

De vloek van het winnende bod

Therapieën zonder

effect, resultaten die

wegsmelten. Soms

schiet onderzoek tekort.

Uit enthousiasme.

Margriet van der Heijden

bevestigd – of het nou om symmetrische haren van fruitvliegjes ging, om boerenzwaluwstaarten of om mensengezichten. Maar daarna, zo blijkt uit reconstructies van de Canadese bioloog Richard Palmer en de Australische Leigh Simmons, begon de theorie uit elkaar te vallen. Van de veertien publicaties uit 1994 vonden er maar acht het verband terug; in 1997 was er van het oorspronkelijk zo sterke effect zelfs nog maar twintig procent over.

Nog meer voorbeelden haalt Lehrer uit het werk van de Amerikaanse epidemioloog John Ioannidis, verbonden aan de universiteit van Stanford en auteur van venijnige artikelen als *'Why most published research findings are false'*. In 2005 liet Ioannidis zien dat later nogal eens bijgesteld of weerlegd worden.



Van de 49 spraakmakende klinische onderzoeken die Ioannidis analyseerde, werden er zelfs 7 later weerlegd (16 procent dus). Want: "De filosoof Francis Bacon (...) leerde ons dat we met experimenten de natuur kunnen ondervragen, maar tegenwoordig lijkt het wel alsof de natuur ons telkens andere antwoorden geeft."

Zou het? Is de natuur nukkig geworden? Heeft ze fratsen gekregen? Verderop in zijn stuk komt Lehrer met meer voor de hand liggende verklaringen. Zoals die voorliefde voor verrassende resultaten die bruisen en prikkelen en aanzetten tot meer onderzoek.

Lehrer heeft tegen die tijd het decline effect al onderbouwd met een trits voorbeelden uit de psychologie, de medische wetenschap en de biologie. Zoals het werk van de Deense zoöloog Anders Møller. In 1991 ontdekte Møller dat boerenzwaluwmannetjes met lange en symmetrische staartveren meer succes hadden bij de boerenzwaluwvrouwtjes – en meer nakomelingen kregen. En dat aansprekende resultaat werd in de drie jaar daarna in het ene na het andere onderzoek

over zulke doorbraken fantaseren we daarna graag nog wat verder." Neurofysioloog Mansvelter geeft toe dat hij zelf ook wel eens bang is dat menselijke trekjes als overenthousiasme hem parten spelen. "Een jaar of tien geleden had ik een paar resultaten die voor het eerst lieten zien dat zelfs het roken van één sigaret al tot veranderingen in de chemie van het brein leidt. Het werd zo groot opgepakt, ook door populaire bladen en kranten, dat ik er een beetje van schrok. En ik was opgelucht toen mijn werk na tien jaar nog overeind stond."

Dat Lehrer zijn voorbeelden uit de biologie, psychologie en medische wetenschap haalt, verbaast Mansvelter niet. "Onderzoekers in de harde bètavakken hebben het veel makkelijker", verzucht hij. "Scheikundigen werken al snel met miljarden moleculen. Deeltjesfysici maken in hun deeltjesbotsingen vele duizenden deeltjes. Maar in ons soort onderzoek werk je met mensen en met proefdieren en daarvan heb je hooguit tientallen, honderden en af en toe een paar duizend tot je beschikking." Dat maakt de kans op toevallige statistische uitschieters veel groter. Zeg maar, zoals je bij 18 worpen met een dobbelsteen ook wel eens stomtoevallig zeven keer zes kan gooien in plaats van ongeveer drie keer zes (een zesde deel van 18).

"Zo'n toevallige uitschieter kun je met beperkte groepsgroottes nooit helemaal uitsluiten", zegt Mansvelter, "Je moet natuurlijk wel proberen of je een vondst op meerdere manieren kunt onderbouwen. Of je het verschijnsel ook kunt verklaren op het niveau van de moleculen en van de cellen. En je kunt nagaan of het bij de literatuur aansluit. Toch blijft er altijd een zekere onzekerheid over."

RECHTGETROKKEN Maar: toevallige uitschieters kunnen in goed ontworpen vervolgentoetsen toch vrij snel rechtgetrokken worden? Waarom gebeurt dat dan in het decline effect aanvankelijk juist helemaal niet – en daarna alleen stapje voor stapje? Bioloog Leigh Simmons heeft het decline effect van dichtbij meegemaakt. Hij was een van die biologen die begin jaren negentig meteen dook op de vondst dat zwaluwen met symmetrische veren meer voortplantings succes hadden. Maar het lukte Simmons niet om het resultaat te repliceren. Hij slaagde er evenmin in om zijn 'negatieve' onderzoeksresultaat gepubliceerd te krijgen.

In een analyse achteraf ontdekte Simmons dat hij daarin niet alleen stond. In die eerste jaren na de ontdekking werden alle negatieve studies (vier) geweigerd. De vijf positieve studies die het effect bevestigden, haalden wel allemaal de bladen. Ofwel: de bladen holden achter de vondst aan en hun enthousiasme hield die vondst een paar jaar lang in stand.

Pas toen de nieuwe ideeën over symmetrie en voortplanting rond 1995 ingeburgerd waren, begonnen vakbladen de negatieve resultaten aantrekkelijk – want afwijkend – te vinden, zo laat Simmons overtuigend zien [grafiek uiterst rechts]. En toen pas werd het oorspronkelijke geloof in de theorie stapsgewijs getemperd. Zo komt

het decline effect neer op het compenseren van overenthousiasme. Maar dat is niet het hele verhaal, denkt Richard Gill. Hij is hoogleraar wiskundige statistiek in Leiden en werd onder meer bekend door zijn bemoeienis met de zaak van Lucia de Berk.

"Dit effect is een mooi onderwerp", zegt hij vergenoegd aan de telefoon vanaf zijn vakantieadres. Hij haalt de befaamde Britse fysicus Ernest Rutherford aan. Als je moet vertrouwen op statistiek, zei die, dan heb je het verkeerde experiment gedaan. Gill: "Maar ja, psychologen, biologen en medici kunnen meestal niet anders dan op ingewikkelde statistische analyses vertrouwen. Helaas zijn ze amateurs op het gebied van statistiek."

ETHISCHE KANTEN Het zit hem niet alleen in die kleine aantallen, zegt Gill. Een van de lastige kanten van het werken met mensen en dieren is dat ze allemaal verschillend zijn, en niet volslagen identiek zoals elementaire deeltjes of moleculen. Mensen en dieren hebben een eigen genetische opmaak en soms heel veel verschillende achtergronden.

Gill: "Spelen die mee, en hoe dan? Dat is niet zo makkelijk uit te vissen. En dan hebben we nog niet eens de ethische kanten genoemd. Als je het effect van meerooken op longkanker echt goed wilt bepalen, zou je eigenlijk een grote groep mensen jarenlang schone lucht moeten laten inademen, en een vergelijkbare, even grote groep juist jarenlang moeten laten meerooken. Maar ja, dat kan niet." Daarom wordt vaak retrospectief onderzoek gedaan, waarbij onderzoekers achteraf gegevens van grote groepen mensen analyseren. "Dan heb je wel grote aantallen, maar ook snel grote verwarring."

Neem weer dat onderzoek naar meerooken en longkanker dat inderdaad op die manier is uitgevoerd, zegt Gill. "Aanvankelijk werd een groot effect gevonden, veel groter zelfs dan wat je op grond van de meegerookte dosis verwacht. Pas later wezen andere onderzoekers op verborgen variaties die het effect konden uitvergroten, zoals de sociaal-economische klasse." Anders gezegd: roken komt vaker voor in lagere sociale klassen met een ongezonde levenswijze – en niet alleen meerooken, maar ook die levenswijze leidt tot een toename van kanker. Gill: "Zelf denk ik dat decline effecten heel vaak op een dergelijke wijze te verklaren vallen. Dus doordat vervolgonderzoeken verwarrende en complicerende factoren steeds verder uit elkaar rafelen. En dat het zo gaat is best logisch: een onderzoek op exact dezelfde wijze herhalen heeft geen zin. Je probeert het altijd te verfijnen en te verbeteren." En soms verklaar je zo dus na vele jaren en studieboeken het oorspronkelijke effect weg.

HOOG INZETTEN Is het erg dat het zo werkt? Schei toch uit, helemaal niet, schreef de Amerikaanse molecuul medicus David States in een reactie aan de *New Yorker*. Het is gewoon de *winner's curse*, 'de vloek van het winnende bod'. Dat is het idee dat bij een veiling van voorwerpen van onzekere waarde de winnaar die met een object naar huis gaat, grote kans loopt dat

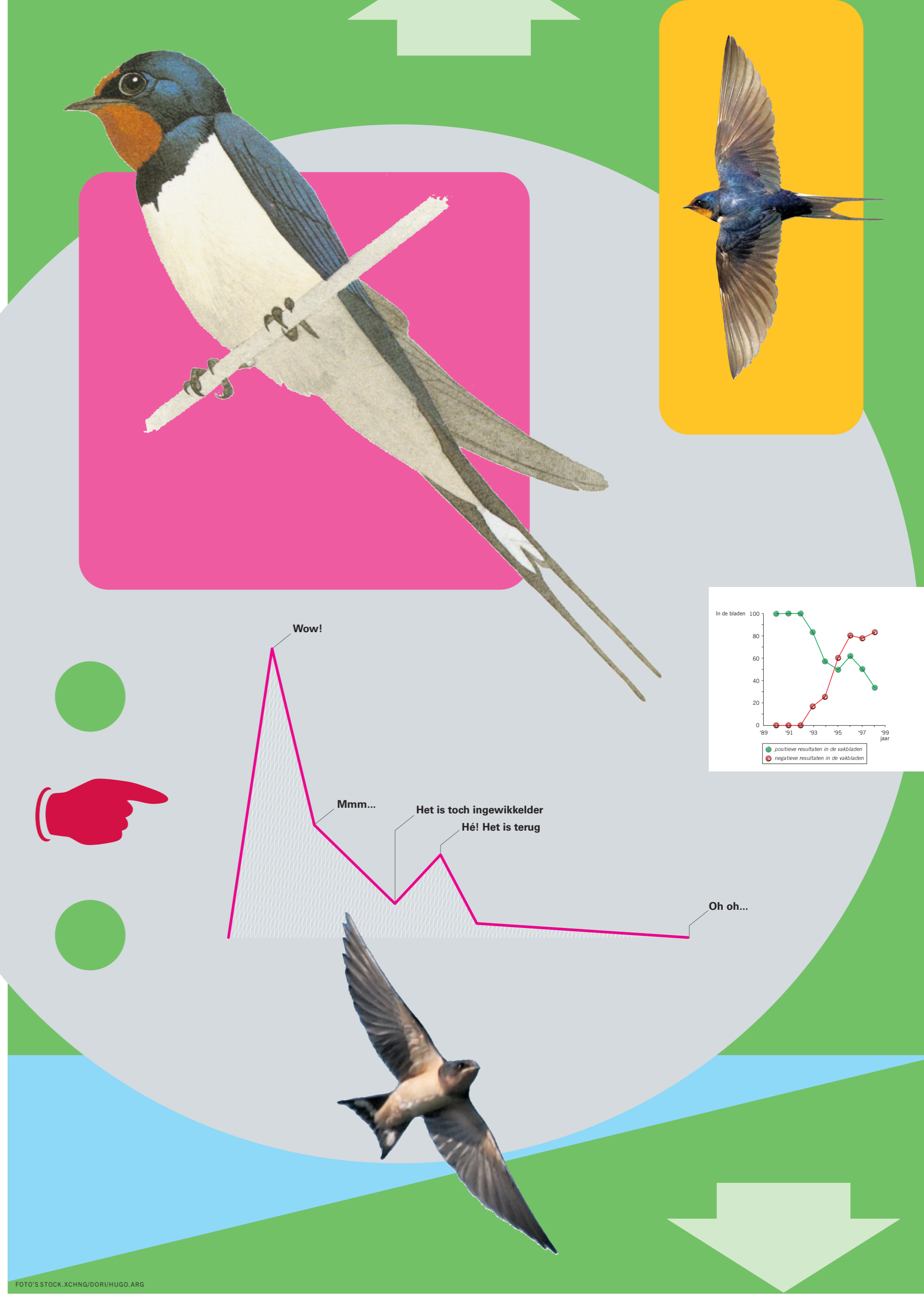
hij veel te veel betaald heeft. Zo werkt het in de wetenschap ook, schreef States. Wie snel werkt en hoog inzet, wordt veel geciteerd, maar loopt daarna ook het risico onderuit gehaald te worden. Collega's springen nu eenmaal bovenop sterke nieuwe claims. Helemaal niet erg, schreef States, want juist competitie stuwt de wetenschap voort.

"Ik denk dat het inderdaad zo werkt en dat het decline effect daarvan een bijverschijnsel is", zegt Mansvelter. "En ik denk ook niet dat je dit kunt veranderen. Deze werkwijze sluit aan bij hoe mensen in elkaar zitten. Je moet je er alleen van bewust zijn. Maar dat is iedereen in het onderzoek ook wel. Onderzoekers die over een nieuw resultaat horen, wachten vaak eerst maar eens af of het niet weerlegd wordt. Soms hoor je er dan niets meer over. Nou, dan weet je ook genoeg."

Gill moet aan de telefoon een beetje grinniken om de term. "Een vloek zou ik het niet noemen, want tegen de tijd dat het resultaat onderuitgehaald wordt, heeft de onderzoeker er al lang carrière mee gemaakt en subsidies mee binnengehaald. Een vloek is het hooguit voor mensen die voor niks een bepaald middel hebben geslikt. Die zich voor niks zorgen hebben gemaakt over de onbewezen ongezonde effecten van koffie. Of die zich voor niks verheugd hebben over onbewezen gezonde effecten van rode wijn."

Maar het is waar, zegt ook Gill, dat er weinig vooruitgang zou zijn als elk onderzoek bij voorbaat heel groot-schalig en totaal doordacht zou worden opgezet. "Bovendien is er best veel verbeterd. In 1981 constateerde Doug Altman nog dat de helft van de statistische resultaten in de medische topbladen niet deugde. Nu spreken vooraanstaande onderzoekers op dit terrein, zoals de Oostenrijker Alexander Strasak, nog maar van zo'n 15 procent. Al is er dan in de psychologie en delen van de medische biologie helaas weer nauwelijks iets veranderd."

KRONKELPADEN Kom, zegt KNAW-president Robbert Dijkgraaf, wordt er ook niet een beetje overenthousiast op het decline effect gereageerd? "Kronkelpaden en misleunen horen er bij – als je maar de onzekerheden en foutmarges aangeeft." Achteraf lijkt het zo makkelijk om de juiste conclusie te trekken, maar in de praktijk moeten onderzoekers, vaak onder druk, uit beperkte informatie een beeld zien te reconstrueren. "Het lijkt op zo'n computerspelletje waarbij steeds meer punten op het scherm verschijnen en waarbij je zo snel mogelijk moet raden welk beeld die samen vormen. Als je het goede antwoord geeft, maar wel veel te laat, tja, dan word je niet meer gehoord." Dat is in elk geval iets dat ook de buitenstaander maar beter in het achterhoofd kan houden. Wordt er een spannend nieuw effect gerapporteerd op grond van één beperkt experiment? Gaat het om een vervolgentoets dat helemaal in de heersende wetenschappelijke mode past? Iets spannends om achter aan te hollen? Mm, dan eerst maar eens een paar jaar wachten tot de *winner's curse* is uitgeverkt en het bijbehorende decline effect zijn werk heeft gedaan. ●



FOTO'S STOCK.XCHNG/DORI/HUGO.ARG