



Universiteit  
Leiden  
The Netherlands

## Possibility days

Rooij, M.J. de

### Citation

Rooij, M. J. de. (2015). *Possibility days*. Leiden: Universiteit Leiden. Retrieved from <https://hdl.handle.net/1887/35766>

Version: Not Applicable (or Unknown)  
License: [Leiden University Non-exclusive license](#)  
Downloaded from: <https://hdl.handle.net/1887/35766>

**Note:** To cite this publication please use the final published version (if applicable).

Prof.dr. M.J. de Rooij

## Possibility days



Universiteit  
Leiden

Bij ons leer je de wereld kennen

# Possibility days

Oratie uitgesproken door

**Prof.dr. M.J. de Rooij**

bij de aanvaarding van het ambt van hoogleraar op het gebied van  
Methodologie en Statistiek van Psychologisch Onderzoek  
aan de Universiteit Leiden  
op vrijdag 9 oktober 2015.



**Universiteit  
Leiden**



*Mijnheer de Rector Magnificus, zeer gewaardeerde toehoorders,*

Om een docent, lector of meer in het algemeen een andere persoon, goed te begrijpen, is het van belang de persoon een beetje te kennen. Laat me daarom beginnen met een titelverklaring. De titel van deze oratie, "Possibility days", is tevens de titel van een nummer van mijn favoriete band Counting Crows op hun laatste album getiteld "Somewhere under wonderland". Voor diegenen die nog nooit van Counting Crows gehoord hebben: het is een folk-rock-band uit Berkeley, de Verenigde Staten. Veel van de teksten van de band zijn gebaseerd op depressies van de lead zanger - Adam Duritz. Ik kan me herinneren dat Arine en ik eens, het moet zo rond 1998 geweest zijn, kaartjes hadden voor een zogenaamde twee meter sessie in Paradiso met Counting Crows waar Duritz er zo slecht aan toe was dat hij het na enkele nummers voor gezien hield en pas na lang aandringen van de presentator, Jan Douwe Kroeske, nog even terugkwam om het nummer "A long december", al een beetje neerslachtig nummer van zichzelf, op een uiterst depressieve manier ten gehore te brengen. Ik denk dat het de kortste twee meters in de geschiedenis waren. Gelukkig ben ik ook bij concerten geweest van Counting Crows waar ze wel gewoon tot het eind bleven doorspelen.

Nu u weet wat mijn favoriete band is, kunt u misschien ook een inschatting maken van mijn persoonlijkheid, ik hou blijkbaar van wat depressieve nummers, her en der afgewisseld met olijker dingen. Ik hou blijkbaar van een mix van rock en folk. Verder kunt u uit de titel van dit nummer achterhalen dat ik misschien iets heb met "possibilities". De Engelse term possibility kan je op verschillende manieren vertalen: mogelijkheid, kans, gelegenheid. Dus possibility days zijn kansrijke dagen of dagen vol mogelijkheden.

Door deze relatief korte introductie hebt u veel kernbegrippen van mijn vakgebied, methodologie en statistiek van psychologisch onderzoek, al gehoord: ten eerste zijn we het begrip kans

tegengekomen; ten tweede depressie, een psychologische kwaal; en ten derde persoonlijkheid, een van de grondbegrippen uit de psychologie. In deze rede zal ik mijn vakgebied met u bespreken. *Methodologie en Statistiek van Psychologisch Onderzoek* is een hele mond vol. Een vriend van mij heeft er in het voorjaar een heel weekend over gedaan om de titel van mijn leerstoel vloeiend uit te spreken. Wat behelst mijn leeropdracht? In simpele termen gaat mijn leeropdracht over de manier waarop psychologisch onderzoek wordt uitgevoerd, of uitgevoerd zou moeten worden. Mijn vakgebied kijkt dus naar het proces van wetenschap binnen de psychologie: van onderzoeksidee tot publicatie. Psychologie is een empirische wetenschap, wat wil zeggen dat we gegevens verzamelen en deze gegevens gebruiken om tot conclusies te komen. In dit proces gebruiken we statistische methoden om de gegevens te analyseren en op basis daarvan uitspraken te doen. Een belangrijk element van deze uitspraken is dat we ze met een bepaalde zekerheid doen. Geen enkele conclusie in de psychologie is definitief noch universeel. In de bespreking van mijn vakgebied zal ik verschillende elementen aanwijzen waar mogelijkheden zijn tot verbetering van huidige praktijken. Ik ben van mening dat we elke dag moeten nadenken of dingen beter kunnen en als dat het geval is dit zo spoedig mogelijk aan te pakken. Wat dat betreft is iedere dag een dag vol mogelijkheden, oftewel iedere dag is een possibility day.

In mijn rede zal ik veelal gebruik maken van het afstandsbe-grip. Afstanden zijn heel intuïtief. Om dat te illustreren wil ik een klein ( $N=1$ ) proefje doen. Mam, waar ben je? Kan je even gaan staan? Gerrie waar ben jij? Kan je ook even gaan staan? Bas, waar ben jij? Wil je ook even gaan staan? Bas, jij bent de proefpersoon voor dit moment. Kan je me zeggen welke oma het dichtst bij je staat? ... Goed zo. Kan je me ook vertellen welke oma het dichtst bij woont? ... Wederom goed. Dank jullie wel, jullie mogen weer gaan zitten.

Zoals u zojuist gezien hebt, heeft een kind van 9 jaar, ik denk de jongste onder de toehoorders vandaag, al een goed begrip

van afstanden: hij weet wat dichterbij is, en wat verder weg is. Afstand is dus een begrip dat we al op jonge leeftijd leren kennen en dat we makkelijk kunnen begrijpen. Ik zou Bas ook kunnen vragen over correlatie, of over samenhang, dat zou hij veel moeilijker vinden. Afstanden hebben ook implicaties, zo komt de oma die dichtbij woont vaker eventjes langs. Afstand is wat dit betreft dus gekoppeld aan frequentie. Daarnaast is de reistijd naar de andere oma langer, wat voor Bas betekent dat hij langer kan gamen achter in de auto. Afstanden worden in de psychologie vaak gekoppeld aan gelijkenis: twee objecten die sterk op elkaar lijken, hebben een kleine, psychologische afstand.

### Meten

Een van de grootste invloeden van mijn vakgebied op het dagelijks leven van veel mensen is die van de psychologische- of onderwijstest. Psychologische tests zijn meetinstrumenten, ze meten een bepaalde eigenschap, een construct, zoals intelligentie, rekenvaardigheid of emotionele stabiliteit. U denkt misschien dat u niet veel te maken heeft gehad met dergelijke metingen maar dat is een misverstand. Bijna alle kinderen, en dus ook hun ouders, zijn bekend met de CITO-toets, een onderwijskundige test. Ook zijn er veel mensen bekend met personeelsselectie toetsen of studiekeuzetests. De psychometrie heeft richtlijnen opgesteld waar tests aan moeten voldoen. Zo moet een test betrouwbaar zijn, dat wil zeggen dat wanneer we herhaald meten we grofweg dezelfde uitkomst krijgen. De test moet valide zijn, dat wil zeggen dat de test meet wat we ermee beogen te meten. En een test moet precies zijn. Stel je voor dat we lichaamslengte alleen kunnen meten in hele meters (dus niet erg precies), dan zouden bijna alle mannen en vrouwen even lang zijn, namelijk twee meter.

Voor ik verder zal uitwijden over mogelijkheden voor psychologische tests wil ik iets vertellen over het grootschalig toepassen van psychologische tests in de maatschappij.<sup>1</sup> Aan het begin van de 20<sup>e</sup> eeuw maakten de Fransen Binet en Simon wat tegenwoordig gezien wordt als de eerste intelligentietest.

Goddard, een Amerikaans onderzoeker (expert) op het gebied van zwakzinnigheid was er al snel van overtuigd dat een gestandaardiseerde methode, een test, ideaal zou zijn om iets te zeggen over zwakzinnigheid. Toen hij las over deze Binet test, importeerde hij deze naar de Verenigde Staten en vertaalde hem in het Amerikaans. Goddard was overtuigd van het nut van de Binet test en probeerde de test te valoriseren, zoals we dat tegenwoordig noemen. Valoriseren staat voor: wetenschappelijke kennis tot maatschappelijke waarde brengen. Goddard was uitermate succesvol in zijn acties en werd in 1910 verzocht naar Ellis Island te komen om daar de immigranten psychologisch te testen.

Goddard had de Franstalige Binet test in het Amerikaans vertaald. Deze Amerikaanse test werd door een tolk, dus zonder standaardisatie of validatie, afgenomen vlak na aankomst van de immigranten op Ellis Island. Ik weet niet precies hoe lang een dergelijke tocht over de oceaan duurde, maar ik kan me voorstellen dat de net aangekomen immigranten niet geheel fris waren. De norm scores van de Franse test werden zonder verder onderzoek direct overgenomen. Goddard is uiteindelijk het meest bekend geworden door zijn artikel in 1917 waarin hij schrijft dat 80% van de immigranten zwakzinnig is. Ondanks dat hij in dit artikel nadrukkelijk aangeeft dat dit getal op basis van geselecteerde proefpersonen tot stand is gekomen en dus niet in het algemeen generaliseerbaar is, zijn Goddards resultaten door de politiek wel in die zin overgenomen.

Volgens Gelb<sup>2</sup> had Goddard nooit de intentie iets slechts te zeggen over immigranten; het enige wat Goddard wilde laten zien, was dat deze intelligentietest goed bruikbaar was in de praktijk. In een sociaal culturele context van racisme en achterdocht werden zijn papers gevormd en wordt Goddard tegenwoordig gezien als een racist. Wat mijns inziens belangrijk is aan deze anekdote is dat Goddard goede bedoelingen had met de Binet test. Hij wilde een wetenschappelijke methode toepassen op een maatschappelijk probleem. Nu echter, ruim honderd jaar later, weten we dat testen gestandaardiseerd moeten worden,

dat wanneer ze vertaald zijn ze opnieuw genormeerd moeten worden. Verder weten we dat het moment van testen een grote invloed kan hebben op de uitkomst. Het direct valoriseren van net opgedane kennis kan dus zeer negatief uitvallen.

Veel psychologische en onderwijskundige tests zijn zogenaamde dominantietesten, waarin een “wedstrijdje” wordt gespeeld tussen de opgave en de ondervraagde. Stel, we willen de *rekenvaardigheid* van Joris, de oudere broer van Bas, meten. Daarvoor maken we een stel rekenopgaven, van makkelijk naar moeilijk. Eén van die opgaven is bijvoorbeeld: hoeveel is 12 keer 14. Als de vaardigheid van Joris groter is dan de benodigde vaardigheid voor de opgave dan zal de opgave waarschijnlijk goed gemaakt worden; als de gevraagde vaardigheid echter groter is dan de vaardigheid van Joris dan zal de opgave waarschijnlijk fout gemaakt worden. We hebben dus dat of de opgave het kind domineert of het kind domineert de opgave. Het is echter niet zo dat wanneer de vaardigheid van Joris groter is, Joris de vraag per definitie correct beantwoordt, toevalligheden spelen hier ook een rol. Wanneer we de vaardigheid van het kind willen vaststellen, zit er daarom een kanselement in het model dat we gebruiken.

Niet alle menselijke kenmerken kunnen met dominantietests worden gemeten, al proberen we dat wel. Dominantietesten zijn minder goed bruikbaar voor het meten van denkwijzen, attitudes of voor het meten van persoonlijkheid. Deze kenmerken zijn van belang voor bijvoorbeeld personeelsselectie en studiekeuze. Hier kan men niet spreken van dominantie, er is geen goed of fout antwoord, het ene is niet per se beter dan het ander. Daarvoor zijn er andere meetmodellen nodig, de zogenaamde enkeltoppige meetmodellen waarin afstand een cruciale rol speelt. Deze meetmodellen gaan uit van een positie van de ondervraagde op de eigenschap die we willen meten en als de positie van de vraag daarbij in de buurt ligt dan zegt de ondervraagde waarschijnlijk ja, als de afstand groot is dan zegt de ondervraagde waarschijnlijk nee. Wanneer we bijvoorbeeld *properheid* willen meten dan kan dat met een stel vragen. Neem

het item “ik vind dat een bed dat dagelijks beslapen wordt eens in de twee weken verschoond moet worden”. Iemand die heel proper is kan het hier mee oneens zijn omdat hij vindt dat een bed toch wel minimaal elke week (of misschien elke dag) verschoond moet worden; iemand anders die niet zo proper is kan het ermee oneens zijn omdat zij vindt dat één keer per maand ook wel genoeg is. Men kan het dus om twee verschillende redenen oneens zijn met deze stelling; alleen als je in de buurt ligt van eens per twee weken dan zal je het eens zijn met de stelling. De mensen die het eens zijn, vormen in die zin een homogene groep, de mensen die het oneens zijn echter niet.

In de geschiedenis van de psychometrie hebben dominantiemodellen veruit het meest aandacht gekregen. Voor dominantietesten zijn veel methoden ontwikkeld, zowel binnen de klassieke testtheorie als de moderne. Methoden voor tests gebaseerd op afstanden zijn echter veel lastiger; zo heb je niks aan een somscore of correlaties tussen item en totaalscores. Dit heeft er toe geleid dat er trucs zijn bedacht om kenmerken die eigenlijk niet vanuit een dominantie perspectief gemeten kunnen worden toch in dit raamwerk te plaatsen, denk bijvoorbeeld aan het meten van persoonlijkheid. De bekendste van deze trucs is het maken van een Likert schaal. Voor het vormen van een Likert schaal stellen we alleen vragen die extreem zijn voor het denkbeeld dat we willen meten. Door alleen de extremen te bevragen kunnen we methoden gebruiken die voor dominantie vragen gemaakt zijn.

De Likert schaal wordt bijvoorbeeld gebruikt voor het meten van de vijf grote persoonlijkheidsschalen of dimensies: extraversie, mildheid, ordelijkheid, emotionele stabiliteit en autonomie. Deze “Big-five” theorie wordt veel gebruikt in psychologisch onderzoek en er zijn instrumenten die metingen geven van ieder van de vijf dimensies. Echter, pogingen om een bevestigende analyse te doen van deze vijf dimensies resulteren in een afwijzing van het model.<sup>3</sup> Dat wil zeggen dat de theorie niet bevestigd kan worden met empirische data. Voor de bevestiging worden echter modellen gebruikt uit de dominantie

hoek, terwijl het voor de hand zou liggen modellen te gebruiken die op het afstandsprincipe gebaseerd zijn.<sup>4</sup> Er zijn echter op dit moment geen meerdimensionale bevestigende meetmodellen voor afstandsdata.

Het ontwikkelen van methodologie gebaseerd op het afstandsprincipe zal leiden tot een grotere precisie in de metingen. De huidige methode, de Likert schaal, geeft voornamelijk informatie over personen in het midden van de schaal, terwijl een model gebaseerd op afstanden ook goede informatie kan geven over personen aan de uiteinden. Dit is van belang voor bijvoorbeeld het diagnosticeren of in het kader van selectie.

Het verder ontwikkelen van meettheorie op basis van een afstandsmodel is dus nodig om betere metingen te krijgen. Hier is nog veel werk te doen, hier zijn nog elke dag mogelijkheden.

6 Naast psychologische tests maken we in psychologisch onderzoek steeds vaker gebruik van nieuwe meetmethoden. Zo kunnen we met MRI scanners tegenwoordig een kijkje in de hersenen nemen en met zogenaamde eye-trackers kijken we naar oogbewegingen tijdens bijvoorbeeld het lezen van een tekst of het maken van een rekenopgave. De psychometrische eigenschappen van dit soort metingen verdienen nog meer onderzoek. Hoe betrouwbaar zijn ze? Hoe precies zijn ze? Dit zijn vragen waarop het antwoord nog niet geheel is uitgekristalliseerd.

### **Modelleren**

Wanneer we eenmaal onze constructen/variabelen gemeten hebben, willen we daar ook wat mee doen. Het meten staat nooit op zich. Veelal gaan we op zoek naar relaties van de variabelen met andere variabelen. In hoeverre hangt emotionele stabiliteit samen met depressie of andere psychologische kwalen? Kunnen we op basis van extravertie verklaren waarom het ene kind meer vriendjes heeft dan het andere? Kunnen we op basis van ontwikkelingspatronen voorspellen welke kinderen de school vroegtijdig verlaten?

Traditioneel bestaan er twee raamwerken waarin we naar dergelijke relaties kunnen kijken: exploratief en confirmatief. In exploratief of verkennend onderzoek hebben we een aantal tests en eventueel andere metingen afgenomen bij een groep mensen. We zijn dan geïnteresseerd in de vraag in hoeverre er samenhang is tussen de verschillende testcores. Van tevoren hebben we geen sterke ideeën over de samenhang, maar willen de data in die zin laten spreken. Uit dit soort onderzoek komen de meeste nieuwe ideeën voort. De exploratieve analyse levert de meeste verwondering op.<sup>5</sup> In exploratieve analyse worden zaken gevonden die we niet al van tevoren zagen aankomen.

In confirmatief, bevestigend of bekrachtigend, onderzoek stelt de onderzoeker van tevoren hypotheses vast om een theorie te toetsen. Dit gebeurt veelal in experimenten. Hierin wordt een groepje mensen over twee condities verdeeld. In de eerste conditie worden de mensen aan een behandeling blootgesteld, in de tweede conditie, de zogenaamde controle conditie, niet. Na de behandeling wordt het effect gemeten op een uitkomstvariabele. De mensen die deelnemen in de studie zijn de zogenaamde proefpersonen. Deze proefpersonen zijn natuurlijk verschillend, zoals alle mensen verschillend zijn. Als je pech hebt, zitten alle proefpersonen die hoog scoren op de uitkomstvariabele in de ene conditie en zij die laag scoren in de andere. Door ze willekeurig (“at random”) in de twee condities te plaatsen hopen we, ja hopen we, dat de individuele verschillen wegmiddelen en dat we een goede schatting krijgen van het behandelingseffect.

Wat er vervolgens standaard in de statistiek gedaan wordt, is ervan uitgaan dat de behandeling geen enkel effect heeft. Zelfs wanneer de behandeling geen effect heeft, vinden we toch een gemiddeld verschil in de uitkomstvariabele tussen de twee condities omdat mensen nu eenmaal verschillend zijn. Doordat we de proefpersonen willekeurig ingedeeld hebben in de twee condities kunnen we de kans bepalen dat het gevonden verschil in de uitkomstvariabele, gegeven dat er geen behandelings-effect is, minimaal zo groot is als het verschil gevonden in de data. Dit

noemen we de p-waarde: de kans dat het effect op toeval berust. Als deze kans klein is, concluderen we dat het dan wel niet de toevallige groepsindeling geweest is die voor het effect zorgt maar juist de behandeling. Dit weten we echter niet zeker: er is een bepaalde mate van onzekerheid. In de statistische procedure zitten lacunes die ertoe kunnen leiden dat we de verkeerde conclusie trekken. We moeten daarom voorzichtig zijn met het trekken van al te stevige conclusies. Dit is ook de reden waarom er al jaren geroepen wordt dat studies gerepliceerd moeten worden. Pas wanneer een studie een groot aantal keren gerepliceerd wordt, kunnen we gaan spreken van kennis. Met één replicatie zijn we er niet.<sup>6</sup> Vooralsnog zijn replicatiestudies zeldzaam en hebben we dus nog maar weinig kennis.

Eén van de grote problemen van huidig psychologisch onderzoek is dat voor exploratieve studies de technieken gebruikt worden van confirmatief onderzoek.<sup>7</sup> Onderzoekers zoeken dus relaties en doen daarna alsof ze daar al hypothesen over hadden. Wanneer het onderscheid niet kritisch bewaakt wordt, wordt al het onderzoek minder waard!

### Recht toe recht aan

In de psychologie, maar meer in het algemeen de sociale wetenschappen, ligt de focus op lineaire relaties, dat wil zeggen in de vorm  $Y = a + bX + cZ$ ; als  $X$  met één omhoog gaat dan gaat  $Y$  met  $b$  omhoog. Lineaire relaties hebben een aantal voordelen, ze zijn namelijk makkelijk te interpreteren en ze zijn makkelijk te berekenen. Dit laatste voordeel was voornamelijk van belang voordat we de beschikking hadden over computers. Het enige echte voordeel is dus de makkelijke interpretatie. Wanneer we echter in de wereld om ons heen kijken zijn er nauwelijks lineaire relaties. Wij mensen groeien vanaf de geboorte eerst snel en daarna steeds langzamer totdat we volwassen zijn en dan veel later krimpen we weer iets. Eenzelfde groeiproces zie je ook wanneer kinderen leren lezen of rekenen. Bepaalde dieren leven op plekken waar er voor hen genoeg water is, dat wil zeggen niet te weinig, maar ook weer niet teveel. Wederom

een afstand tussen een ideale situatie voor dit dier en de werkelijke situatie die bepaalt of de soort daar overleeft. Kijken we in mijn favoriete boekje van de middelbare school, BINAS, dan vinden we daar veel wetten. Geen van de wetten in dit boekje heeft de vorm  $Y = a + bX + cZ$ .

Tagepeera<sup>8</sup> onderzocht hoe sociale wetenschappers hun data analyseren. Daartoe trok hij 75 willekeurige getallen uit het telefoonboek tussen 2 en 99 en ordende ze in 3 kolommen. In de sociale wetenschappen zeggen we dan dat we voor 25 proefpersonen gegevens hebben op 3 variabelen. Met de drie getallen in iedere rij berekende hij een vierde getal op basis van Newton's zwaartekrachtwet. Deze wet zegt dat de aantrekkingskracht tussen twee objecten proportioneel is aan het product van de massa's gedeeld door de gekwadrateerde afstand:

$$Y = m_1 \times m_2 / d^2,$$

niet lineair dus.

In iedere rij berekende hij dus het product van de eerste twee en deelde dat vervolgens door het kwadraat van de derde. Dit vierde getal werd afgerond en deze afronding is de enige vorm van zogenaamde "error". Een onderzoeker zou dus tot een bijna perfect passend model moeten kunnen komen. Dit data setje gaf hij aan een stel sociale wetenschappers, waarbij hij vertelde welke variabele de afhankelijke was (de vierde) en welke de verklarende variabelen (de eerste drie). Geen van de onderzoekers vond het werkelijke, ware, model. Vele pasten een lineair regressie model en waren tevreden met de uitkomst met 70 procent verklaarde variantie. Slechts één onderzoeker nam eerst de logaritme van de afhankelijke variabele en gebruikte deze in de lineaire regressie waardoor zij dichter bij het ware model kwam. Tagepeera concludeert dat ondanks enkele tekortkomingen in de opzet van zijn studie, er aanwijzingen zijn dat sociale wetenschappers snel tevreden zijn, alleen naar lineaire relaties kijken en niet op zoek zijn naar het ware model dat ten grondslag ligt aan de data.

In de psychologie ligt volgens mij teveel de nadruk op lineaire relaties waardoor we belangrijke patronen missen. Wanneer er bijvoorbeeld een enkeltoppige relatie is tussen de persoonlijkheidsdimensie mildheid en het krijgen van een depressie, zullen we die nooit en te nimmer vinden met een lineair model. Veel relaties blijven dan dus voor ons verborgen en de ware aard zullen we niet vinden.

### **Verklaren = Voorspellen**

Naast een preoccupatie met lineaire modellen ligt de nadruk in de psychologie ook op verklaren, op dingen begrijpen. Wetenschap heeft echter twee doelen.<sup>9</sup> Het eerste doel is het verklaren van gebeurtenissen; het tweede doel is om voorspellingen te maken. In de psychologische wetenschap ligt de nadruk zoals gezegd op verklaren. Daarvoor worden theorieën ontwikkeld die eigenschappen aan elkaar koppelen en relaties blootleggen. Om kennis echter in de praktijk toe te passen, moeten we voorspellen. Ik zal dit illustreren met drie voorbeelden.

Wanneer een selectiepsycholoog moet bepalen of zij meneer Bakker of mevrouw De Boer aanneemt voor een functie bij bedrijf X zal zij een inschatting moeten maken hoe goed beide kandidaten het gaan doen in het bedrijf. Daarvoor moet zij als het ware in de toekomst kijken oftewel een voorspelling maken. Om die voorspelling te maken, kan zij gebruik maken van de kenmerken van meneer Bakker: man, 34 jaar, gehuwd, geen kinderen, IQ score van 98, enzovoort en de kenmerken van mevrouw De Boer: vrouw, 41 jaar, gescheiden, 3 kinderen, IQ score 112, enzovoort. Op basis van deze gegevens en een statistisch voorspellingsmodel waarin ook de kenmerken van het bedrijf zijn opgenomen kan nu voorspeld worden welke kandidaat het het best gaat doen. Niet alleen deze voorspelling is van belang om uiteindelijk een beslissing te nemen, ook de accuraatheid en de onzekerheid van de voorspelling zijn van belang.

Hetzelfde geldt voor het selecteren van studenten voor een opleiding. Op basis van een aantal tests willen we een voorspelling doen over het aantal studiepunten dat de potentiële

student in het eerste jaar gaat halen. Deze voorspelling moet getoetst worden aan observaties. Blijkt bijvoorbeeld dat we een voorspelling maken dat de student 45 studiepunten gaat halen maar blijkt ook dat we er met onze voorspelling gemiddeld 15 punten naast zitten, dan is die foutenmarge te groot om een betrouwbare beslissing te nemen. In dat geval is het beter om te loten dan te selecteren.

Nog een voorbeeld: Wanneer een psycholoog een vrouw met depressieve klachten in zijn praktijk krijgt, moet er na de intake besloten worden of therapie A of therapie B gegeven zal worden. De vrouw is 45 jaar, gehuwd, heeft volwassen kinderen, geen werk, en heeft een ordelijkheidsscore van 24 en een autonomiescore van 12, etc. Om een beslissing te nemen welke therapie er gegeven zal worden, moet er een voorspelling gemaakt worden van de status van deze cliënt over zeg 3 maanden wanneer zij therapie A gekregen heeft en wanneer zij therapie B gekregen heeft. De beslissing voor één van de twee therapieën moet komen uit de voorspelling welke therapie het beste resultaat heeft.

Deze drie voorbeelden laten zien dat we voor het maken van een beslissing een voorspelling nodig hebben. In de huidige praktijk doet de beslisser (de selectiepsycholoog of klinische psycholoog) dat impliciet op basis van theorieën. Een expert maakt dus voorspellingen op basis van een verklarend model. Meehl<sup>10</sup> liet zien dat statistische modellen over het algemeen betere voorspellingen geven dan experts. Daarnaast zijn verklaren en voorspellen niet hetzelfde.<sup>11</sup> Ik zal trachten dit duidelijk te maken aan de hand van twee voorbeelden.

In een steriele ruimte waar geen wind is of andere onbekende invloeden zijn, kunnen we het traject van een tennisbal uit een kanon precies berekenen, gebruik makend van de wetten van Newton. Omdat we dit precies kunnen berekenen, kunnen we onder deze omstandigheden ook een perfecte voorspelling maken van het traject van de tennisbal en het punt waar de tennisbal neerkomt. Gaan we nu echter naar buiten, waar de

weersomstandigheden variabel zijn, dan is deze gelijkenis verdwenen. We kunnen, als we eenmaal de tennisbal afgeschoten hebben en de bal is geland, het hele traject wel verklaren op basis van het standaard traject en de invloed van bijvoorbeeld wind maar het is moeilijk om het precieze traject te voorspellen, aangezien we niet weten of die windvlaag nu wel of niet komt. Dit levert onvoorspelbare variatie op. In de steriele ruimte hebben we een zogeheten deterministisch systeem, buiten hebben we een probabilistisch systeem waar kansen een rol spelen. In een deterministisch systeem zijn voorspellen en verklaren aan elkaar gelijk. Zodra we echter naar een kanssysteem gaan, verdwijnt deze gelijkheid.

Het verschil tussen verklaren en voorspellen zien we ook elke avond in het weerbericht. Het eerste deel van het weerbericht kijkt terug op het weer van de afgelopen dag en geeft een verklaring van de weersomstandigheden; deze verklaring is perfect, weermannen en -vrouwen kunnen achteraf goed verklaren waarom het weer was zoals het was. Het tweede deel van het weerbericht kijkt vooruit, het geeft een voorspelling van het weer voor morgen en de dagen daarna. Hoe vaak worden we niet geconfronteerd met foute voorspellingen?

Psychologisch onderzoek en de methoden die we daarvoor gebruiken, richten zich voornamelijk op verklaren en begrijpen. In de artikelen die daarover geschreven worden, staat soms de term predictie; dit is echter een postdictie, namelijk een voorspelling van datgene wat we al weten. Of beter gezegd, het statistische model maakt in dat geval een voorspelling voor gegevens waarop het statistisch model gebaseerd is. We zijn dan achterafwaarzeggers.

Met de theorieën die we vormen op basis van dit soort verklarende modellen wordt de praktijk van de psycholoog gevormd. De theorieën liggen ten grondslag aan de beslissingen die de psycholoog neemt. Een probleem is dat we echter nooit gemeenten hebben hoe goed dergelijke voorspellingen zijn. De beslissingen die we nemen, zijn dus niet gevalideerd.

In principe kunnen voorspellende en verklarende statistische modellen dezelfde generieke vorm hebben. Wat is dan het cruciale verschil? Het is de manier waarop we tot een optimaal model komen en op welke aspecten van het statistische model we de nadruk leggen. Voor verklaren is het van belang dat we zogenaamde zuivere schatters hebben. Dat zijn schatters die het gemiddeld goed doen. Bedreigingen voor de zuiverheid worden gevormd door vergeten variabelen, door missende waarden, hoge samenhang tussen verklarende variabelen en meetfouten in de variabelen. In verklarende modellen is het daarnaast van belang de goede paden te vinden, A beïnvloedt B, waarna B invloed heeft op C. In de verklaringen zijn we vaak op zoek naar dergelijke mediators. Voor voorspellingsmodellen doet dit er allemaal veel minder toe. We zijn niet direct geïnteresseerd in de precieze effecten van variabelen, hooguit in de vraag welke variabelen bijdragen aan een goede voorspelling. Het enige wat echt telt, is hoe goed de voorspellingen zijn, al het andere is van secundair belang.

De calibratie van een voorspellingsmodel moet gebruik maken van twee onafhankelijke data sets: een training set en een test set. Op de eerste data set, de training set, vinden we een optimaal model welke we op de tweede data set testen. In de test kijken we hoe goed de voorspellingen daadwerkelijk zijn. Wanneer we de voorspellingen voldoende accuraat vinden, kunnen we ermee naar de buitenwereld. Wanneer we de voorspellingen nog niet voldoende accuraat vinden, moet de aanbeveling zijn dat we op basis van de huidige gegevens nog geen goede beslissingen kunnen nemen.

Hier dient zich een ander belangrijk verschil aan tussen verklaren en voorspellen. Voor verklarende modellen maken we vaak gebruik van een beperkt domein. Zo kunnen we wanneer we depressie willen verklaren bijvoorbeeld kijken naar persoonlijkheidsvariabelen. Dit levert een theorie op over de relatie tussen persoonlijkheid en depressie. In een andere theorie kijken we naar de relatie tussen depressie en grote gebeurtenissen in het leven zoals het krijgen van kinderen, scheiden, het

verliezen van een naaste, een huwelijk. Dit levert een tweede theorie op. Een derde onderzoeker kijkt misschien naar gen expressies en depressie en de vierde naar de werking van de hersenen in relatie met depressie. Zo vormen zich veel theorieën over het ontstaan van een depressie die ieder op zich redelijk verklaren. Wanneer we echter kijken naar het voorspellende vermogen van ieder van de theorieën is die waarschijnlijk niet hoog. Waarschijnlijk geeft een van de domeinen de beste voorspelling, er is altijd een beste. Dit wil echter niet zeggen dat we met dat enkele domein een goede voorspelling kunnen geven. Om een goede voorspelling te krijgen of iemand een depressie krijgt, hebben we de persoonlijkheidsvariabelen, de levensgebeurtenissen, DNA gegevens, hersenscans en misschien nog wel veel meer informatie nodig. Sommige van deze domeinen zullen bijdragen aan de voorspelling, andere niet.

### **Psychologie voor voorspellende statistiek**

Voor het maken van voorspellende statistische modellen kunnen we gebruik maken van psychologische kennis. Hier worden statistische modellen dus gevoed met ideeën uit de psychologie. Een voorbeeld daarvan is hoe we als mens dingen herkennen. Wanneer wij over straat lopen, krijgen we continue visuele input. We kijken om ons heen. Als er iets over ons hoofd vliegt, kijken we omhoog en kunnen direct bepalen of het een vogel, een vliegtuig, een drone, Super Grover, of iets anders is. Dit proces noemen we categoriseren, en mensen zijn daar heel goed in. Er is veel literatuur in de psychologie over hoe we dit categoriseren eigenlijk doen. De meeste theorieën maken gebruik van gelijkenissen of ongelijkenissen, dat is van psychologische afstanden. Het object dat we zien, vergelijken we met objecten in ons hoofd, zogenaamde exemplaren of prototypes. De kans dat we het object in een bepaalde categorie plaatsen is direct gerelateerd aan de psychologische afstand tussen het geziene object en het exemplaar van die categorie: hoe kleiner de afstand des te groter de kans. Deze theorie is goed bruikbaar om statistische modellen voor voorspellen te maken. Standaard gebruiken we een aantal variabelen om te voorspellen. We kunnen echter ook eerst de afstanden tussen

proefpersonen berekenen op basis van deze variabelen en vervolgens deze afstanden gebruiken om te voorspellen. Dit levert zeer flexibele voorspellingsmodellen op.

Het tweede element dat we kunnen gebruiken voor voorspellingsmodellen is hoe we als mens beslissingen nemen. Wanneer wij thuis een vakantie boeken, bekijken we verschillende websites, we vergelijken de opties en we kijken hoe andere mensen geoordeeld hebben over de opties. Al die informatie wegen we op een bepaalde manier en zo komen we tot een conclusie. Ook op andere gebieden maken mensen gebruik van verschillende informatiebronnen: een second en third opinion van andere doktoren voordat een behandeling gestart wordt, meerdere referenties van een sollicitant voordat hij of zij wordt aangenomen.

Wanneer we echter in psychologisch onderzoek gegevens hebben verzameld, analyseren we die in het algemeen met maar één expert, bijvoorbeeld een lineair regressiemodel of een regressie boom of een neuraal netwerk model. De keuze voor welk model we gebruiken, wordt vaak impliciet gemaakt zonder kenmerken van de data mee te nemen in de beslissing. De mening van deze ene expert nemen we daarna als een voldongen feit. Zou het niet veel beter zijn om meerdere modellen op dezelfde dataset los te laten en dan uiteindelijk een gewogen gemiddelde van de conclusies te bepalen?

### **Psychologie**

Zoals ik aan het begin van mijn rede heb gezegd, bestudeert de methodologie en statistiek het proces van wetenschap binnen de psychologie. Zonder methodologie en statistiek zou psychologie een wetenschapsgebied zijn van gevalsbeschrijvingen. Dat levert wel lekker leesvoer maar uiteindelijk geen kennis op. Methodologie en statistiek is dus de ruggengraat van de psychologie als wetenschap. Zoals onze oud rector Wagenaar ooit zei: “Een grote sprong vooruit in de psychologie wordt vaak veroorzaakt door een verbetering van de methode en analyse”. Tegelijkertijd werkt het ook andersom: door samenwerkings-

projecten met onderzoekers in de psychologie worden nieuwe methodologische vragen geboren.

We hebben dus een duale rol: methodologie en statistiek is een wetenschapsgebied met een eigen onderzoekslijn, daarnaast bieden we ondersteuning aan onderzoekers binnen de psychologie op methodologisch en statistisch terrein. Dit laatste element, de consultatie, zou ik graag verder willen uitbouwen in de toekomst zowel binnen het Instituut Psychologie als binnen de Graduate School van de Faculteit Sociale Wetenschappen. Tegelijkertijd moet ik ervoor waken dat dit niet ten koste gaat van de eigen onderzoekslijnen.

### Onderwijs

Tot op dit moment heb ik voornamelijk onderzoek met u besproken. Het grootste deel van de tijd en energie van de wetenschappelijke staf in een universiteit gaat echter uit naar onderwijs. De Universiteit Leiden legt een directe link tussen onderwijs en onderzoek in haar beleidsplannen. Ik ben ook van mening dat nieuwe inzichten een plaats moeten krijgen in het onderwijs. Daarbij moeten we echter de oorsprong niet vergeten.

In het onderwijs aan psychologie studenten geven we veel methodologie en statistiek. In het onderwijzen van statistiek presenteren we vaak recepten. Wanneer je groentesoep wilt eten, neem dan deze ingrediënten en bereid ze op deze manier. Dit is makkelijk en het neemt onzekerheid weg. Onderzoek werkt echter niet zo. In empirisch onderzoek moeten er veel keuzes gemaakt worden gedurende het proces, van de opzet van het onderzoek tot en met het schrijven van het artikel. Voor veel van de keuzes weten we op dat moment niet wat de gevolgen daarvan zijn. Ieder van de keuzes leidt echter wel tot net iets beter of net iets slechter onderzoek. Daardoor ligt ieder onderzoek uiteindelijk op een continuüm van bijna wit naar bijna zwart (al kan je zeggen dat de Stapel affaire heeft laten zien dat pikzwart ook mogelijk is). Het is daarom van belang de onzekerheid sterker terug te laten komen in onze cursussen

en de studenten ervan te doordringen dat er geen definitieve conclusies zijn, slechts aanwijzingen. Studenten zullen dat niet leuk vinden, zij willen graag zekerheid. Dat is echter niet iets wat onze methoden kunnen bieden.

### Wetenschap in de maatschappij

Zoals u uit mijn betoog over voorspellen hebt kunnen opmerken, ben ik voor wetenschap die aansluit bij de praktijk en maatschappij. Aansluiten bij de maatschappij is echter wat anders dan gestuurd worden door de maatschappij. De laatste jaren is er vanuit de politiek maar ook vanuit de universiteit meer nadruk komen te liggen op de toepassing van kennis in de maatschappij. Dit is natuurlijk mooi, maar zoals ik aan het begin van de rede heb laten zien, de lessen van Goddard, schuilt er ook een gevaar in. Ik ben er geen voorstander van om net nieuwe inzichten direct toe te passen. Pas als we kennis van zaken hebben, moeten we die kennis gebruiken. Zoals eerder betoogd, vraagt kennis om een serie onderzoeken waarin stapje voor stapje meer aanwijzingen worden gekregen. Wanneer we echter kijken naar het financieringsmodel van wetenschap moet geconcludeerd worden dat kennis momenteel niet wordt gefinancierd. Alleen nieuwe innovatieve ideeën krijgen geld voor onderzoek. Daarmee stuurt men op hypes in plaats van op kennis.

### De mogelijkheden op een rijtje

Ik heb voor u een verhaal gehouden over de inhoud van mijn leerstoel en mogelijkheden voor de toekomst. De mogelijkheden die ik schets zijn natuurlijk Leids. In vergelijkbare groepen in Utrecht, Amsterdam of elders in Nederland zullen ze ook elke dag mogelijkheden zien voor verbetering, echter op andere terreinen. Samenvattend zijn hier in kreten de mogelijkheden:

1. Leer van de geschiedenis;
2. Meet preciezer door gebruik te maken van afstandsmodellen;
3. Bedenk dat herhaling de basis is van kennis;
4. Denk krom;
5. Durf te voorspellen.

Dit zijn wat mij betreft de mogelijkheden om elke dag aan te werken en hier zullen we als sectie M&S in de nabije toekomst aandacht aan besteden.

### **Dankwoord**

Mijnheer de Rector Magnificus, zeer gewaardeerde toehoorders, beste familie en vrienden. Aan het eind van mijn rede wil ik u graag allen bedanken dat u vandaag bent gekomen om naar mijn oratie te luisteren. Ik hoop dat ik u een beeld heb kunnen geven van mijn vakgebied en de verdere mogelijkheden voor ontwikkeling.

Een aantal mensen wil ik in het bijzonder bedanken.

Ik wil allen danken die aan de totstandkoming van mijn benoeming hebben bijgedragen. Het college van bestuur van de Universiteit Leiden en het bestuur van de Faculteit Sociale Wetenschappen. Dank voor het in mij gestelde vertrouwen dat ik op een waardige manier invulling zal geven aan deze leerstoel.

Het bestuur van het Instituut Psychologie en mijn collega hoogleraren in de psychologie. Binnen de Faculteit is Psychologie het enige instituut met een leerstoel en sectie op het gebied van de methodologie en statistiek. Ik wil jullie hartelijk danken dat jullie mijn mening delen dat de methodologie en statistiek een centraal onderdeel is van de psychologie en dat elk zichzelf serieus nemend onderzoeksinstituut een zelfstandige onafhankelijke sectie moet hebben die het proces van onderzoek bestudeert. Ik kijk uit naar de verdere samenwerking waar moderne methodologie in het psychologisch onderzoek wordt toegepast en naar de nieuwe methodologische vragen die daaruit voortkomen.

Ook de leden van de benoemingsadviescommissie voor deze leerstoel wil ik bedanken voor het in mij gestelde vertrouwen. Ik heb het vandaag al eerder gehad over personeelsselectie. Jullie hebben mij geselecteerd maar hebben naar mijn weten daar niet een statistisch voorspellingsmodel voor gebruikt. Wel

is het oordeel gekomen van meerdere experts, een element dat veelal leidt tot betere beslissingen. Aangezien het grootste deel van de benoemingsadvies commissie uit statistici bestond heb ik nog een kleine puzzel voor jullie. Binnen de Faculteit Sociale Wetenschappen zijn er tot mijn benoeming 12 hoogleraren geweest op het terrein van de methodologie en statistiek.<sup>12</sup> Ik ben de 13e die in dit rijtje is gevoegd en ik ben geboren op 19 december. Twee eerdere hoogleraren, professor Van der Kamp en Professor De Leeuw, hadden ook 19 december als geboortedag. Wat is precies de kans dat er drie van de dertien hoogleraren in een specifiek vakgebied dezelfde geboortedag hebben?

Leden van de Sectie Methodologie en Statistiek. Voor sommige van jullie was ik ooit de jonge nieuwe promovendus in de groep. Voor anderen was ik al de hoogleraar toen jullie aangenomen werden. Elke dag voel ik me bevoorrecht met jullie te mogen werken. Ik hoop nog op vele vruchtbare jaren van samenwerking waarin we de studenten psychologie leren leven met onzekerheid en hen vormen tot kritische denkers.

Docenten en stuurgroep van de masteropleiding “Statistical Science for the Life and Behavioral Sciences”. Zeven jaar geleden zijn we gestart met deze masteropleiding, in den beginne bijna in onze vrije tijd aangezien er destijds geen financiering voor was. Op dit moment wordt er gewerkt aan de officiële accreditatie van dit programma. Ik hoop en ga ervan uit dat dit lukt. Naast deze gezamenlijke inspanning op onderwijsgebied zou ik het mooi vinden ook op onderzoekgebied de krachten verder te bundelen.

Van veel mensen heb ik dingen geleerd. Ik kan hen niet allemaal bedanken. Toch wil ik een aantal leermeesters bedanken, in chronologische volgorde.

Hooggeleerde Hoijtink, beste Herbert. Onder jouw hoede deed ik mijn eerste onderzoek op het gebied van statistiek. Je leerde me programmeren in Fortran en ik herinner me nog dat we samen uren achter je computer zaten waarin je me leerde wat

‘debuggen’ was. Daarna ging je op fietsvakantie in Ierland (regen en asfalt) en stond je computer op je werkkamer 3 weken achter elkaar te werken aan mijn programmaatje. Je enthousiasme en de ruimte die je me bood om mijn eigen ideeën in het onderzoek te stoppen, heeft me destijds doen beseffen dat dat precies is wat ik wilde. Dank je wel.

Hooggeleerde Van Mechelen, beste Iven en hooggeleerde De Boeck, beste Paul. Voor mijn doctoraalscriptie ging ik naar Leuven en deed daar onderzoek onder jullie leiding op de kruising van statistiek en personeelsselectie. Nog vaak denk ik terug aan mijn tijd in Leuven. Samen met jullie promovendi bestudeerde ik het boek “Introduction to the theory of statistics” en in een groter statistisch gezelschap bestudeerden we het boek “Bayesian Data Analysis” van Andrew Gelman en collega’s. Kernwoorden die ik bij jullie heb geleerd en waaraan ik nog elke dag waarde hecht, zijn precisie en structuur.

Hooggeleerde Heiser, beste Willem. Jij bent degene die me de schoonheid van afstanden heeft laten inzien, zowel gewone afstanden als afstanden tussen drie punten (triadische afstanden). De geometrische benadering van de data analyse en van statistische modellen die je keer op keer tijdens besprekingen op een kladblaadje ontvouwt, geven zoveel inzicht en zijn ronduit fascinerend. Dank voor die mooie lessen!

Ten slotte familie. Lieve papa en mama: hadden jullie het gedacht dat jullie oudste zoon die in zijn puberjaren veel meer interesse had voor sport dan voor school hier nu zou staan; hadden jullie het gedacht toen ik met de hakken over de sloot mijn HAVO diploma haalde; hadden jullie het gedacht toen bleek dat ik regelmatig spijbelde en jullie mij daarover ’s avonds nog uit bed haalden om er over te praten; hadden jullie het gedacht toen ik in VWO 5 bleef zitten en jullie me dat moesten vertellen tijdens mijn EK volleybal; hadden jullie het gedacht toen ik in mijn eerste studiejaar maar weinig studiepunten haalde en niet zo goed wist wat ik met die studie informatica moest. Ik weet niet wat jullie dachten, wel weet ik dat

jullie me altijd het gevoel gaven dat het wel goed zou komen met mij. Dank voor dat gevoel!

Thijs, Joris, Bas, lieve jongens van me. Door een van de meesters op de Lorentzschool werden jullie vorig jaar “de criminele bende van De Rooij” genoemd. Ik denk niet dat jullie echt een criminele bende zijn, hooguit een beetje baldadig met een ondeugende blik in de ogen. Toen ik benoemd werd als hoogleraar vroegen jullie me of ik dan ook zo’n jurk aanmoest. Op mijn ja barstten jullie in lachen uit. En of ik dan ook zo’n gek hoedje op moest; weer ja - nog meer gelach. De laatste vraag was of jullie dat dan ook mochten zien. Wel, hier sta ik dan en zoals je kan zien, ben ik niet de enige. Wat vinden jullie ervan? Morgen is het voor mij weer langs-de-lijn-dag met volleybal en voetbal. Dan trek ik ook weer mijn gewone kloffie aan, spijkerbroek, t-shirt en gympen. Ik verwacht dat dat jullie beter zal bevallen. Mij trouwens ook. Ik wil jullie danken voor de hectische uurtjes thuis, voor jullie mooie enthousiaste verhalen, en voor de nieuwste trends op taalgebruik.

Lieve Arine: wat moet een hoogleraar zonder iemand die thuis de zaken op orde houdt. Gekscherend noem ik je af en toe “manager-thuiszaken” - dit moet je echter zien als geuzennaam. Je bent ook eigenlijk niet zo’n goede manager, zaken delegeren is niks voor jou. Je lost het liever zelf op. Psychologen zouden je karakteriseren als een “sjouwer”. Ik ben je op zoveel terreinen dankbaar, bijvoorbeeld op het terrein van de Nederlandse taal. Van jongs af aan is mijn exacte hersendeel iets beter ontwikkeld dan het taalcentrum. Ik kan bijvoorbeeld maar niet onthouden wat een lange ei of een korte ij is. Ikzelf word niet langer of korter wanneer ik ga liggen of wanneer ik weer ga staan en heb de logica van taalkundigen om het ene lang en het ander kort te noemen nooit begrepen. Jij gelukkig wel. Op veel andere gebieden heb je ook zoveel meer kennis dan ik en leer ik elke dag een beetje van je. Dank voor al je steun de afgelopen jaren waarin ik soms in afwezige, dromerige buien over algoritmes en onderzoeksideeën niet goed luisterde naar wat je me vertelde. Dank voor alle uren die je investeert wanneer ik

op reis ben en jij alle taken op je neemt. Zeker toen de kinderen kleiner waren, was dit af en toe een heidens karwei. Uiteindelijk komt het er op neer dat zonder jou steun ik hier nooit had gestaan. Dank, dank, dank!

Ik heb gezegd.

## Noten

- 1 Gregory, R.J. (2000). *Psychological Testing: history, principles and applications*. Needham heights, MA: Allyn & Bacon, Inc.
- 2 Gelb, S.A. (1986). Henry H. Goddard and the immigrants, 1910-1917: The studies and their social context. *Journal of the History of the Behavioral sciences*, 22, 324- 332.
- 3 McCrae, R.R., Zonderman, A.B., Costa, P.T., Jr., Bond, M.H. and Paunonen, S.V. (1996). Evaluating replicability of factors in the Revised NEO Personality Inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, 70, 552-566.
- 4 Stark, S., Chernyshenko, O.S., Drasgow, F. and Williams, B.A. (2006), Examining Assumptions about Item Responding in Personality Assessment: Should Ideal Point Models be Considered for Scale Development and Scoring? *Journal of Applied Psychology*, 91, 25-39.
- 5 Meulman, J.J. (1999). *De toegepaste datatheorie en de verwondering*. Oratie, Universiteit Leiden.
- 6 Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349, aac4716.
- 7 Simmons, J.P., Nelson, L.D. and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, 22, 1359-1366.
- 8 Taguepeera, R. (2008). *Making social sciences more scientific: the need for predictive models*. Oxford University Press.
- 9 Helmer, O. and Rescher, N. (1959). On the epistemology of the inexact sciences. *Management Science*, 6, 25-52.
- 10 Meehl, P. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minneapolis Press.
- 11 Shmueli, G. (2010). To explain or to predict. *Statistical Science*, 25, 289-310.
- 12 Met dank aan Willem Heiser heb ik het volgende lijstje gecreëerd: Van de Geer, De Leeuw, Van der Kamp, Heiser, Meulman, Mooijaart, Kroonenberg, Speckman, Philipsen, Irwin, Rombouts, Bethlehem.



## PROF.DR. MARK DE ROOIJ



- 1996      Doctoraal Psychologie aan de Rijksuniversiteit Groningen
- 2001      Promotie aan de Universiteit Leiden (cum laude)
- 2000-2002   Postdoc, Sectie Methodologie en Statistiek, Instituut Psychologie, Universiteit Leiden
- 2002-2011   UD, Sectie Methodologie en Statistiek, Instituut Psychologie, Universiteit Leiden
- 2011-2014   UHD, Sectie Methodologie en Statistiek, Instituut Psychologie, Universiteit Leiden
- 2014-heden   Hoogleraar Methodologie en Statistiek van Psychologisch Onderzoek, Instituut Psychologie, Universiteit Leiden

Methodologie en Statistiek van Psychologisch Onderzoek bestudeert het proces van wetenschap binnen de psychologie: van onderzoeks idee tot publicatie. Belangrijke elementen in dit proces zijn het meten en het statistisch modelleren.

Hoe meten we eigenschappen zoals intelligentie, emotionele stabiliteit en rekenvaardigheid en welke meetmodellen gebruiken we daarvoor? Hoe kunnen we metingen preciezer, betrouwbaarder, of meer valide maken? In het modelleren bestudeert de methodologie en statistiek samenhang tussen eigenschappen. Hiervoor worden statistische modellen gebruikt. Wat zijn de eigenschappen van deze modellen? Wat zijn de vooronderstellingen, en passen de modellen bij de vragen die psychologen werkelijk hebben?

Een belangrijk onderscheid voor zowel het meten als het modelleren wordt gecreëerd door het doel. Willen we gedrag verklaren of willen we gedrag voorspellen? Afhankelijk van het doel zijn andere meetmethoden en andere statistische modellen nodig.



Universiteit  
Leiden