijd is het aannemelijk dat rechters in toenemende mate statistisch bewijsmateriaal onder ogen zullen krijgen in onze steeds meer door data en kunstmatige intelligentie gedreven maatschappij.¹ In het strafrecht is al een ontwikkeling gaande waarbij in de opsporing gebruik wordt gemaakt van op algoritmen en *big data* gebaseerde voorspellingen over de kans dat iemand een strafbaar feit zal plegen.² Hoewel het gebruik van algoritmen en *big data* voor het voorspellen van toekomstige (strafbare) feiten een andere afweging vereist dan voor het bewijzen van (civiele) feiten uit het verleden,³ kan het gebruik van dergelijke statistieken in beide contexten van belang zijn. Dat de rechter zich zal moeten buigen over door kunstmatige intelligentie geproduceerd statistisch bewijs, is allang geen futuristische gedachte meer.⁴ Vooral wanneer het (betwiste) feit waarop de aansprakelijkheid van de gedaagde is gebaseerd, louter of hoofdzakelijk steunt op statistisch bewijs, is het gebruik hiervan geen sinecure.

Een van de problemen die kunnen spelen bij het gebruik van statistieken als bewijs in het civiele recht, wordt geïllustreerd door de *Blue Bus*-paradox.⁵ De paradox draait om het (fictieve) geval waarin een blauwe bus een aanrijding heeft veroorzaakt en het enige bewijsstuk dat ter discussie staat statistisch van aard is. Hoewel onder omstandigheden de civielrechtelijke bewijsdrempel wordt behaald en aansprakelijkheid daarvan een rechtsgevolg kan zijn, is niet algemeen geaccepteerd dat uitsluitend of hoofdzakelijk statistisch bewijs toereikend is voor het bewijsoordeel. De paradox heeft in de Angelsaksische juridische literatuur aanzienlijk meer aandacht gekregen dan in de Nederlandse, terwijl de problematiek ook relevant is voor de Nederlandse civielrechtelijke bewijspraktijk en daarom hier eveneens een kritische beschouwing verdient.⁶ In dit artikel probeer ik daartoe slechts een aanzet te bieden door, zonder naar volledigheid te streven, enkele belangrijke afwegingen te bespreken die spelen bij het verklaren en oplossen van de paradox.

In het vervolg wordt eerst de *Blue Bus*-casus beschreven en van context voorzien (par. 2). Daarna worden drie strategieën besproken om de paradox op te lossen. Als eerste strategie wordt het afwijzen van het

* Mr. K.A.F.M. Rasul is student aan het Hoger Instituut voor Wijsbegeerte van de KU Leuven en is tevens afgestudeerd jurist.

1. Zie bijvoorbeeld: A.-R. Mackor, 'Veroordelen met "naakte" statistieken', *RMThemis* 2019, afl. 3, p. 96.
2. Zie bijvoorbeeld: A. Das & M. Schuilenburg, 'Predictive policing: waarom bestrijding van criminaliteit op basis van algoritmen vraagt om aanpassing van het strafprocesrecht', *Strafblad* 2018/33, afl. 4, p. 19-26.
3. Bijvoorbeeld omdat de gevolgen en kosten van fout-positieve voorspellingen bij de strafrechtelijke opsporing (zoals het onterecht inzetten van opsporingsbevoegdheden) verschillen van die bij de civiele bewijsvoering (zoals wanneer gedaagden onterecht aansprakelijk worden gesteld). Hetzelfde geldt voor de kosten van fout-negatieve voorspellingen, en de baten van juist-positieve en juist-negatieve voorspellingen. Zie bijvoorbeeld: M. de Vries e.a., 'Al-risicotaxatie: nieuwe kansen en risico's voor statistische voorspellingen van recidive', *Boom Strafblad* 2021/2, afl. 2, p. 58-66.
4. Zie over de toelaatbaarheid van door kunstmatige intelligentie geproduceerd bewijs in de Verenigde Staten: P.W. Grimm, M.R. Grossman & G.V. Cormack, 'Artificial intelligence as evidence', *Northwestern Journal of Technology and Intellectual Property* (19) 2021, afl. 1, p. 9-106.
5. Voor het eerst beschreven in H. Tribe, 'Trial by mathematics: Precision and ritual in the legal process', *Harvard Law Review* (84) 1971, afl. 6, p. 1341.
6. Alleen Mackor schreef hierover in Nederland, zie bijvoorbeeld: Mackor 2019, p. 93-96, en A.-R. Mackor, 'Different ways of being naked: A scenario approach to the naked statistical evidence problem', *Journal of Applied Logics* (8) 2021, afl. 9, p. 2407-2432.

statistische bewijs vanwege het referentieklassen-probleem behandeld (par. 3). De tweede strategie richt zich op het verwerpen van een benadering van de bewijsstandaard als een kwantificeerbare, absolute waarschijnlijkheid waaraan het bewijs getoetst moet worden (par. 4). De laatste strategie draait om het aanvaarden van de aansprakelijkheid die voortvloeit uit het statistische bewijs, met het besef dat de reden dat dit ongemakkelijk voelt, voortkomt uit een (irrationele) intuïtie (par. 5). Tot slot volgt een conclusie (par. 6).

2. De Blue Bus-paradox: beschrijving en context

Er zijn de afgelopen decennia veel varianten verschenen van de *Blue Bus*-paradox. Hoewel de namen en details kunnen verschillen, draaien ze alle om de civiele aansprakelijkheid van het vervoersbedrijf dat een aanrijding zou hebben veroorzaakt. Hierbij is de vaststelling dat dit vervoersbedrijf de aanrijding heeft veroorzaakt louter of hoofdzakelijk gebaseerd op statistisch bewijs. In dit artikel wordt de oorspronkelijke, hypothetische,⁷ *Blue Bus*-casus die afkomstig is van Tribe, als vertrekpunt genomen. Deze luidt als volgt:

‘Eiser is aangerezen door grove nalatigheid van een blauwe bus. De vraag is of de bus toebehoorde aan de gedaagde. De eiser kan bewijzen dat de gedaagde vier-vijfde van alle blauwe bussen in de stad exploiteert. Welke waarde dient aan dit bewijs te worden toegekend?’⁸

Voor de beantwoording van de vraag uit deze casus in de context van het Nederlandse civiele recht, dient eerst de bewijsstandaard te worden vastgesteld. In het Nederlandse civiele recht geldt gewoonlijk dat de partij die – door de wederpartij voldoende betwiste – feiten stelt die noodzakelijk zijn voor het gewenste rechtsgevolg, in de regel deze feiten dient te bewijzen (art. 149 jo. 150 Wetboek van Burgerlijke Rechtsvordering (Rv)). Hoewel de bewijsmaatstaf niet in de wet is geregeld – de wet laat de rechter vrij in de waardering van het bewijs krachtens artikel 152 Rv – wordt in de rechtspraak en

literatuur aangenomen dat het te bewijzen feit (naar het oordeel van de rechter) met ‘een redelijke mate van zekerheid’ waar is.⁹ Dit criterium is verder niet gekwantificeerd in de rechtspraak.¹⁰

Voor een mogelijke kwantificering van het bewijscriterium van ‘een redelijke mate van zekerheid’, kan (niet meer dan) ter oriëntatie worden gekeken naar Angelsaksische rechtssystemen. Hier wordt ten aanzien van de civiele bewijsstandaard vaak gesproken van ‘*the balance of probabilities*’ of van ‘*preponderance of the evidence*’.¹¹ Het dient bij die bewijsstandaard waarschijnlijker te zijn dat het feit wel heeft plaatsgevonden dan niet. Uitgedrukt in kansen levert dit op dat het te bewijzen feit met een kans van méér dan 50% heeft plaatsgehad (dus: $Pr > 0,5$; waar Pr voor de kansfunctie staat).¹²

In de literatuur is gedebatteerd over de vraag of de Nederlandse civielrechtelijke bewijsstandaard moet worden gekwantificeerd en, zo ja, wat dan de juiste kwantificering is.¹³ De literatuur is eensgezind in de opvatting dat de bewijsstandaard lastig kwantificeerbaar is. Degenen die een poging wagen, komen vaak uit op een percentage niet hoger dan 75%.¹⁴ Een belangrijke bijdrage aan dit debat is van Giesen, die de voorkeur geeft aan het gelijkstellen van de civielrechtelijke bewijsstandaard van ‘een redelijke mate van zekerheid’ met $>75\%$.¹⁵ Enkele overwegingen die daarbij een rol hebben gespeeld zijn dat verschil moet bestaan met de lagere civielrechtelijke bewijsstandaard van ‘aannemelijkheid’ die hij gelijkstelt met $>50\%$ en de hogere strafrechtelijke bewijsstandaard van ongeveer $>90\%$, alsook dat bij een hogere bewijsstandaard dan $>75\%$ het bewijsoordeel te vaak zou afwijken van de werkelijkheid (te veel fout-negatieven).¹⁶

Als de bewijsstandaard van $>75\%$ wordt aangenomen en toegevoegd aan de *Blue Bus*-casus, dan zou de casus, na enige aanpassing, als volgt kunnen worden weergegeven:

7. De casus van Tribe is fictief, maar wel geïnspireerd op een werkelijke casus, namelijk de Amerikaanse zaak *Smith v. Rapid Transit, Inc.*, 317, Mass. 469, 58 N.E.2d 754 (1945). In deze zaak daagde mevrouw Smith het busbedrijf *Rapid Transit* voor de rechter omdat zij letsel zou hebben opgelopen door grove nalatigheid van een bus van *Rapid Transit*. Het bewijs van mevrouw Smith tegen het busbedrijf bestaat uit het feit dat dit busbedrijf als enige een vergunning heeft om een buslijn in de straat van het ongeval te exploiteren. De rechtbank oordeelde: ‘The most that can be said of the evidence in the instant case is that perhaps the mathematical chances somewhat favor the proposition that a bus of the defendant caused the accident. This was not enough.’, Tribe 1971, p. 1341.
8. Tribe 1971, p. 1340-1341 (eigen vertaling).
9. W.D.H. Asser, *Mr. C. Assers Handleiding tot de beoefening van het Nederlands Burgerlijk Recht. Procesrecht. 3. Bewijs*, Deventer: Wolters Kluwer 2023, aant. 264; G. de Groot, in: *GS Burgerlijke Rechtsvordering*, art. 149 Rv, aant. 5.2.
10. Wel worden bij uitzondering waarschijnlijkheidspercentages gehanteerd bij het bewijzen van causaal verband en bij schadebegroting. Dit uit zich bijvoorbeeld in het leerstuk van de proportionele aansprakelijkheid (als oplossing voor situaties waarin schade is geleden maar niet kan worden vastgesteld of die is veroorzaakt door een normschending) en de leer van het verlies van kans (als oplossing voor situaties waarin onzekerheid bestaat over de vraag of een normschending schade heeft veroorzaakt). Zie hierover: ECLI:NL:PHR:2019:1382.
11. Asser *Procesrecht*/Asser 3 2023/264. Zie bijvoorbeeld de Amerikaanse zaken: *Grogan v. Garner*, 498 U.S. 279, 286 (1991) en *Herman & Maclean v. Huddleston*, 459 U.S. 375, 391 (1983).
12. R. Urbaniak & M. Di Bello, ‘Legal probabilism’, in: E.N. Zalta (red.), *The Stanford Encyclopedia of Philosophy* (herfst 2021 editie), par. 5.2 en 6.1, plato.stanford.edu/archives/fall2021/entries/legal-probabilism.
13. Giesen besteedt in zijn proefschrift uitgebreid aandacht aan deze discussie, zie: I. Giesen, *Bewijs en aansprakelijkheid: Een rechtsvergelijkend onderzoek naar de bewijslast, de bewijsvoeringslast, het bewijsrisico en de bewijsrisico-omkering in het aansprakelijkheidsrecht* (diss. Tilburg), Den Haag: Boom Juridische uitgevers 2001, hoofdstuk 2.4. Zie ook: R. de Bock, *Tussen waarheid en onzekerheid: Over het vaststellen van feiten in de civiele procedure* (diss. Tilburg), Deventer: Kluwer 2011, p. 206-210.
14. C.J.M. Klaassen, ‘Bewijs(perikelen) in letselschadeprocedures’, *VR* 2022/94, afl. 7/8, p. 200.
15. Giesen 2001, p. 52-53.
16. Giesen 2001, p. 59.

De Blue Bus-casus

Eiser is aangereden door grove nalatigheid van een blauwe bus. Eiser kan bewijzen dat Blauwe Bus B.V. 80% van de blauwe bussen in de stad van het ongeval exploiteert. Civielrechtelijke aansprakelijkheid ontstaat als de rechter 'een redelijke mate van zekerheid' heeft dat het feit zich heeft voorgedaan, wat kan worden gelijkgesteld met een waarschijnlijkheid van groter dan 75%. Blauwe Bus B.V. kan (dus) civielrechtelijk aansprakelijk worden gehouden voor het feit.

Uit de casus kan de volgende redenering worden gedestilleerd:

De Blue Bus-redenering

(Premisse 1) P	De blauwe bus die, door grove nalatigheid, eiser aanrijdt, wordt met een waarschijnlijkheid van 80% ($Pr = 0,80$) geëxploiteerd door Blauwe Bus B.V. ¹⁷ (statistisch bewijs)
(Premisse 2) Als P, dan Q	Indien de rechter overtuigd is dat het feit met 'een redelijke mate van zekerheid' waar is, te kwantificeren als een waarschijnlijkheid van groter dan 75% ($Pr > 0,75$), kan de gedaagde civielrechtelijk aansprakelijk worden gehouden voor het feit. (bewijsdrempel-implicatie)
(Conclusie) Dus Q	Blauwe Bus B.V. kan civielrechtelijk aansprakelijk worden gehouden voor het feit. (aansprakelijkheid)

De Blue Bus-redenering komt in de kern neer op een *modus ponens*; het schoolvoorbeeld van een geldige redenering.¹⁸ Als aangenomen wordt dat de premissen waar zijn, dan is de conclusie noodzakelijk waar. Mochten de premissen inderdaad waar zijn, dan wordt niet alleen van een 'geldige' maar ook van een 'correcte redenering' gesproken.¹⁹

Vervolgens rijst de vraag waarom de Blue Bus-redenering kan worden beschouwd als een paradox. Een gangbare definitie van het fenomeen 'paradox' is die van Sainsbury: 'an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises'.²⁰ Hieruit kunnen drie vereisten worden gedestilleerd:

1. ogenschijnlijk aanvaardbare (ook wel: ware) premissen;
2. een ogenschijnlijk aanvaardbare (ook wel: geldige) redenering;
3. een ogenschijnlijk *onaanvaardbare* conclusie.

Als de Blue Bus-redenering naast deze definitie van een paradox wordt gehouden, dan lijkt inderdaad van een Blue Bus-paradox te kunnen worden gesproken.²¹ Allereerst is sprake van ogenschijnlijk aanvaardbare premissen (vereiste 1), immers: premisse 1 (statistisch bewijs) en premisse 2 (bewijsdrempel-implicatie) lijken te corresponderen met respectievelijk de statistische en juridische werkelijkheid. Belangrijk is om op te merken dat absolute zekerheid niet wordt vereist, volstaan wordt met de aanvaardbaarheid die zich bij een oppervlakkigere beschouwing voordoet. Ten tweede is sprake van een ogenschijnlijk aanvaardbare redenering, immers: als de premissen waar zijn, volgt hieruit logisch de conclusie (aansprakelijkheid). De redenering betreft namelijk – zoals hiervoor beschreven – een *modus ponens*; een geldige deductieve redeneervorm.

Ten derde is sprake van een ogenschijnlijk *onaanvaardbare* conclusie. Hier zit dan ook de crux: hoewel de conclusie logisch volgt uit de premissen, is deze ogenschijnlijk onacceptabel, in ieder geval wordt de conclusie niet *onverdeeld* acceptabel gevonden. Een indicatie hiervoor is dat er een gebrek aan rechtspraak is waarbij het (betwiste) feit waarop de aansprakelijkheid van de gedaagde is gebaseerd, louter of hoofdzakelijk steunt op statistisch bewijs. Ook lijken de literatuur en empirische studies een breed gedeelte afkeer onder juristen te laten zien van zodanig gebruik van statistisch bewijs.²² Deze afkeer wordt soms aangeduid als het Wells-effect, vernoemd naar een empirische studie van Wells onder juridische professionals en studenten die (een variant van) de Blue Bus-casus moesten beoordelen, waaruit deze afkeer naar voren kwam.²³ De reden waarom

17. Er zijn meerdere redenen waarom men niet direct zou kunnen concluderen dat, omdat Blauwe Bus B.V. 80% van de blauwe bussen in de stad exploiteert, er ook 80% zekerheid is dat Blauwe Bus B.V. de blauwe bus van het ongeval exploiteert. Denk bijvoorbeeld aan de mogelijkheid dat de locatie van het ongeval ook toegankelijk was voor bussen van buiten de stad. Ik zal niet al deze omstandigheden bespreken, omdat dit de casus onnodig complex zou maken en het niet nodig is om toch trouw te blijven aan de essentie van de paradox. In dit artikel ga ik er daarom in algemene zin van uit dat de redenen die deze redeneersprong zouden kunnen verhinderen, zijn uitgesloten. Bij gebrek aan nadere informatie wordt bijvoorbeeld aangenomen dat de locatie van het ongeval niet toegankelijk is voor bussen van buiten de betreffende stad, en dat iedere blauwe bus die in de stad rijdt evenveel kans heeft om de bus te zijn die het ongeval heeft veroorzaakt.

18. P. Smith, *An Introduction to formal logic*, Cambridge: Cambridge University Press 2020, p. 149.

19. Smith 2020, p. 14.

20. R.M. Sainsbury, *Paradoxes*, Cambridge: Cambridge University Press 2009, p. 1.

21. Zie voor het paradoxale karakter van de Blue Bus-casus ook: M.S. Pardo, 'The paradoxes of legal proof: A critical guide', *Boston University Law Review* (99) 2019, afl. 1, p. 253.

22. M. Spottswood, 'Paradoxes of proof', in: C. Dahlman, A. Stein & G. Tuzet (red.), *Philosophical Foundations of Evidence Law*, Oxford: Oxford University Press 2021, p. 318-319; K.J. Heller, 'The cognitive psychology of circumstantial evidence', *Michigan Law Review* (105) 2006, afl. 2, p. 244-247. Beide auteurs verwijzen hierbij naar meerdere empirische studies die deze afkeer onderbouwen. Anders: R.J. Allen & C.K. Smiciklas, 'The law's aversion to naked statistics and other mistakes', *Legal Theory* (28) 2022, afl. 3, p. 179-209, zij menen dat van een dergelijke afkeer tegen statistisch bewijs geen sprake is.

23. G.L. Wells, 'Naked statistical evidence of liability: Is subjective probability enough?', *Journal of Personality and Social Psychology* (62) 1992, afl. 5, p. 739-752.

aansprakelijkheid op basis van louter of hoofdzakelijk statistiek, in tegenstelling tot klassiek bewijs zoals getuigenbewijs, vaak onacceptabel wordt bevonden, blijkt lastig onder woorden te brengen. In de literatuur is hierover geen consensus, maar er zijn enkele verklaringen aan te wijzen, zoals het referentieklassen-probleem (bespreking volgt in par. 3) en het kennisprobleem (bespreking volgt in par. 5).²⁴

3. Oplossingsstrategie I: verwerp premisse 1 (reden: het referentieklassen-probleem)

De eerste manier om de *Blue Bus*-paradox op te lossen is het verwerpen van de eerste premisse. Uit onware premissen kan immers – ook in geval van een op zichzelf geldige redenering – geen ware conclusie worden afgeleid, zodat niet meer van een paradox kan worden gesproken. Toegepast op de *Blue Bus*-redenering is hetgeen dat wordt bestreden (de eerste premisse): het statistische gegeven dat de blauwe bus van de aanrijding met een waarschijnlijkheid van 80% wordt geëxploiteerd door Blauwe Bus B.V.

De waarschijnlijkheid van 80% uit de casus is eenvoudigweg gebaseerd op de verhouding:

$$\frac{\text{Aantal blauwe bussen in de stad die door Blauwe Bus B.V. worden geëxploiteerd}}{\text{Aantal blauwe bussen in de stad}}$$

De keuze voor deze breuk (en geen andere) is echter betrekkelijk willekeurig. Uit de noemer blijkt bijvoorbeeld dat de stad (van de aanrijding) als referentieklassen is gekozen, maar er zijn ook andere referentieklassen die hadden kunnen worden gekozen. Gedacht kan worden aan de specifieke wijk binnen de stad (plaats) waarin, of het dagdeel (tijd) waarop, de aanrijding plaatsvond. Dat er op verschillende manieren kan worden gekwalificeerd, komt omdat een concrete gebeurtenis behoort tot oneindig veel verzamelingen, oneindig veel (referentie) klassen van gebeurtenissen.²⁵

Wat dit vervolgens problematisch maakt, is dat geen dwingend antwoord bestaat op de vraag welke referentieklassen zou moeten worden geselecteerd.²⁶ Dit wordt ook wel het referentieklassen-probleem genoemd.²⁷ Ter illustratie: stel dat blijkt dat de aanrijding uit de casus 's nachts is veroorzaakt, en de nacht als (onderdeel van de) referentieklassen wordt gekozen. En stel dat ook blijkt dat de blauwe bussen van Blauwe Bus B.V. nooit 's nachts rijden, omdat al haar blauwe bussen schoolbussen zijn. De waarschijnlijkheid van de propositie dat de blauwe bus die de aanrijding heeft veroorzaakt afkomstig is van Blauwe Bus B.V., keldert dan naar 0%.

Het referentieklassen-probleem lijkt niet alleen te spelen als de waarschijnlijkheid wordt opgevat als een relatieve frequentie. Het Bayesiaanse kansbegrip, een subjectieve interpretatie van waarschijnlijkheid die in het recht aan belang wint, worstelt met een soortgelijk probleem.²⁸ Dit werkt als volgt. De stelling van Bayes luidt:

$$Pr(A|B) = \frac{Pr(B|A) Pr(A)}{Pr(B)}$$

Met deze formule wordt $Pr(A|B)$, ook wel: de *posterior*, berekend. Dit is de kans op een hypothese A (in casu: dat Blauwe Bus B.V. de aanrijding heeft veroorzaakt) gegeven het bewijs B (in casu: dat de bus van de aanrijding blauw was en Blauwe Bus B.V. 80% van de blauwe bussen exploiteert). Hoe deze kans precies wordt berekend hoeft hier niet in al zijn details te worden behandeld. Slechts van belang is in te zien dat deze kans afhankelijk is van $Pr(A)$, ook wel: de *prior probability* (in casu: de waarschijnlijkheid dat Blauwe Bus B.V. de aanrijding heeft veroorzaakt voordat het bewijs in aanmerking wordt genomen). De *prior probability* is – en hieruit bestaat de overeenkomst met het referentieklassen-probleem – afhankelijk van de vraag onder welke referentieklassen de gestelde aanrijding door Blauwe Bus B.V. wordt ondergebracht.²⁹ Enkele van de vele mogelijkheden zijn: een aanrijding door een busbedrijf (breed), door een busbedrijf in de betreffende stad (smaller) of door een busbedrijf in de betreffende stad en – indien bekend – op het betreffende dagdeel, uur of zelfs tijdstip (nog smaller).

In de literatuur zijn verschillende pogingen ondernomen om het referentieklassen-probleem op te lossen. Althans, om betere van slechtere referentieklassen te onderscheiden, zonder dat dé juiste referentieklassen hoeft te bestaan. Daarbij zijn meerdere criteria voorgesteld – soms in combinatie toepasbaar – die bij het selecteren van een geschikte(re) referentieklassen kunnen worden gebruikt. Een van de criteria is de grootte van de referentieklassen; er zou gestreefd moeten worden naar de smalst mogelijke referentieklassen ('*the narrowest class*') waarvoor betrouwbare statistieken voorhanden zijn.³⁰ Als bijvoorbeeld blijkt dat de blauwe bus die de aanrijding heeft veroorzaakt, over niet-functionerende verlichting beschikte, dan zou de referentieklassen (de noemer) kunnen betreffen: blauwe bussen *met niet functionerende verlichting*. Ook ten aanzien van de plaats en tijd zal – voor zover hierover data bestaan – eerder voor een kleinere geografisch of temporeel afgebakende referentieklassen worden gekozen.³¹

Bij het bepalen van een geschikte referentieklassen wordt behalve de grootte van de referentieklassen eveneens

24. Zie hierover bijvoorbeeld M. Redmayne, 'Exploring the proof paradoxes', *Legal Theory* (14) 2008, afl. 4, p. 281-309 en L.J. Cohen, *The dialogue of reason: an analysis of analytical philosophy*, New York: Oxford University Press 1986.

25. A. Hájek, 'The reference class problem is your problem too', *Synthese* (156) 2007, afl. 3, p. 564-565.

26. R.J. Allen & M.S. Pardo, 'The problematic value of mathematical models of evidence', *The Journal of Legal Studies* (36) 2007, afl. 1, p. 112.

27. Hájek 2007, p. 563.

28. M. Colyvan, H.M. Regan & S. Ferson, 'Is it a crime to belong to a reference class', *Journal of Political Philosophy* (9) 2001, afl. 2, p. 174-175.

29. Urbaniak & Di Bello, par. 7.3.

30. Hájek 2007, p. 580.

31. M.S. Pardo, 'The field of evidence and the field of knowledge', *Law and philosophy* (24) 2005, afl. 4, p. 376.

van belang geacht dat alleen eigenschappen of gebeurtenissen worden meegenomen die *relevant* zijn voor de te berekenen waarschijnlijkheid.³² Niet functionerende verlichting kan hier bijvoorbeeld onder vallen, terwijl het feit dat er in de blauwe bus een Feyenoordvaandel zou hangen, niet hoeft te worden meegenomen. Al is zelfs dit te problematiseren: doorgaans zijn bestuurders met een Feyenoordvaandel (fanatieke) voetbalsupporters, en wat als uit onderzoek zou blijken dat (fanatieke) voetbalsupporters (of specifiek: die van Feyenoord) relatief vaker roekeloos zijn en (daardoor) vaker verkeersongelukken veroorzaken? Het is dus ook van belang of er een causaal verband of zelfs al een correlatie bestaat tussen de eigenschap of gebeurtenis en de waarschijnlijkheid.³³

Een ander criterium dat vaker in de literatuur wordt genoemd heeft betrekking op de homogeniteit van de referentieklassse.³⁴ Stel dat blijkt dat Blauwe Bus B.V. inderdaad 80% van de blauwe bussen exploiteert, maar dat dit schoolbussen betreffen die elk slechts twee keer per dag rijden (net voor en na schooltijd) en slechts in een paar specifieke wijken (waar de scholen zich bevinden), niet zijnde de wijk waarin het ongeval plaatsvond. En stel dat blijkt dat de overige 20% van de blauwe bussen wordt geëxploiteerd door Turbosnel Bezorgd B.V., een bezorgservice die belooft binnen 10 minuten te kunnen leveren en daardoor structureel te hard rijdt, waarbij elk van haar blauwe bussen honderden ritten per dag maakt, in het bijzonder in de (woon)wijk waarin het ongeval plaatsvond. De referentiegroep van blauwe bussen in de stad is dan niet homogeen en geeft een zeer vertekend beeld; dit moet voorkomen worden door een homogeniteitseis.

In de forensische literatuur wordt de meest ‘representatieve’ referentieklassse vaak aangewezen als de geschikte referentieklassse.³⁵ Het referentieklassse-probleem kan dan worden opgevat als een modelselectie-probleem: hoe kiezen we het statistische model dat het beste aansluit bij de data en zo de werkelijkheid het nauwkeurigst weerspiegelt en voorspelt?³⁶ Voor de keuze van het juiste model bestaan weer verschillende statistische methoden, elk met hun eigen voor- en nadelen en uitdagingen.³⁷ Twee ongewenste verschijnselen bij het afstemmen van het model op de data zijn *underfitting* en *overfitting*.³⁸ Bij een te brede referentieklassse, zoals in de *Blue Bus*-casus zou kunnen zijn: ‘aanrijdingen door een voertuig’, ontstaat *underfitting*. Hierbij wordt

onvoldoende gebruikgemaakt van relevante, onderscheidende informatie, wat ten koste gaat van de nauwkeurigheid. Aan de andere kant kan een te specifieke referentieklassse, zoals: ‘aanrijdingen door een blauwe bus in een bepaalde straat op een exact tijdstip’, resulteren in *overfitting*. Hierbij worden toevalligheden en irrelevante overeenkomsten meegenomen, wat ook weer de nauwkeurigheid schaadt.

Hoewel de criteria voor het selecteren van een geschikte referentieklassse waardevol zijn, gaan ze (in hun toepassing) niet zonder problemen gepaard. Een van de problemen is het vaak beperkte aantal beschikbare data om de gebeurtenis in de geschikte referentieklassse te plaatsen.³⁹ De beschikbaarheid van data is cruciaal, aangezien altijd slechts de geschikte(re) referentieklassse kan worden geselecteerd, *gegeven de beschikbare data*. Daarbij speelt in de (rechts)praktijk dat de referentiegroep nooit echt ‘af’ zal zijn; na verder onderzoek (bijvoorbeeld in hoger beroep) kunnen meer feiten naar boven komen die de gedaagde in een geschiktere referentiegroep kunnen plaatsen waardoor de waarschijnlijkheidsberekening verandert.⁴⁰ De mate waarin de waarschijnlijkheid die volgt uit het statistische bewijs, gevoelig is voor veranderingen, is mede afhankelijk van het ‘Keynesiaanse gewicht’ van het bewijs.⁴¹ Het Keynesiaanse gewicht van het bewijs is de hoeveelheid informatie (de kwantiteit) waarop het is gebaseerd. Dat betekent dus dat als het oorspronkelijke bewijs gebaseerd is op een beperkte hoeveelheid informatie, zoals bij de *Blue Bus*-casus, de waarschijnlijkheid gevoeliger is voor aanvullende informatie (minder stabiel of robuust).

In de context van de civiele zitting kan worden opgemerkt dat, nadat de eiser belastend statistisch bewijs heeft geleverd, de rechter de gedaagde kan toelaten tot het leveren van tegenbewijs (art. 151 lid 2 Rv).⁴² De gedaagde kan bijvoorbeeld de door eiser aangevoerde data ter discussie stellen, of nieuwe data leveren die de gebeurtenis in een geschiktere, voor hem ontlastende referentieklassse plaatsen. Als dit in de *Blue Bus*-casus succesvol zou gebeuren, zou daarmee de statistische bewijspremisse kunnen worden ontkracht, waarmee de paradox zou verdwijnen. Dit zal echter niet altijd mogelijk zijn, waardoor er moet blijven worden gezocht naar een oplossing voor dergelijke gevallen.

Concluderend kan worden gesteld dat het referentieklassse-probleem de waarde van bepaald statistisch be-

32. Urbaniak & Di Bello, par. 7.4.2.

33. Zie uitgebreider over dit causaal verband: Redmayne 2008, p. 289-291.

34. Allen & Pardo 2007, p. 112.

35. Zie bijvoorbeeld J.S. Buckleton, K.A.J. Walsh & I.W. Evett, ‘Who is “random man”?’, *Journal of the Forensic Science Society* (31) 1991, afl. 4, p. 466.

36. Zie voor deze opvatting bijvoorbeeld E.K. Cheng, ‘A practical solution to the reference class problem’, *Columbia Law Review* (109) 2009, afl. 8, p. 2090-2095.

37. Zo kan hier bijvoorbeeld weer de vraag rijzen welke database het geschiktst is voor het model, zie bijvoorbeeld: C. Champod, I.W. Evett & G. Jackson, ‘Establishing the most appropriate databases for addressing source level propositions’, *Science & Justice* (44) 2004, afl. 3, p. 153-164.

38. Cheng 2009, p. 2095.

39. Cheng 2009, p. 2097.

40. Urbaniak & Di Bello, par. 6.4.

41. C. Dahlman, ‘Naked statistical evidence and incentives for lawful conduct’, *The International Journal of Evidence & Proof* (24) 2020, afl. 2, p. 168.

42. HR 29 juni 2018, ECLI:NL:HR:2018:1045, r.o. 3.3.4, NJ 2018/396, m.nt. E. Verhulp.

wijs kan ondermijnen. Hoewel dé juiste referentieklassie niet bestaat, bieden de criteria voor het oplossen van dit probleem wel handvatten om betere van slechtere referentieklassen te onderscheiden, wat op zich al waardevol is. Slechts in gevallen zoals de *Blue Bus*-casus, waarin zeer beperkte informatie beschikbaar is om een geschikte referentiegroep vast te stellen, zou dit kunnen leiden tot de verwerping van de concrete statistische bewijs-premissie.

4. Oplossingsstrategie II: verwerp premisse 2 (reden: alternatieve benadering bewijsstandaard)

Een tweede oplossingsstrategie voor de *Blue Bus*-paradox is het verwerpen van de tweede premisse. Deze premisse, de bewijsdrempel-implicatie, houdt in dat als aan de bewijsstandaard van 'een redelijke mate van zekerheid' is voldaan – gekwantificeerd door een waarschijnlijkheid van groter dan 75% ($Pr > 0,75$) – er civielrechtelijke aansprakelijkheid ontstaat.

De tweede premisse kan allereerst verworpen worden door te stellen dat de bewijsdrempel te laag is en verhoogd moet worden, bijvoorbeeld van 75% naar 85%. Bij deze verhoogde bewijsstandaard leidt het feit dat de blauwe bus met 80% waarschijnlijkheid van Blauwe Bus B.V. is (premissie 1) niet tot aansprakelijkheid, waardoor de paradox verdwijnt. Dit is echter geen duurzame oplossing. De paradox kan namelijk behouden blijven (zonder af te wijken van de geest van de paradox) door de eerste premisse te wijzigen naar een waarschijnlijkheid van 85%. Zelfs als een meer strafrechtelijke bewijsstandaard van 90%, 95% of zelfs (een theoretische) 100% wordt gehanteerd, en eiser statistisch bewijs levert dat daaraan beantwoordt (bijvoorbeeld de statistiek dat Blauwe Bus B.V. 100% van de blauwe bussen exploiteert), dan nog wordt de paradox niet opgelost. Dit komt omdat aansprakelijkheid louter of hoofdzakelijk gebaseerd op statistisch bewijs niet onverdeeld acceptabel blijft worden gevonden, bijvoorbeeld vanwege de redenen uit de vorige paragraaf (referentieklassie-probleem) en volgende paragraaf (kennisprobleem).

Een duurzamere manier om de paradox op te lossen door middel van het verwerpen van de tweede premisse, houdt in dat de bewijsstandaard niet kan worden beschouwd als een kwantificeerbare, absolute waarschijnlijkheid waartegen het bewijs moet worden afgewogen. Het standpunt dat de bewijswaarderingsmaatstaf niet kan worden gekwantificeerd tot een waarschijnlijkheidspercentage, is bijvoorbeeld verdedigd door De Bock.⁴³

De bewijsbeslissing kan volgens haar niet als een (objectieve) frequentistische kansbepaling worden gezien, omdat de concrete bewijsbeslissing uniek is en betrekking heeft op een niet-repliceerbaar geval.⁴⁴ Ook met toepassing van de (subjectieve) Bayesiaanse statistiek is – in deze door De Bock verdedigde opvatting – een gekwantificeerde bewijswaarderingsmaatstaf niet bruikbaar. Dit komt niet omdat de Bayesiaanse statistiek intrinsiek ongeschikt is om de bewijsbeslissing te kwantificeren, maar omdat deze op haar beurt kampt met andere fundamentele problemen. Zo is volgens De Bock het bepalen van de kwantitatief uitgedrukte bewijskracht (diagnostische waarde) die aan bewijsmiddelen moet worden toegekend vaak problematisch.⁴⁵ Dit geldt niet zozeer bij een DNA-vaderschapstest of schrift-expertise, maar bij bijvoorbeeld een getuigenverklaring. In zo'n geval is het onmogelijk om te bepalen hoe vaak een specifieke getuigenverklaring iets zegt over de waarheid ervan.⁴⁶ Een ander probleem is dat de *prior probability* (de waarschijnlijkheid van het te bewijzen feit volgens de rechter nog voordat hij het bewijs in aanmerking neemt) moeilijk kwantitatief is vast te stellen.⁴⁷

Als alternatief stelt De Bock een 'intersubjectieve bewijswaarderingsmaatstaf' voor, waarbij een feit bezeugen is als 'uit de beschikbare bewijsmiddelen redelijkerwijs kan worden afgeleid dat het feit zich heeft voorgedaan, en uit die bewijsmiddelen niet even goed kan worden afgeleid dat hetgeen de wederpartij met betrekking tot dat feit stelt, zich heeft voorgedaan, terwijl zich evenmin de situatie voordoet dat bewijsmateriaal dat redelijkerwijs verwacht mocht worden, ontbreekt'.⁴⁸ Dit alternatief sluit aan bij de tweede grote stroming van bewijstheorieën naast de probabilistische, namelijk die van de scenariovergelijking (ofwel: relatieve plausibiliteit of *explanationism*).⁴⁹ Hierbij worden verschillende scenario's die een bepaalde gebeurtenis (bepaald bewijsmateriaal) verklaren tegen elkaar afgewogen; de beste verklaring wordt vervolgens als juist aangenomen (*inference to the best explanation*). De kracht en kwaliteit van de verklaringen kunnen dan worden afgewogen aan de hand van criteria zoals consistentie, coherentie en eenvoud.⁵⁰ Behaalt een procespartij de civielrechtelijke bewijsstandaard, dan betekent dit (slechts) dat zij een betere verklaring biedt voor het bewijs dan de wederpartij (relatief in plaats van absoluut).⁵¹

Het is niet noodzakelijk een probabilistische benadering van de bewijsstandaard te verlaten om de tweede premisse te verwerpen en zo de paradox op te lossen. Zo zou bijvoorbeeld de rechter kunnen beslissen op

43. De Bock 2011, p. 206-210 en 234.

44. De Bock 2011, p. 207.

45. Opgemerkt kan worden dat het ontbreken van een vaste, dwingende diagnostische waarde niet noodzakelijk betekent dat er in zijn geheel geen diagnostische waarde, zij het afhankelijk van meerdere (interpretatieve) keuzes, kan worden vastgesteld.

46. De Bock 2011, p. 209.

47. De Bock 2011, p. 208-209.

48. De Bock 2011, p. 234.

49. Hoewel ook voor het civiel recht van belang, is de meeste literatuur over bewijstheorieën meer toegesneden op het strafrecht. Zie voor een overzicht van de drie meest voorkomende benaderingen van bewijs in het (straf)recht (de argumentatieve, de scenariovergelijking en de probabilistische) bijvoorbeeld: H. Prakken, 'Strafrechtelijk bewijzen: met Bayes of met verhalen? Of is er een derde weg?', *EeR* 2014, afl. 1, p. 4-19.

50. Zie bijvoorbeeld: R.J. Allen & M.S. Pardo, 'Relative plausibility and its critics', *The International Journal of Evidence & Proof* (23) 2019, afl. 1-2, p. 16.

51. Allen & Pardo 2019, p. 15.

basis van niet de *posterior* maar de *posterior odds*: de kansverhouding tussen de kans dat de hypothese waar is (in casu: dat Blauwe Bus B.V. de aanrijding heeft veroorzaakt), gegeven het bewijs, en dat deze onwaar is, gegeven het bewijs. Hierbij worden dus twee scenario's vergeleken. De *posterior odds* kunnen worden berekend met een variant van de stelling van Bayes:

$$\frac{Pr(A|B)}{Pr(\neg A|B)} = \frac{Pr(B|A)}{Pr(B|\neg A)} \times \frac{Pr(A)}{Pr(\neg A)}$$

In woorden: *posterior odds* (a posteriori-kansverhouding) = *likelihood ratio* (aannemelijkheidsquotiënt) × *prior odds* (a priori-kansverhouding).⁵² De *prior odds* representeren dezelfde kansverhouding als de *posterior odds*, zij het *voordat* het bewijs in aanmerking wordt genomen. De *likelihood ratio* (ook wel: bewijskracht of diagnostische waarde) drukt de kansverhouding uit tussen de kans op het bewijs, gegeven dat de hypothese waar is, en de kans op het bewijs, gegeven dat de hypothese onwaar is. Overigens stellen sommige auteurs voor om de focus te verleggen van de *posterior odds* naar de *likelihood ratio*.⁵³ Een mogelijke bewijsstandaard zou dan zijn dat de *likelihood ratio* groter moet zijn dan 1,

in formulevorm: $\frac{Pr(B|A)}{Pr(B|\neg A)} > 1$.⁵⁴

In de *Blue Bus*-casus kan worden betoogd dat de *likelihood ratio* gelijk is aan 1, omdat de kans op het bestaan van het statistische bewijs even groot is ongeacht of Blauwe Bus B.V. de aanrijding heeft veroorzaakt of niet, wat zou betekenen dat de toets niet wordt doorstaan.⁵⁵

Om van *posterior odds* naar een beslissing te komen over de aansprakelijkheid van gedaagde, kan de *Bayesian decision theory* worden gebruikt.⁵⁶ Hierbij worden twee beslissingen overwogen: (1) de beslissing om de gedaagde aansprakelijk te stellen gegeven het bewijs en (2) de beslissing om gedaagde *niet* aansprakelijk te stellen gegeven het bewijs. Voor beide beslissingen wordt ingeschat wat de (mogelijke) kosten zijn.⁵⁷ De kosten van beslissing 1 (aansprakelijkheid) omvatten bijvoorbeeld de negatieve gevolgen van een onterechte vaststelling van aansprakelijkheid (fout-positieven), terwijl bij beslissing 2 (niet-aansprakelijkheid) de kosten van een onterechte vaststelling van niet-aansprakelijkheid (fout-negatieven) worden meegenomen. Deze kosten worden gekwantificeerd, bijvoorbeeld op een schaal van 0 tot 1. Vervolgens worden de (subjectieve) kansen be-

rekend dat deze kosten zich voordoen. Het product van de kosten en de kans hierop, wordt het 'te verwachten verlies' (*expected loss*) genoemd.⁵⁸ De theorie stelt dat de beslissing met het laagst te verwachten verlies moet worden verkozen. De bewijsstandaard die hieruit voortvloeit, luidt dan: aansprakelijkheid (beslissing 1) volgt, als en slecht als

$$\frac{Pr(A|B)}{Pr(\neg A|B)} > \frac{K(C_{12})}{K(C_{21})}.$$

Hier staat K voor de kosten, C_{12} voor de consequentie als beslissing 1 tot een fout(-positief) leidt omdat beslissing 2 had moeten worden genomen (i.e. juist was), en C_{21} voor de consequentie als beslissing 2 tot een fout(-negatief) leidt omdat beslissing 1 had moeten worden genomen.

Er zijn verschillende uitdagingen bij het toepassen van dit model, zoals de vraag hoe kosten van beslissingen kunnen worden gekwantificeerd. Voor dit artikel is het voldoende om te zien dat door het gebruik van de *Bayesian decision theory* de vraag of het statistische bewijs in de *Blue Bus*-casus voldoet aan de bewijsstandaard geen absoluut, maar een relatief of comparatief antwoord kent. De *posterior odds* worden vergeleken met de kostenratio, en aangezien deze waarden per zaak kunnen verschillen, leidt dit steeds tot een andere afweging en beslissing.

Concluderend, door de civiele bewijsmaatstaf niet te interpreteren als een kwantificeerbare, absolute waarschijnlijkheid waaraan het bewijs getoetst moet worden, wordt de tweede premisse ondermijnd, wat mogelijk een oplossing zou kunnen bieden voor de *Blue Bus*-paradox.

5. Oplossingsstrategie III: accepteer de 'onaanvaardbare' conclusie (reden: het kennisprobleem is onproblematisch)

Een laatste belangrijke strategie voor het oplossen van de *Blue Bus*-paradox is het (alsnog) accepteren van de (aanvankelijk) 'onaanvaardbare' conclusie. De conclusie dat Blauwe Bus B.V. aansprakelijk kan worden gehouden op basis van louter of hoofdzakelijk statistisch bewijs komt ons wellicht verrassend of ongemakkelijk voor, maar deze conclusie volgt wel uit de ware premissen. In zo een geval zou kunnen worden gezegd dat we simpelweg door de zure appel (de conclusie) heen dienen te bijten (*biting the bullet*).⁶⁰

52. NFI, *De reeks waarschijnlijkheidstermen van het NFI en het Bayesiaanse model voor interpretatie van bewijs*, Den Haag: NFI 2017, par. 5.

53. Voorgesteld in bijvoorbeeld: E.K. Cheng, 'Reconceptualizing the burden of proof', *Yale Law Journal* (122) 2023, p. 1254-1279; M. Di Bello & B. Verheij, 'Evidential reasoning', in: G. Bongiovanni e.a. (red.), *Handbook of Legal Reasoning and Argumentation*, Dordrecht: Springer 2018, p. 458-460 en 471-473; M. Di Bello, 'Trial by statistics: Is a high probability of guilt enough to convict?', *Mind* (128) 2019, p. 1045-1084.

54. Cheng 2023, p. 1269.

55. Cheng 2023, p. 1269-1271.

56. Zie bijvoorbeeld: A. Biedermann, S. Bozza & F. Taroni, 'Analysing and exemplifying forensic conclusion criteria in terms of Bayesian decision theory', *Science & Justice* (58) 2018, afl. 2, p. 159-160.

57. Sommige toepassingen van de theorie kijken naar zowel kosten als baten (tezamen: utiliteit), andere alleen naar kosten; ik richt mij hier ter vereenvoudiging van de theorie enkel op de kosten.

58. Biedermann, Bozza & Taroni 2018, p. 160.

59. Biedermann, Bozza & Taroni 2018, p. 160 en 164.

60. Deze oplossingsstrategie van paradoxen wordt besproken in R.A. Sorensen, *Thought experiments*, New York: Oxford University Press 1992, p. 147-148. Hij spreekt in dit verband van een conclusie die 'pseudo absurd' is.

Veel van de redenen waarom aansprakelijkheid op basis van louter of hoofdzakelijk statistisch bewijs onaanvaardbaar wordt gevonden, kunnen worden herleid tot kwesties die op enige manier verband houden met het concept 'kennis'. Dit vraagt om enige filosofische context. In de filosofie wordt kennis vaak gedefinieerd als '*justified true belief*' (JTB), wat inhoudt dat (1) iemand de overtuiging heeft dat een bewering waar is (*belief*), (2) de bewering daadwerkelijk ook waar is (*true*),⁶¹ en (3) de persoon een gegronde reden voor deze overtuiging heeft (*justified*). In zeldzame gevallen, bekend als *Gettier*-gevallen,⁶² is een extra voorwaarde nodig om 'toevallige kennis' uit te sluiten, maar er is geen consensus over de formulering hiervan. Een bekend voorbeeld van zo'n *Gettier*-geval is dat iemand de overtuiging heeft dat het vier uur is (*belief*) omdat hij dit afleest van zijn doorgaans betrouwbare horloge (*justified*), en het daadwerkelijk vier uur is (*true*), maar deze overtuiging slechts toevallig waar is omdat het horloge precies een etmaal geleden is gestopt met werken.

De vraag rijst vervolgens waarom het statistische bewijs uit de *Blue Bus*-casus zich problematisch verhoudt tot dit kennisbegrip. Hiervoor zijn verschillende redenen aan te voeren.⁶³ Ik zal een veelvoorkomende verklaring bespreken: er ontbreekt een (geschikte) causale link tussen het bewijs en het te bewijzen feit. In dat geval wordt het bewijs soms ook wel 'niet-geïndividualiseerd' of 'niet-specifiek' genoemd.⁶⁴ In de *Blue Bus*-casus ontbreekt deze causale link, omdat de aanrijding door Blauwe Bus B.V. (te bewijzen feit) geenszins verklaart dat 80% van de blauwe bussen in de stad van het ongeval wordt geëxploiteerd door Blauwe Bus B.V. (bewijs). Van een (geschikte) causale link zou wel sprake zijn als een getuige had gezien dat de aanrijding door Blauwe Bus B.V. werd veroorzaakt, bijvoorbeeld door het logo van Blauwe Bus B.V. te hebben waargenomen. Dan zou de aanrijding door Blauwe Bus B.V. (te bewijzen feit) immers een zeer goede verklaring zijn voor het feit dat de getuige meent te hebben waargenomen dat Blauwe Bus B.V. de aanrijding heeft veroorzaakt (bewijs). Het getuigenbewijs is dan een gevolg dat is veroorzaakt door het feit; dit wordt wel het '*backwards-looking*'-element

genoemd dat het statistische bewijs in de *Blue Bus*-casus ontbeert.⁶⁵

In aansluiting hierop, als blijkt dat Blauwe Bus B.V. het ongeval *niet* heeft veroorzaakt, verandert dit niets aan het bestaan van de statistiek dat 80% van de blauwe bussen in de stad door deze onderneming wordt geëxploiteerd.⁶⁶ Het statistische bewijs mist hetgeen wat ook wel 'epistemische sensitiviteit' wordt genoemd; het bewijs bestaat los van de waarheid van het feit waarvan wordt getracht het te bewijzen.⁶⁷ Dit is anders voor het eerdergenoemde getuigenbewijs: de kans op het bestaan van dit bewijs is (veel) hoger als het te bewijzen feit daadwerkelijk is gebeurd (waar is), ten opzichte van het geval waarin het feit niet is gebeurd (onwaar is).

Vanwege het ontbreken van een (geschikte) causale link tussen bewijs en het te bewijzen feit en van epistemische sensitiviteit, kan dus worden bepleit dat statistisch bewijs niet tot kennis over het vermeende feit leidt, omdat onvoldoende rechtvaardiging (*justified* in JTB) bestaat en/of het onvoldoende de factor geluk uitsluit, zoals wel vereist.⁶⁸ Voor iets zo belangrijks als een rechterlijke uitspraak wordt dit wel noodzakelijk geacht. Daarom dient aansprakelijkheid op basis van louter of hoofdzakelijk statistiek achterwege te worden gelaten, luidt dan het argument.⁶⁹

Hiertegen kunnen verschillende argumenten worden ingebracht. Een belangrijk argumentatiepad is gericht op het betogen dat statistisch bewijs wél, net als getuigenbewijs, tot kennis kan leiden.⁷⁰ Dit is een meer puur (kennis)filosofische discussie die ik hier verder niet zal uitdiepen.⁷¹ Als we meegaan in het standpunt dat statistisch bewijs zoals in de *Blue Bus*-casus geen (echte) kennis oplevert, kan een overtuigende denkrichting zijn dat we ons in het recht – en misschien zelfs ook buiten het recht – niets aan hoeven te trekken van het kennisconcept.

In deze context pleit Papineau in een prikkelend artikel voor het verlaten van het kennisconcept dat in zich draagt dat we alleen iets 'weten' als er een causaal pad

61. Wat uiteraard een andere fundamentele vraag oproept: wat is 'waarheid'? Het gangbaarste antwoord, gebaseerd op de correspondentietheorie, is dat een bewering waar is als deze correspondeert met de werkelijkheid. Hoe 'corresponderen' en 'werkelijkheid' moeten worden ingevuld, is een discussie die hier verder buiten beschouwing kan worden gelaten.

62. Vernoemd naar de naamgever die dit probleem in zijn baanbrekende artikel aan de orde stelde: E.L. Gettier, 'Is justified true belief knowledge?', *Analysis (Oxford)* (23) 1963, afl. 6, p. 121-123.

63. Zie hierover bijvoorbeeld: Redmayne 2008, p. 296-307 en L.D. Ross, 'Recent work on the proof paradox', *Philosophy Compass* (15) 2020, afl. 6, par. 3.

64. Zie bijvoorbeeld: Thomson 1986, p. 203.

65. Thomson 1986, p. 203; C. Littlejohn, 'Truth, knowledge, and the standard of proof in criminal law', *Synthese* (197) 2020, afl. 12, p. 5268.

66. Zoals in de vorige paragraaf besproken, kan dit zich uiten in een *likelihood ratio* van 1.

67. Zie bijvoorbeeld: D. Enoch & T. Fisher, 'Sense and "sensitivity": Epistemic and instrumental approaches to statistical evidence', *Stanford Law Review* (67) 2015, afl. 3, p. 575. Zie voor het begrip 'epistemische sensitiviteit' bijvoorbeeld R. Nozick, *Philosophical explanations*, Cambridge Massachusetts: Harvard University Press 1981, hoofdstuk 3.3.

68. Al zijn de opvattingen hierover verdeeld afhankelijk van de epistemische stroming die wordt aangehangen. Zie bijvoorbeeld: Littlejohn 2020, par. 6, waarin ingegaan wordt op drie stromingen ten aanzien van deze kwestie.

69. Littlejohn 2020, p. 297.

70. Overigens kan bij wijze van preliminair verweer al worden gewezen op de onzuiverheid van het conceptueel tegenover elkaar plaatsen van getuigenbewijs en statistisch bewijs. Getuigenbewijs kan namelijk als een vorm van statistisch bewijs worden beschouwd, omdat de bewijswaarde afhangt van de kans dat de getuige foutief verklaart, wat kan worden berekend door vergelijking met een referentieklaas van getuigen. Zie hierover bijvoorbeeld: Thomson 1986, p. 218 en Dahlman 2020, p. 166.

71. Zie hiervoor bijvoorbeeld: Littlejohn 2020, par. 5 en 6 en S. Moss, *Probabilistic knowledge*, Oxford: Oxford University Press 2018.

is van het feit naar de kenner van het feit.⁷² In de *Blue Bus*-casus zou van een dergelijk causaal pad sprake zijn geweest, indien de (vermeende) aanrijding door Blauwe Bus B.V. (het feit), zou zijn waargenomen door een getuige (de kenner van het feit). Bij statistische gegevens ontbreekt een dergelijk causaal pad. Papineau bekritiseert de menselijke intuïtie om vanwege het ontbreken van het causale pad meer waarde te hechten aan getuigenbewijs dan aan statistisch bewijs, zelfs als het statistische bewijs niet minder betrouwbaar is.⁷³

Het is belangrijk te beseffen dat het bestaan van een causale link in het geval van getuigenbewijs, de waarschijnlijkheid van het (vermeende) feit dat ermee tracht te worden bewezen, om die reden niet in hogere mate doet toenemen dan in geval van bewijs zonder deze causale link, zoals de statistiek uit de *Blue Bus*-casus.⁷⁴ En als bepaald statistisch bewijs de waarschijnlijkheid van een (vermeend) feit evenveel of zelfs meer⁷⁵ verhoogt dan getuigenbewijs, kan worden betoogd dat er geen reden meer is om het minder epistemische waarde toe te kennen en/of buiten beschouwing te laten.

Kortom: de *Blue Bus*-paradox kan worden opgelost door de conclusie te aanvaarden, ook al voelt dat misschien ongemakkelijk. Betoogd kan worden dat het ongemak veroorzaakt wordt door misleidende intuïties die voortvloeien uit een (voor het recht) ontoereikend kennisconcept.⁷⁶ In plaats daarvan zouden we ons moeten laten leiden door onze ratio, die erkent dat de epistemische waarde van getuigenbewijs niet bij voorbaat kan worden verkozen boven die van statistisch bewijs.

6. Conclusie

Aan de hand van de *Blue Bus*-paradox is de vraag onderzocht of in gevallen waarin het bewijs louter of hoofdzakelijk bestaat uit een statistisch gegeven, een (betwist) feit waarop de aansprakelijkheid van de gedaagde is gebaseerd kan worden bewezen. De vraag is te veelomvattend om enkel vanuit de overwegend epistemische invalshoek van dit artikel te beantwoorden. Zo zijn in dit artikel bijvoorbeeld geen morele overwegingen betrokken, zoals het bezwaar dat een individu door aansprakelijkheid op basis van louter of hoofdzakelijk statistisch bewijs aangesproken wordt op groepskenmerken (bijvoorbeeld etniciteit), wat in strijd lijkt met het idee dat personen als vrije en autonome individuen moeten worden behandeld.⁷⁷ Niettemin kunnen enkele afrondende opmerkingen worden gemaakt over de drie

mogelijke oplossingsrichtingen voor de paradox die hier wel zijn besproken.

Het verwerpen van de statistisch bewijs-premissie vanwege het referentieklassen-probleem (oplossingsstrategie 1) vormt een doeltreffende manier om de paradox op te lossen in het fictieve voorbeeld van de *Blue Bus*-casus. Het referentieklassen-probleem legt de potentiële zwakte van het statistische bewijs bloot. In de *Blue Bus*-casus is de referentieklassen bijzonder instabiel; zelfs een bescheiden onderzoeksbevinding, bijvoorbeeld dat de aanrijding 's nachts heeft plaatsgevonden terwijl Blauwe Bus B.V. schoolbussen exploiteert die niet 's nachts rijden, zou al voldoende zijn om de waarde van de statistiek te ondermijnen. Dit betekent echter niet dat deze oplossingsstrategie kan worden toegepast in alle gevallen van aansprakelijkheid op basis van louter of hoofdzakelijk statistisch bewijs, met een categorisch verbod van dit bewijs als gevolg. Niet elke statistiek is zo zwak als die in de *Blue Bus*-casus. Indien er in een concrete zaak voldoende (betrouwbare en valide) data beschikbaar zijn om een geschikte referentiegroep te bepalen, is deze oplossingsstrategie niet toepasselijk. Wel kan de gedaagde, nadat de rechter hem heeft toegelaten tot het leveren van tegenbewijs, data aandragen die het (betwiste) feit in een geschiktere referentieklassen plaatsen, waardoor het statistische bewijs van de eiser zijn waarde kan verliezen. Verder is het denkbaar dat met de opkomst van kunstmatige intelligentie en *big data* het eenvoudiger wordt tot sterkere statistieken te komen waarbij de waarschijnlijkheid minder gevoelig is voor veranderingen bij nieuwe feiten.

De paradox kan ook worden opgelost door de civiele bewijsstandaard niet te zien als een kwantificeerbare, absolute waarschijnlijkheid waartegen het bewijs moet worden afgewogen (oplossingsstrategie 2). Dit kan bijvoorbeeld door de bewijsstandaard te benaderen vanuit een scenariovergelijkingsperspectief. Dit hoeft niet noodzakelijk te betekenen dat probabilistische methoden moeten worden losgelaten, zoals bijvoorbeeld de *Bayesian decision theory* – waarin denken in kansen, kosten en scenario's samenkomen – aantoont. De vrijheid die de wet de civiele rechter biedt bij de bewijswaardering, verschaft ruimte voor een dergelijke alternatieve benadering van de bewijsstandaard. Of de civiele bewijsstandaard ook op die manier benaderd zou moeten worden, is een andere vraag die een diepgaander en

72. D. Papineau, 'The disvalue of knowledge', *Synthese* (198) 2019, afl. 6, p. 5320.

73. Papineau 2019, p. 5327-5328.

74. Thomson 1986, p. 205-206; Papineau 2019, p. 5313 en 5322.

75. Dit komt niet alleen door het statistische bewijs, maar ook het getuigenbewijs. De afgelopen decennia heeft (recht)psychologisch onderzoek reden gegeven te twijfelen aan de betrouwbaarheid van getuigenverklaringen vanwege de fouten die kunnen optreden bij het waarnemen, zich herinneren en het trekken van conclusies door de getuige. Zie hierover bijvoorbeeld: M.J. Dubelaar, *Betrouwbaar getuigenbewijs. Totstandkoming en waardering van strafrechtelijke getuigenverklaringen in perspectief* (diss. Leiden), Deventer: Kluwer 2014, p. 67, 73 en 162-165.

76. Vgl. F. Schoeman, 'Statistical vs direct evidence', *Noûs* (21) 1987, afl. 2, p. 179, die het wantrouwen jegens statistische inferentie als een irrationele bias beschouwt.

77. Zie bijvoorbeeld: D. Wasserman, 'The morality of statistical proof and the risk of mistaken liability', *Cardozo Law Review* (13) 1991, p. 941. Zie voor een kort overzicht van enkele morele overwegingen die een rol kunnen spelen bij de paradox, bijvoorbeeld: Ross 2020, par. 4.

breder onderzoek vergt dat niet slechts gemotiveerd is vanuit de behoefte de *Blue Bus*-paradox op te lossen.⁷⁸

De laatste oplossingsstrategie voor de *Blue Bus*-paradox is het accepteren van de aanvankelijk 'onaanvaardbare' conclusie van aansprakelijkheid op basis van louter of hoofdzakelijk statistisch bewijs (oplossingsstrategie 3). Veel mensen lijken een (intuïtieve) afkeer te hebben van statistisch bewijs, mogelijk omdat het zich moeilijk verhoudt met ons traditionele kennisbegrip, vanwege het ontbreken van een causaal pad van het (vermeende) feit naar de kenner van het feit. Toch kan worden betoogd dat deze afkeer onterecht is, zeker in situaties waarin statistisch bewijs de waarschijnlijkheid van het (vermeende) feit evenveel of zelfs meer verhoogt dan getuigenbewijs. In zulke gevallen zou statistisch bewijs

epistemisch gezien zelfs beter te verdedigen kunnen zijn dan getuigenbewijs. Daar moet weer bij worden opgemerkt dat het bewijsoordeel niet alleen epistemische, maar ook morele componenten bevat. Zelfs als de weerszin tegen statistisch bewijs gestoeld is op irrationaliteit, dwingt het normatieve karakter van het recht om deze intuïtie serieus te nemen. Het recht is er voor de mens, en niet andersom, wordt weleens gezegd. Een van (menselijke) irrationaliteit ontdaan bewijsstelsel kan zijn legitimiteit verliezen, als het contra-intuïtief is voor een (nog) te grote groep burgers. Rechters kunnen hier een sleutelrol vervullen door het gebruik van statistisch bewijs in hun uitspraken extra zorgvuldig te motiveren. Zo kan het publiek worden geholpen eventuele irrationele intuïties bij zichzelf te herkennen als cognitieve biases.

78. Zie hiervoor bijvoorbeeld: Allen & Pardo 2019, par. 1 en 2, alsook: A.-R. Mackor, H. Jellema & P.L. van Koppen, 'Explanation-based approaches to reasoning about evidence and proof in criminal trials', in: B. Brozek, J. Hage & N. Vincent (red.), *Law and mind: A survey of law and the cognitive sciences*, Cambridge: Cambridge University Press 2021, par. 21.3 en 21.4.