KWANTIFICEREN EN METEN IN DE SOCIALE EN GEDRAGSWETENSCHAPPEN

door

Jan de Leeuw

1. Inleiding

Het dikwijls gehanteerde onderscheid tussen 'zachte' en 'harde' wetenschappen is gebaseerd op verschillende criteria. We noemen er een aantal, niet noodzakelijk in volgorde van belangrijkheid. In de eerste plaats de mate van mathematische formalisering van de desbetreffende wetenschap, in de tweede plaats de mate van overeenstemming die bestaat tussen verschillende onderzoekers, in de derde plaats het gemak waarmee experimenten gedaan kunnen worden, en in de vierde plaats de veelvuldigheid waarmee kwantitatieve uitspraken voorkomen. We zullen ons hier vooral bezig houden met het laatste criterium, in het bijzonder met de vraag of en in hoeverre in de sociale en gedragswetenschappen gemeten kan worden. En met de verwante vraag of meten in de sociale en gedragswetenschappen in de één of andere zin fundamenteel verschilt van het meten in de natuurwetenschappen of levenswetenschappen.

Negentiende eeuwse fysici als Maxwell en Kelvin benadrukten dat we een fysisch proces of een fysische eigenschap niet goed of onvoldoende begrepen, wanneer we dat proces of die eigenschap niet konden meten. 'Aan elke grootheid die een fysische eigenschap of toestand van een systeem pretendeert te beschrijven, stelt men de eis, dat in principe een meetprocedure moet kunnen worden aangegeven, waardoor deze grootheid op ondubbelzinnige wijze in een getal kan worden uitgedrukt.' (Alkemade, Hoogenboom, en Smit, 1979, p.1-2). In de dertiger jaren besliste een comité van de British Association for the Advancement of Science, na veel discussie, dat de term meten alleen van toepassing kon zijn op objecten, die, in de één of andere zin, bij elkaar opgeteld konden worden. Een standpunt dat al eerder verdedigd werd door Campbell, in een aantal invloedrijke boeken (1920, 1928). Met name betekende dit voor de psychologie, dat de sterkte van de ervaring van een prikkel niet op een verantwoorde manier numeriek vastgelegd kon worden. Dat was wat jammer voor de psychofysici, die op dat moment al bijna honderd jaar bezig waren met juist die sterkte van de ervaring te meten. En het was ook pijnlijk voor een groot aantal kwantitatief ingestelde psychologen, die de positivistische wetenschapsopvatting van de fysici als Maxwell en Kelvin overgenomen hadden. De differentiële psychologie was al rond 1880 ontstaan uit het werk van Francis Galton, die en passant ook als één der eersten de mathematische biologie en de kwantitatieve genetica beoefende. Galton was idolaat van meten en tellen, en al zijn wetenschappelijk activiteiten waren ingesteld op kwantificeren. De befaamde Amerikaanse psycholoog E.L. Thorndike stelde rond 1920, dat

Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs, Vol. 5, nr. 4, 1989

'everything that exists, exists in a certain amount, and can consequently be measured.' De leidinggevende figuren in de intelligentietest beweging beschouwden zich als harde wetenschappers, en iemand als bijvoorbeeld H.J. Eysenck begint steevast zijn boeken en artikelen met de relevante citaten van Maxwell, Kelvin, Thorndike, of Kepler waarin de lof van het meten gezongen wordt.

Onze belangrijkste stelling in dit artikel zal zijn, dat de verschillen tussen de diverse hardheidsgraden van de wetenschappen niet zozeer een kwestie zijn van verschillen in *meetmethoden*, maar eerder een kwestie van verschillen in *meetresultaten*. Dat wil zeggen van precisie van de meting, van inpassing van de resultaten in een bestaand theoretisch kader, en van invariantie van de resultaten bij variatie van bijvoorbeeld onderzoeker, politiek klimaat, jaartal, enzovoorts. We lichten deze stelling toe aan de hand van een aantal voorbeelden uit de sociale en gedragswetenschappen. Maar eerst enige algemeenheden over meten.

2. *Definitie van kwantificatie en meten*

Bij de bespreking van wat meten eigenlijk is, moeten we een aantal begrippen duidelijk onderscheiden. We volgen daarbij vooral de definities van Bunge (1967, hfdst 13, 1973), maar we houden ook rekening met de in de fysica gebruikelijke omschrijvingen, en met de bijdragen uit de recente meettheorie. In het eerder genoemde leerboek van Alkemade e.a. wordt gesteld dat aspecten van de werkelijkheid voor meting toegankelijk zijn, wanneer ze zich op *reproduceerbare* wijze in een *getal* laten uitdrukken (l.c., p.1). In een zeer invloedrijk artikel definieert Stevens (1946) op overeenkomstige wijze meten als het toekennen van getallen aan objecten volgens regels, willekeurig wat voor regels. Dit is een wat al te ruime omschrijving. In het voetbalspel bijvoorbeeld is, of liever was, er een afspraak dat we het feit dat een persoon linksbuiten speelt aangeven door op zijn shirt het nummer elf te bevestigen. Met deze afspraak hebben we een aspect van de werkelijkheid op reproduceerbare wijze in een getal uitgedrukt, maar het lijkt toch niet helemaal vanzelfsprekend om hier van meten te spreken.

Bunge onderscheidt in dit verband *kwantificatie* van *symbolisatie*. Bij symbolisatie gebruiken we getallen als etiketten. Bij de kodering van een vragenlijst kunnen we bijvoorbeeld besluiten om de protestantse respondenten aan te duiden met het cijfer één, de katholieken met het cijfer twee, en de overigen met het cijfer drie. Bij een sociometrisch onderzoek laten we de 15 kinderen in een schoolklas van al hun klasgenootjes zeggen of ze ze aardig vinden of niet. De gegevens verzamelen we in een tabel met 15 rijen en kolommen. In rij 'Pietje' en kolom 'Jantje' staat een 1 als Pietje Jantje aardig vindt, anders staat er een 0. De zo ontstane matrix heet een sociogram. Bunge spreekt in dit soort gevallen niet van kwantificatie, omdat de getallen slechts als codes gebruikt worden, en omdat er verder geen operaties op deze getallen uitgevoerd zouden kunnen worden die een fysische betekenis hebben. Bunge gebruikt zelfs het woord symbolisatie wanneer de etiketten op een bepaalde manier *geordend* zijn. Het is bijvoorbeeld zo, dat de linksbuiten niet toevallig nummer elf op zijn shirt heeft. Wanneer we de voetballers ordenen van achteren naar voren en van rechts naar links, dan komen we vanzelf op het getal elf uit. Op dezelfde manier wordt de aardbevingsschaal van Richter gebruikt, of de hardheidsschaal van Mohs. Wanneer iets een hardheid gelijk aan vier heeft, betekent dit niet dat het in de één of andere fysisch welomschreven zin twee maal zo hard is als een object met hardheid twee. Het lijkt overigens verstandig om voor het begrip symbolisatie in dit soort gevallen het begrip *ordening* te gebruiken, omdat er in ieder geval meer wiskundige eigenschappen van getallen empirisch zinvol te interpreteren zijn. Immers hardheid vier is wel meer dan hardheid twee, en hardheid drie zit wel tussen twee en drie in.

Van *kwantificatie* is volgens Bunge slechts sprake wanneer een aanvankelijk kwalitatief begrip op theoretisch niveau met een numerieke variabele in verband gebracht wordt. Het kwalitatieve begrip 'omvang' wordt gekwantificeerd als 'volume', het kwalitatieve begrip 'moeilijkheid om te tillen' wordt gekwantificeerd als 'gewicht', 'helling' wordt 'hoek', en 'geloof in' of 'vertrouwen in' wordt 'waarschijnlijkheid'. Het is duidelijk dat kwantificatie, in deze zin, duidelijk onderscheiden moet worden van meten. Kwantificatie is een theoretische handeling, iets wat achter het bureau uitgevoerd wordt, terwijl meten een empirische handeling is, uitgevoerd bijvoorbeeld in een laboratoruim of via een telefonische enquête. Bij kwantificatie stellen we vast dat het begrip 'lang zijn' overeenkomt met de variabele 'lengte', bij meten stellen we van mijnheer X vast hoe lang hij nu eigenlijk is. Of we hebben 'slimheid' gekwantificeerd met 'intelligentietestscore', en we meten iemands slimheid door hem de desbetreffende test af te nemen. Bunge, en ook Ellis (1966), gebruiken het onderscheid tussen kwantificatie en meten om zich af te zetten tegen extreem operationalistische standpunten. Zo houdt Dingle (1950) bijvoorbeeld vol dat grootheden niet bestaan onafhankelijk van de operaties die uitgevoerd worden om ze te meten, maar dat ze door die operaties juist gedefiniëerd worden. Meten gaat daarom vooraf aan kwantificatie. Het standpunt van Bunge en Ellis, dat theorie vooraf gaat aan het meten, past beter bij de moderne Popperiaanse wetenschapsfilosofie, en heeft een aantal aanzienlijke voordelen.

Het onderscheid tussen kwantificatie en meten is duidelijk, en zeer nuttig, maar het onderscheid tussen kwantificatie en symbolisatie (of ordening) is gedeeltelijk gebaseerd op een toch wat verouderde verering van de getallenrechte. Beter lijkt het ons het algemene begrip *mathematisering* in te voeren. We veronderstellen hierbij dat een oorspronkelijk kwalitatief begrip overeenkomt of zelfs isomorph is met een bepaalde mathematische structuur, of, anders gezegd, we veronderstellen dat het object of het proces dat we bestuderen beschreven kan worden met een mathematisch *model*. Dat model kan een geordende verzameling zijn, een matrix met elementen gelijk aan nul en één zoals bij een sociogram, het complexe vlak, of de getallenrechte. Op deze modelmatige representatie van de werkelijkheid kunnen de in de mathematische structuur gedefiniëerde operaties uitgevoerd worden. Of de resultaten van die operaties ook overeenkomen met zinvolle en interpreteerbare elementen van de gemodelleerde werkelijkheid hangt af van de toepasbaarheid van het model. Zo kunnen we in het geval van de hardheidsschaal de empirische uitspraak 'x krast y' representeren door de mathematische uitspraak '$f(x) > f(y)$', waarbij $f(x)$ de schaalwaarde is van materiaalsoort x. Bij gewichten komt de empirische uitspraak 'als x even zwaar is als y, dan zullen x en z samen zwaarder zijn dan y' overeen met 'als $f(x) = f(y)$, dan $f(x) + f(z) > f(y)$'. We zien dat de mathematische relaties die weergegeven worden met de symbolen +, =, en > zinvolle empirische interpretaties hebben in termen van krassen, en van het gedrag van bijvoorbeeld een jukbalans.

Bunge (1967, p.202) bespreekt ook de *voordelen* van kwantificatie. Hij noemt er een aantal, gedeeltelijk ontleend aan Hempel (1952, p.56-57). Met behulp van kwantificatie kunnen begrippen precieser gedefinieerd, beschreven, en geclassificeerd worden. Bovendien kunnen we gemakkelijker hypothesen en theorieën opstellen, en deze hypothesen en theorieën omtrent gekwantificeerde begrippen kunnen op preciese en eenvoudige manier getoetst worden. Het spreekt vanzelf dat deze voordelen voor de sociale en gedragswetenschappen op precies dezelfde manier van toepassing zijn als voor de natuur- en levenswetenschappen. Interessanter is misschien de vraag in hoeverre de begrippen in de sociale en gedragswetenschappen gemathematiseerd of zelfs gekwantificeerd kunnen worden. Bunge (1973, p.111) noemt het een methodologisch principe van de moderne fysica dat iedere ordening uiteindelijk omgezet kan worden in een kwantificatie, wanneer er maar voldoende theorie over de betreffende verschijnselen bestaat. De hardheidsschaal van Mohs is hier een voorbeeld van. Of dit in de 'zachte' wetenschappen ook tot methodologisch beginsel verheven kan worden is voorlopig nog niet te overzien. Voorlopig is het zeker zo, dat voor veel belangrijke theoretische begrippen nog geen eenduidige kwantificaties beschikbaar zijn, terwijl in sommige gevallen zelfs de ordeningen ontbreken.

Twee belangrijke begrippen die bij discussie van kwantificatie en meten een rol spelen zijn *grootheid* en *schaal*. Zie bijvoorbeeld Bunge (1967, hfdst 13, 1973) of Ellis (1966, hfdst 2 en 3). Een grootheid (ook wel een *variabele*, in het Engels *magnitude* of *quantity*) is een kwantitatief predikaat, een numerieke functie. De grootheid kent aan alle elementen in zijn domein een numerieke waarde toe, dat wil zeggen de grootheid is de conceptuele pendant van het feitelijke meten. In veel gevallen kan de gekwantificeerde eigenschap op verschillende manieren zinvol numeriek afgebeeld worden. Denk maar aan temperatuur, bijvoorbeeld. De aard van de representatie die we kiezen heet de *schaal*, en een speciaal gekozen interval op de numerieke schaal heet de *eenheid*. In sommige gevallen bewaren we of definiëren we een materiële of empirische versie van die conceptuele eenheid. Dit heet een *standaard*. Ook bij deze door Bunge besproken begrippen maken we een aantekening. Het is natuurlijk redelijk om woorden als 'grootheid' uitsluitend te gebruiken voor numerieke variabelen, als men maar inziet dat de definitie van een variabele op zichzelf niets numerieks heeft, net zo min als de definitie van het begrip functie in de wiskunde iets numerieks heeft. Een variabele kan religieuze overtuiging afbeelden in de verzameling [protestant, katholiek, overig], en kan voetballers afbeelden in de verzameling [doelman, rechtsback,..., linksbuiten]. Of, nogmaals, het kwalitatieve begrip 'standpunt ten opzichte van de kruisraket' kan gemathematiseerd worden als 'mate waarin men het eens is met de stelling dat zo spoedig mogelijk in Nederland 48 kruisraketten geplaatst moeten worden'. De vraag kan aan alle Nederlanders, of een groep Nederlanders, gesteld worden, en definieert daarmee een variabele die Nederlanders afbeeldt in [helemaal niet mee eens, niet mee eens, ... , volkomen mee eens]. Het feitelijke stellen van de vraag is dan de meting.

3. Soorten kwantificatie

Bij het bespreken van diverse soorten kwantificaties en metingen onderscheiden we in de eerste plaats, met Campbell (1920, 1928), *directe* en *indirecte* kwantificatie. Deze termen ontlenen we aan Ellis (1966) en Roskam (1981), Campbell zelf gebruikt *fundamental* en *derived measurement*. We spreken van directie kwantificatie wanneer er aan de definitie van de desbetreffende grootheid geen eerdere kwantificatie vooraf gaat. Het is bijvoorbeeld mogelijk iemand op te sluiten in een kamer met uitsluitend een groot aantal stokken, en hem op te dragen niet te voorschijn te komen voor hij alle stokken gemeten heeft. Die opgave kan hij volbrengen door één stok als eenheid te nemen, en vervolgens de andere stokken te meten door ze af te passen aan de eenheid. Dat zal in het algemeen geen al te preciese metingen opleveren, en de schaal zal afhangen van de keuze van de eenheid, maar niettemin gaat het wel degelijk om meting en komt er geen getal aan te pas. Hetzelfde kunnen we doen wanneer we iemand een stel gewichten geven en een jukbalans. In beide gevallen gaat het om *extensieve* grootheden, dat wil zeggen dat er een empirische operatie bestaat die lijkt op de mathematische optellingsoperatie (en die ook als zodanig gemodelleerd wordt). De empirische optelling, dikwijls *concatenatie* genoemd, kunnen we aanduiden met het symbool $\triangle$. De kwantificatie die we bestuderen heeft dan de *additieve* eigenschap dat voor alle x en y geldt $f(x \triangle y) = f(x) + f(y)$. Bij *intensieve* grootheden komt geen concatenatie operatie voor. Voorbeelden zijn temperatuur, dichtheid, enzovoorts. Bunge wijst er op dat extensieve grootheden in de fysica relatief zeldzaam zijn, en dat de grootheden van theoretisch het meeste belang in het algemeen niet precies additief en vaak zelfs intensief zijn. Massa bijvoorbeeld is slechts bij benadering additief, en kan misschien beter *kwasi-extensief* genoemd worden. Volgens Campbell, en ook volgens het in de inleiding genoemde comité van de British Association, kunnen alleen extensieve grootheden direct gekwantificeerd worden. Dit laatste idee is, zoals we verderop zullen zien, volslagen achterhaald door de ontwikkelingen in de moderne meettheorie (volgens de definities van Bunge is het overigens aanzienlijk juister te spreken van kwantificatietheorie).

In de sociale en gedragswetenschappen komen concatenatie operaties niet of nauwelijks voor. Dat wil echter in het geheel niet zeggen dat er niet direct gemeten kan worden. We zullen verderop voorbeelden van direct meten tegenkomen, waarbij de concatenatie operatie niet empirisch gegeven is, maar aan de hand van andere empirische relaties gedefinieerd kan worden. De zo geconstrueerde structuur is dan extensief. Voorlopig wijzen we er overigens op dat direct meten van bijvoorbeeld lengte in de praktijk niet uitgevoerd wordt doordat fysici in laboratoria met stokken in de weer zijn. Op basis van eerder verrichte fundamentele metingen zijn geijkte meetinstrumenten geconstrueerd die veel snellere en veel preciesere metingen mogelijk maken. Dat wil zeggen, in de terminologie van Suppes en Zinnes (1963, p.20-22), dat het meten in veel gevallen *pointer measurement* is. We vertalen dit met *instrumentmeten*. Instrumentmeten is, uiteindelijk, gebaseerd op direct meten. Wanneer de ijking van het instrument niet op directe meting gebaseerd is, dan spreken we met Suppes en Zinnes van *pseudo-instrumentmeten*. In de sociale en gedragswetenschappen bestaan veel meetinstrumenten die, in deze zin, pseudo-instrumenten zijn.

We wijzen er ook op dat de maat voor bijvoorbeeld lengte die alom gebruikt wordt gedeeltelijk bepaald is door overwegingen die samenhangen met gemak en eenvoud. Wanneer we iemand een volmaakt ronde kamer insturen met zijn verzamelingen stokken, met op de muur van die kamer een centimeterschaal, dan kan hij de kamer als instrument gebruiken en de stokken meten door ze met beide uiteinden tegen de muur te leggen en de lengte af te lezen. Dit geeft een definitie van lengte, gerelateerd aan de gebruikelijke volgens $g(x) = \arcsin f(x)$. Zelfs de concatenatie operatie, en daardoor het directe meten, is niet gebaseerd op fysische noodzakelijkheid, maar op conventies. Wanneer we twee stokken x en y hebben, dan kunnen we ze een rechte hoek met elkaar laten maken en als $x \triangle y$ de schuine zijde van de zo ontstane rechthoekige driehoek nemen. Dit is een soort optelling die tot gevolg heeft dat (volgens Pythagoras) directe meting van lengte nu voldoet aan $g(x) = f(x)^2$. Veel formules in de fysica zouden aanzienlijk ingewikkelder worden bij gebruik van deze alternatieve grootheden om lengte af te beelden, maar iets wezenlijks zou er natuurlijk niet veranderen.

Als voorbeeld van indirecte of afgeleide kwantificatie noemen we bijvoorbeeld de definitie van weerstand als spanning gedeeld door stroom of de definitie van dichtheid als massa gedeeld door volume, maar in principe kan het aantal voorbeelden natuurlijk willekeurig uitgebreid worden. In ieder geval is voor indirecte kwantificatie zowel eerder direct meten als berekening nodig. De aard van de berekeningen wordt dikwijls gedicteerd door de fysische wetten die voor de betreffende verschijnselen gelden. In de sociale en gedragswetenschappen zijn deze vormen van indirect meten zeldzaam, omdat ze op direct meten gebaseerd moeten zijn, en omdat er voor de meeste terreinen geen wetten van enige betekenis beschikbaar zijn. Wat we wel dikwijls tegenkomen is het meten van niet-observeerbare grootheden via *indicatoren*. Dit komt ook in de fysica voor. De grootheid temperatuur, dat wil zeggen de bewegingssnelheid van de moleculen van een stof, wordt gemeten via de lengte van een kwikkolom in een thermometer. Deze lengte *indiceert* de bewegingssnelheid van de moleculen. Op dezelfde manier indiceert het EEG hersenactiviteit, indiceert inkomen economische status, en indiceert de IQ-test intelligentie. De formele overeenkomst tussen al deze indicatoren is echter minder belangrijk dat de feitelijke verschillen die er tussen bestaan. In het geval van temperatuur kennen we een *wet* die aangeeft hoe de relatie tussen de grootheid temperatuur en de lengte van de kwikkolom is. Met andere woorden: de thermometer is een instrument in de zin van Suppes en Zinnes. Wat de preciese relatie is tussen EEG en 'hersenactiviteit' of tussen IQ en 'intelligentie' kunnen we niet in een wet uitdrukken. En inderdaad, er zijn dan ook veel discussies gevoerd over de 'betekenis' van het EEG en de IQ-test. Deze discussie staat overigens in principe los van de vraag in hoeverre pseudo-instrumenten van deze aard gebruikt kunnen worden bij *voorspellen* van allerlei medisch of maatschappelijk belangrijke zaken. In de 'zachte' wetenschappen, waar we hier voor het gemak ook even de medische wetenschap bijrekenen, komen veel indicatoren of pseudo-instrumenten voor die samen blijken te hangen met zaken als gezondheid en succes, terwijl het verre van duidelijk is hoe die samenhang precies of zelfs maar ongeveer in elkaar zit.

Er is één bepaalde vorm van afgeleid meten, die van het begin af aan een grote rol gespeeld heeft in het sociaal wetenschappelijk kwantificeren. Dit is meten op *waarschijnlijkheid*. Zoals we verderop, bij de voorbeelden, zullen zien is de psychofysica van Fechner en Thurstone gebaseerd op modellen voor waarschijnlijkheden. Deze waarschijnlijkheden zijn fundamenteel gemeten, door frekwenties van oordelen of uitspraken te tellen, en ze worden via een model gerelateerd aan de parameters van het systeem dat we bestuderen. Dit soort modellen is van uitnemend belang in de sociale en gedragswetenschappen. We zien hier overigens dat de begrippen meten en kwantificeren op de achtergrond raken, of in ieder geval opgenomen worden in het model als geheel. Het model moet bovendien uitgebreid worden met een *foutentheorie*, net als in de fysica, die verklaart waarom bij replicatie van het experimenten niet precies hetzelfde resultaat gevonden wordt. Model en foutentheorie leveren samen een stochastisch model op, waarin de parameters niet zo zeer gemeten worden als wel *geschat*.

4. Typen schalen

We hebben in de vorige paragraaf verschillende typen kwantificaties behandeld, die verschillende typen meetmethoden suggereren. In de indeling zoals die gehanteerd wordt door Bunge en Ellis hebben we naast de kwantificatie ook nog de *schaal* waarop de grootheid afgebeeld wordt als belangrijke component. We weten dat temperatuur op verschillende schalen afgebeeld kan worden, genoemd naar de onderzoekers Celsius, Fahrenheit, en Reaumur. Al deze schalen staan in een lineaire relatie tot elkaar, en hebben een nulpunt en schaaleenheid die op puur conventionele basis gekozen is. De temperatuurschaal van Kelvin daarentegen is gebaseerd op de gaswet, en neemt de theoretisch laagst mogelijke temperatuur als nulpunt. De eenheid blijft willekeurig. Hetzelfde geldt voor de verschillende stelsels die gebruikt worden om lengte te meten. Ze zijn allemaal gebaseerd op de fundamentele concatenatie operatie 'naast elkaar leggen', en de schalen die ze opleveren hebben hetzelfde nulpunt ('geen lengte') maar een verschillende eenheid. Zoals we eerder hebben laten zien, kunnen we andere concatenatie operaties verzinnen (de hypothenusa-meting) die tot wezenlijk andere schalen voor lengte leiden. Om deze verschillen tussen schalen nader te specificeren geven we kort enige begrippen uit de moderne kwantificatietheorie (of, als men wil, meettheorie).

Bij het meettheoretisch onderzoek van een bepaalde grootheid proberen we een representatiestelling te vinden en een uniekheidsstelling. Laten we ons even beperken tot het directe meten. De representatiestelling vertelt ons bijvoorbeeld dat een verzameling stokken X door middel van de grootheid lengte afgebeeld kan worden in de reële getallen R op zo'n manier dat 'x is langer dan y' als en alleen als $f(x) > f(y)$ en 'x en y zijn samen even lang als z' als en alleen als $f(x \triangle y) = f(z)$. Om een dergelijke stelling af te leiden moeten de stokken aan een aantal eisen voldoen. We moeten ze naast elkaar kunnen leggen, het mag niet zo zijn dat de ene keer x langer is dan y en de andere keer y langer dan x, als x langer is dan y en y is langer dan z dan is x langer dan z, x en y zijn samen langer dan x en ook langer dan y, enzovoorts. De stokken mogen dus niet al te krom zijn, en niet al te buigzaam, en niet al te veel varieren in lengte bij temperatuurwisseling, enzovoorts. Wanneer de empirische begrippen 'langer zijn' en 'achter elkaar leggen' maar genoeg eigenschappen gemeen hebben met de getallenrechte, dan bewijst men in de representatiestelling dat stokken en de getallenrechte *isomorph* zijn, en dat de gezochte afbeelding dus inderdaad bestaat. De uniekheidstelling levert daarnaast het resultaat op, dat als f en g twee afbeeldingen zijn met de gewenste eigenschappen, dat dan $f = \alpha g$ voor één of andere $\alpha > 0$. We zeggen dan dat lengte gedefinieerd is op een *ratioschaal*, evenals bijvoorbeeld absolute temperatuur. Wanneer objecten niet geconcateneerd kunnen worden, maar alleen vergeleken (zoals de elkaar krassende stoffen in de hardheidsschaal), dan vinden we een *ordinale* schaal. Twee verschillende afbeeldingen f en g in de getallenrechte zijn dan slechts *monotoon* aan elkaar gerelateerd, dat wil zeggen dat de twee schalen alleen maar op dezelfde manier geordend zijn. De meettheoretische analyse van situaties zoals de gebruikelijke temperatuurmeting levert een *intervalschaal*, waarbij zowel nulpunt als eenheid willekeurig zijn, en waarbij iedere positieve lineaire transformatie van de schaal weer een andere toelaatbare schaal oplevert. Let wel: we zeggen niet dat temperatuur op een intervalschaal afgebeeld wordt. Een dergelijke uitspraak heeft weinig zin, omdat we immers weten dat er ook zoiets bestaat als absolute temperatuur op een ratioschaal en de temperatuur van Dalton is logarithmisch aan de gebruikelijke gerelateerd (zie bijvoorbeeld Ellis, 1966). Evenmin kunnen we zeggen dat lengte op een ratioschaal afgebeeld wordt. Met het hypothenusa-voorbeeld kunnen we laten zien dat dit alleen opgaat voor lengte met de gebruikelijke concatenatie operatie.

Het trio ratioschaal, intervalschaal, ordinale schaal wordt gewoonlijk aangevuld met de *nominale* schaal. Deze vierdeling is vooral gepopulariseerd door Stevens (1946). De nominale schaal wordt gebruikt voor symbolisatie, dat wil zeggen voor voorbeelden zoals protestant = 1, katholiek = 2, en overig = 3. Het enige dat hiermee aangegeven wordt, is dus dat sommige objecten gelijk zijn, en andere niet. Zoals eerder gezegd is het nogal riskant bij een nominale schaal te spreken van meting, maar dat wil in het geheel niet zeggen dat nominale begrippen niet gemathematiseerd kunnen worden. Het is, na het klassieke artikel van Stevens, overigens duidelijk geworden dat er tussen de diverse schaalniveaus nog vele overgangsgevallen voor kunnen komen, en dat verschillende meetprocedures ook dergelijke tussenvormen opleveren. We zullen hierop niet nader ingaan, maar we verwijzen de geinteresseerde toehoorder naar de moderne literatuur over schaalmethoden (Torgerson, 1958, Coombs, 1964) of meettheorie (Bezembinder, 1970, Krantz, Luce, Supppes, en Tversky, 1971). Het is voor ons doel belangrijk om te onthouden dat het schaalniveau geen intrinsieke eigenschap van het fysische systeem is, maar een uitkomst van de meettheoretische analyse van het systeem, dat wil zeggen een onderdeel van het model dat we voor dat systeem aannemen.

5. Voorbeelden

In deze paragraaf bespreken we enige belangrijke voorbeelden van sociaal of gedragswetenschappelijk kwantificeren en meten. De voorbeelden zijn in de eerste plaats historisch van belang, maar gedeeltelijk ook maatschappelijk. In ieder geval zijn ze zo gekozen dat ze een grote variëteit van situaties laten zien.

We beginnen met het werk van Fechner, rond 1870. Het ging hem erom de psychofysica, dat wil zeggen de leer van de relaties tussen sterkte van prikkels en sterkte van ervaringen, tot een exacte wetenschap te ontwikkelen. Daar had Fechner wat obscure bedoelingen mee, maar die doen hier niet ter zake. Het spreekt min of meer vanzelf dat de formulering van Fechner's theorieën gebruik maakt van de in die tijd gebruikelijke terminologie. We zullen daarom een moderne vertaling presenteren, gebaseerd op Falmagne (1985). In een typisch psychofysisch experiment vergelijken we bijvoorbeeld een aantal gewichten. Stel X is de verzameling van gewichten, en x en y zijn elementen van die verzameling. We vragen onze proefpersoon of hij denkt dat x zwaarder is dan y, en we vragen dat een groot aantal malen (of we vragen het aan een groot aantal proefpersonen). Stel π (x,y) is de kans dat x groter gevonden wordt dan y. Fechner nu veronderstelde dat er een f bestond, en een verdelingsfunctie H, zodanig dat π (x,y) = H [f(x) – f(y)]. De functie f beeldt de gewichten uit X af op de getallenrechte, de functie H die monotoon van 0 naar 1 stijgt maakt van de verschillen f(x) – f(y) waarschijnlijkheden. We kunnen nu f(x) de sterkte van de ervaring (of sensatie) noemen veroorzaakt door gewicht x, en we kunnen nagaan hoe f(x) samenhangt met x. Meettheoretische analyse van de Fechneriaanse psychofysica onderzoekt aan welke voorwaarden de kansen π (x,y) moeten voldoen om een representatie in termen van verschillen tussen schaalwaarden mogelijk te maken, en hoe uniek een dergelijke representatie is. We vinden bijvoorbeeld dat de conditie π (x,y) $\leqslant$ π (u,v) als en alleen als π (x,u) $\leqslant$ π (y,v) op moet gaan voor alle viervouden van gewichten (u,v,x,y), en dat de psychofysische schaal f een intervalschaal is. Dat is één kant van de zaak. Een meer praktische aanpak is om de waarschijnlijkheden te schatten met behulp van relatieve frekwenties p(x,y), en de statistiek te gebruiken om na te gaan of het model op de gegevens past, en wat de schattingen van de parameters zijn. De twee aanpakken sluiten elkaar niet uit, maar vullen elkaar aan. De meettheoretische analyse van Falmagne laat zien, dat het Fechneriaanse model equivalent is met een eenvoudige kwalitatieve uitspraak over de waarschijnlijkheden, die gemakkelijker statistisch te toetsen is dan de oorspronkelijke functionele formulering. Het Fechneriaanse model H[f(x) – f(y)] speelt ook een grote rol in de analyse van preferentiegedrag, waarbij we bijvoorbeeld bestuderen hoe groot de kans is dat iemand cigarettenmerk x boven cigarettenmerk y prefereert. Dikwijls gebruiken we dan een veel minder algemene versie van het model, geïntroduceerd door Thurstone (1927), waarin de vorm van H vastgelegd wordt als de cumulatieve normaalverdeling. Dit maakt de statistiek een stuk eenvoudiger, maar het model een stuk willekeuriger.

Een tweede voorbeeld dat we bespreken is van geheel andere aard. Het gaat hierbij om de schaalmethode van Guttman, die rond 1940 ontwikkeld werd. Een groot aantal soldaten in WO II geeft antwoord op een aantal vragen naar angstverschijnselen die ze hadden tijdens gevechtshandelingen. We coderen de antwoorden in een matrix M, en element m(i,j) = 1 als soldaat i verschijnsel j had. Anders m(i,j) = 0. Het model van Guttman zegt nu dat er afbeeldingen f en g van respectievelijk de soldaten en de verschijnselen in de reële getallen zijn zodanig dat m(i,j) = 1 als en alleen als f(i) > g(j). Een soldaat heeft dus een verschijnsel wanneer zijn schaalwaarde rechts van die van het verschijnsel op de getallenrechte ligt. Dat betekent dat verschijnselen en soldaten gezamenlijk geordend moeten kunnen worden in termen van angstigheid. Een soldaat heeft uitsluitend de verschijnselen die meer angstig zijn dan hij. Het model van Guttman is als fundamentele meetmethode geanalyseerd door Falmagne en Doignon (zie Falmagne, 1985, hfst 1). Bijna vanzelfsprekend levert het slechts een ordinale schaal op. Het model kan op een zeer grote hoeveelheid verschijnselen toegepast worden, maar het past zelden goed. De probabilistische versie, waarin π (i,j) de kans is dat soldaat i verschijnsel j heeft, is praktisch van meer belang. We kunnen, geheel analoog aan de Fechnerse psychofysica, veronderstellen dat π (i,j) van de vorm H[f(i) – g(j)] is, waarbij de betekenis van de symbolen nu wel duidelijk is. Dit model speelt ook een grote rol in de testtheorie, waarbij we geinteresseerd zijn in leerlingen i die opgaven j al dan niet goed beantwoorden. Dikwijls wordt hierbij weer de vorm van H nader vastgelegd.

We bekijken nog een ander voorbeeld van fundamenteel meten, al als zodanig bestudeerd sinds Euclides. Het is gebruikelijk de eigenschappen van de driedimensionale Euclidische ruimte af te leiden uit een aantal eenvoudige axiomas en postulaten waaraan de geometrische grootheden zoals punten en lijnen moeten voldoen. In het begin van deze eeuw was er een ware hausse van axiomatische systemen, waarbij getracht werd de systemen eenvoudig te maken in de zin dat er zo weinig mogelijk ongedefinieerde termen moesten bestaan. Het systeem van Pieri (1908) is een goed voorbeeld. De enige ongedefinieerde termen zijn 'punten' en de drievoudige relatie 'x en z liggen even ver van y'. Op basis van zo'n twintig postulaten wordt de gehele Euclidische meetkunde in deze termen gedefinieerd. Nu zijn er in de psychologie verschillende gebieden waarin gelijkenisoordelen een grote rol spelen, bijvoorbeeld het onderzoek naar de cognitieve structuur van kleuren, tonen, geuren, politieke partijen, enzovoorts. Stel dat we onze proefpersonen vragen stellen over kleuren, in de vorm 'lijkt rood even veel op blauw als op groen'. Wanneer de zo verzamelde oordelen aan de axiomas van Pieri voldoen, dan volgt daaruit dat de cognitieve struktuur van de kleuren isomorph is met de driedimensionele Euclidische ruimte, en dus hebben we de kleuren gekwantificeerd. Het voorbeeld is wat eenvoudig gekozen, maar de gedachten erachter zijn hopenlijk duidelijk. Wanneer oordelen van mensen voldoen aan axiomastelsels voor de een of andere meetkunde, dan kunnen ze isomorph in die meetkunde afgebeeld worden. En dat is een voorbeeld van direct of fundamenteel meten. In dit voorbeeld komen we ook, bij de meettheoretische analyse, de techniek tegen om op basis van de primitieven een concatenatie operatie te construeren. Dit gebeurt hier door eerst de relatie te definiëren 'z ligt tussen x en y', mathematisch te representeren als 'f(z) ligt op de lijn die f(x) en f(y) verbindt'. Vervolgens kunnen lijnstukken opgeteld worden. We zien hier dus een voorbeeld van direct meten zonder concatenatie, of liever, zonder concatenatie operatie als primitief gegeven.

Een ander belangrijk voorbeeld van een dergelijke geconstrueerde concatenatie operatie vinden we bij additief conjunct meten. Deze techniek is van belang in evaluatie studies, in marketing, en in multicriteria besliskunde. Het gaat er hierbij om het vergelijken van samengestelde objecten, zoals bijvoorbeeld [3000 gulden in de maand, veel vergaderen, weinig onderzoek] en [1500 gulden in de maand, weinig vergaderen, veel onderzoek]. Het idee bij additief conjunct meten is dat de beoordelaars de diverse aspecten van de objecten, in dit geval salaris, vergaderfrekwentie, en onderzoekstaak, apart schalen en vervolgens bij elkaar optellen om een eindevaluatie van het samengestelde object te verkrijgen. De enige primitieven zijn hier de vergelijkingen van de samengestelde objecten, maar door gebruik te maken van de factoriële structuur kan weer een concatenatie operatie ingevoerd worden. We verwijzen naar bijvoorbeeld Bezembinder (1970), Krantz e.a. (1971), of Narens (1985) voor de details.

We bekijken nu de meting van intelligentie, zowel maatschappelijk als methodologisch van groot belang. De twee belangrijkste bijdragen zijn die van Binet en Spearman, beiden rond 1900. Binet kreeg het belangwekkende idee om de items van een test en de personen die de test invulden gezamenlijk te ordenen, ongeveer zoals bij een Guttman schaal, maar minder expliciet. Op basis van de test resultaten werd de leeftijd van een persoon geschat. Iemand kreeg de mentale leeftijd van 15 jaar als hij zijn test invulde op de manier die typisch was voor vijftienjarigen. Op deze manier konden allerlei soorten tests op een gezamenlijke schaal gebracht worden, een schaal die bovendien boven alle kritiek verheven was omdat het gewoon de leeftijdsschaal was. Hoewel het concept van de mentale leeftijd later losgelaten werd, blijft het idee van het standariseren op leeftijdsnormen bestaan (het wordt ook gebruikt in vele medische groeistudies). De test is een indicator voor intelligentie, maar natuurlijk een pseudo-instrument omdat er geen directe meting van intelligentie mogelijk is. Er is ook geen theorie die zegt hoe de samenhang tussen intelligentie en testresultaat is, tenzij men het extreme operationalisme van sommige psychometrici overneemt, en intelligentie definieert als 'dat wat de intelligentietest meet'. Een poging tot theorievorming werd ondernomen door Spearman, om aan het bezwaar tegemoet te komen dat iedere test een nieuw intelligentiebegrip met zich meebrengt. Volgens Spearman meten alle intelligentietests hetzelfde, te weten intelligentie, maar doen ze dat meer of minder precies en meer of minder zuiver. Door een groot aantal tests uit te kiezen, kan men echter datgene wat ze gezamenlijk hebben steeds beter benaderen. De techniek om intelligentie indirect te meten heet *factor analyse*. Tegenwoordig is het gebruikelijk te beginnen met een probabilistisch model dat de structuur van de testbatterij beschrijft, en de statistiek te gebruiken om de intelligentie te meten. Daar komen veel problemen bij kijken, en het gaat eigenlijk helemaal niet zo goed, maar dat doet hier niet ter zake. Waar het om gaat is dat we, via allerlei omwegen, uitkomen bij een vorm van indirect meten die veel weg heeft van de door Guttman voorgestelde vorm.

Indicatoren zijn buitengewoon populair in de sociologie. Dat komt misschien omdat sociologen wat ongaarne achter hun bureau vandaan komen, en zeer gehecht zijn aan theorieën die ze achter dat bureau uitgedacht hebben. Om de nare discrepantie te verklaren die soms optreedt tussen theorie en uitkomsten van onderzoek, is het handig om een wat losse relatie aan te nemen tussen de theoretische begrippen, die vanzelfsprekend niet observeerbaar zijn, en de gemeten grootheden. Ook in de onderwijskunde, de sociale psychologie, en de differentiële psychologie worden indicatoren voor dit doel overvloedig aangewend. De gemeten grootheden zijn slechts imperfecte indicatoren van de begrippen, voorzien van meetfouten en andere soorten storingen. Op dezelfde manier als intelligentie tests allemaal slechts imperfecte indicatoren zijn voor de intelligentie. Zo zijn er de laatste tientallen jaren indicatoren ontwikkeld voor welbevinden, welzijn, beroepsprestige, sociaal economische status, cultureel kapitaal, prestatiemotivatie, arbeidssatisfactie, inkomenssatisfactie, enzovoorts. Deze indicatoren worden in het algemeen opgenomen in een statistisch model, verwant aan het factor analyse model, dat uitspraken doet over hun onderlinge relaties. Dat model wordt vervolgens aangepast aan de gegevens en getoetst op zijn mate van aanpassing. We zien hier duidelijk indirect meten via pseudo-instrumenten, waarbij de relatie tussen begrip en indicator speculatief is. Evenals bij factor analyse brengt deze manier van meten van de onderliggende grootheden grote gevaren met zich mee.

Als laatste voorbeeld bekijken we nog even het sociogram, volgens de criteria van Bunge een voorbeeld van symbolisatie en niet van meten. We hebben gezien dat de gegevens ondergebracht worden in een binaire matrix, waarbij een één 'ja' betekent en een nul 'nee'. Zo'n matrix kan ook opgevat worden als de cijfermatige representatie van een mathematische structuur die wel een netwerk, graaf, of pijldiagram genoemd wordt. Er is een uitgebreide literatuur die zich met de analyse van dit soort structuren bezig houdt, en met name kan met de sociogrammen gewoon gerekend worden om klieken in de klas te ontdekken, om indices voor populariteit af te leiden, om cognitieve dissonanties te ontdekken, enzovoorts. Of we dit kwantificatie noemen, is niet erg belangrijk. Het is in ieder geval een voorbeeld van mathematisatie, waarbij een andere structuur dan de reële getallen ons goed van pas komt.

6. Samenvattend

We zouden het aantal voorbeelden nog op verschillende manieren uit kunnen breiden. Directe meting op ordinaal niveau wordt in de archeologie gebruikt om vondsten bij opgravingen in de tijd te dateren. Indirecte meting op basis van waarschijnlijkheden wordt in de literatuurwetenschap gebruikt om bijvoorbeeld de werken van Plato of Mark Twain in een betere tijdsvolgorde te zetten. In de experimentele psychologie wordt via directe (en zelfs extensieve) meting bepaald in hoeverre mensen van bepaalde zaken overtuigd zijn. Maar ook zonder deze extra voorbeelden zal de boodschap van deze lezing waarschijnlijk duidelijk zijn.

In de sociale en gedragswetenschappen wordt op zeer veel manieren gekwantificeerd en gemeten. Met veel inventiviteit worden zaken direct of indirect gemeten, op ordinaal of ratio niveau, waarvan de meetbaarheid op het eerste gezicht verre van evident is. Dat levert veel technische problemen op, en in nogal wat gevallen metingen van een belabberde kwaliteit die nauwelijks reproduceerbaar zijn, en die zeker niet via theorie vastzitten aan in principe direct meetbare grootheden. Niettemin is er geen enkele reden om met meten op te houden, of om te veronderstellen dat sommige zaken niet meetbaar zijn.

We hebben geprobeerd in het verloop van deze lezing meetbaarheid steeds te formuleren in termen van een mathematisch model voor de klasse verschijnselen die bestudeerd wordt. Extensief direct meten is een uitermate krachtig en eenvoudig model, en bij lengte en soortgelijke begrippen is er eigenlijk nauwelijks een foutentheorie nodig. Kwantificatie kan in bijna puur algebraïsche termen beschreven worden, zonder statistiek. In de sociale en gedragswetenschappen spelen meetfouten en storingen een dusdanige rol, dat ze direct in de modellen betrokken moeten worden. Hoewel analyse van de algebraïsche component belangrijk blijft, verschuift het accent van meten naar schatten en van algebraïsche condities naar de analyse van aanpassing en residuen. En daardoor van de moderne meetheorie naar de statistiek en data analyse.

Literatuur

Alkemade, C.Th.J., Hoogenboom, A.M. & Smit, J.A. (1979). *Inleiding in de Fysische Meetmethoden*. Utrecht: Bohn, Scheltema & Holkema.

Bezembinder, Th.G.G. (1970). *Van Rangorde naar Continuum*. Deventer: Van Loghum Slaterus.

Bunge, M. (1969). *Scientific Research*. Berlin: Springer.

Bunge, M. (1973). On Confusing 'Measure' and 'Measurement' in the Methodology of Behavioural Science. In M. Bunge (ed), *The Methodological Unity of Science*. Dordrecht: Reidel.

Campbell, N.R. (1920). *Physics: the Elements*. Cambridge: Cambridge University Press.

Campbell, N.R. (1928). *An Account of the Principles of Measurement and Calculation*. London: Longmans-Green.

Coombs, C.H. (1964). *A Theory of Data*. New York: Wiley.

Dingle, H. (1950). A theory of Measurement. *Brit. J. Phil. Sc.*,1,5-26.

Ellis, B. (1966). *Basic Concepts of Measurement*. Cambridge: Cambridge University Press.

Falmagne, J-C. (1985). *Elements of Psychophysical Theory*. Oxford: Oxford University Press.

Hempel, C.G. (1952). Fundamentals of Concept Formation in Empirical Science. *Encyclopedia of Unified Science*, 2, no 7. Chicago: University of Chicago Press.

Krantz, D.H., Luce, R.D., Suppes, P. & Tversky, A. (1971). *Fundations of Measurement*. New York: Academic Press.

Narens, L. (1985). *Abstract Measurement Theory*. Cambridge, Mass.: MIT Press.

Pieri, M. (1908). La Geometria Elementare Instituta sulle Nozioni di 'Punta' e 'Sfera'. *Mem. Mat. Fis. Soc. Ital. Sz*. 15,345-450.

Roskam, E.E.Ch.I. (1981). Methodenleer. In H.C.J. Duijker & P.A. Vroon (red), *Codex Psychologicus*. Amsterdam: Elsevier.

Stevens, S.S. (1946). On the Theory of Scales of Measurement. *Science*, 103, 677-680.

Suppes, P. & Zinnes, J.L. (1963). Basic Measurement Theory. In R.D. Luce, R.R. Bush & E. Galanter (eds), *Handbook of Mathematical Psychology*. New York: Wiley.

Thurstone, L.L. (1927). A law of Comparative Judgement. *Psych. Rev.*, 34, 273-286.

Torgerson, W.S. (1958). *Theory and Methods of Scaling*. New York: Wiley.

Adres van de auteur: Dr.J. de Leeuw, Departments of Psychology and Mathematics University of California Los Angeles.