

## Significant of niet?

### Denkfout in neurowetenschappelijk onderzoek

Neurowetenschappers schrijven opvallend vaak dat onderzoeksresultaten significant zijn, zonder dat goed te onderbouwen. Ze maken een redeneerfout die cognitief psycholoog Sander Nieuwenhuis opspoorde in verschillende wetenschappelijke tijdschriften. **door Geertje Dekkers**

**'H**ET

is verbazingwekkend dat het zo vaak gebeurt.' Cognitief psycholoog Sander Nieuwenhuis (Universiteit Leiden) heeft het over een statistische fout die hij regelmatig tegenkomt in neurowetenschappelijke artikelen. Samen met twee andere auteurs schreef hij erover in *Nature Neuroscience*.

De fout werkt als volgt: een wetenschapper bestudeert, bijvoorbeeld, het effect van een specifieke genmutatie bij muizen op leergedrag. Daarvoor heeft hij twee groepen muizen tot zijn beschikking: een onderzoeksgroep met de mutatie en een controlegroep zonder. Van beide groepen test hij aan het begin van het experiment hoe ze reageren op een bepaalde prikkel. Vervolgens ondergaan beide groepen een training die hun reactie op die prikkel verandert. Daarna bepaalt de onderzoeker opnieuw hun reactie.

Dan begint de analyse van de resultaten. Stel: de controlegroep lijkt enigszins te reageren op de prikkel maar het effect is niet significant. Bij de onderzoeksgroep is het effect wat groter en daardoor wel significant. De wetenschapper concludeert – en daar gaat het fout – dat het verschil tussen de effecten van de training in beide groepen significant is.

Maar dat hoeft niet het geval te zijn. Het verschil tussen wel en niet significant is geen scherpe scheidslijn. Vaak leggen wetenschappers de grens bij  $P < 0,05$ . Dat



Sander Nieuwenhuis.

0,06 kunnen liggen. De enige reden dat 0,05 zo vaak wordt gekozen is dat het zo'n mooi rond getal is, denk ik. Dat wetenschappers die grens zo absoluut interpreteren wordt het *cliff effect* genoemd. Aan de ene kant van die grens is het fantastisch en kun je mooie resultaten presenteren; aan de andere kant is het vreselijk leven.'

Nieuwenhuis en zijn collega's gingen na hoe vaak de fout voorkwam in artikelen over neurowetenschappen en aan-

verwante disciplines in vijf toptijdschriften: *Nature*, *Science*, *Nature Neuroscience*, *Neuron* en *The Journal of Neuroscience*. In 79 van de 513 onderzochte artikelen kwamen ze de fout tegen. In andere wetenschappen komt de fout uiteraard ook voor, maar die bleven buiten het onderzoek.

Het grote aantal fouten wil nog niet zeggen dat de conclusies van de artikelen niet kloppen, zo benadrukt Nieuwenhuis.

**ERRONEOUS ANALYSES OF INTERACTIONS IN NEUROSCIENCE: A PROBLEM OF SIGNIFICANCE**  
In: *NATURE NEUROSCIENCE* (2011) 14/9: 1105-1107.  
door Sander Nieuwenhuis, Birte U. Forstmann en Eric-Jan Wagenmakers.

Het verschil tussen de effecten kan significant zijn, ook als de auteurs dat slechts op basis van een redeneerfout hebben geconcludeerd. 'Als ik een schatting moet maken, is in een derde tot de helft van alle artikelen waarin we die fout hebben gevonden, echt iets mis', zegt Nieuwen-

huis: 'Als er eenmaal een publicatie is in *Nature of Science*, wordt het vaak moeilijk om een niet-replicatie te publiceren.'

Daarom is het zaak dat reviewers beter op gaan letten, vindt Nieuwenhuis. 'De schuld ligt echt bij hen. Ik kan me voorstellen dat auteurs in de verleiding komen om die fout te maken, bijvoorbeeld omdat de resultaten er mooi uitzien als je ze op deze manier presenteert. Maar die verleiding moet worden afgestraft door de reviewers. Ik ga ervan uit dat veel van hen ons artikel zullen lezen en in de toekomst alerter zijn op deze fout.'

Daarnaast moet een deel van de neurowetenschappers volgens Nieuwenhuis

**'De verleiding om deze fout te maken moet worden afgestraft door de reviewers.'**

huis. 'Dat is gebaseerd op de afstand tussen het significante effect en het niet-significante. Soms liggen die zo dicht bij elkaar dat je weet: het verschil is niet significant. Dat kan gewoon bijna niet.'

'We hebben het aantal fouten met belangrijke consequenties niet echt kunnen tellen omdat dat veel te veel werk zou zijn en in sommige gevallen inhoudelijke kennis vereiste die we niet hebben. Bovendien willen we geen mensen irriteren. Het gaat erom dat we het in de toekomst beter doen.'

Nieuwenhuis maakt zich geen heel grote zorgen over de gevolgen van de fout: 'In de meeste artikelen zijn de conclusies gebaseerd op heel veel resultaten. Misschien is de fout slechts gemaakt bij een van de twintig resultaten waarop de conclusies zijn gebaseerd.'

En als de redeneerfout toch tot verkeerde conclusies leidt, heeft Nieuwenhuis veel vertrouwen in het zelfreinigend vermogen van de wetenschap. 'Resultaten die in een blad als *Nature of Science* worden gepubliceerd trekken zo veel aandacht dat er altijd mensen zijn die proberen de resultaten te reproduceren; ofwel om het onderzoek onderuit te halen, ofwel om erop voort te bouwen. Dat systeem van replicatie werkt redelijk om dit soort fouten te corrigeren, denk ik.'

Maar perfect is het systeem niet. Publicatie in een toptijdschrift kan ook een tegengesteld effect hebben, erkent Nieu-

een betere statistiekopleiding krijgen: 'Voor de biologen onder de neurowetenschappers maken deze fout vaak omdat ze in hun opleiding te weinig statistiek krijgen. Bij psychologen komt het minder vaak voor doordat ze er tijdens hun studie uitgebreid voor worden gewaarschuwd.'

Het effect van een goede opleiding vermindert overigens in de loop van de tijd, als een onderzoeker regelmatig artikelen leest waarin de fout gemaakt wordt: 'Daardoor treedt gewenning op: 'blijkbaar is de fout niet zo heel erg', zegt Nieuwenhuis. 'Het taboe gaat eraf.' Nog een reden voor reviewers om beter op te letten en de redeneerfout in de toekomst uit wetenschappelijke publicaties te houden.

**'Aan de ene kant van de grens van 0,05 is het fantastisch en kun je mooie resultaten presenteren; aan de andere kant is het vreselijk leven.'**

wil zeggen: de kans dat de resultaten van een experiment op toeval berusten (en dus niet op de verklaring die onderzoekers geven) moet kleiner zijn dan 5%. Stel nu dat de resultaten in de onderzoeksgroep *nét* significant zijn ( $P = 0,049$ ) en in de controlegroep net niet ( $P = 0,051$ ). Dan ligt het voor de hand dat het verschil tussen de twee effecten niet significant is.

'In die grens van 0,05 zit enige willekeur', zegt Nieuwenhuis: 'Hij had ook bij

### ABG = Open Access

De ABG is een Open Access tijdschrift. U kunt alle artikelen op de website kosteloos downloaden en vrij gebruiken in het onderwijs.

Gelieve bij bronvermelding de volgende gegevens te vermelden: auteursnaam, nummer waarin het artikel is gepubliceerd en de uitgever Amsterdam University Press.

[www.academischeboekengids.nl](http://www.academischeboekengids.nl)