Statistische fouten in een logisch dialoogsysteem

Wouter Bouvy

16 juni 2008

Inhoudsopgave

1 Inleiding **2**

2 *Prosecutor's fallacy* en *Defense attorney's fallacy* **4**
- 2.1 Situatieschets 4
- 2.2 *Prosecutor's fallacy* 4
- 2.3 *Defense attorney's fallacy* 5

3 Explaining anchors van Floris Bex[1] **6**
- 3.1 Dialoogsystemen 6
- 3.2 Het systeem van Bex 7
- 3.3 De actie *explain* 9
- 3.4 Voorbeelden van gebruik 10
- 3.5 Opmerkingen over notatie 12

4 Testen van het systeem **14**
- 4.1 *Prosecutor's fallacy* 14
- 4.2 *Defense attorney's fallacy* 18

5 Conclusie **21**

1 Inleiding

In het artikel 'Explaining Anchors'[1] beschrijft Floris Bex een logisch dialoogsysteem, om argumentatie in rechtszaken te modelleren voor beter inzicht. Met zijn systeem kunnen onderliggende aannamen, zogenaamde 'common sense generalisations', duidelijk worden gemaakt. Deze 'common sense generalisations' zijn algemene regels waarvan we uit gaan dat ze waar zijn, zoals 'Experts zijn over het algemeen betrouwbare bronnen' of 'Getuigen spreken meestal de waarheid'. Bij elke zaak worden deze aannamen gebruikt, maar ze kloppen niet altijd. Met het systeem van Bex kan men evalueren of deze wel correct zijn.

Deze scriptie onderzoekt of het systeem ook breder gebruikt kan worden. Fouten die vaak in rechtszaken worden gemaakt, en waar het artikel van Bex het niet over heeft, zijn fouten met betrekking tot statistisch bewijs.

Voor een goede beschrijving van een aantal statistische fouten kijken we naar een artikel van Huygen. In ' "Bayesian Belief Networks" voor redeneren over juridische bewijsvoering'[3] bespreekt Huygen een aantal fouten in het gebruik van statistiek in de rechtszaal. De belangrijkste statistische fouten zijn de *Prosecutor's fallacy* en de *Defense attorney's fallacy* .

Dit leidt ons naar de volgende vraag: Kan het systeem ook gebruikt worden om deze statistische fouten te detecteren?
Om dit te onderzoeken gaan we een situatie waar een *Prosecutor's fallacy* in voor komt (en een situatie met een *Defense attorney's fallacy*) in het systeem proberen te modelleren. Vervolgens kijken we in hoeverre het ontstane schema voldoet aan de eisen van het systeem, in hoeverre het de fout aan het licht brengt en of dit voldoende is om de fout af te vangen. De fout kan afgevangen worden doordat de gebruiker het ziet of doordat het systeem het niet toe laat. Er zijn verschillende mogelijke resultaten:

- Het systeem laat de fouten toe en maakt ze niet zichtbaar. In dat geval kijken we of dat met een kleine aanpassing is op te lossen.
- Het systeem laat de fouten toe maar laat de fouten wel duidelijk zien. In dat geval kan het systeem in ieder geval helpen met het afvangen van de fouten. Het is dan wel belangrijk dat de gebruiker van het systeem de nodige kennis heeft van de gemaakte statistische fout.
- Het systeem laat de fouten helemaal niet toe, omdat bepaalde stappen in het systeem niet worden geaccepteerd. In dat geval hoeft de gebruiker geen kennis te hebben van de statistische fouten.

In sectie 2 volgt een voorbeeld van een situatie waarin de *Prosecutor's fallacy* en *Defense attorney's fallacy* op kunnen spelen. Aan de hand van het voorbeeld zullen de beide fouten worden uitgelegd. In sectie 3 wordt het systeem van Bex in detail uitgelegd.

Bayesiaanse netwerken

De oplossing die Huygen[3] in zijn artikel stelt is het inzetten van een "Bayesian Belief Network". Een "Bayesian Belief Network", of Bayesiaans netwerk, is een netwerk bestaande uit een verzameling waarschijnlijkheden en de afhankelijkheden tussen deze waarschijnlijkheden.

Het gebruik van een Bayesiaans netwerk om de *Prosecutor's fallacy* en *Defense attorney's fallacy* af te vangen is een logische keuze, het netwerk is gebaseerd op statistiek en het correct verwerken van statistiek. Specifieker, de *Prosecutor's fallacy* gaat over het gebruik van de regel van Bayes nalaten waar dit wel moet, iets wat in een Bayesiaans netwerk gegarandeerd goed gaat.

Waarom dan niet gewoon een Bayesiaans netwerk gebruiken om alles op te lossen? Er zijn een paar nadelen die in het gebruik in combinatie met rechtszaken naar voren komen. In een Bayesiaans netwerk moeten alle waarschijnlijkheden bekend zijn. Dit betekent vaak dat een rechter de waarschijnlijkheden af moet schatten, op basis van eerdere ervaring. Dit is geen makkelijke taak voor een rechter, en voor veel situaties is het gebruik van zulke gedetailleerde waarschijnlijkheden overdreven, en kan de rechtszaak zonder deze complicatie ook opgelost worden.

Bayesiaanse netwerken zijn geenszins onbruikbaar, ze zijn in veel situaties erg nuttig en kunnen goed worden gebruikt. Echter, juristen houden niet zo van waarschijnlijkheden en werken liever met een minder complex systeem van waarheden en onwaarheden, wat een dialoogsyteem geschikter maakt voor algemeen gebruik.

Maar dit alles terzijde, het doel in dit artikel is niet om een waardeoordeel of vergelijking te geven tussen dialoogsystemen en Bayesiaanse netwerken, maar om te kijken of (een deel) van de gestelde problemen in het artikel van Huygen ook in het systeem van Bex gedetecteerd kunnen worden.

2 *Prosecutor's fallacy* en *Defense attorney's fallacy*

We zullen nu met behulp van een voorbeeld uitleggen wat de *Prosecutor's fallacy* en de *Defense attorney's fallacy* inhouden.

2.1 Situatieschets

Neem aan dat een misdaad is gepleegd in een stad. De politie weet dat de dader een man is, en er is DNA materiaal gevonden.[1]

20.000 mannen in deze stad hebben hun DNA vergeleken met een sample van de misdaad. Eén van deze mannen heeft een overeenkomend DNA profiel, en tijdens zijn rechtszaak wordt door een expert getuigd dat de kans dat twee DNA profielen matchen slechts 1 op 10.000 is. Dat wil zeggen dat de kans dat het profiel overeenkomt terwijl het bloed toch niet van de verdachte is, 1 op de 10.000 bedraagt. De aanklager betoogt dat daarom de kans de verdachte onschuldig is, 1 op de 10.000 bedraagt.

Er zijn in de stad naast de verdachte nog 100.000 andere mannen. Hiervan zullen waarschijnlijk 10 personen hetzelfde DNA profiel hebben als de verdachte. De verdedigende advocaat betoogt daarom dat de kans dat het bloed dat is gevonden op de plek van de misdaad van de verdachte is, slechts 1 op de 11 bedraagt, dus ongeveer 9%, en dat de kans dat de verdachte onschuldig is, daarom 91% is.

2.2 *Prosecutor's fallacy*

De *Prosecutor's fallacy* is een denkfout waarbij conditionele kansen met elkaar worden verward. Het gaat hier om het onderscheid tussen

- de waarschijnlijkheid van bewijs, gegeven schuldig zijn of onschuldig zijn $P(bewijs|schuldig)$

  tegenover

- de waarschijnlijkheid van schuldig of onschuldig zijn, gegeven bewijs $P(schuldig|bewijs)$

In het eerder gestelde voorbeeld heeft de aanklager een DNA match verzameld, en een expert laten getuigen dat de kans dat dit gevonden was als de verdachte onschuldig was heel erg klein is. Deze kans wordt genoteerd als $P(match|onschuldig)$, de kans op een match, gegeven de informatie dat de verdachte onschuldig is. Hierna betoogt de aanklager dat zijn bewijs de verdachte met hoge waarschijnlijkheid schuldig maakt. Dit is de kans $P(schuldig|match)$,

[1] Dit voorbeeld komt grotendeels uit [4]

de kans dat iemand schuldig is, als deze matcht met het gevonden DNA. Hij doet dit door te stellen dat als de kans op een match gegeven onschuldig zo klein was, dan moet de kans dat de verdachte schuldig is, gegeven de match, wel heel groot zijn.

Dit laatste is de denkfout. De aanklager stelt dat de kans $P(schuldig|match)$ heel hoog is, omdat hij denkt dat de kans $P(onschuldig|match)$ heel laag is. Hij maakt hier de fout door te denken dat $P(match|onschuldig)$ gelijk is aan $P(onschuldig|match)$, maar deze twee kansen zijn niet gelijk.

Bayes' Theorem stelt de volgende functie:

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

Met andere woorden, de a priori kans op A en de a priori kans op B moeten meegenomen worden in de omzetting van $P(A|B)$ naar $P(B|A)$.

Als we het bovenstaande voorbeeld in de regel invullen krijgen we:

$$P(onschuldig|match) = \frac{P(match|onschuldig) \times P(onschuldig)}{P(match)}$$

(we hebben $P(match|onschuldig)$ en we willen $P(onschuldig|match)$)
$P(onschuldig)$ is hier de a priori kans dat verdachte P onschuldig is, deze is waarschijnlijk erg groot. $P(match)$ is de kans dat een DNA match wordt gevonden met een willekeurig persoon, schuldig of onschuldig.

De aanklager heeft dus twee factoren overgeslagen bij zijn aanname, en deze moeten berekend en verwerkt worden voordat de kans van $P(match|onschuldig)$ omgezet kan worden naar de kans van $P(onschuldig|match)$ (en daarna naar $P(schuldig|match)$).

2.3 *Defense attorney's fallacy*

De fout in het voorbeeld in sectie 2.1 zien we dat de verdedigende advocaat 10 willekeurige andere (hypothetische) mannen uit de stad pikt en als mogelijke schuldigen aanwijst op basis van de DNA test.

De fout die de verdedigende advocaat hier maakt: hij gaat er van uit dat deze tien personen dezelfde mogelijkheden en eigenschappen hebben als de huidige verdachte. Hij gaat uit van dezelfde a priori kans voor elke man in de stad. Hierbij houdt hij geen rekening met de mogelijke leeftijden die deze nieuwe verdachten kunnen hebben, van baby tot bejaard. Ook toegang tot de locatie van het misdrijf wordt genegeerd. Ook is er vaak ander bewijs dat naar de verdachte wijst, en/of ander personen zou uitsluiten.

3 Explaining anchors van Floris Bex[1]

In het artikel 'Explaining Anchors'[1] beschrijft Floris Bex een logisch dialoogsysteem, om argumentatie in rechtszaken te modelleren voor beter inzicht.
Het artikel werd geschreven op basis van een boek van Crombag, van Koppen and Wagenaar, genaamd 'Dubieuze Zaken: De Psychologie van Strafrechtelijk Bewijs'[2], en het opvolgende boek 'Anchored Narratives: The Psychology of Criminal Evidence'. Hierin werd een theorie beschreven voor beslissingen in rechtszaken.

Het belangrijkste element in deze theorie is dat beslissingen vaak gemaakt worden door middel van zogenaamde 'common sense generalisations'.

Deze 'common sense generalisations' liggen aan de basis van alle redenatie in de rechtszaak. Beslissingen liggen in een hiërarchie met een zogenaamd anchor-regel die de verbinding legt met de 'common sense generalisations' onderaan de hiërarchie.

Voorbeelden van anchor-regels zijn 'Experts zijn over het algemeen betrouwbare bronnen' en 'Getuigen onder ede spreken meestal de waarheid'.

Door duidelijk te maken welke archor-regels aan de grond van een verhaal liggen, kan een rechter kijken of deze niet dubieus of te algemeen zijn.

3.1 Dialoogsystemen

Dialoogsystemen hebben allemaal een algemene vorm:
Een dialoogsysteem modelleert de dialoog tussen twee partijen. Er is een 'proponent'(voorstander) en een 'opponent'(tegenstander), beide argumenteren over een bepaalde stelling. De proponent heeft een bepaalde stelling gedaan en wil aantonen dat deze waar is, terwijl de opponent wil aantonen dat de stelling onwaar is.

De twee spelers kunnen verschillende acties uitvoeren om de ander tegen te gaan:
argue om een stelling toe te voegen.
why om te stellen dat je het niet met een stelling eens bent en bewijs wilt,
concede om een stelling waar de speler het niet mee eens was toe te geven,
retract om een stelling waar de speler achter stond op te geven,

Terwijl *why*, *concede* en *retract* maar over een enkele stelling gaan, kan een speler met *argue* een heel argument geven, bestaande uit meerdere stellingen en regels.
Als het de eerste stelling is wordt de stelling waar de discussie over gaat gesteld. In latere zetten wordt vaak een regel gesteld samen met de premissen voor deze regel. Zo'n regel heeft dan als conclusie een eerder genoemde stelling waar de

tegenstander bewijs om vroeg, of een negatie van een stelling van de tegenstander, om te bewijzen dat deze niet waar is.

Dit alles wordt beregeld door het systeem. Stellingen waar de verschillende speler voor en tegen zijn worden bijgehouden, en het systeem bepaald welke zetten wel en welke niet gedaan kunnen worden.

Bex stelt in zijn artikel voor om deze basis uit te bereiden met de mogelijkheid om deze anchor regels verder te definiëren. Denk aan het volgende voorbeeld: 'Als een expert iets getuigt is dit waar' zal in de veel zaken waar zijn, maar in sommige zaken zou deze regel moeten worden gespecificeerd naar bijvoorbeeld 'Als een expert iets getuigd en de expert is betrouwbaar is het waar'. Op deze manier kunnen onderliggende assumpties zoals de betrouwbaarheid van de expert aan de kaart gebracht worden.

3.2 Het systeem van Bex

Het systeem van Bex bestaat uit een tupel:
$(L, Players, Acts, Moves, PlayerToMove, Comms, Disp, G, Admissible, Winner)$, waarbij geldt:

- L is de uitgebreide logica van Pollock.[2]
  Met deze logica worden de verzameling van correct geformuleerde fomulae ($wff(L)$) en de verzameling van correct geformuleerde argumenten ($Args(L)$) gedefinieerd.

- $Players$ zijn de twee spelers, de *proponent* en de *opponent*.

- $Acts$ zijn de mogelijk zetten.
  Dit zijn de eerdergenoemde *argue* A, *why* ϕ, *concede* ϕ en *retract* ϕ, en een nieuwe actie: *explain* ($\phi = A$). Met de *explain* actie kan men een regel specificeren, bijvoorbeeld de regel 'Als een expert iets getuigt is dit waar' omzetten naar 'Als een expert iets getuigt en de expert is betrouwbaar is dit waar'. De explain wordt uitgelegd in sectie 3.3

- $Moves$ is een verzameling met alle zetten die gedaan zijn, welke speler deze gedaan heeft en wanneer. $M_i = (Player, Act)$ en elke dialoog van lengte i is een lijst van $M_1, ..., M_i$.

[2] Deze laten we verder buiten beschouwing, het systeem is zonder uitleg over de logica van Pollock ook duidelijk. Voor meer informatie over de logica van Pollock, zie het originele artikel van Pollock[5] en het artikel van Floris Bex[1] voor een korte samenvatting en uitleg over zijn aanpassingen.

- $PlayerToMove$ is de speler die aan de beurt is.
De verliezende speler is altijd aan de beurt. Zodra deze speler een actie heeft gedaan waardoor deze winnend is (een bewijs waardoor het bewijs van de tegenstander niet meer klopt), is de tegenstander aan de beurt en kan deze acties gaan doen.

- $Comms$ zijn de commitments van de spelers, aan welke uitspraken ze zich gebonden hebben. De verschillende acties hebben een effect op commitments:
argue $Pso\phi$: voegt zowel P als ϕ aan de commitments van de speler toe,
why ϕ: heeft geen effect op commitments,
concede ϕ: voegt ϕ aan de commitments van de speler toe,
retract ϕ: verwijdert ϕ uit de commitments van de speler,
explain $\phi = (Pso\psi)$: Bij elke speler die ϕ in zijn commitments heeft, worden P en ψ toegevoegd.

- $Disp$ zijn de disputations van de spelers, welke uitspraken ze het niet mee eens zijn. De verschillende acties hebben een effect op commitments:
argue $Pso\phi$: heeft geen effect op disputations,
why ϕ: voegt ϕ toe aan de disputations van de speler,
concede ϕ: verwijdert ϕ uit de disputations van de speler,
retract ϕ: heeft geen effect op disputations,
explain $\phi = (Pso\psi)$: Bij elke speler die ϕ niet in zijn commitments heeft, wordt ϕ toegevoegd aan de disputations.

- G is de argument graph. In deze graph zijn de verschillende stellingen en regels als nodes weergegeven. Nodes zijn verbonden met support en defeat links, die het volgende doen: Als x een support link het naar y, is x een reden voor y. Als x een defeat link heeft naar y, is x een stelling die y onwaar maakt.

- $Admissible$ zijn de regels waar een zet aan moet voldoen.
argue A als het de eerste zet is, als de conclusie van A in de disputations van de tegenstander zit of als de conclusie van A een eerder argument verslaat,
why ϕ als ϕ niet in de disputations van de speler zit, de speler geen commitments heeft die leiden tot ϕ en ϕ nog niet uitgelegd is,
concede ϕ als ϕ niet al in de commitments van de speler zit en ϕ eerder gesteld is,
retract ϕ als ϕ in de commitments van de speler zit,
explain $(\phi = A)$ als ϕ eerder gesteld is en de nieuwe regel iets nieuws toe voegt.

Als de proponent in de eerste zet ϕ stelde en ϕ zit niet meer in de commitments van de proponent, dan zijn geen zetten meer mogelijk (dan heeft de opponent gewonnen, want het eerste argument is opgegeven).
Als de proponent in de eerste zet ϕ stelde en de commitments van opponent geven ϕ, dan zijn er geen zetten meer mogelijk (de proponent heeft dan gewonnen, aangezien deze toegeeft dat ϕ waar is).

- $Winner$ bepaalt wie winnende is. Als de speler aan zet geen zetten meer kan doen, heeft $Winner$ gewonnen.
  Dit wordt bepaald door de verzameling argumenten die niet bij een van de spelers in de disputations zitten. Als de proponent in de eerste zet ϕ heeft gesteld, en uit deze verzameling blijkt dat ϕ waar is, dan is de proponent aan het winnen. Blijkt niet uit de verzameling dat ϕ waar is, dan is de opponent aan het winnen.

Verder gelden de volgende eisen:
Spelers mogen geen stellingen hebben die in zowel de commitments als de disputations voorkomen.
Spelers mogen zichzelf niet tegen spreken.
Spelers mogen niet tweemaal dezelfde actie doen.

3.3 De actie *explain*

Om regels te kunnen specificeren moet er naast de eerdergenoemde acties een actie bij komen; de actie om een regel te specificeren. Dit is de $explain(\phi = A)$ actie.

Met een *explain* zet kan een speler een eerder gestelde regel specifieker maken.

Bex geeft de volgende definitie van een explain act:
An explanation is of the form $(\phi = A)$, where ϕ is a generalization of the form $\psi \Rightarrow \chi$ and A is an argument in $Args(L)$ such that ψ is a premise of A and χ is the conclusion of A.
(Bex [1, p. 26])

De *explain* regel wordt het snelst duidelijk aan de hand van een voorbeeld:

Stel de proponent heeft 'Expert e zegt $x \Rightarrow x$ is waar' gesteld. De opponent kan nu de regel specificeren met de explain actie:

Con:	explain	
	(ϕ)	Expert e zegt $x \Rightarrow x$ is waar
	$=$	
	(1)	Expert e zegt x
	(3)	Expert e is betrouwbaar
	r_{nieuw}	Als Expert e zegt $x \wedge$ Expert e is betrouwbaar $\Rightarrow x$ is waar
	so	
	(2)	x is waar

Merk op dat van de $explain(\phi = A)$, de ϕ hier de (ϕ) is en de A het argument na de =, dus ((1),(3),r_{nieuw} so (2)).
De nieuwe regel vervangt ϕ en komt dus als commitment van de proponent in het systeem. Dit betekent dat de opponent nu kan proberen te bewijzen dat de expert niet betrouwbaar is, of met een *why* de proponent kan laten bewijzen dat de expert wel betrouwbaar is.

3.4 Voorbeelden van gebruik

Hier volgen een aantal voorbeelden om het gebruik van het systeem duidelijk te maken. We beginnen met een simpel voorbeeld zonder *explain*. Dit voorbeeld komt uit het artikel van Floris Bex[1] op pagina 31.

Om verwarring te voorkomen: pro_1, con_2, pro_3, etc zijn de beurten en de speler aan de beurt, in de vorm $speler_{beurt-nummer}$. Meer over dit bij de opmerkingen over notatie, sectie 3.5.

pro begint met de eerste *argue*.

		$Comm_1(pro) = \{ (1) \quad q\}$
		$Comm_1(con) = \{\}$
Pro_1:	argue (1) q	$Disp_1(pro) = \{\}$
		$Disp_1(con) = \{\}$

con kan nu reageren met een argument voor $\neg q$ of met *why* q. Hij kan ook *concede* q doen, maar dan zou hij opgeven.

		$Comm_2(pro) = \{ (1) \quad q\}$
		$Comm_2(con) = \{\}$
Con_2:	why (1) q	$Disp_2(pro) = \{\}$
		$Disp_2(con) = \{ (1) \quad q\}$

Nu moet *pro* zijn q verdedigen door een argument te geven.

$$
\begin{array}{llll}
Pro_3: & \text{argue} & (2) & p \\
& & r_1 & p \Rightarrow q \\
& & \text{so} & \\
& & (1) & q
\end{array}
\qquad
\begin{array}{lrl}
Comm_3(pro) = & \{\ (1) & q, \\
& (2) & p, \\
& r_1 & p \Rightarrow q\} \\
Comm_3(con) = & \{\} & \\
Disp_3(pro) = & \{\} & \\
Disp_3(con) = & \{\ (1) & q\}
\end{array}
$$

con kan nu een *why* op r_1 doen.

$$
\begin{array}{lll}
Con_4: & \text{why } r_1 & p \Rightarrow q
\end{array}
\qquad
\begin{array}{lrl}
Comm_4(pro) = & \{\ (1) & q, \\
& (2) & p, \\
& r_1 & p \Rightarrow q\} \\
Comm_4(con) = & \{\} & \\
Disp_4(pro) = & \{\} & \\
Disp_4(con) = & \{\ (1) & q, \\
& r_1 & p \Rightarrow q\}
\end{array}
$$

Nu moet *pro* zijn $p \Rightarrow q$ verdedigen door een argument te geven.

$$
\begin{array}{llll}
Pro_5: & \text{argue} & (3) & u \\
& & r_2 & u \Rightarrow (p \Rightarrow q) \\
& & \text{so} & \\
& & r_1 & p \Rightarrow q
\end{array}
\qquad
\begin{array}{lrl}
Comm_5(pro) = & \{\ (1) & q, \\
& (2) & p, \\
& r_1 & p \Rightarrow q, \\
& (3) & u, \\
& r_2 & u \Rightarrow (p \Rightarrow q)\} \\
Comm_5(con) = & \{\} & \\
Disp_5(pro) = & \{\} & \\
Disp_5(con) = & \{\ (1) & q, \\
& r_1 & p \Rightarrow q\}
\end{array}
$$

Nu ziet *con* zijn kans en weet p onwaar te maken.

$$
\begin{array}{llll}
Con_6: & \text{argue} & (3) & u \\
& & r_3 & u \Rightarrow \neg p \\
& & \text{so} & \\
& & (4) & \neg p
\end{array}
\qquad
\begin{array}{lrl}
Comm_6(pro) = & \{\ (1) & q, \\
& (2) & p, \\
& r_1 & p \Rightarrow q, \\
& (3) & u, \\
& r_2 & u \Rightarrow (p \Rightarrow q)\} \\
Comm_6(con) = & \{\ (3) & u, \\
& r_3 & u \Rightarrow \neg p, \\
& (4) & \neg p\} \\
Disp_6(pro) = & \{\} & \\
Disp_6(con) = & \{\ (1) & q, \\
& r_1 & p \Rightarrow q\}
\end{array}
$$

pro heeft geen andere keus dan opgeven.

Pro_7: retract (2) p

Pro_7: retract r_1 $p \Rightarrow q$

Pro_7: retract (1) q

Alternatief, *pro* had in beurt 5 ook een explain actie kunnen doen:

Pro_5: explain

r_1	$p \Rightarrow q$
$=$	
(2)	p
r_2	$p \Rightarrow s$
so	
(3)	s
r_3	$s \Rightarrow q$
so	
(1)	q

$Comm_5(pro) =$	$\{$ (1)	$q,$
	(2)	$p,$
	(3)	$s,$
	r_2	$p \Rightarrow s,$
	r_3	$s \Rightarrow q\}$
$Comm_5(con) =$	$\{\}$	
$Disp_5(pro) =$	$\{$ r_1	$p \Rightarrow q\}$
$Disp_5(con) =$	$\{$ (1)	$q,$
	r_1	$p \Rightarrow q\}$

Er zijn uiteraard nog veel meer mogelijke zaken te modelleren, en de verschillende stellingen p, q, etc kunnen een stuk complexer worden, maar dit is de algemene vorm van een dialoog in dit dialoogsysteem.

3.5 Opmerkingen over notatie

Bij het schrijven van dit artikel kwamen enkele opmerkingen over de opmaak en symboolkeuze van Floris Bex[1] naar voren. Om dit artikel zo veel mogelijk bij het artikel van Bex aan te laten sluiten, worden dezelfde notaties gebruikt, en zal hier kort melding worden gemaakt van een paar mogelijk verwarrende notaties.

- pro_1, con_2, pro_3, etc.
  Het lijkt hier alsof elke beurt een andere pro of con aan zet is, omdat deze genummerd zijn in plaats van de beurten. Een duidelijkere notatie zou iets zijn als:
  1_{pro}
  $1 : Pro :$ of
  $1.Pro :$

- $\wedge$ in combinatie met $\Rightarrow$
  De merkwaardigheid hier is dat de gebruikte symbolen niet in dezelfde categorie vallen:
  Object-taal: $\wedge$ en $\rightarrow$
  Meta-taal: ‘en’ en $\Rightarrow$

Hier zou dus moeten worden gekozen voor één van de twee talen en de notatie uit die taal zou gebruikt moeten worden.

- Er is een 'proponent', *pro*, en een 'opponent', *con*. Deze laatste benaming kan verwarrend zijn, al zijn de *pro* en *con* zonder benaming wel duidelijk.

4 Testen van het systeem

In deze sectie proberen we de *Prosecutor's fallacy* en de *Defense attorney's fallacy* in het systeem van Bex te zetten, en kijken we hoe deze met de fouten omgaat.

4.1 *Prosecutor's fallacy*

De *Prosecutor's fallacy* , zoals uitgelegd in sectie 2.2, betreft een fout bij het omzetten van twee conditionele kansen. We voeren nu het voorbeeld uit sectie 2.1 in in het systeem van Bex en kijken hoe het systeem hier mee om gaat.

In het systeem ziet de *Prosecutor's fallacy* er als volgt uit:

Pro_1:	argue (1)	P is schuldig
Con_2:	why (1)	P is schuldig
Pro_3:	argue (2)	P matcht met gevonden DNA
	(3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
	r_1	(ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is ψ
	so	
	(4)	Kans dat P is onschuldig is 1 op de 10.000
	(5)	Kans van 1 op de 10.000 is heel klein
	r_2	(Kans op ϕ is onschuldig is ψ $\wedge$ ψ is heel klein) $\Rightarrow$ ϕ is schuldig
	so	
	(1)	P is schuldig
Con_4:	why (3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
Pro_5:	argue (6)	Deze expert zegt dat de kans op overeenkomt terwijl de persoon toch niet de verdachte is: 1 op de 10.000
	r_3	(E is een expert $\wedge$ E zegt ϕ) $\Rightarrow$ ϕ
	so	
	(3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
Con_6:	explain	
	r_1	(ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is ψ

	=	
	(2)	P matcht met gevonden DNA
	(3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
	r_4	(ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is een functie op ψ : $f(\psi)$
	so	
	(7)	Kans dat P is onschuldig is f(1 op de 10.000)
	(8)	Kans van f(1 op de 10.000) is heel klein
	r_2	(Kans op ϕ is onschuldig is $\psi \wedge \psi$ is heel klein) $\Rightarrow \phi$ is schuldig
	so	
	(1)	P is schuldig
Con_7:	why (8)	Kans van f(1 op de 10.000) is heel klein
Pro_8:	argue (9)	f(1 op de 10.000) $\approx$ 1 op de 10.000
	(5)	Kans van 1 op de 10.000 is heel klein
	r_5	(Kans op ϕ is klein $\wedge$ kans $\phi \approx$ kans χ) $\Rightarrow$ Kans op χ is klein
	so	
	(8)	Kans van f(1 op de 10.000) is heel klein
Con_9:	explain	
	r_4	(ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is een functie op ψ : $f(\psi)$
	=	
	(2)	P matcht met gevonden DNA
	(3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
	r_6	(ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is een functie op ψ : $f(\psi) = \psi \times \frac{P(onschuldig)}{P(match)}$
	so	
	(10)	Kans dat P gegeven deze match onschuldig is is (1 op de 10.000) $\times \frac{P(onschuldig)}{P(match)}$
	(11)	Kans van (1 op de 10.000) $\times \frac{P(onschuldig)}{P(match)}$ is heel klein
	r_2	(Kans op ϕ is onschuldig is $\psi \wedge \psi$ is heel klein) $\Rightarrow \phi$ is schuldig
	so	
	(1)	P is schuldig
Con_{10}:	why (11)	Kans van (1 op de 10.000) $\times \frac{P(onschuldig)}{P(match)}$ is heel klein

Pro_{11}:	argue (12)	1 op de 10.000 $\times \frac{P(onschuldig)}{P(match)} \approx$ 1 op de 10.000
	(5)	Kans van 1 op de 10.000 is heel klein
	r_5	(Kans op ϕ is klein $\wedge$ kans $\phi \approx$ kans χ) $\Rightarrow$ Kans op χ is klein
	so	
	(11)	Kans van (1 op de 10.000) $\times \frac{P(onschuldig)}{P(match)}$ is heel klein
Con_{12}:	why (12)	1 op de 10.000 $\times \frac{P(onschuldig)}{P(match)} \approx$ 1 op de 10.000
Pro_{13}:	argue (13)	$\frac{P(onschuldig)}{P(match)} \approx 1$
	(10)	Kans dat P gegeven deze match onschuldig is is (1 op de 10.000) $\times \frac{P(onschuldig)}{P(match)}$
	r_7	$\phi \times 1 \Rightarrow \phi$
	so	
	(12)	1 op de 10.000 $\times \frac{P(onschuldig)}{P(match)} \approx$ 1 op de 10.000
Con_{14}:	why (13)	$\frac{P(onschuldig)}{P(match)} \approx 1$

Nu moet de aanklager naar de kansen gaan kijken en zorgen dat de regel van Bayes goed uitgerekend wordt. Op deze manier kan de verdediging deze fout tegengaan.

Dit is hoe de zaak ongeveer zou moeten lopen om de fout aan het licht te brengen. Het nadeel is dat de verdediging zich bewust moet zijn van de fout om deze aan te kaarten. Maar, er moet nog met iets anders rekening worden gehouden: Worden alle regels wel in het systeem toegelaten?

Het systeem stelt dat de regels uit kennis van de wereld voort moeten komen, dat ze op basis van kennis van de wereld waar moeten zijn. We kunnen dit bij elke regel bekijken:

De regels gebruikt in het *Prosecutor's fallacy* voorbeeld

Gebruikte algemene regels

r_2 (Kans op ϕ is onschuldig is $\psi \wedge \psi$ is heel klein) $\Rightarrow \phi$ is schuldig
Deze regel is een correcte regel, bij een bepaalde kleine kans op χ, kan worden aangenomen dat deze verwaarloosbaar klein is en dat χ kan worden uitgesloten. Echter kan dit in combinatie met andere kansen niet meer kloppen, en dan zou de speler de regel uit moeten bouwen met de nieuwe kansen. Ook is niet duidelijk waar de grens ligt voor 'heel klein'.

r_3 (E is een expert $\wedge$ E zegt ϕ) $\Rightarrow \phi$
Floris Bex[1] : "when an expert says something about his subject domain, we can believe what he says most of the time"

r_5 (Kans op ϕ is klein $\wedge$ kans $\phi \approx$ kans χ) $\Rightarrow$ Kans op χ is klein
Als twee kansen dicht bij elkaar liggen en één ervan wordt een bepaalde eigenschap van grootte toegewezen, kan men redelijkerwijs stellen dat deze ook toepasbaar is op de andere kans.

Gebruikte *Prosecutor's fallacy* regels

r_1 (ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is ψ
Het probleem is dat mensen die niet bekend zijn met *Bayes' Theorem* of met de *Prosecutor's fallacy* dit zo als waar accepteren. Het voordeel is dat het systeem deze regel zou kunnen weigeren, en duidelijk zou kunnen maken dat deze incompleet is (zie r_4 en r_6).

r_4 (ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is een functie op ψ : $f(\psi)$
Deze regel zou het systeem wel toe moeten laten, met de opmerking dat de functie de waarde flink kan veranderen. Deze regel, met

$$f(\psi) = \psi \times \frac{P(onschuldig)}{P(match)}$$

kan direct uit de werkelijkheid worden gehaald. Deze regel is waarschijnlijk toegankelijker dan r_6, omdat hier alleen wordt gesteld dat de kans gemuteerd wordt, zonder berekeningen in het argument te verwerken (als een expert zijn mening geeft, geeft deze ook niet alle achterliggende berekeningen en consideraties).

r_6 (ϕ matcht met DNA $\wedge$ kans op match als ϕ onschuldig was is ψ) $\Rightarrow$ kans dat ϕ onschuldig is (gegeven deze match) is een functie op ψ : $f(\psi) = \psi \times \frac{P(onschuldig)}{P(match)}$
Dit is de complete correcte regel. Deze regel is geheel uit wereldkennis af te leiden. Het systeem zou deze regel zonder problemen toe moeten laten. De *Prosecutor's fallacy* is hier erg onwaarschijnlijk, omdat de gehele berekening bloot ligt en niet genegeerd kan worden.

Aangezien het systeem eist dat de gestelde regels een anchor hebben in de wereldkennis / common knowledge, is het te betwijfelen of een regel als r_1 in het systeem toegelaten wordt.

Zoals gezien in het voorbeeld kan de *Prosecutor's fallacy* nog steeds van toepassing zijn op r_4, hoewel deze duidelijker zichtbaar wordt voor de verdediging in statement (9) f(1 op de 10.000) $\approx$ 1 op de 10.000

Regel r_4 is misschien incompleet, maar wel correct, en zou net zo goed in het systeem toegelaten moeten worden als een regel als r_3, die ook incompleet is (deze bevat bijvoorbeeld niets over de betrouwbaarheid van de expert), maar wel correct en bruikbaar.

Conclusie:
De *Prosecutor's fallacy* zal waarschijnlijk door het systeem aan het licht gebracht worden. Hoewel de verdediging zich misschien bewust moet zijn van de fout om deze aan te kaarten, maakt het systeem de fout goed zichtbaar.

Omdat de regel waar $P(match|onschuldig)$ gelijk is aan $P(onschuldig|match)$ niet in het systeem geaccepteerd zou mogen worden, moet de regel worden aangepast naar $P(match|onschuldig)$ gelijk is aan $f(P(onschuldig|match))$. Dit maakt duidelijk dat er een verandering in de kans is, waardoor de aanklager moet verklaren waarom er geen verandering zou zijn. Als de verdediging de fout niet kent, is de enige manier waarop de verdediging het kan accepteren is als ze een stelling als 'f(1 op de 10.000) $\approx$ 1 op de 10.000' zonder verdere vragen accepteren, en dit is niet waarschijnlijk.

Het is natuurlijk wel belangrijk dat het systeem hier de gestelde eisen aan regels ook echt afdwingt en een regel als r_1 weigert. Hoe het systeem de regels test, en hoe het aangeeft waarom een regel fout is, is niet beschreven in het artikel van Bex.

Samenvattend, bij het afvangen van de *Prosecutor's fallacy* is geen kennis van de fout vereist, omdat de fout duidelijk wordt gemaakt door de eisen die het systeem aan regels stelt; het systeem is dus geschikt om de *Prosecutor's fallacy* duidelijk te maken en af te vangen, zolang de verdediging niet blindelings een regel als $\phi \approx f(\phi)$ accepteert.

4.2 *Defense attorney's fallacy*

Bij de *Defense attorney's fallacy* , zoals uitgelegd in sectie 2.3, wordt de verdachte gelijk gesteld aan een aantal andere personen, om het minder waarschijnlijk de dader te maken, terwijl deze niet dezelfde a priori kansen hebben om de dader te zijn. We voeren nu het voorbeeld uit sectie 2.1 in in het systeem van Bex en kijken hoe het systeem hier mee om gaat.

Pro_1: argue (1) P is schuldig

Con_2:	argue (2)	P matcht met gevonden DNA
	(3)	Kans op overeenkomt terwijl de persoon toch niet de dader is: 1 op de 10.000
	(4)	In de stad wonen nog 100.000 andere mannen
	r_1	Een greep uit een populatie van ϕ met een kans van $\psi \Rightarrow \phi \times \psi$ matches
	so	
	(5)	In de stad wonen nog 10 andere mannen die matchen met het DNA
	r_2	Een persoon is 1 van de ϕ verdachten $\Rightarrow$ Kans dat de persoon de juiste verdachte is is $\frac{1}{\phi}$
	so	
	(6)	De kans dat P de juiste verdachte is is $\frac{1}{11} = 9\%$
Pro_3:	explain	
	r_2	Een persoon is 1 van de ϕ verdachten $\Rightarrow$ Kans dat de persoon de juiste verdachte is is $\frac{1}{\phi}$
	=	
	(5)	In de stad wonen nog 10 andere mannen die matchen met het DNA
	(7)	Alle mannen in de stad hebben dezelfde mogelijkheden om de misdaad te zijn begaan
	r_3	(Een persoon is 1 van de ϕ verdachten $\wedge$ alle verdachten hebben dezelfde a priori kans om schuldig te zijn) $\Rightarrow$ Kans dat de persoon de juiste verdachte is is $\frac{1}{\phi}$
	so	
	(6)	De kans dat P de juiste verdachte is is $\frac{1}{11} = 9\%$
Pro_4:	why (7)	Alle mannen in de stad hebben dezelfde mogelijkheden om de misdaad te zijn begaan

Nu moet de verdedigende advocaat inzien dat niet alle door hem gestelde verdachten dezelfde a priori kans op schuldig zijn hebben.

Dit is hoe de zaak ongeveer zou moeten lopen om de *Defense attorney's fallacy* aan het licht te brengen. Net als bij de *Prosecutor's fallacy* is het nadeel hier dat de aanklager zich bewust moet zijn van de fout om deze aan te kaarten. Ook hier kijken we of alle regels aan de eisen van het systeem voldoen:

De regels gebruikt in het *Defense attorney's fallacy* voorbeeld

r_1 Een greep uit een populatie van ϕ met een kans van $\psi \Rightarrow \phi \times \psi$ matches
Deze regel is correct, niets op aan te merken.

r_2 Een persoon is 1 van de ϕ verdachten $\Rightarrow$ Kans dat de persoon de juiste verdachte is is $\frac{1}{\phi}$
Deze regel is incompleet, omdat de assumptie dat de a priori kans voor alle verdachten gelijk is niet in de regel is verwerkt. Maar het incompleet zijn van een regel is geen reden voor het systeem om de regel te weigeren. Bijna alle regels hebben onderliggende assumpties.
Als we de regel vergelijken met de regel (E is een expert $\wedge$ E zegt ϕ) $\Rightarrow \phi$, die geaccepteerd wordt door het systeem, zien wat dat deze net zo een grote assumptie verborgen houdt (namelijk, alle experts zijn betrouwbaar).
Echter, het systeem is geschikt om dit soort verborgen assumpties duidelijk te maken. Men kan de regel uitbreiden door de betrouwbaarheid-eis toe te voegen. Ditzelfde kan gedaan worden met r_2, dit leidt tot r_3.

r_3 (Een persoon is 1 van de ϕ verdachten $\wedge$ alle verdachten hebben dezelfde a priori kans om schuldig te zijn) $\Rightarrow$ Kans dat de persoon de juiste verdachte is is $\frac{1}{\phi}$
Dit is de correcte versie van r_2. Deze regel maakt ook meteen de *Defense attorney's fallacy* duidelijk.

Conclusie:
De *Defense attorney's fallacy* wordt gewoon in het systeem toegelaten. Dat is niet slecht, aangezien in dit geval de oplossing is om de regel te specificeren, en het systeem gemaakt is om verborgen assumpties in regels (in dit geval de 'alle verdachten hebben dezelfde a priori kans om schuldig te zijn') duidelijk te maken. Het systeem is dus uitermate geschikt om de *Defense attorney's fallacy* aan het licht te brengen.

Het nadeel is dat kennis van de fout bij de *Defense attorney's fallacy* wel vereist is, tegenover het redelijk duidelijk maken van de *Prosecutor's fallacy* in de vorige sectie. Als de speler geen kennis heeft van de *Defense attorney's fallacy* is de kans klein dat hij de regel op de juiste manier uit kan bereiden om de fout af te vangen. Daar staat tegenover dat de fout over het algemeen makkelijker te doorzien is dan de *Prosecutor's fallacy* . Het feit dat niet alle verdachten dezelfde mogelijkheden hebben is redelijk intuïtief.

In één zin: Het systeem is gemaakt om een fout als de *Defense attorney's fallacy* af te vangen, maar kennis van de fout is wel vereist.

5 Conclusie

We hebben in dit artikel gekeken naar het dialoogsysteem van Floris Bex[1] in combinatie met een aantal statistische fouten uit het artikel van Huygen[3]. We hebben de fouten besproken aan de hand van een voorbeeld. Ook hebben we het dialoogsysteem besproken, met een paar opmerkingen.

Als laatste hebben we het voorbeeld, dat we bij de uitleg van de *Prosecutor's fallacy* en *Defense attorney's fallacy* gebruikten, omgezet naar de vorm van het dialoogsysteem. Dit hebben we voor zowel de *Prosecutor's fallacy* als de *Defense attorney's fallacy* gedaan. We hebben bij beide fouten gekeken of de gebruikte regels in het systeem toegelaten zouden worden, of het systeem de fout duidelijk zou maken en of de gebruiker kennis van de fouten nodig heeft om deze af te vangen in het systeem.

De volgende dingen kwamen naar voren. Het systeem is geschikt om de *Prosecutor's fallacy* duidelijk te maken en af te vangen, zolang de verdediging niet blindelings een regel als $\phi \approx f(\phi)$ accepteert. De *Defense attorney's fallacy* bleek een vorm van een regel met verborgen assumptie te zijn. Het systeem is gemaakt om een fout als deze af te vangen, maar kennis van de fout is wel vereist.

Al met al presteerde het dialoogsysteem goed. Er zijn geen aanpassingen nodig geweest om de fouten in het systeem te zetten, en beide fouten kunnen worden afgevangen. Dit betekent dat het waarschijnlijk is dat een dialoogsysteem als dat van Bex goed zou kunnen functioneren in rechtszaken waar verschillende statistiek in voor komt.

Referenties

[1] Floris Bex. Explaining anchors. 2005.

[2] P. J. van Koppen en W. A. Wagenaar H. F. M. Crombag. Dubieuze zaken: De psychologie van strafrechtelijk bewijs. 1992 & 1994.

[3] P.E.M. Huygen. "bayesian belief networks" voor redeneren over juridische bewijsvoering. 2004.

[4] Jenneke IJzerman. Bayesiaanse statistiek in de rechtspraak. 2004.

[5] J. Pollock. Cognitive carpentry: A blueprint for how to build a person. 1995.