

## Hoe in wetenschap de waarheid wordt gekneed

NRC, oktober 2012

Miquel Ekkelenkamp Bulnes

Enkele jaren geleden liep ik om onduidelijke redenen rond op een beurs met buitenlands onroerend goed, waar ik in gesprek raakte met een makelaar die me een driekamerappartement in de buurt van de Turkse badplaats Alanya wilde verkopen. Het appartement zou tevens voor me worden verhuurd en deze huur zou me een gegarandeerd jaarlijks rendement van tien procent opleveren. Ik vroeg hem hoe hij die garantie kon geven? Wat als er helemaal geen toeristen kwamen? De makelaar glimlachte hartelijk, schudde zijn hoofd alsof een van ons tweeën debiel was en pakte er een klapper bij: „Zou ik zeggen dat we een rendement garanderen van tien procent als dat niet helemaal geregeld was?” Het klonk me als een strikvraag. Hij sloeg met de vlakke hand tegen de klapper: „Het is allemaal netjes vastgelegd bij een notaris.”

Ander voorbeeld: Dion Graus, Kamerlid voor de PVV, zat een tijdje in de parlementaire enquêtecommissie voor de bankensector. Toen hij uiteindelijk voormalig topman van ABN Amro Jan Peter Schmittmann mocht verhoren, bleek hij niet de ondervragingsterriër die hij meende te zijn (bij mijn weten wist hij Schmittmann tot niet één zinvol antwoord te bewegen). De dag na dit verhoor stapte Graus uit de commissie met in essentie de volgende verklaring voor zijn afhaken: „Geloof me nou maar dat ik een hele goede reden heb gehad.”

Ik denk dat u in allebei de voornoemde gevallen pas bereid bent de conclusie te geloven, nadat u de volledige onderbouwing ervan hebt gehoord. Waarschijnlijk zou u zelfs extra, externe bronnen willen raadplegen.

Nog een voorbeeld, een bericht dat meerdere kranten onlangs op hun website publiceerden: Amerikaanse onderzoekers claimen dat „de breinen van kinderen zijn ingeprent met voedingslogo's, waardoor ze bijzonder vatbaar zijn om slechte beslissingen te nemen over wat ze eten”. Wilt u bij deze conclusie ook precies weten hoe hij is onderbouwd? En zo

niet: waarom? Vanwege uitgesproken vertrouwen in de wetenschapper in kwestie? Omdat het bericht iets bevestigt dat u geloofwaardig acht?

De laatste tijd is er weer eens ophef over de betrouwbaarheid van de wetenschap. Er zou veel worden gefraudeerd: in de Verenigde Staten schat men dat bij 1,5 procent van de studies de data worden gemanipuleerd. Dat deze 1,5 procent als teleurstellend wordt ervaren komt door misverstanden over de precieze betekenis van 'wetenschap'. Sommigen menen dat het een kwaliteitskeurmerk is, anderen zien er een machine in die feiten produceert. Beide zijn onjuist: wetenschap is een proces, een systematische zoektocht naar de waarheid, in het besef dat deze slechts kan worden benaderd. Wetenschap kent geen feiten, maar metingen en observaties (feiten zijn eigenlijk slechts afspraken die we met elkaar maken over wat we 'waar' vinden). Het kent geen meningen, slechts hypotheses en interpretaties.

Aan dit proces, de wetenschappelijke methode, is ruim duizend jaar gesleuteld. In de elfde eeuw bedachten we kennis te vergaren uit experimenten en niet door te redeneren vanuit 'universele waarheden'. Sinds de dertiende eeuw eisen we reproduceerbaarheid van onze experimenten. Sinds de achttiende zijn we overtuigd van de noodzaak van controles, zo nodig gerandomiseerd (1926), placebo-gecontroleerd (1937) en dubbelblind (1950). En al in 1934 bedacht Karl Popper dat wetenschappelijke stellingen falsifieerbaar moeten zijn: uitspraken waarvan de onjuistheid niet kan worden aangetoond zijn per definitie betekenisloos. De ontwikkeling van de wetenschappelijke methode staat nog niet stil: een recente vereiste is bijvoorbeeld het aanmelden van bepaalde studies voordat ze worden verricht, om te voorkomen dat ze na afloop in een diepe lade verdwijnen omdat de uitkomst niet bevalt.

Wetenschap houdt zich aan de volgorde: hypothese → proef → resultaten → interpretatie/conclusie. De proef moet volgen uit de hypothese, de resultaten moeten voortkomen uit de proef en de conclusie moet worden ondersteund door de resultaten. Volg dit schema, met de eerdergenoemde randvoorwaarden (reproduceerbaarheid enzovoorts) en er kan bijna niets fout gaan. Men (u dus ook) zou hier vaker op moeten letten, want een veel grotere plaag dan 1,5 procent onderzoeken met datavervalsing zijn de onderzoeken waarin de volgorde is gemanipuleerd.

Veelvoorkomend is bijvoorbeeld het nalaten een hypothese te formuleren. Meestal betreft dit het zogenaamde datamining: in een grote database op zoek gaan naar willekeurige relaties tussen gegevens. Dit leidt tot bizarre onzinconclusies als: duikers krijgen meer dochters, visolie beschermt tegen Alzheimer, verveling leidt tot een hogere sterftkans. Voor de resultaten is geen aparte proef verricht, en een hypothese is er al helemaal niet. We hebben dus te maken met: resultaten → interpretatie.

Een andere frequente deviatie van de wetenschappelijke methode is het trekken van conclusies uit onbewezen modellen: reageerbuismodellen, diermodellen, epidemiologische modellen, financiële modellen. Tijdens de Mexicaanse griepepidemie ontstond paniek bij een groep wetenschappers omdat hun fretten doodgingen aan het virus. ‘Mexicaanse griep is echt gevaarlijk’, meldden de kranten. Intussen, in de echte wereld, was al uit zo’n honderdduizend humane gevallen bekend dat er bijna geen mens aan overleed. Het model is altijd ondergeschikt aan de werkelijkheid en moet hier voortdurend aan worden getoetst, dus de juiste interpretatie was: ‘Frettenmodel waardeloze voorspeller voor griep’. Net zulke overschatte modellen zijn verkiezingspeilingen. Peilers toetsen immers (bijna) nooit de conclusies uit hun steekproeven aan een echte uitkomst. De uitspraak ‘als er vandaag verkiezingen zijn, haalt de SGP dertig zetels’ is juist noch onjuist. Hij is niet falsifieerbaar – er zijn ‘vandaag’ immers helemaal geen verkiezingen – dus eenvoudigweg betekenisloos. Conclusies op basis van onbewezen modellen zijn als volgt weer te geven: hypothese → proef → resultaten → extra ongetoetste hypothese → interpretatie.

Hierna wordt het pas echt schimmig. Niet al te lang geleden verrichte een Nederlandse universiteit een onderzoek naar yoghurt met probiotica. Vrijwilligers kregen probiotica of een placebo en daarna werden stukjes dunne darm (biopten) bij hen afgenomen om de mate van ontsteking te bepalen. Men verwachtte dat de ontsteking minder zou worden door de probiotica en speculeerde dat zo’n ontstekingsremmend effect gunstig zou zijn voor bepaalde darmziekten. Echter: de ontsteking werd juist méér. Dat was een probleem, want de conclusie (‘probiotica zijn goed voor u’) stond al vast. Daarom werd de hypothese veranderd in ‘probiotica zijn goed als ze de ontsteking doen toenemen’, ze zouden dan beschermen tegen infectie of zo... U kunt het inmiddels zelf wel invullen, we hadden hier te maken met: ((proef → resultaten) + conclusie) → hypothese.

Wil je met een enquête bevestigen wat jij vindt – als overheidsinstelling (‘wij doen belangrijk werk’), als bedrijf (‘mensen houden van ons product’) of als politieke partij (‘mensen zijn het met ons eens’) – dan bestaat daarvoor een methode die bekend staat als framing. Je stelt vragen die geen vragen zijn, om uiteindelijk een vastgesteld antwoord te forceren. ‘Wist u dat Barack Obama homo-erotische kunst wil subsidiëren?’ Gevolgd door de vraag op wie de geënquêteerde wil gaan stemmen. ‘Vindt u onderzoek naar meegroeïende kunsthartkleppen voor zieke kinderen belangrijk?’ Gevolgd door de vraag of er meer geld naar wetenschap moet. De proef wordt bepaald door de conclusie en een hypothese bestaat niet, dus wat je overhoudt is: Conclusie → proef → resultaten

Sommige resultaten geven in eerste instantie de verkeerde conclusie en moeten daarom worden opgepoetst. Hiertoe bestaan diverse methodes, maar bijzonder populair in medisch onderzoek was het veranderen van de uitkomstparameters (de ‘eindpunten’). Nu het steeds vaker verplicht is studies vóór uitvoering tot in detail te beschrijven, wordt dit gelukkig lastiger. Alle manipulaties zijn namelijk na afloop terug te zien. Zo valt op de belangrijkste aanmeldingsite, [clinicaltrials.gov](http://clinicaltrials.gov), te lezen hoe de fabrikant van Tamiflu tracht te rommelen met studie nr. 412737, de enige goede studie ooit verricht met deze griepremmer. Het middel had hierin geen beschermende werking op de vooraf bepaalde eindpunten ‘diagnose influenza’ en ‘griepverschijnselen’. Een jaar nadat de studie volledig was afgerond besloot het bedrijf daarom éxtra eindpunten in te brengen. Kennelijk is daar nog niemand ingetrapt, want vier jaar later is hij nog altijd niet gepubliceerd. De gehanteerde volgorde: (conclusie + resultaten + hypothese) → proef.

Dit kan nog een graadje onzinniger: als zelfs de algemene hypothese nog kan worden verzonnen. Neem het PVV-onderzoek naar de financiële gevolgen van een terugkeer naar de gulden; daar bestaat geen enkele plausibele rekenmethode voor, dus onderzoeksbureau Lombard Street kon ongeremd fantaseren met alle financiële data die ooit waren vastgelegd. Oftewel: (conclusie + resultaten) → (hypothese + proef).

De trucs doorzien is eenvoudig zolang je drie zaken helder formuleert: welke vraag stelde men in de studie, wat is de beste en meest logische manier om deze te beantwoorden, en hoe hebben de onderzoekers dat uiteindelijk gedaan? Zodra er een discrepantie bestaat tussen de drie antwoorden, is er iets niet in de haak.

De bovenstaande manipulaties leveren net zo’n onzin op als de methode Diederik Stapel, die metingen doen maar vermoeiend vond en ze dus maar zelf verzon (met andere woorden: conclusie → resultaten); ze komen vaker voor, ze worden nauwelijks verhuld en ze leiden tot veel grotere missers. Bij twee van de grootste medische schandalen van het afgelopen decennium – de Vioxx-affaire en de Mexicaanse griep – waren helemaal geen data vervalst. In de Vioxx-studie, een pijnstillertje met als bijwerking patiënten die doodgingen aan hartinfarcten, waren de eindpunten veranderd en waren wat bizarre extra hypothesen geformuleerd. Voor de keuze 340 miljoen euro te verkwisten aan vaccins tegen de Mexicaanse griep bestond de wetenschappelijke onderbouwing uit een paar zelfbenoemde experts die erop los fantaseerden, op een manier die het meest weg had van de eeuwenlang door de kerk gepropageerde kennisvergaring door deductie uit axioma’s.

De met verzonnen gegevens tot stand gebrachte publicaties van Stapel hadden daarentegen helemaal geen consequenties. Kent u iemand die is benadeeld door de

‘uitkomsten’ van zijn werk? Heeft zijn ‘rotzooi-maakt-mensen-racistisch’-studie u soms racistischer gemaakt? Het onderzoek werd door niemand ooit belangwekkend genoeg geacht om te reproduceren, dus de resultaten waren hooguit ‘voorlopig’. De verontwaardiging over zijn geval was in hoge mate gecultiveerd.

Een commissie wil nu vanwege een paar irrelevante oplichters 4.000 studies gaan doorlichten om te bepalen hoeveel er wordt gefraudeerd. Zo’n onderzoek is moeilijk, duur en behoorlijk arbeidsintensief; het zal dus op volstrekt inadequate wijze worden gefinancierd en uitgevoerd, met als noodgedwongen conclusie ‘wetenschappelijk onderzoek is ondoorzichtig’ (beter te lezen als: ‘we komen er niet uit’). Laten we daarom, zoals het hoort, vooraf de methode en de eindpunten (wat noemen we precies fraude?) goed definiëren, en vooral ook een hypothese opstellen: bij hoeveel procent van de studies mogen er data zijn gemanipuleerd? Mijn voorstel is 5 procent voor álle studies samen en 0 procent voor de gerandomiseerde trials in de geneeskunde.

Wetenschap heeft historisch gezien duizend keer meer te lijden gehad onder regels, dogma’s en verboden dan onder fraude. Zij is immers – wat sommigen met een kortetermijnvisie ook beweren – zelfreinigend. Onzin wordt weggefilterd: meteen door een co-auteur, iets later door een redacteur bij een wetenschappelijk tijdschrift, veel later door een lezer, of soms pas eeuwen later door een nieuwe generatie wetenschappers, maar uiteindelijk altijd. Al kan ik me voorstellen dat een eeuw lang is als het geduld je ontbreekt om je af en toe in een experiment te verdiepen.