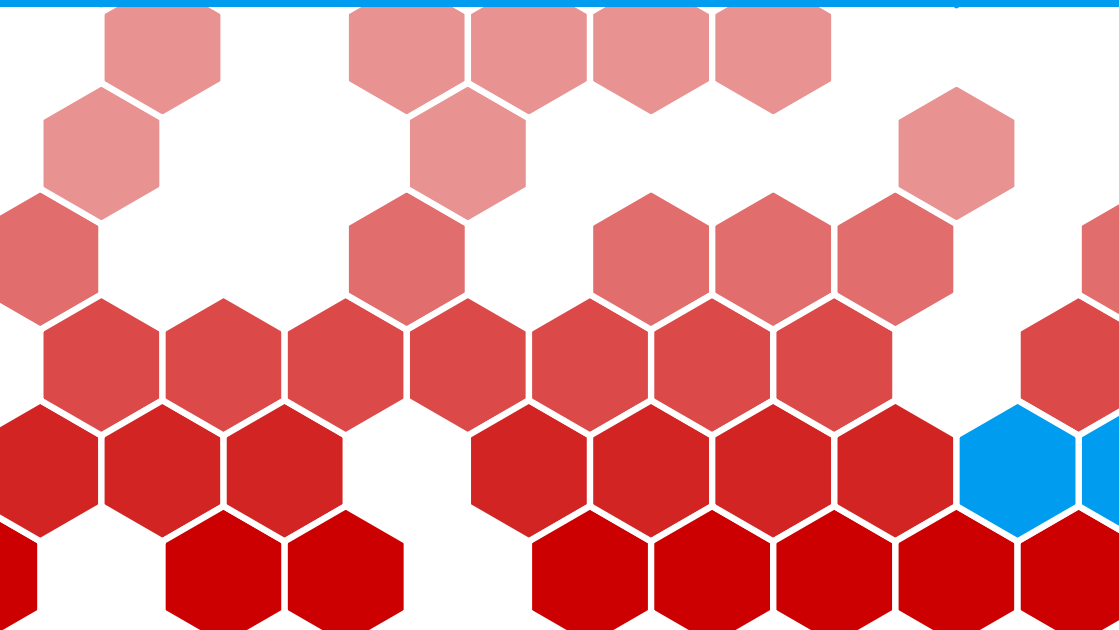


# Toegepaste Statistiek: Van Ontwikkeling tot Communicatie

Prof. dr. Casper Albers





## **Toegepaste Statistiek: van Ontwikkeling tot Communicatie**

Rede uitgesproken bij de aanvaarding van het ambt van hoogleraar in de Toegepaste Statistiek en Datavisualisatie aan de Rijksuniversiteit Groningen op dinsdag 19 maart 2019 door Casper J. Albers.



**rijksuniversiteit  
groningen**

## Dankwoord

*You brought me fame and fortune and everything that goes with it,  
I thank you all ☺*

— Freddie Mercury, 1978

Deze oratie zou niet mogelijk zijn geweest zonder de samen- en medewerking van veel geweldige mensen. Mijn dank hiervoor is groot. Ik wil mijn dank uitspreken aan Marijn, Emma, Lara, familie en vrienden, de (oud)collega's bij de Psychometrie- en Statistiekbasiseenheid en de (oud)collega's van de Faculteits- en Universiteitsraad.

Dank ook aan iedereen met wie ik de afgelopen tien jaar samen wetenschap heb mogen bedrijven: Angélique Cramer, Anja Boevé, Anja Ernst, Bertus Jeronimus, Carlien Vermue, Cathelijne Stoof, Christien Slofstra, Daniel Lakens, Don van Ravenzwaaij, Elliot Sharpe, Fionneke Bos, Gemma Venhuizen, Gert Stulp, Hans Beldhuis, Henk Kiers, Henriët Scholts, Ionica Smeets, Jan Gerard Hoendervanger, Jan Willem Bolderdijk, John Gower, Jorge Tendeiro, Jorien Vugteveen, Laura Bringmann, Leonie Kreuze, Li JiYao, Lieke Voncken, Linda Steg, Lise Jans, Maaïke Nauta, Maarten Derksen, Maliheh Namazkhan, Mandy van der Gaag, Marieke Schuppert, Marieke Timmerman, Marieke Wichers, Marije aan het Rot, Mariska Barendse, Mark Moch, Mark Verschoor, Matt Williams, Nick Brown, Nico van Yperen, Nicola Klein, Nina Schwarzbach, Nitin Bhushan, Peter Kuppens, Pia Tio, Renske Bosman, Rink Hoekstra, Rob Meijer, Roel Bosker, Rolf Hut, Sanne Willems, Saskia Kunnen, Sjoukje van Dellen, Stuart Ritchie, Tanja Krone, Thijs Bouman, Veronika Cheplygina, Willem Schaafsma en Ybe Meesters.



Leden van het College van Bestuur,

Zeer geachte aanwezigen,

Mijn promotor, Willem Schaafsma, heeft mij met grote regelmaat meegegeven dat de wetenschapper zich moet laten leiden door drie platonische principes, te weten de waarheid, schoonheid en goedheid. Met die waarheid, schoonheid en goedheid zit het wel goed binnen de wiskunde en de wiskundige statistiek. Dat wiskundige bewijzen 'waar' zijn, staat buiten kijf. Dat de wiskunde ons veel goeds heeft gebracht, moge ook duidelijk zijn. De schoonheid van de wiskunde en wiskundige statistiek is mijns inziens ook duidelijk.

Neem als voorbeeld de welbekende normale verdeling. Carl Friedrich Gauss bestudeerde astronomische data en stelde vier criteria op waaraan de verdeling van meetfouten van die data moest voldoen:<sup>2</sup>

- (i) Kleine meetfouten zijn waarschijnlijker dan grote meetfouten;
- (ii) De verdeling van meetfouten is symmetrisch;
- (iii) De verdelingscurve moet een vloeiende vorm hebben;
- (iv) Als je een bepaalde kwantiteit meermalen gemeten hebt, is het gemiddelde van de metingen de meest voor de hand liggende schatter.

Het elegante van de normale verdeling is dat deze vier ogenschijnlijk eenvoudige criteria compleet voldoende zijn om de vorm van de verdeling exact te bepalen. Dus ook de schoonheid is aanwezig.

Maar dan de toegepaste statistiek. De waarheid, schoonheid en goedheid lijken overboord gekieperd te worden. Je weet als toegepast statisticus immers nooit zeker of je daadwerkelijk het juiste antwoord geeft, je werkt doorgaans met aannames waarvan je weet dat ze eigenlijk niet waar zijn en helaas zijn de *questionable research practices* aan de orde van de dag. Dit zijn onderzoekspraktijken waarin een loopje genomen wordt met statistische technieken, met het doel de resultaten mooier weer te geven dan ze zijn. Hier later meer over.

U zult zich wellicht afvragen waarom ik dan de overstap van de wiskundige naar de toegepaste statistiek gemaakt heb. Welnu, dat die waarheid, schoonheid en goedheid binnen de toegepaste statistiek zo veel moeilijker te bereiken zijn, maakt het naar mijn mening juist een uitdagender vakgebied.

De toepassing in de psychologie is een extra uitdaging. Voordat ik bij psychologie kwam werken, werkte ik bij de Open University in Engeland aan modellen om de verkeersstromen op Britse snelwegen te voorspellen.<sup>52,13,14</sup> Per rijstrook wisten we per minuut van de dag, en dat maandenlang, hoeveel voertuigen er overheen reden. Tellen hoeveel auto's ergens rijden is simpel. Zo simpel dat we computers het werk kunnen laten doen. Dat gaat binnen de psychologie een stuk uitdagender. Emoties, attitudes en zelfs gedrag laten zich een stuk lastiger langs de meetlat leggen. En net zoals dat een automobilist mogelijk de snelheid aanpast als er een verkeerscamera op de auto gericht is, past een mens zijn gedragingen aan als hij of zij gemeten wordt (het bekende Hawthorne-effect). Die extra complexiteit is tevens een extra uitdaging.

Een tweede reden waarom ik de statistiek graag toepas, is dat ik daarmee een bijdrage kan leveren aan het oplossen van grote vraagstukken waar de maatschappij voor staat. Daar kom ik graag mijn wiskundige safe zone voor uit. Vandaag wil ik twee toepassinggebieden waarin ik speciale interesse heb met u bespreken.

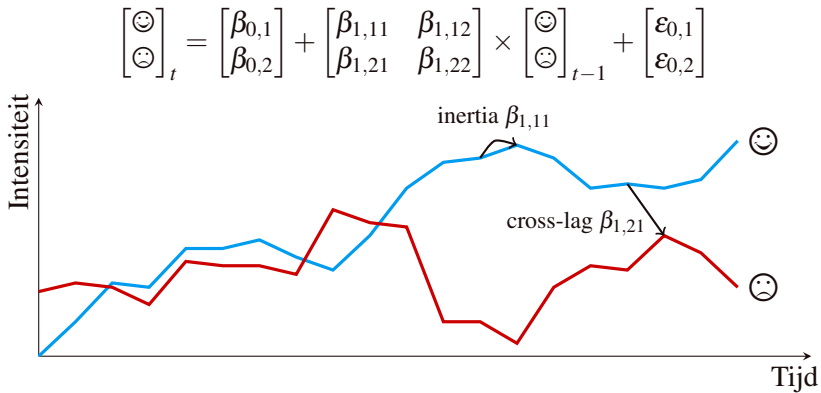
## Statistiek toegepast in de psychopathologie

Ten eerste, de psychopathologie en ontwikkelingspsychologie. Een op de vijf volwassenen krijgt psychische problemen. Depressie en angststoornissen zijn hiervan de meestvoorkomende problemen.<sup>60</sup> Suicide is doodsoorzaak nummer 1 onder Nederlandse jongeren. Empirisch onderzoek op dit gebied is enorm belangrijk maar ook ingewikkeld.

Er is namelijk een ingewikkeld samenspel van veel variabelen. Zo is het om goed te voorspellen hoe ernstig iemands winterdepressie is onvoldoende om alleen te kijken naar de hoeveelheid zonlicht. Een combinatie van maar liefst negen meteorologische variabelen is nodig om een goed beeld te krijgen van de situatie.<sup>54</sup>

Veel variabelen tegelijk meten is voor dit soort modellen niet eens voldoende. Je moet ze ook nog eens met grote regelmaat meten. Πάντα ρεῖ (panta rhei), een uitspraak toegeschreven aan de Griekse filosoof Heraclitus, betekent ‘alles stroomt’. Alles verandert continu, zo ook de mens. Om goed de processen die ten grondslag aan menselijk gedrag liggen te begrijpen, is het dus van vitaal belang om regelmatig dezelfde variabelen bij dezelfde personen te meten. Dit is nodig om inzicht te verkrijgen in de gecompliceerde aspecten van menselijk gedrag. Die complexiteit uit zich door fluctuaties in gedrag en emotionele gesteldheid over de tijd. Deze fluctuaties hangen af van de context, van interindividuele verschillen, en van toevallige verstoringen. Het begrijpen van de dynamiek van een psychologisch proces is een essentiële voorwaarde om het proces zelf te begrijpen.

Tot voor kort was het om praktische redenen onmogelijk of, op z'n minst, enorm moeilijk om bijvoorbeeld bij depressie-onderzoek met voldoende regelmaat de mentale toestand van proefpersonen te meten. Dankzij technologische vooruitgang is dit nu wel mogelijk. Sinds een jaar of tien wordt de *experience sampling method* en de *ecological momentary assessment* methode steeds meer gebruikt. Via deze methodes vullen deelnemers aan een onderzoek meerdere keren per dag via hun smartphone een korte vragenlijst in. Er wordt bijvoorbeeld vijf keer per dag gevraagd om een viertal emoties te kwantificeren. Op die manier wordt intensieve longitudinale data verkregen. Vervolgens wordt via een regressiemodel bekeken hoe de emotie op tijdstip  $t$  samenhangt met die op het vorige tijdstip,  $t - 1$ . Bij een enkele emotie verkrijg je het zogenaamde AR(1) model en bij het modelleren van meerdere variabelen tegelijk het VAR(1) model (zie Figuur 1).



*Figuur 1: Het bivariate VAR(1) model*

Met Tanja Krone en Marieke Timmerman zijn de prestaties van verschillende varianten van deze modellen vergeleken.<sup>41,42</sup> Voor gebruik in de praktijk zijn (V)AR modellen doorgaans te elementair en wordt er gebruik gemaakt van ingewikkeldere modellen, maar het AR en VAR model liggen wel aan de basis van deze methoden. Deze modellen voor intensieve longitudinale data laten zien dat niet alleen hoe iemand zich gemiddeld gesproken voelt van belang is om iets over iemands welzijn te kunnen zeggen; de temporele dynamiek geeft ons inzichten die we niet uit de gemiddeldes kunnen halen.<sup>26,31,39</sup> Tevens geeft het bestuderen van trajecten over de tijd de onderzoeker de mogelijkheid om bijvoorbeeld te voorspellen of en wanneer een depressieve patiënt een terugval kan verwachten,<sup>44,56</sup> zodat de therapeut pro-actief kan optreden.

De kracht van deze modellen zit hem in de mogelijkheid om tegelijkertijd interindividuele en intraindividuele uitspraken te doen, oftewel zowel iets te zeggen over de vergelijking tussen personen als de dynamiek van een individu. Als het aantal deelnemers aan een studie te groot wordt, wordt de hoeveelheid individuele modellen onoverzichtelijk en biedt het meerwaarde om



personen met gelijksoortige dynamiek te clusteren. Een ad hoc aanpak<sup>40</sup> is uitgegroeid tot het lopende promotietraject van Anja Ernst.<sup>28,29</sup>

De meeste tijdsreeksmodellen maken een onrealistische aanname van stationariteit, namelijk dat de zogenaamde emotionele inertia niet kan wijzigen over de tijd. Zeker wanneer de metingen tot doel hebben een patiënt te volgen die onder behandeling staat, is dit een groot probleem. Immers, het hele doel van de behandeling is om de interne dynamiek van de patiënt te wijzigen. Die dynamiek kan zowel geleidelijk<sup>21</sup> wijzigen als abrupt.<sup>32</sup> Met Laura Bringmann wordt momenteel de laatste hand gelegd aan een nieuw model<sup>7,20</sup> dat beide types verandering tegelijk aan kan. De komende jaren dient dit model uitgebreid te worden om ook multivariate veranderingen aan te kunnen, alsmede om bruikbaar te zijn in de context van psychologische netwerken en clusteringsmethoden.

De afgelopen jaren zijn de modellen voor intensieve longitudinale psychologische data als paddestoelen uit de grond geschoten. Doorgaans zijn deze modellen gebaseerd op enkele verstandige aannames waarna via uitgebreide computersimulaties is aangetoond dat die modellen, binnen een bepaalde context, goed werken. Dat is uiterst nuttig maar uiteindelijk gaat het er om dat de modellen in de praktijk ook bruikbare antwoorden opleveren. Benchmarking-onderzoeken zijn een standaardmethode binnen ondermeer de informatica, maar nog niet binnen de psychologie. Wij hebben een studie opgezet<sup>15</sup> waarbij twaalf geselecteerde teams, bestaande uit zowel methodologische als klinische experts, elk exact dezelfde dataset moesten analyseren. Momenteel liggen de antwoorden van de verschillende aanpakken nog zodanig ver uit elkaar dat ze slechts van beperkt praktisch nut zijn. Tot nu toe wint de statistische predictie het dus niet van de klinische predictie<sup>46</sup> en het is nog maar de vraag of dit in de toekomst anders kan zijn.

Het is ook de vraag of dit überhaupt wenselijk is. Mijns inziens kunnen dit soort statistische modellen het beste gebruikt worden om het klinische

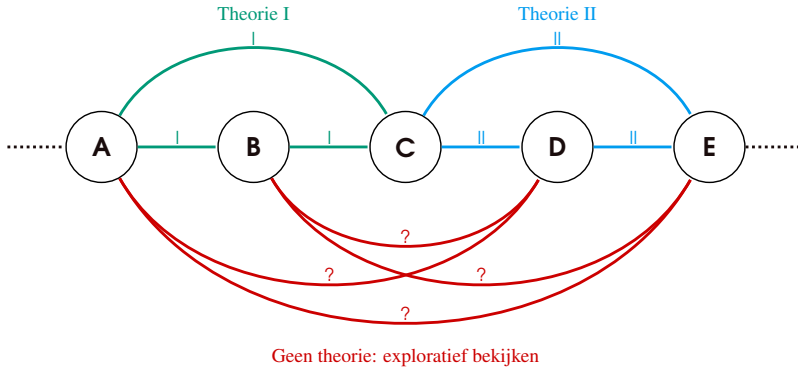
oordeel te verbeteren en niet om deze vervangen. Dit verbeterde oordeel zal leiden tot verbeterde klinische diagnostiek en dus tot verbeterde behandeling van de meest kwetsbare personen in de samenleving.

## ◆ Statistiek toegepast in de omgevingspsychologie

Het tweede toepassinggebied dat ik met u wil bespreken, is de energietransitie en dan met name de omgevingspsychologische kant er van. Rapport na rapport<sup>35</sup> laat zien dat er enorme veranderingen moeten plaatsvinden, willen we de opwarming van de aarde beperken tot 1,5°C. Met alleen technologische innovaties en extra windmolens redden we het niet; een gedragsverandering is nodig. Voordat op gedragsverandering aangestuurd kan worden, is het van belang dat gedrag te begrijpen. Samen met Linda Steg, en promovendi Nitin Bhushan, Maliheh Namazkhan en Mark Verschoor bestudeer ik het gedrag van huishoudens rond hun energieverbruik en wat de determinanten zijn van opvattingen rond klimaatverandering.

Hierbij spelen meerdere sociaal-psychologische en economische theorieën een rol.<sup>45,51,65</sup> Elke theorie afzonderlijk bestudeert een onderdeel van het geheel, bijvoorbeeld hoe attitudes en zelfeffectiviteit invloed hebben op intenties en hoe dit weer invloed heeft op gedrag. Je kan de theorieën afzonderlijk bestuderen, in een energiecontext, maar het is veel interessanter ze gelijktijdig te bestuderen (zie Figuur 2). Je kan dan zowel confirmatief kijken naar de reeds bestaande getheoretiseerde verbanden als exploratief kijken naar verbanden die individuele theorieën overstijgen. Zo'n geïntegreerde aanpak leidt dus tot inzichten die zonder deze grootschalige aanpak niet verkregen kunnen worden. Omdat voor deze methoden veel data tegelijk geanalyseerd dient te worden, zijn goede *data warehousing* faciliteiten essentieel.

Uiteindelijk is het doel om de psychologische variabelen te koppelen aan het energieverbruik. Omdat energieverbruik vanzelfsprekend ook afhangt van tal van andere variabelen, zoals bouwtechnische aspecten van de woning,



*Figuur 2: Schematisch overzicht van de koppeling tussen meerdere theorieën. Theorie I postuleert de onderlinge verbanden tussen **A**, **B** en **C**; Theorie II die tussen **C**, **D** en **E**. Over de in rood weergegeven verbanden, bijvoorbeeld die tussen **A** en **D**, is (nog) geen theorie.*

sociaal-economische variabelen en het weer<sup>49</sup> is dit geen gemakkelijke klus. Het is als het op elkaar proberen te zetten van twee LEGO-blokjes van verschillend formaat en hiertoe dienen op maat gemaakte statistische modellen ontwikkeld te worden.

Omdat met deze aanpak het aantal gemeten variabelen snel stijgt, raakt het model snel onoverzichtelijk. Het psychologische netwerkmodel, tot nu toe met name binnen de psychopathologie gebruikt, biedt een visuele uitkomst om door de bomen toch een bos te zien.

Eerste analyses, waarin we hebben gekeken naar een geïntegreerd beeld van relevantie theorieën, maar nog niet naar de koppeling met individueel energieverbruik, zijn veelbelovend. Zowel de analyse van een grootschalig Nederlands buurtinitiatief<sup>16</sup> als die van Ronde 8 van de European Social Survey<sup>62</sup> laten zien dat dergelijk netwerkmodellen van toegevoegde waarde zijn binnen

de omgevingspsychologie. Tevens zien we dat verbanden tussen variabelen in de meeste Europese landen zeer overeenkomstig zijn. Dit laatste is van belang omdat dit impliceert dat beleidswijzigingen in alle landen een vergelijkbaar effect zullen hebben.<sup>19</sup>

De vervolgstap in dit onderzoek is om de dynamiek binnen huishoudens te onderzoeken. Wie en wat binnen een huishouden kan er voor zorgen dat het energiegedrag groener wordt? Momenteel verzamelen we data in verschillende Europese landen bij verschillende typen huishoudens, zoals gezinnen en studentenhuisen, om antwoord op deze vragen te krijgen.

## Communicatie

Ik heb twee toepassingsgebieden met u besproken. Ik heb verteld wat de toegepaste statistiek daar de afgelopen jaren betekend heeft en de komende jaren kan betekenen. Het ontwikkelen van een statistisch model is maar het halve verhaal. De andere helft is er voor zorgen dat het model, en haar uitkomsten, door de belanghebbenden goed begrepen wordt. Deze belanghebbenden zijn doorgaans zelf geen statisticus, maar sociaal-wetenschapper, beleidsmaker, therapeut of patiënt. Zonder goede statistische communicatie geen toegepaste statistiek. Juist in deze tijd van *fake news* en *framing*<sup>6</sup> kan een goede communicatie van wetenschappelijke resultaten niet meer gezien worden als een fijne bonus, maar dient het een essentieel ingrediënt van de wetenschappelijke methode te zijn.

En statistiek is niet altijd even makkelijk uit te leggen. Nobellaureaat Daniel Kahneman onderscheidt twee denksystemen:<sup>38</sup> *thinking fast*, waarbij moeiteloos en intuïtief wordt gedacht, en *thinking slow*, waarbij analytisch wordt gedacht. Bij statistische vraagstukken is uit de context niet altijd duidelijk dat er analytisch gedacht moet worden, waardoor het brein in de verkeerde modus staat. En dan is statistiek inderdaad moeilijk.

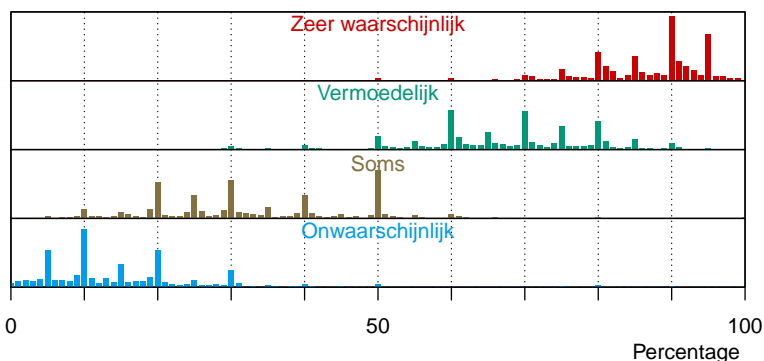
Deze moeite met statistiek, die zelfs door hoogopgeleide wetenschappers wordt ervaren, is een van de hoofdoorzaken van de replicatiecrisis<sup>50</sup> die sinds een aantal jaar de psychologie, en daarna ook de andere empirische wetenschappen, in haar greep heeft. Kort samengevat behelst de crisis dat onderzoekers vaak methodologisch onverantwoorde stappen zetten in de wetenschaps-cyclus, waardoor de gevonden resultaten vertekend en daarmee onjuist zijn. Deze onverantwoorde stappen, of *questionable research practices*,<sup>36</sup> worden in uitzonderlijke gevallen – gevallen van fraude – opzettelijk gemaakt, maar doorgaans komen zijn door een onvoldoende methodologisch inzicht van de wetenschapper.

Hoewel methodologen als Paul Meehl<sup>47,48</sup> en Jacob Cohen<sup>24,25</sup> een halve eeuw geleden al waarschuwden voor weeffouten in het wetenschappelijk proces, lijkt er het afgelopen decennium pas echt momentum te zijn om de crisis aan te pakken. Het is de taak van de statisticus om een actieve rol hierbij te spelen. Dit kan op een meerdere manieren, waaronder

- (i) Het onder de aandacht brengen van minder bekende bestaande methodologische kennis;<sup>27,43</sup>
- (ii) Het wijzen op methodologische fouten in gepubliceerde *high-impact* artikelen;<sup>22,1</sup>
- (iii) Het actief meewerken aan de introductie van technieken die de weeffouten in de wetenschap kunnen verkleinen, zoals pre-registratie<sup>69</sup> en *open science*;<sup>4,61</sup>
- (iv) Het ontwikkelen en aanbieden van open access methodologisch onderwijs aan wetenschappers en het algemeen publiek.

Deze taken leiden niet direct tot nieuwe wetenschappelijke kennis en zouden door de purist dus niet als taak van de onderzoeker bestempeld worden. Echter, de wetenschapper heeft ook een onderwijstaak en die dient zich zeker niet tot alleen het klaslokaal te beperken.

Terug naar de moeilijke statistiek. Er is veel onderzoek gedaan naar welke



Figuur 3: Verdeling van de interpretatie van vier kanswoorden.<sup>68</sup>

statistische vraagstukken mensen moeilijk vinden. Het driedeurenprobleem is een overbekend voorbeeld. Het blijkt dat zelfs duiven beter te trainen zijn in het oplossen van dit probleem dan mensen.<sup>5,33</sup> Er is echter nog verrassend weinig onderzoek gedaan naar *waarom* mensen dit soort vragen zo moeilijk vinden, noch naar welk type mens er extra moeite mee heeft of wat men kan doen om de moeilijkheden weg te halen. Het is bijvoorbeeld erg interessant om te kijken vanaf welke leeftijd de intuïtie men de verkeerde kant op stuurt bij vragen als het driedeurenprobleem. Als duiven het beter kunnen dan mensen, omdat duiven niet gehinderd worden door menselijke intuïtie, dan kunnen kinderen het mogelijk beter dan volwassenen; maar tot welke leeftijd?

Samen met Sanne Willems en Ionica Smeets heb ik onlangs uitgezocht hoe het zit met de interpretatie van Nederlandstalige kanswoorden<sup>68</sup> (zie Figuur

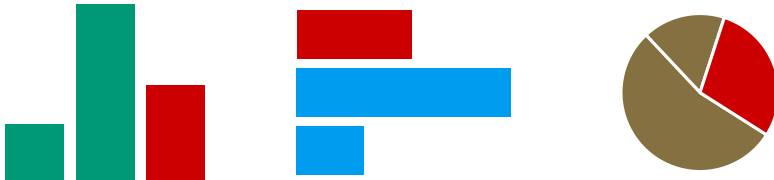
3). Net als bij soortgelijk Engelstalig onderzoek vonden wij dat de ene persoon een woord als ‘meestal’ totaal anders interpreteert dan de ander. Dit heeft vergaande gevolgen voor risicocommunicatie, bijvoorbeeld in flyers met informatie over ziekten. Wat wij ook onderzocht hebben is of statistici, die beroepsmatig met kanswoorden bezig zijn, homogener zijn in hun interpretaties van kanswoorden dan leken. Dit bleek niet het geval. Uitgebreide kanstraining zorgt dus niet voor een eenduidiger kansbegrip, althans niet wanneer het kansbegrip omgezet wordt in woorden. Welke variabelen dan wel een rol spelen bij deze heterogeniteit is iets dat ik de komende jaren zal onderzoeken.

## Visualisatie

Een belangrijk onderdeel van statistische communicatie is datavisualisatie. Met John Gower heb ik verschillende bijdragen geleverd aan de visualisatie van zogenaamde triadditieve tabellen.<sup>8,9</sup> Hoewel dit een zeer interessante exercitie in de lineaire algebra was, was het toepasbaar nut beperkt: zelfs met geavanceerde visualisaties, blijven interpretaties van drieweg interacties ingewikkeld.

Het gebruik van visualisaties om kwantitatieve informatie inzichtelijk te maken, gaat eeuwen terug.<sup>30</sup> Door technologische vooruitgang, de beschikbaarheid van softwarepakketten zoals *R*<sup>53</sup> en *ggplot2*<sup>66</sup> en de mogelijkheid tot animaties en interactieve visualisaties maar ook door de opkomst van ingewikkeldere statistische datasets en modellen zijn datavisualisaties aan een opmars bezig in de sociaal-wetenschappelijke literatuur.

Om te weten of de boodschap die via een statistische grafiek wordt uitgedragen goed overkomt, moeten we weten hoe mensen grafieken interpreteren. Er zijn uitgebreide richtlijnen omtrent goede datavisualisatie.<sup>67</sup> Er is ook enig onderzoek gedaan naar hoe grafieken geïnterpreteerd worden.<sup>55,23,58</sup>



*Figuur 4: Drie verschillende visuele representaties van dezelfde verhoudingen.*

Echter, er is verrassend weinig onderzoek gedaan naar waarom het op die manier geïnterpreteerd wordt. Zo schatten mensen kansen anders in wanneer deze via horizontale staafdiagrammen worden gepresenteerd dan wanneer deze via verticale staafdiagrammen worden gepresenteerd, ook al is de informatie wiskundig gezien equivalent (zie Figuur 4). We weten niet goed waarom. Er is literatuur over de psychologische aspecten van het verwerken van grafische informatie.<sup>64</sup> Maar hoe dit in een statistische context in zijn werk gaat, is onvoldoende bekend.

Welke visualisatietechniek optimaal is, hangt vanzelfsprekend ook van de context af: moet snel een oppervlakkig inzicht verkregen worden, of moet het juist een boodschap zijn die goed blijft hangen? Is de boodschap vrij eenvoudig of juist gecompliceerd en genuanceerd? De keuze hangt ook af van de doelgroep die men wil bereiken, bijvoorbeeld medewetenschappers of patiënten.<sup>57</sup>

Neem het toepassingsgebied van de psychopathologie, waarin netwerkvisualisaties, en zelfs netwerkanimaties, een prominente rol innemen. Er worden technisch gezien vernuftige visualisaties gemaakt, maar of de boodschap uit de grafiek ook goed overkomt bij de therapeut en de patiënt is nog maar de vraag.<sup>18</sup> Indien er structureel verkeerde interpretaties gemaakt worden, kan





dit leiden tot suboptimale behandelingen of zelfs verkeerde klinische diagnoses. Kortom, het maken van de visualisatie is de ene helft van het verhaal, zorgen dat de visualisatie goed begrepen wordt de andere helft.

De komende jaren wil ik me richten op het slaan van een brug tussen het statistische model en het gebruik ervan. Dit houdt in dat ik wil bijdragen aan het verbeteren van de wetenschappelijke cultuur om *questionable research practices* tegen te gaan. Tevens wil ik me storten op de vragen rond de interpretatie en communicatie van kansen, onzekerheden en risico's. Tenslotte wil ik middels nieuw onderzoek meewerken aan een cultuurverschuiving richting het gebruik van visualisaties die bewezen effectief zijn. Dit zal er voor zorgen dat de toegepaste statistiek van waarde blijft voor de toepassing.

## ◆ **Het academisch onderzoeksklimaat**

Om goed onderzoek te kunnen doen is een goede onderzoeksomgeving onontbeerlijk. Hoewel er ontzettend veel goed gaat aan deze universiteit, wil ik er toch op wijzen dat de academische omgeving, zowel voor studenten als medewerkers, beter kan en beter moet.

De focus ligt nu te veel op rendement, wat een nadelig effect heeft. Het klopt dat de karige bijdrage van de overheid aan wetenschap en hoger onderwijs niet helpt, maar dit is zeker niet de enige schuldige. Om tegenwoordig promovendus te kunnen worden, dien je eigenlijk wel een research masteropleiding gedaan te hebben. En om zo'n opleiding te kunnen volgen, is het zeker een voordeel als je in de bacheloropleiding deelnam aan een *honours college*. En daar kom je eigenlijk niet binnen zonder uitstekende cijfers en extra-curriculaire activiteiten gedurende de middelbare school.

Wanneer je een academische carrière ambieert moet je hier dus al rond je zestiende mee bezig zijn. Laatbloeiers, zij-instromers en twijfelaars wordt het moeilijk gemaakt. Dit systeem sorteert erop voor dat er te veel homogeniteit

is aan de universiteit, zowel qua type persoonlijkheden als qua achtergrond. Dit zie je het duidelijkst terug in de vertekende genderbalans. Los van de evidente wenselijkheid van diversiteit van ideeën, is het voor het draagvlak van het hoger onderwijs en de wetenschap van belang dat de maatschappij zich vertegenwoordigd ziet binnen de academie. Hoewel er qua diversiteit aan de universiteit voortgang geboekt wordt, gaat het te traag.

Het zogenaamde Tenure Track-systeem heeft zeker goede punten – het is een van de redenen dat ik vandaag hier kan staan – maar ook minder goede kanten. De lijst met Tenure Track-eisen staat vol met targets die behaald moeten worden en dan geldt natuurlijk de wet van Goodhart:<sup>59</sup> “*als een maatstaf een doel wordt, is het geen goede maatstaf meer*”. De succeskans bij beursaanvragen van NWO is erg laag. Dit komt omdat het aantal aanvragen het afgelopen decennium flink gestegen is terwijl de financiële middelen niet meebewogen. Die vele aanvragen komen niet uit een sterk gegroeide intrinsieke motivatie maar vooral omdat men beurzen moet binnenhalen van het systeem. Het niet binnenhalen van beurzen wordt als een mislukking gezien, terwijl het gezien de succeskans juist de *default* is dat dit niet lukt.<sup>3</sup> Bijkomend probleem is dat ook dit systeem de homogenisering in de hand werkt: iedereen moet exact dezelfde targets halen en moet dus dezelfde kwaliteiten in huis hebben. De targets zijn ook nog grotendeels solo-targets, samenwerking wordt onvoldoende beloond. Dit leidt tot een toxisch competitief systeem.

Een alternatief systeem, waarbij de medewerker punten kan verdienen met bijvoorbeeld publicaties en media-optredens, en vervolgens vanaf een bepaald aantal punten op kan gaan voor bevordering, zal al helpen die homogenisering tegen te gaan. Immers, de ene academicus zal punten verzamelen omdat zij succesvolle beursvoorstellen schrijft, de ander omdat hij een brug slaat tussen academie en maatschappij.

Het Tenure Track-systeem dwingt medewerkers om continu nét harder te bewegen dan wenselijk. Immers, de medewerker moet niet alleen aantonen dat

hij of zij goed werk levert, het moet ook in een korte tijdsperiode aangetoond. Deze dwang zorgt niet alleen voor onnodige werkstress onder de medewerkers, maar remt ook de creativiteit af. Zoals de Jedi-meester Yoda zei, “*The greatest teacher, failure is.*”<sup>37</sup> De ruimte om fouten te kunnen maken is essentieel voor succes, zowel voor studenten als voor medewerkers. Die ruimte moet je dan ook krijgen.

Zeer gewaardeerde toehoorders,

In het voorgaande heb ik u meegenomen in het statistische universum.<sup>12</sup> Haar ongrijpbare zekerheid<sup>11</sup> kan worden ingezet bij tal van maatschappelijke problemen. Dit vereist niet alleen goede wiskunde en goede methodologie, maar ook goede communicatie.

Ik heb u uitgebreid verteld over de statistische toepassing in de psychopathologie en omgevingspsychologie. Ik heb het voorrecht gehad om aan nog veel meer toepassingsgebieden te werken, zoals de effectiviteit van innovaties in het hoger onderwijs,<sup>17</sup> de normering van psychologische tests,<sup>63</sup> en de optimale inrichting van kantoorruimtes.<sup>34,10</sup> Door die toepassingen leer ik dagelijks bij over nieuwe onderwerpen buiten de statistiek. Ik sta vandaag voor u als hoogleraar maar ook als student en dat is een dubbele eer.

John Tukey zei ooit “*The best thing about being a statistician is that you get to play in everyone’s backyard.*” John Tukey sloeg de spijker op z’n kop.

*Ik heb gezegd.*

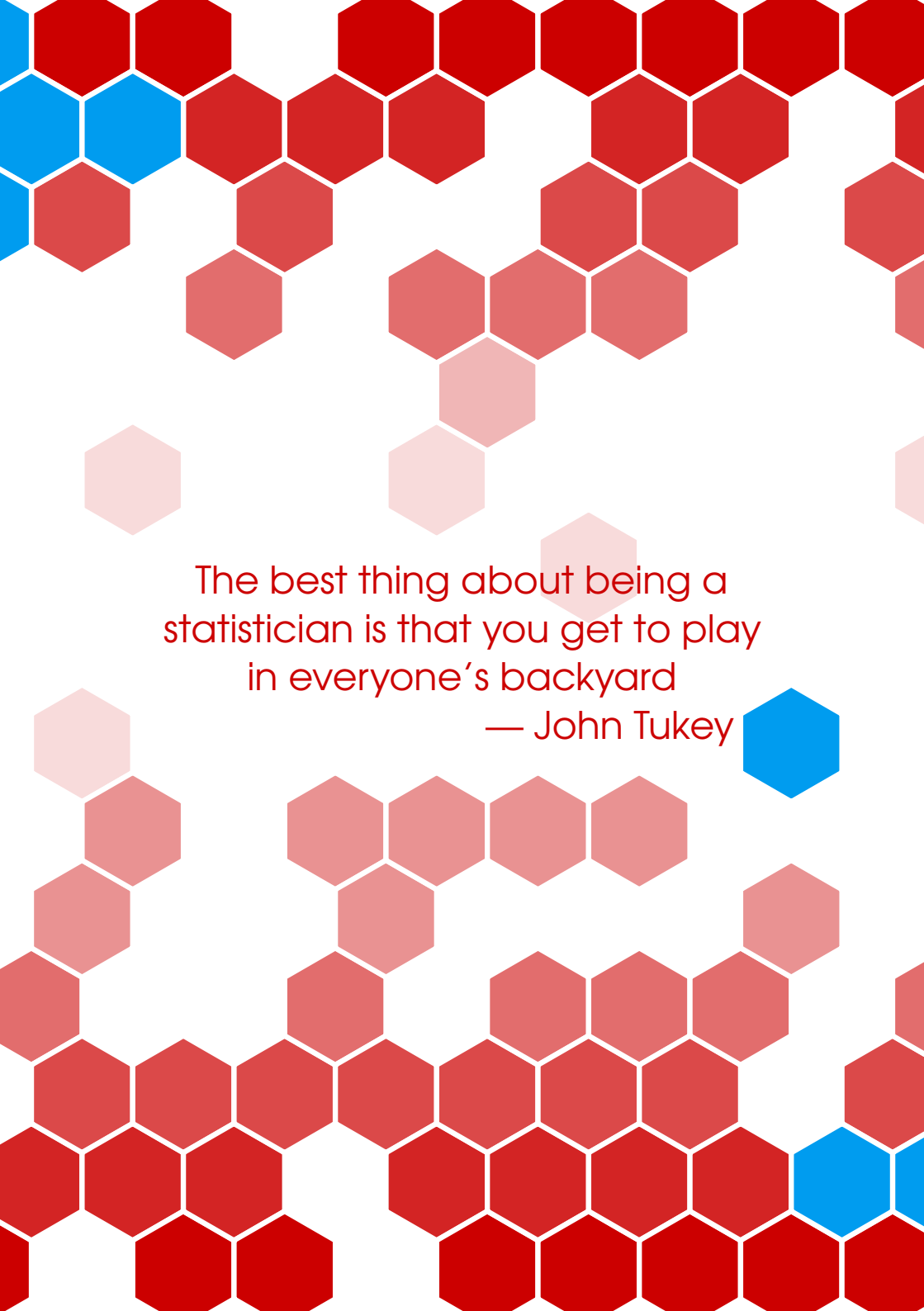
## Referenties

1. Albers, C. J. (2015). Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences*, 112(50), E6828–E6829.
2. Albers, C. J. (2018a). De Moivre–Gauss–Laplace: extraordinarily normal. *Nieuw Archief voor Wiskunde, Vijfde serie, deel 19*(1), 37–38.
3. Albers, C. J. (2018b). Mislukking. *UKrant Groningen*, 20 november.
4. Albers, C. J. (2018c). Valid reasons not to participate in open science practices. *Twitter*, 21 februari.
5. Albers, C. J. (2019a). Aan intuïtie heb je weinig bij het oplossen van statistische puzzels. *De Volkskrant*, 22 januari.
6. Albers, C. J. (2019b). Klimaatceptici geven antwoord op niet-relevante vragen. *De Volkskrant*, 20 februari.
7. Albers, C. J. & Bringmann, L. F. (2019). The time-varying change point autoregressive model. *Ingediend voor publicatie*.
8. Albers, C. J. & Gower, J. C. (2014). A contribution to the visualisation of three-way arrays. *Journal of Multivariate Analysis*, 132, 1–8.
9. Albers, C. J. & Gower, J. C. (2017). Visualising interactions in bi- and triadditive models for three-way tables. *Chemometrics and Intelligent Laboratory Systems*, 167, 238–247.
10. Albers, C. J., Vermue, C. E., Wolff, T. de & Beldhuis, H. J. A. (2018). Model-based academic dismissal policies; a case-study from the Netherlands. *PsyArXiv preprints*, osf.io/6a9cz.
11. Albers, W. (1985). De ongrijpbare zekerheid. Oratie. *Rijksuniversiteit Limburg*, 27 juni.
12. Albers, W. (2012). Het statistische universum. Afscheidsrede. *Universiteit Twente, Memorandum 1987*. 21 september.
13. Anacleto, O., Queen, C. M. & Albers, C. J. (2013a). Forecasting multivariate road traffic flows using Bayesian dynamic graphical models, splines and other traffic variables. *Australian and New Zealand Journal of Statistics*, 55(2), 69–86.
14. Anacleto, O., Queen, C. M. & Albers, C. J. (2013b). Multivariate forecasting of road traffic flows in the presence of heteroscedasticity and measurement errors. *Journal of the Royal Statistical Society, Series C: Applied Statistics*, 62(2), 251–270.
15. Bastiaansen, J. A., Kunkels, Y. K., . . . , Albers, C. J. & Bringmann, L. F. (2019). Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology. *Ingediend voor publicatie*.
16. Bhushan, N., Mohnert, F., Sloot, D., Jans, L., Albers, C. J. & Steg, E. M. (2019). Using a Gaussian graphical model to explore relationships between items and variables in environmental psychology research. *Ingediend voor publicatie*.
17. Boevé, A. J., Meijer, R. R., Bosker, R. J., Vugteveen, J., Hoekstra, R. & Albers, C. J. (2017). Implementing the flipped classroom: an exploration of study behaviour and student performance. *Higher Education*, 74(6), 1015–1032.

18. Bos, F. M., Snippe, E., Bruggeman, R., Wichers, M. E. & Krieken, L. van der. (2018). Will the experience sampling methodology deliver on its promise for psychiatric care? *Ingediend voor publicatie*.
19. Bouman, T., Verschoor, M., Steg, E. M., Böhm, G., Fisher, S., Poortinga, W., ... Albers, C. J. (2019). When worry about climate change leads to climate action and policy support. *Ingediend voor publicatie*.
20. Bringmann, L. F. & Albers, C. J. (2019). Inspecting gradual and abrupt changes in emotion dynamics with the time-varying change point autoregressive model. *Ingediend voor publicatie*.
21. Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D. & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, 22(3), 409–425.
22. Brown, N. J. L., Albers, C. J. & Ritchie, S. J. (2017). Contesting the evidence for limited human lifespan. *Nature*, 546, E6–E7.
23. Cleveland, W. S. & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), 531–554.
24. Cohen, J. (1962). The statistical power of abnormal-social psychological research: a review. *Journal of Abnormal Social Psychology*, 65, 145–153.
25. Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997–1003.
26. Cramer, A. O. J., Borkulo, C. van, Giltay, E. J., Maas, H. L. J. van der, Kendler, K. S., Scheffer, M. & Borsboom, D. (2016). Major depression as a complex dynamical system. *PLoS ONE*, 11(12), e0167490.
27. Ernst, A. F. & Albers, C. J. (2017). Regression assumptions in clinical psychology research practice — a systematic review of common misconceptions. *PeerJ*, 5, e3323.
28. Ernst, A. F., Timmerman, M. E., Jeronimus, B. F. & Albers, C. J. (2018). Probabilistic time series clustering by vector autoregressive metric – a filtering method for dynamic clustering. *Ingediend voor publicatie*.
29. Ernst, A. F., Timmerman, M. E., Jeronimus, B. F. & Albers, C. J. (2019). Inter-individual differences in multivariate time series: dynamic adaptive cluster modelling based on finite mixtures of vector-autoregressive processes. *Ingediend voor publicatie*.
30. Friendly, M. (2006). A Brief History of Data Visualization. In C. Chen, W. Härdle & A. Unwin (Red.), *Handbook of Computational Statistics: Data Visualization* (Deel 3). Heidelberg: Springer-Verlag.
31. Hamaker, E. L., Ceulemans, E., Grasman, R. & Tuerlinckx, F. (2015). Modeling affect dynamics: State of the art and future challenges. *Emotion Review*, 7(4), 316–322.
32. Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57, 357–384.
33. Herbranson, W. T. & Schroeder, J. (2010). Are birds smarter than mathematicians? Pigeons (*Columba livia*) perform optimally on a version of the Monty Hall dilemma. *Journal of Comparative Psychology*, 124, 1–13.
34. Hoendervanger, J. G., Been, I. de, Yperen, N. W. van, Mobach, M. P. & Albers, C. J.

- (2016). Flexibility in use: Switching behaviour and satisfaction in activity-based work environments. *Journal of Corporate Real Estate*, 18(1), 48–62.
35. IPCC. (2018). *Global warming of 1.5°C*. Geneva, Switzerland: IPCC.
36. John, L. K., Loewenstein, G. & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23, 524–532.
37. Johnson, R. (regie/scenario). (2017). *Star Wars: Episode VIII – The Last Jedi*. San Francisco: Lusasfilm.
38. Kahneman, D. (2011). *Thinking Fast and Slow*. New York: Farrar, Straus en Giroux.
39. Koval, P. & Kuppens, P. (2012). Changing emotion dynamics: Individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, 12(2), 256–267.
40. Krone, T., Albers, C. J., Kuppens, P. & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, 18, 739–754.
41. Krone, T., Albers, C. J. & Timmerman, M. E. (2015). A comparative simulation study of AR(1) estimators in short time series. *Quality & Quantity*, 1–21.
42. Krone, T., Albers, C. J. & Timmerman, M. E. (2016). Comparison of estimation procedures for multilevel AR(1) models. *Frontiers in Psychology: Quantitative Psychology and Measurement*, 7.
43. Lakens, D., Adolfs, F. G., Albers, C. J., ... & Zwaan, R. (2018). Justify your alpha. *Nature Human Behaviour*, 2, 168–171.
44. Leemput, I. van de, Wichers, M., Cramer, A. O. J., Borsboom, D., Tuerlinckx, F., Kuppens, P., ... Scheffer, M. (2014). Critical slowing down as early warning for the onset and termination of depression. *Proceedings of the National Academy of Sciences*, 111(1), 87–92.
45. Linden, S. L. van der. (2017). Determinants and measurement of climate change risk perception, worry, and concern. In *The Oxford encyclopedia of climate change communication*. Oxford, UK: Oxford University Press.
46. Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and review of the evidence*. University of Minnesota press.
47. Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald and the slow progress of soft psychology. *Journal of consulting and clinical psychology*, 46, 806–834.
48. Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports*, 66, 195–244.
49. Namazkhan, M., Albers, C. J. & Steg, E. M. (g.d.). The role of environmental values, socio-demographics and building characteristics in setting room temperatures in winter. *Energy*, 171, 1183–1192.
50. Open Science Foundation. (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.
51. Poortinga, W., Whitmarsh, L., Steg, E. M., Böhm, G. & Fisher, S. (2019). Climate change perceptions and their individual-level determinants: a cross-European analysis. *Global environmental change*, 55, 25–35.

52. Queen, C. M. & Albers, C. J. (2009). Intervention and causality: Forecasting traffic flows using a dynamic Bayesian network. *Journal of the American Statistical Association*, 104(486), 669–681.
53. R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
54. Sarran, C., Albers, C. J., Sachon, P. & Meesters, Y. (2017). Meteorological analysis of symptom data for people with seasonal affective disorder. *Psychiatry Research*, 257, 501–505.
55. Simkin, D. & Hastie, R. (1987). An information-processing analysis of graph perception. *Journal of the American Statistical Association*, 82, 454–465.
56. Slofstra, C., Nauta, M. H., Bringmann, L. F., Klein, N. S., Albers, C. J., Batalas, N., ... Bockting, C. L. H. (2018). Individual negative affective trajectories can be detected during different depressive relapse prevention strategies. *Psychotherapy and Psychosomatics*, 87, 243–245.
57. Smeets, I. (2014). *Het exacte verhaal. Wetenschapscommunicatie voor bèta's*. Amsterdam: Nieuwezijds.
58. Spence, I. (1990). Visual psychophysics of simple graphical elements. *Journal of Experimental Psychology: Human Perceptions and Performance*, 16, 683–692.
59. Strathern, M. (1997). Improving ratings: audits in the British university system. *European Review*, 5, 305–321.
60. Trimbos Instituut. (2016). *Depressie. Enkele cijfers. Factsheet, INFO16*.
61. Universiteitsbibliotheek Groningen. (2018). Open peer review. An interview with Casper Albers. *Open Science Newsletter*, 28 maart.
62. Verschoor, M., Albers, C. J., Poortinga, W., Böhm, G. & Steg, E. M. (2019). Attributes to climate change and energy: A network analysis of the European Social Survey. *Inge-dient voor publicatie*.
63. Voncken, L., Albers, C. J. & Timmerman, M. E. (2018). Improving confidence intervals for normed test scores: Include uncertainty due to sampling variability. *Behavior Research Methods*, geaccepteerd voor publicatie.
64. Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M. & Heydt, R. von der. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figureground organization. *Psychological bulletin*, 138, 1172–1217.
65. Werff, E. van der & Steg, E. M. (2015). One model to predict them all: Predicting energy behaviours with the norm activation model. *Energy Research & Social Science*, 6, 8–14.
66. Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.
67. Wilkinson, L. (2005). *The Grammar of graphics* (2de ed.). New York: Springer.
68. Willems, S. W., Albers, C. J. & Smeets, I. (2019). Variability in the interpretation of Dutch probability phrases – a risk for miscommunication. *arXiv preprints*, 1901.09686.
69. Williams, M. & Albers, C. J. (2018). Dealing with distributional assumptions in preregistered research. *PsyArXiv preprints*, osf.io/ab5tk.



The best thing about being a  
statistician is that you get to play  
in everyone's backyard

— John Tukey