

EEN ZOGENAAMDE DENKFOUT

Frank Veltman*

1. Het Probleem

Het volgende testmateriaal komt uit Hoofdstuk 6 van *Judgment under uncertainty: Heuristics and biases*, geredigeerd door D. Kahneman, P. Slovic, and A. Tversky, Cambridge University Press, 1982.

Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Please rank the following statements by their probability, using 1 for the *most* probable and 8 for the least probable.

- (a) Bill is a physician who plays poker for a hobby
- (b) Bill is an architect
- (c) Bill is an accountant
- (d) Bill plays jazz for a hobby
- (e) Bill surfs for a hobby
- (f) Bill is a reporter
- (g) Bill is an accountant who plays jazz for a hobby
- (h) Bill climbs mountains for a hobby (op. cit., 92)

Allerlei proefpersonen, met en zonder statistische scholing, zijn aan deze test onderworpen. Meer dan 85%, ook van de statistisch geschoolden, vindt (c) waarschijnlijker dan (d), en plaatst (g) daar tussen in.

Dat is vreemd. Sterker, het is in strijd met een van de basisprincipes van de waarschijnlijkheidsrekening. De conjunctieregel stelt dat de kans dat een bepaalde combinatie van gebeurtenissen zich voordoet, nooit groter kan zijn dan de kans op elk van die gebeurtenissen afzonderlijk. Denk aan roulette: geen van de proefpersonen zou ooit zo dom zijn om de kans op een zwart + even nummer groter in te schatten dan de kans op een zwart nummer of de kans op een even nummer. Wat brengt hen er dan toe om de kans dat Bill een accountant is die in zijn vrije tijd aan jazz doet groter in te schatten dan de kans dat hij een willekeurig iemand met die hobby is?

* ILLC/Faculteit der Wijsbegeerte, Universiteit van Amsterdam, Nieuwe Doelenstraat 15, 1012 CP Amsterdam.

Tverskys en Kahnemans verklaring voor dit verschijnsel is de volgende: De proefpersonen ordenen (a) t/m (h) niet naar waarschijnlijkheid, maar naar de mate waarin ze Bill *representatief* achten voor de genoemde categorieën. Wat de proefpersonen over Bill is verteld, past vermoedelijk aardig bij het beeld dat ze hebben van een accountant, maar niet bij het beeld dat ze hebben van iemand die in zijn vrije tijd jazzmuzikant is. Dat accountants ook die hobby kunnen hebben, is wellicht zelfs nooit bij ze opgekomen, maar door de vraagstelling gedwongen zich een beeld van zo iemand te vormen — een keurige meneer met een bolhoed op in dixielandbandje, zoiets — zullen ze tot een soort compromis komen tussen (c) en (d). Uitkomst: op grond van de informatie die ze over Bill hebben geldt: Bill is representatiever voor (c) dan voor (g), en representatiever voor (g) dan voor (d).

Het begrip *representativiteit* heeft een eigen logica, die systematisch verschilt van de logica van het waarschijnlijkheidsbegrip, zo stellen Tversky en Kahneman. Jammer genoeg doen ze geen poging de logica van dat begrip in kaart te brengen. Integendeel, ze suggereren dat die zich aan elke formalisering onttrekt:

“human reasoning cannot be adequately described in terms of context independent formal rules.” (op.cit., 499)

In het onderstaande hoop ik te laten zien dat het wel meevalt met die ongrijpbaarheid van het ‘menselijk’ redeneren. Contextafhankelijkheid sluit formaliseerbaarheid niet uit. Weliswaar is het zo dat we om een adekwate beschrijving te geven van wat ik nu maar even ‘redeneren in context’ zal noemen enkele karakteristieken van het ‘contextvrije’ redeneren op moeten geven, maar de resulterende theorie is als het om mathematische vormgeving gaat niet minder formeel dan de meer klassieke logische theorieën.¹

2. *Vuistregels*

Accountants zijn gewoonlijk nogal saai, en keurig gekleed, en man. Saaië mensen spelen gewoonlijk geen saxofoon. Mannen met een paardenstaart zijn gewoonlijk geen accountant.

¹ Overigens is het niet het doel van dit stuk die theorie tot in alle formele details te ontwikkelen. Zie daarvoor Veltman, F., ‘Defaults in Update Semantics’, *Journal of Philosophical Logic*, 25 (1996), 221-261. Hier gaat het om de ideeën die in die formalisering gestalte krijgen.

Wie dat allemaal vindt, zal bij eerste kennismaking met de nieuwe accountant een nogal saaie, keurige man verwachten. Zeker niet de sjofel geklede man met een paardestaart die zojuist de vraag “Wat kan ik voor u betekenen?” beantwoord heeft met “Aangenaam, Willemse, ik ben de nieuwe accountant”. Willemse ziet er totaal anders uit dan verwacht. En wat moet hij trouwens met die saxofoonkist? Speelt hij nog saxofoon ook?

Iemand die van mening is dat accountants gewoonlijk keurig gekleed gaan, hoeft nog niet te vinden dat *alle* accountants keurig gekleed gaan. Daarom kan zo iemand na kennismaking met Willemse bij zijn mening blijven, en bij kennismaking met een volgende accountant opnieuw verwachten een keurige geklede mijnheer te zullen zien. Willemse is voor zo iemand de uitzondering die de regel bevestigt.

Met andere woorden: een zin van de vorm ‘A's zijn gewoonlijk B’ betekent niet hetzelfde ‘Alle A's zijn B’. Misschien verrassender: een zin van de vorm ‘A's zijn gewoonlijk B’ is ook niet synoniem met ‘*De meeste* A's zijn B’. De volgende redeneervorm is immers ongeldig:

premissie 1: De meeste A's zijn B
premissie 2: De meeste A's zijn C
conclusie: De meeste A's zijn B en C

(Statistisch gezien kan het best zo zijn dat van de in honderd A's er zestig de eigenschap B hebben, en zestig de eigenschap C, terwijl er maar twintig beide eigenschappen hebben).

Echter, met de analoge redeneervorm is op het eerste gezicht niets mis.

premissie 1: A's zijn gewoonlijk¹ B
premissie 2: A's zijn gewoonlijk C
conclusie: A's zijn gewoonlijk B en C

De theorie die in paragraaf 4 behandeld wordt, ondersteunt deze intuïtie.

Ook de volgende redeneervorm is interessant:

¹ Ik heb lang getwijfeld tussen ‘gewoonlijk’ enerzijds en ‘normaliter’ of ‘normaal gesproken’ anderzijds als vaste vertaling van het Engelse ‘normally’. De laatstgenoemde term is het best, maar bij herhaald gebruik wordt het al vlug teveel van het goede.

premissie 1: De meeste A's zijn B
premissie 2: De meeste B's zijn C
conclusie: De meeste A's zijn C

Het is niet moeilijk tegenvoorbeelden tegen deze redeneervorm te verzinnen. Ook voor zinnen met 'gewoonlijk' in plaats van 'de meeste' gaat hij niet altijd op, getuige het feit dat de volgende verzameling zinnen consistent is.

- Studenten zijn gewoonlijk ouder dan 18 jaar
- Wie ouder is dan 18 jaar, heeft gewoonlijk een rijbewijs
- Studenten hebben gewoonlijk geen rijbewijs

Toch maken we dikwijls gebruik van een redeneervorm die veel op de bovenstaande lijkt. Hierboven bijvoorbeeld staat — impliciet — de volgende gedachtengang:

premissie 1: Accountants zijn gewoonlijk saai
premissie 2: Saaie mensen spelen gewoonlijk geen saxofoon
premissie 2: Willemse is accountant
conclusie: Willemse speelt vermoedelijk geen saxofoon

Waar anders komt de verrassing vandaan, wanneer blijkt dat Willemse dat wel doet?

Zinnen van de vorm 'A's zijn gewoonlijk B' verwoorden een 'default'-regel — zo heet het niet alleen in de Engelstalige maar ook in de Nederlandstalige vakliteratuur — ik prefereer het woord 'vuistregel'. Wie zo'n regel accepteert zal telkens als hij een A tegenkomt aannemen dat die A ook de eigenschap B heeft, en daar eventueel ook naar handelen, totdat blijkt dat die veronderstelling niet gehandhaafd kan worden. Met andere woorden: de volgende gevolgtrekking is vrij natuurlijk — misschien zelfs geldig te noemen — tenminste zolang men over *x* niet veel meer weet dan in de tweede premissie gesteld wordt.

premissie 1: A's zijn gewoonlijk B
premissie 2: x is A
conclusie: x zal wel B zijn

Soms komt men zoveel over x te weten dat men de veronderstelling dat x de eigenschap B heeft in moet trekken. De volgende redeneervorm bijvoorbeeld is zeker niet geldig:

premissie 1: A's zijn gewoonlijk B
premissie 2: x is A
premissie 3: x is niet B
conclusie: x zal wel B zijn

Ook andere informatie dan de directe informatie dat x een uitzondering op de regel is (premissie 3 hierboven), kan ervoor zorgen dat het niet geoorloofd is te concluderen dat x wel de eigenschap B zal hebben. Bekijk bijvoorbeeld:

premissie 1: A's zijn gewoonlijk B
premissie 2: C's zijn gewoonlijk niet B
premissie 3: x is A en C

Concreet: Accountants zijn gewoonlijk saai. Mensen die saxofoon spelen zijn dat gewoonlijk niet. Willemse is accountant en hij speelt saxofoon. Dit betekent dat Willemse een uitzondering moet zijn op een van beide regels, maar op grond van de ter beschikking staande informatie is niet uit te maken op welke.

Bovenstaande voorbeelden verduidelijken misschien wat auteurs als Tversky en Kahneman bedoelen als ze het hebben over de 'contextafhankelijkheid' van het 'menselijk' redeneren. Zo blijkt er bijvoorbeeld uit dat we herhaaldelijk op grond van zeer gebrekkige informatie allerlei allemaal even aannemelijke conclusies trekken, die we vervolgens net zo makkelijk weer intrekken als er meer informatie tot onze beschikking komt. Daaruit blijkt dat het geldigheidsbegrip dat in de bovenstaande voorbeelden figureert, niet het standaard geldigheidsbegrip kan zijn. Blijkbaar is dit nieuwe geldigheidsbegrip *niet monotoon*: door het toevoegen van premissen kan een geldige redenering omslaan in een ongeldige. Het standaardbegrip is wel monotoon. Immers, als een bepaalde conclusie waar is in alle gevallen waarin de premissen waar zijn,

dan is die conclusie ook waar in alle gevallen waarin die premissen plus nog een aantal andere premissen waar zijn.

3. *Dynamische Semantiek*

Vergelijk¹ de volgende twee zinnen:

- Willemse is saai
- Willemse zal wel saai zijn

De tweede zin is zwakker dan de eerste. Om de eerste te kunnen bevestigen zou je Willemse eigenlijk persoonlijk moeten kennen. Voor de tweede heb je wellicht genoeg aan de wetenschap dat Willemse accountant is.

Vergelijk ook de volgende paren van zinnen.

- Ik heb honger/ Je zal wel honger hebben
- Ik zal wel honger hebben/ Je hebt honger

Klinkt het eerste paar niet een stuk natuurlijker dan het tweede? Dat komt omdat je over je eigen honger directe evidentie kan hebben, en over andermans honger alleen maar indirecte evidentie.

Als je het logische gedrag van vuistregels in kaart wilt brengen, dan wordt het hier geconstateerde verschil belangrijk. Vuistregels hebben uitzonderingen; daarom moet je, wanneer je ze toepast, altijd een slag om de arm houden. In een goede theorie voor vuistregels is het dan ook niet toegestaan om uit ‘A's zijn gewoonlijk B’ en ‘ x is A’ zomaar de conclusie ‘ x is B’ te trekken. Verder dan ‘ x zal wel B zijn’ of ‘Vermoedelijk is x B’ kun je niet gaan.

Als je het verschil tussen ‘ x is B’ en ‘ x zal wel B zijn’ wilt verantwoorden in een logische theorie dan kun je niet volstaan met een theorie waarin de logische kracht van een zin geïdentificeerd wordt met zijn waarheidscondities. Het is tamelijk onzinnig om je bij een zin die met ‘Het zal wel zo zijn dat’ begint, af te vragen onder welke omstandigheden hij waar is. Het gaat bij dit soort zinnen niet om ‘waar’ en ‘onwaar’ zonder meer, maar hoogstens om zoiets als ‘waar op grond van de beschikbare informatie’ en ‘onwaar op grond van de beschikbare informatie’. Hetzelfde kan gezegd worden van andere

¹ Een aantal van de voorbeelden die in deze paragraaf aan de orde komen, heb ik al eerder gebruikt. Bijvoorbeeld in ‘Redelijkheid in het redeneren’. In: H. Parret (ed.), *In alle redelijkheid*, Meppel, Boom, 1989, 201-222.

modale kwalificaties als ‘misschien’, en ‘het zou kunnen’. Als Pieterse weet dat Willemse accountant is, dan zegt hij iets dat onwaar is op grond van de hém ter beschikking staande informatie als hij beweert dat Willemse *misschien* geen accountant is. Jansen, die niet zo goed geïnformeerd is, kan met dezelfde zin best iets zeggen dat waar is: waar op grond van de informatie waarover hij beschikt. Het is onzinnig om je in dit geval af te vragen wie van de twee hier nu dé waarheid spreekt. Anders dan in het geval van de louter descriptieve zin ‘Willemse is accountant’ is waarheid hier geen zaak tussen de volzin aan de ene kant en de feiten aan de andere. Niet dat de feiten er bij ‘Misschien is Willemse accountant’ helemaal niet toe doen. Niet alle feiten doen er toe, alleen de feiten waarvan degene die een dergelijke uitspraak doet, op de hoogte is.

Voor onze doeleinden biedt de dynamische semantiek een geschikter theoretisch raamwerk. De relatie tussen taal, denken en werkelijkheid wordt daarin zo uitgelegd dat de betekenis van modale kwalificaties als ‘vermoedelijk’, ‘het zal wel’, ‘misschien’, ‘het zou kunnen’, ‘het had gekund’ optimaal belicht kan worden.

In een dynamische theorie worden geen *waarheidscondities* uitgespeld, maar *verwerkingsregels*. Bij elke volzin wordt precies aangegeven wat het effect ervan is op de cognitieve toestand van iemand die de boodschap die met die volzin gebracht wordt, verwerkt. ‘Betekenis’ wordt zo een operationeel begrip: de betekenis van een zin wordt opgevat als een functie die aan elke cognitieve toestand een cognitieve toestand toewijst.

In het algemeen zal een cognitieve toestand bij het verwerken van een volzin overgaan in een andere: kennis neemt toe, verwachtingen worden bijgesteld, enzovoorts. Maar het kan ook voorkomen dat een volzin nauwelijks effect heeft. Soms is de houding van het subject ten opzichte van de wereld na verwerking precies hetzelfde als ervoor — kennelijk brengt de volzin oud nieuws¹. In dit geval zullen we zeggen dat de volzin in de cognitieve toestand in kwestie *geaccepteerd* wordt. Diametraal daartegenover staat het geval dat het nieuws *niet acceptabel* is. Dat geval treedt op wanneer de informatie die de volzin bevat, zich niet laat rijmen met de reeds geaccepteerde informatie.

Als s een cognitieve toestand is en ϕ een volzin, dan schrijven we $s + \phi$ voor de toestand die uit s ontstaat bij verwerking van ϕ . Het is misschien misleidend

¹ Ik druk me hier zo omslachtig uit omdat er wel veranderingen kunnen optreden in de ‘discourse’ informatie. In het artikel van Groenendijk en Stokhof wordt dit onderscheid verder uitgewerkt.

om hier het teken voor optelling te gebruiken, want de ‘+’ die hier gebruikt wordt mist een hele hoop eigenschappen de ‘+’ uit de rekenkunde. Zo is het niet altijd zo dat $s + \phi + \psi$ dezelfde toestand is als $s + \psi + \phi$. De volgorde waarin een en ander verwerkt wordt doet er toe.

Een voorbeeld : als het goed is, kost het u geen moeite om de informatie die in de eerste van de volgende twee reeksen zinnen wordt aangeboden, te verwerken, maar stuit u bij de tweede reeks op problemen.

- Er wordt aan de deur geklopt ... Het zal Sinterklaas wel zijn ... Het is Zwarte Piet.
- Er wordt aan de deur geklopt ... Het is zwarte Piet ... Het zal Sinterklaas wel zijn.

Beide reeksen zijn opgebouwd uit dezelfde drie zinnen; alleen de volgorde is anders. Toch maakt dat een enorm verschil. De eerste reeks is consistent, de tweede niet. Het is niets bijzonders wanneer, zoals in de eerste reeks van zinnen, een aanvankelijke verwachting later door de feiten weersproken wordt. Maar het is raar om een eenmaal geconstateerd feit in twijfel te trekken. En dat is wat er in de tweede reeks gebeurt. De zin ‘Het zal Sinterklaas wel zijn’ is niet acceptabel in een toestand waarin het een geaccepteerd gegeven is dat het Zwarte Piet is.

Zinnen die met ‘vermoedelijk’ of ‘het zal wel’ of ‘misschien’ of ‘het zou kunnen’ beginnen, verschillen van puur descriptieve zinnen hierin dat ze niet *persistent* zijn onder kennistoename. ‘Misschien schijnt de zon’ denkt u op grond van de beperkte informatie die u, zojuist wakker geworden, tot uw beschikking heeft, ook al zult u die bewering niet meer willen herhalen als u de gordijnen eenmaal geopend heeft. In de tussentijd is uw feitenkennis toegenomen. Voor puur descriptieve zinnen geldt echter dat ze als ze eenmaal geaccepteerd zijn, ook geaccepteerd blijven zolang men méér leert over de buitentalige werkelijkheid.

We zijn bijna zover dat we een definitie van het dynamische begrip van logische geldigheid kunnen geven. Nog één opmerking over de +-operatie. Het is niet zo dat voor elke willekeurige zin ϕ en elke willekeurige cognitieve toestand s geldt dat $s + \phi$ bestaat. Soms is een volzin niet onbegrijpelijk voor iemand in toestand s . Dit is met name zo wanneer in ϕ een persoonlijk voornaamwoord optreedt, waarvoor in de toestand s geen geschikte referent voorhanden is. In het artikel van Groenendijk en Stokhof (deze aflevering)

wordt dit verschijnsel uitgebreid besproken, maar wellicht kan het geen kwaad hier een voorbeeld te geven waaruit blijkt dat de aanwezigheid van modale operatoren de zaken nog eens compliceert¹.

Vergelijk:

- Willemse heeft een auto. Het is een BMW.
- Willemse heeft een auto. Het zal wel een BMW zijn.
- Willemse zal wel een auto hebben. Het is een BMW.

Bij de eerste twee teksten komt u na het verwerken van de eerste volzin in een toestand waarin de tweede volzin begrijpelijk is. De derde tekst is bijna net zo raar als

- Willemse heeft geen auto. Het is een BMW.

Dat een op zich syntactisch welgevormde volzin soms niet interpreteerbaar is, is niet altijd te wijten aan de aanwezigheid van anaforen. Vergelijk:

- Ik heb geen pudding toe genomen. Ik zou ziek geworden zijn.
- Ik heb pudding toe genomen. Ik zou niet ziek geworden zijn.

De eerste van deze teksten zult u — min of meer automatisch — gelezen hebben als:

- Ik heb geen pudding toe genomen. Als ik dat wel gedaan had, zou ik ziek geworden zijn.

De tweede was, neem ik aan, onbegrijpelijk — in ieder geval niet zonder meer herkenbaar als een ingekorte versie van:

- Ik heb pudding toe genomen. Als ik dat niet gedaan had, zou ik niet ziek geworden zijn.

Kennelijk levert de cognitieve toestand waarin men na verwerking van de *negatieve* eerste zin uit de eerste reeks belandt, wel een goed aangrijpingspunt

¹ Zie verder: J. Groenendijk J., M. Stokhof and F. Veltman, 'Coreference and Modality'. In: S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*, Blackwell, Oxford, 1996, 179-213.

voor de irrealis uit de erop volgende zin, maar de toestand waarin men terecht komt na verwerking van de *positieve* eerste zin uit de tweede teks niet. Logici hebben veel over de irrealis nagedacht, maar alleen voorzover die in conditionele zinnen gebruikt wordt. De dynamische semantiek levert het instrumentarium om naar zinnen in tekstverband te kijken. Dat maakt het mogelijk een wat bredere kijk op de functie van de irrealis te ontwikkelen.¹

Dan nu het geldigheidsbegrip. De meest algemene definitie is deze.

Laat $\phi_1, \dots, \phi_n / \psi$ een redenering zijn met premissen ϕ_1, \dots, ϕ_n en conclusie ψ . De redenering $\phi_1, \dots, \phi_n / \psi$ is geldig dan en slechts dan als voor elke cognitieve toestand s waarvoor $s + \phi_1 + \dots + \phi_n + \psi$ gedefinieerd is, geldt dat

$$s + \phi_1 + \dots + \phi_n + \psi = s + \phi_1 + \dots + \phi_n.$$

Anders gezegd: een redenering is geldig als in alle gevallen waarin de premissen en conclusie interpreteerbaar zijn, geldt dat men na verwerking van de premissen in een cognitieve toestand komt waarin men de conclusie accepteert — liever gezegd: móet accepteren, gegeven de betekenis van de premissen en de conclusie.

In de bovenstaande definitie wordt gekwantificeerd over alle mogelijke informatietoestanden. Daarom is hij voor onze doeleinden te algemeen. De conclusies die men uit vuistregels trekt zijn meestal niet persistent. Dikwijls moeten ze in het licht van nieuwe informatie worden ingetrokken. Met andere woorden: bij de evaluatie van een redenering waarin vuistregels als premissen optreden is het van het grootste belang precies te weten wat de beschikbare informatie is. Zo contextafhankelijk zijn ze.

De definitie die in het volgende een rol speelt is een speciaal geval van de bovenstaande. Het teken ‘**0**’ staat voor de minimale informatietoestand, de toestand waarin men nog volkomen blanco tegenover de wereld staat.

¹ Zie voor een eerste aanzet het proefschrift van R. Kibble, *Anaphora Resolution in Dynamic Semantics: Modality, Subordination and Complementation*, PhD Thesis, Centre for Cognitive Science, University of Edinburgh, 1997.

Laat $\phi_1, \dots, \phi_n / \psi$ een redenering zijn met premissen ϕ_1, \dots, ϕ_n en con-clusie ψ . De redenering $\phi_1, \dots, \phi_n / \psi$ is geldig dan en slechts dan als $\mathbf{0} + \phi_1 + \dots + \phi_n + \psi$ bestaat en identiek is aan $\mathbf{0} + \phi_1 + \dots + \phi_n$.

Overigens: voor louter descriptief taalgebruik leiden beide definities tot dezelfde uitkomsten.

4. *Verwachtingspatronen*

Stelt u zich een taalgemeenschap voor waarin wat het weer betreft maar een beperkt aantal onderscheidingen gemaakt wordt: Het regent (R) of het is droog (D), en het is koud (K) of het is warm (W). Al met al geeft dit vier mogelijke weertypes: RK, RW, DK, en DW. De gebruikte afkortingen spreken voor zich. De leden van deze nogal primitieve taalgemeenschap kunnen niet alleen spreken over wat voor weer het *in feite* is, maar ook over wat voor weer het *gewoonlijk* is. Vuistregels stellen hen in staat hun soms slechts partiële informatie over het feitelijke weer aan te vullen met redelijke hypothesen over wat voor weer het *vermoedelijk* is.

De mechanismen die hierbij een rol spelen kunnen in een eenvoudig model worden beschreven. In het artikel van Van Benthem (deze aflevering) is al uitgelegd hoe feitelijke informatie gerepresenteerd kan worden: identificeer iemands informatie over de feitelijke stand van zaken met de verzameling van alle mogelijke situaties die hij of zij niet van de werkelijke zou kunnen onderscheiden: voorzover de informatie strekt, zou elk van die situaties de werkelijke kunnen zijn. Informatiegroei kan dan worden opgevat als een proces van eliminatie: telkens als de persoon in kwestie iets meer te weten komt, vallen er mogelijkheden af. Iemand die helemaal niets weet moet met alle denkbare situaties rekening houden en iemand die alles denkt te weten hoeft dat nog maar met één situatie.

In ons voorbeeld ziet dat er zo uit: Als iemand helemaal geen idee heeft wat voor weer het is, dan representeren we zijn of haar kennis met de verzameling {RK, RW, DK, DW}. Weet iemand enkel dat het regent dan wordt dat gerepresenteerd met de verzameling {RK, RW}.

Deze eenvoudige representatie van iemands informatie over de feitelijke stand van zaken suggereert een al even eenvoudige representatie van de meer algemene kennis zoals die zijn neerslag vindt in vuistregels. Vuistregels wekken verwachtingen, ze induceren een bepaalde ordening tussen de verschillende logisch mogelijke situaties. Het basisidee is dit: van de verschillende mogelijke

situaties gelden sommige als *normaler* dan andere. Als iemand geen idee heeft wat voor weer het gewoonlijk is, dan zullen alle vier de weertypes als even normaal worden aangemerkt. Wie leert dat het gewoonlijk regent, zal zijn of haar verwachtingen daarop instellen, en de situaties RK en RW als normaler beschouwen dan DK en DW.

In een plaatje:

RK	DK
RW	DW

Uitleg: als twee weertypes als ‘even normaal’ worden aangemerkt dan zijn ze binnen dezelfde cirkel gezet. De situaties in de linkercirkel zijn normaler dan die in de rechtercirkel.

Met het leren van een nieuwe regel wordt het verwachtingspatroon verfijnd. Komt iemand erachter dat het niet alleen normaal is dat het regent, maar dat het bovendien normaliter koud is, dan heeft dat het volgende effect op hierboven weergegeven ordening: Van de vier weertypes is RK het meest normaal en DW het minst normaal. RW en DK liggen daar tussenin, geen van beide is normaler dan de ander. (Toch zijn ze niet *even* normaal; ze zijn elk in een ander opzicht uitzonderlijk).

	DK	
RK		DW
	RW	

Het volgende plaatje geeft de cognitieve toestand weer van iemand die geleerd heeft dat het gewoonlijk regent en gewoonlijk koud is, hoewel het nu toevallig warm is. De verzameling situaties die de feitenkennis weergeven staan binnen de gestippelde lijn.

	DK	
RK		DW
	RW	

Merk op dat de situatie RW een speciale positie inneemt. Het is in de gegeven uitzonderlijke omstandigheden — warm weer — de situatie die het best aan de verwachting beantwoordt. In die situatie regent het en daarom is ‘Het zal wel

regenen' een voor de hand liggende aanname, ook al is de mogelijkheid niet uitgesloten dat het in feite droog is. Situaties als deze — *optimale* situaties genoemd — spelen een cruciale rol in het handelen. Wanneer je een beslissing moet nemen en niet precies weet wat er aan de hand is, dan stem je beslissing af op wat in de gegeven omstandigheden het meest in de lijn der verwachtingen ligt.

Veel meer dan het bovenstaande is er eigenlijk niet nodig om in te zien dat de redenering:

— Gewoonlijk regent het. Gewoonlijk is het koud. Het is niet koud /
Het zal wel regenen.

geldig is (in de dynamische zin van het woord), terwijl de redenering:

— Gewoonlijk regent het. Gewoonlijk is het koud. Het is niet koud /
Het zal wel koud zijn.

dat niet is. Merk nog op dat in het geval blijkt dat het in feite droog is, de hierboven in kaart gebrachte cognitieve toestand overgaat in deze:

	DK	
RK		DW
	RW	

Uit de premissen

— Gewoonlijk regent het. Gewoonlijk is het koud. Het is niet koud. Het regent niet.

volgt dus niet langer dat het wel zal regenen. De logica is niet monotoon.

De zaak wordt een stuk ingewikkelder als we ons realiseren dat vuistregels niet alleen toevallige maar ook systematische uitzonderingen kunnen hebben:

— Gewoonlijk regent het. Maar bij warm weer is het gewoonlijk droog.

En uitzonderingsclausules kunnen zelf ook weer uitzonderingsclausules hebben:

— Bij warm weer en zuidwestenwind regent het gewoonlijk.

Niet alles kan op dit gebied.

— Gewoonlijk regent het. Gewoonlijk is het koud. En als het koud is, is het gewoonlijk droog.

Dit is onacceptabel. En ook in het volgende voorbeeld worden iets teveel uitzonderingen gemaakt:

— Gewoonlijk regent het. Bij warm weer is het gewoonlijk droog. En ook als het niet warm is, is het gewoonlijk droog.

Hier geldt de volgende consistentie-eis: Uitzonderingsclausules zijn toe-gestaan, maar het moet altijd logisch mogelijk blijven dat er zich een volkomen normale situatie voordoet. Sterker: onder alle omstandigheden, hoe uitzonderlijk ook, moet het logisch mogelijk zijn dat er zich een voor die omstandigheden normale situatie voordoet.

Voorbeeld: Bedenk dat iemand die vindt dat het gewoonlijk regent, maar daarbij een uitzondering maakt voor dagen dat het warm is, een warme dag waarop het toch regent moeilijk nog als een normale dag kan zien. Het is eerder een dubbel uitzonderlijke dag: een uitzondering op een uitzonderingsclausule. Als zo iemand ook nog een uitzondering wil maken voor dagen dat het niet warm is, dan tellen ook regendagen waarop het niet warm is, niet meer als normale dagen. Dan zijn er geen dagen meer denkbaar die voor het predikaat ‘normaal’ in aanmerking komen.

Ik zal geen poging doen om de coherentie-eis preciezer te formuleren dan ik hierboven gedaan heb. Ook zal ik niet proberen uit te leggen hoe men, als men naast algemene regels ook uitzonderingsclausules en uitzonderingsclausules op uitzonderingsclausules toelaat, kan bepalen wat de optimale, meest in de lijn der verwachtingen liggende situatie is¹. Voor ons voorbeeld werkt een en ander zo uit dat de volgende redeneringen geldig blijken.

¹ Zie hiervoor het eerder genoemde artikel Veltman(1996).

premissie 1: Gewoonlijk regent het
premissie 2: Bij warm weer is het gewoonlijk droog
conclusie: Het zal wel regenen en koud zijn.

premissie 1: Gewoonlijk regent het
premissie 2: Bij warm weer is het gewoonlijk droog
premissie 3: Het is warm
conclusie: Het zal wel droog zijn

Voor het onderwerp van dit artikel is het verder alleen nog van belang om op te merken dat bij de overstap van dit klimatologische voorbeeld naar het domein waar Willemse de accountant zich bevindt de zaken niet wezenlijk moeilijker worden. In dat domein bevinden zich geen vier weertypes maar een enorm aantal mensentypes. Bij elk subdomein van dat domein — bijvoorbeeld het subdomein der accountants — hoort een verwachtingspatroon, geïnduceerd door de voor dat subdomein geldende vuistregels. Uitzonderingsclausules — bijvoorbeeld voor accountants die in hun vrije tijd saxofoon spelen — zijn toegestaan: bij een subdomein van een subdomein kan een ander verwachtingspatroon horen, zij het dat bovengenoemde consistentie-eis van kracht blijft.

5. Terug naar Bill

Hoe komt het dat 90% van de ondervraagden Bill eerder inschatten als een accountant die in zijn vrije tijd jazzmuziek speelt, dan als zomaar iemand met die hobby. Wie met de volgende gedachtengang in kan stemmen heeft zo'n verklaring niet nodig.

Dit is de informatie over Bill:

Bill is 34 years old. He is intelligent, but unimaginative, compulsive and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Hier zijn zeker eigenschappen bij die normaal zijn voor een accountant en uitzonderlijk voor iemand die in zijn vrije tijd jazzmuziek speelt. Er zijn geen

eigenschappen bij die uitzonderlijk zijn voor een accountant, maar normaal voor iemand die in zijn vrije tijd jazzmuziek spelen.

Accountants die in hun vrije tijd jazzmuziek spelen, zijn geen gewone accountants. Toch zullen sommige van de eigenschappen die normaal zijn voor accountants, ook normaal zijn binnen deze speciale groep. Van de eigenschappen die Bill heeft, is bijvoorbeeld 'goed in rekenen' er een waar ook een accountant die in zijn vrije tijd aan jazz doet, moeilijk buiten kan, ook al is die eigenschap nog zo uitzonderlijk voor jazzhobbyisten in het algemeen.

Samenvattend: Bill heeft enkele eigenschappen die normaal zijn voor accountants met als hobby het spelen van jazzmuziek, maar uitzonderlijk voor jazzhobbyisten in het algemeen. Omgekeerd is geen van de genoemde eigenschappen normaal voor jazzhobbyisten in het algemeen en uitzonderlijk voor accountants met die hobby.

Daarom geldt: voorzover de informatie strekt past Bill beter in het beeld van een accountant die wat aan jazz doet in zijn vrije tijd, dan in het beeld van een jazzhobbyist pur sang.

De theorie voor vuistregels die ik in de voorgaande paragrafen heb geschetst, biedt voldoende aanknopingspunten om de noties die hierboven gebruikt worden, precies te definiëren.

Noem een eigenschap *P* *normaal* binnen de categorie objecten met de eigenschap *A* dan en slechts dan als. 'A's zijn gewoonlijk *P*' een geaccepteerde vuistregel is. Noem een eigenschap *P* *uitzonderlijk* binnen die categorie dan en slechts dan als. 'A's zijn gewoonlijk niet *P*' een geaccepteerde vuistregel is. Merk op: een eigenschap *P* die niet normaal is binnen een bepaalde categorie, is daarmee nog niet uitzonderlijk binnen die categorie. Wie de vuistregel 'Accountants dragen normaliter een bril' niet wil accepteren, zit daarmee nog niet vast aan de regel 'Accountants dragen normaliter geen bril'¹.

Laat *X* een verzameling eigenschappen zijn. De verzameling *X* is *representatiever* voor de categorie objecten met de eigenschap *A* dan voor de categorie objecten met de eigenschap *B* dan en slechts dan als. aan de volgende voorwaarden is voldaan:

¹Merk op dat wie 'De meeste accountants dragen een bril' niet accepteert, vastzit aan 'De meeste accountants dragen geen bril'. Opnieuw een voorbeeld waaruit blijkt dat de logica van vuistregels niet met die van statistische 'meerderheids'-uitspraken samenvalt.

- (i) voor sommige eigenschappen P in X geldt dat P normaal is binnen de categorie A en uitzonderlijk binnen de categorie B.
- (ii) voor geen van de eigenschappen P in X geldt dat P normaal is binnen de categorie B, en uitzonderlijk binnen de categorie A.

Gegeven deze definities levert de bovenstaande gedachtengang als uitkomst: Bills eigenschappen zijn representatiever voor de categorie accountants die in hun vrije jazzmuziek spelen, dan voor de ruimere categorie van jazzhobbyisten in het algemeen.

6. Drogreden of niet?

Tversky en Kahneman vinden dat al degenen die de kans dat Bill een accountant is met als hobby het spelen van jazzmuziek groter inschatten dan de kans dat hij zomaar iemand met die hobby is, een fout maken.

“The reliance on representativeness leads to systematic errors.”
(op.cit., 89)

schrijven ze bijvoorbeeld, en geredeneerd vanuit de waarschijnlijkheidsleer is dat ook zo. In dit verband is het significant om te vermelden dat van de statistisch geschoolde proefpersonen verreweg de meesten (83%) na afloop van de test bereid waren te erkennen dat ze een enorm geblunderd hadden. Daar staat tegenover dat van de statistisch ongeschoolden lang niet iedereen zich op andere gedachten liet brengen: 53% bleef vasthouden aan hun oorspronkelijke antwoord; dit ondanks het feit dat ze best bereid waren om de conjunctieregel *in abstracto* te onderschrijven.

Nu we een idee hebben wat er in de hoofden van de proefpersonen die zondigen tegen de conjunctieregel kan zijn omgegaan, en nu we gezien hebben dat daar heel wat meer systeem in zit dan Tversky en Kahneman bevroedden, is de neiging groot om het bovenstaande als een relevant gegeven in de strijd te werpen.

“Which arguments *should* a logical theory admit as valid/invalid? In the case of nonmonotonic logic the criteria are not the same as in the case of classical logic.[...] In the case of default reasoning the whole notion of what is correct (the notion we are trying to model) is defined in terms of what sort of reasoning people *actually* engage in

and what sorts of reasoning an intelligent agent will have to do to get along in a common sense way. So “mistakes” by common sense reasoners are only possible in the sense that they are different from the majority.[...] Psychologism with respect to default reasoning is correct!”¹

In het verlengde hiervan zou je kunnen stellen dat de logica voor vuistregels die in dit artikel geschetst wordt en de notie van representativiteit die deze met zich meebrengt, te verkiezen is boven de waarschijnlijkheidsleer en wel omdat deze theorie beter aansluit bij de praktijk van alle dag: ze beschrijft hoe mensen werkelijk denken.

Dat lijkt me niet zo'n sterk argument. Om een mogelijk misverstand uit de weg ruimen: zoiets als ‘beschrijven hoe de mensen werkelijk denken’ wil de theorie voor vuistregels die ik in het voorgaande heb gepresenteerd helemaal niet; ze is in dat opzicht niet minder normatief dan de waarschijnlijkheidsleer of de klassieke eerste orde predikatenlogica. De verwerkingsregels die aan de basis ervan liggen, zijn niet descriptiever dan Tarski's waarheidscondities of dan Kolmogorovs axioma's voor de waarschijnlijkheidsleer. Het zijn *regels*; ze vertellen de leden van een taalgemeenschap hoe ze de verschillende uitdrukkingen uit de taal *moeten* interpreteren als ze dezelfde taal willen spreken als de andere leden. Gegeven deze verwerkingsregels zijn bepaalde redeneerwijzen geldig en andere ongeldig. Als een lid van zo'n taalgemeenschap die geldigheid of ongeldigheid om de een of andere reden niet onderkent, dan maakt hij of zij een fout. Kortom, de theorie beschrijft op zijn hoogst hoe mensen *moeten* denken. Het predikaat 'psychologistisch' is daarom niet op zijn plaats.

Aan de andere kant: zowel de waarschijnlijkheidsleer als de default-logica bieden onze taalgemeenschap een instrument dat ons helpt beslissingen te nemen in omstandigheden waarin we eigenlijk te weinig informatie hebben om dat op een absoluut veilige wijze te doen. Het zou best eens kunnen dat een default-logica als hier gepresenteerd het instrument dat we in de praktijk van alle dag gebruiken — de ene keer zonder fouten, de andere keer met — beter benadert dan de waarschijnlijkheidsleer. Anders gezegd: misschien geeft een theorie als de hier gepresenteerde een juistere beschrijving dan de waar-

¹ Dit citaat komt uit het overigens zeer lezenswaardige verslag van een empirisch onderzoek naar het gebruik van vuistregels: Pelletier, F & R. Elio, *Some truths about Non Monotonic Reasoning*. Technical Report TR 93-12, Department of Computing Science, University of Alberta.

schijnlijkheidsleer van de regels waaraan de sprekers van de natuurlijk taal zich in feite gebonden achten. Maar wat dan nog? Dat lijkt me op zichzelf geen reden om de ene theorie *beter* te vinden dan de andere. Misschien zouden we er wel verstandig aan doen ons *en masse* tot de waarschijnlijkheidsleer te bekeren. Immers, de vraag waar het uiteindelijk om gaat is niet welke van de twee theorieën de praktijk van alle dag het best beschrijft, maar welke van de twee theorieën het *meest bruikbare* instrument biedt. Dat is een vraag die in dit relativistische tijdperk wellicht wat ouderwets aandoet. Maar ze lijkt me een ouderwetse filosofische discussie meer dan waard^{1,2}.

Abstract

Psychological experiments have repeatedly shown that in judging the likelihood of uncertain events people do not follow the principles of probability theory. A notorious example is given by the so called conjunction fallacy.

In this paper I argue that this fallacy is not really a fallacy when it is analysed in the light of a dynamic theory of default reasoning. The question that immediately rises is whether the fact that this theory conforms better to the way people actually think provides a reason to prefer it to probability theory.

¹ Een oproep tot zo'n discussie is al eerder in het ANTW gedaan door Yao-Hua Tan. Zie zijn 'Een vergelijking van inductief-statistisch redeneren en default logica; een nieuwe toekomst voor het logisch positivisme?', *ANTW*, 82 (1990), 117-141.

² Ik dank Edit Doron, die het 'conjunctie effect' onder mijn aandacht bracht en suggereerde dat daar vanuit de dynamische semantiek wel iets zinnigs over te zeggen zou zijn.