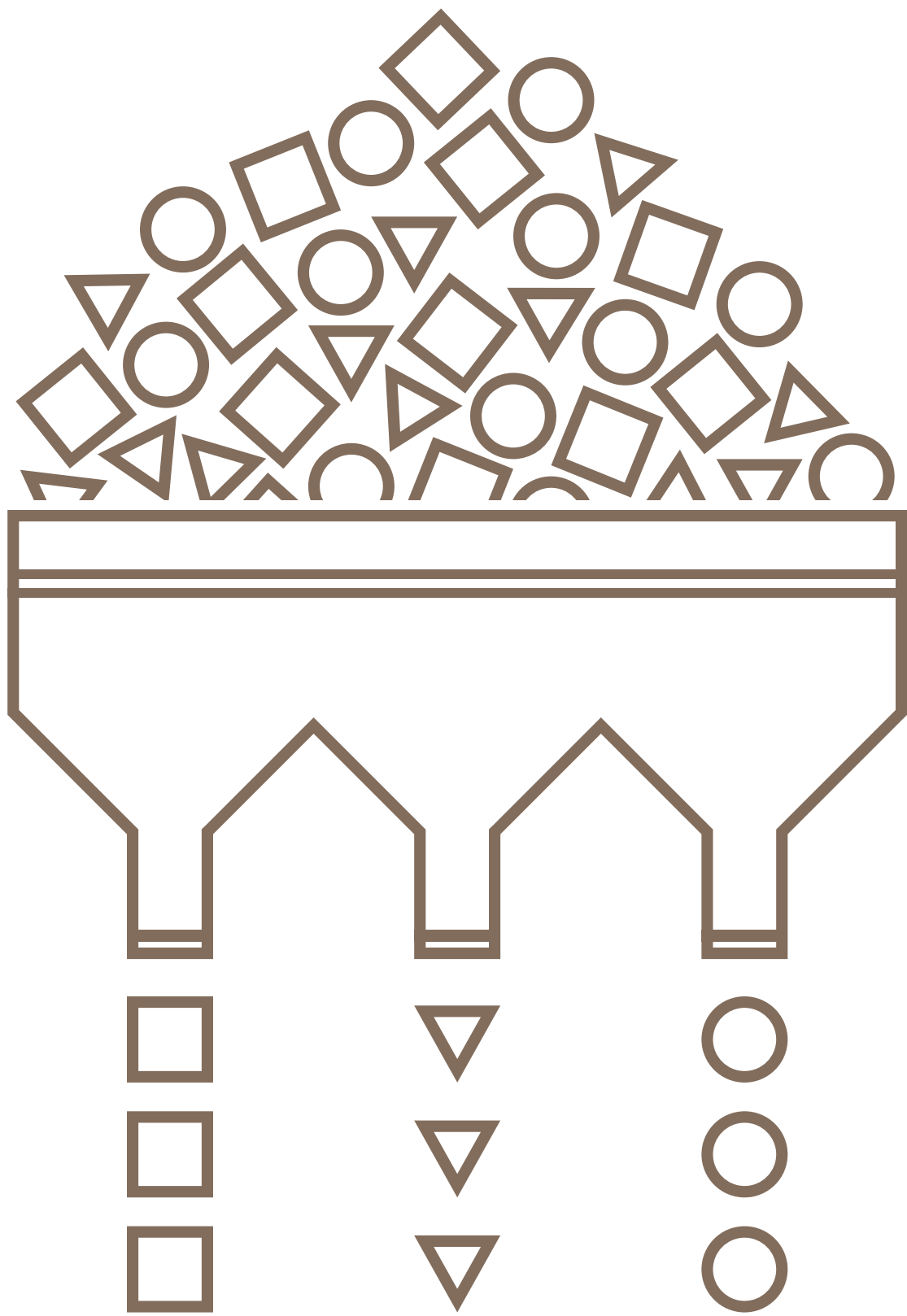




DE RISICO'S VAN CATEGORISCH DENKEN

Bart de Langhe en Philip Fernbach

STRATEGIE

Het is mensen eigen om de wereld in te delen in categorieën. Dat heeft ook een belangrijke functie. Er kleven echter nadelen aan categorisch denken en die worden niet, of te weinig, onderkend. Een verkenning van de risico's en hoe het management de organisatie ertegen kan wapenen.

Zeg eens 'ta'. En zeg nu eens 'da'. Herhaal deze klanken nu, en let er dan steeds op hoe je ze vormt in je mond. Hoe verschilt dat van elkaar?

Dat was een strikvraag! Er is geen verschil. Deze twee klanken klinken niet anders door wat er in je mond gebeurt, maar door de 'inzettijd' van de stem – de tijd tussen het moment waarop je je tong begint te bewegen en het moment waarop je stembanden beginnen te trillen. Als die inzettijd langer duurt dan ongeveer veertig milliseconden, dan horen Engelstaligen 'ta'. Duurt het korter dan die veertig milliseconden, dan horen ze 'da'.

Het verbazingwekkende is dat je nooit iets anders hoort dan 'ta' of 'da'. Wanneer twee Engelstaligen aan dezelfde kant van die scheidslijn van veertig milliseconden zitten, dan maakt het verschil in inzettijd verder niet uit. Bij de één duurt de inzettijd misschien tachtig milliseconden, bij de ander slechts vijftig. Maar in beide gevallen hoor je 'ta'. Zitten ze met hun inzettijd echter aan weerszijden van de scheidslijn, dan kan een verschil van slechts tien milliseconden een wezenlijk

verschil maken. Als de inzettijd van de één 45 milliseconden duurt, hoor je 'ta'. Duurt de inzettijd bij de ander 35 milliseconden, dan hoor je 'da'. Vreemd maar waar.

Op internet was er de laatste tijd nogal wat vertier met de verschillende manieren waarop dergelijke keuzen je voor de gek kunnen houden. Er is bijvoorbeeld een audioclip van een woord dat mensen verstaan als Yanni of als Laurel.¹ Een ander voorbeeld is een jurk die volgens sommigen zwart-blauw gekleurd is, en volgens anderen wit-goud.² In dergelijke gevallen zitten mensen aan de ene of de andere kant van de categorische scheidslijn, net zoals bij 'ta' en 'da', en zijn ze zowat bereid om hun leven in te zetten op het idee dat hun perceptie de 'juiste' is.

De menselijke geest is een categoriseringsmachine. We nemen voortdurend grote hoeveelheden rommelige gegevens in ons op. Die vereenvoudigen en structureren we vervolgens om de wereld om ons heen te kunnen begrijpen. Dit is een van de belangrijkste dingen waartoe onze geest in staat is. Het is ongelooflijk

waardevol om in één oogopslag te kunnen zien of iets een slang is, of een stok.

Om waardevol te kunnen zijn, moet een categorisering echter voldoen aan twee voorwaarden. Ten eerste moet die categorisering geldig zijn. Je kunt een homogene groep niet willekeurig opsplitsen. Geldige categorieën ‘snijden de natuur bij de gewrichten’, zei Plato. Iets is een slang of het is een stok. Ten tweede moet de categorisering nuttig zijn. Twee verschillende categorieën moeten zich ook verschillend gedragen op een manier die voor ons relevant is. Het is nuttig om slangen te onderscheiden van stokken, want dan overleef je een wandeling in het bos.

Tot dusver geen discussie, lijkt ons. Kijken we dan echter naar het bedrijfsleven. Daar onderscheiden we vaak categorieën waarop we vervolgens vertrouwen terwijl ze niet geldig of niet nuttig zijn, of zelfs geen van beide – en dat kan leiden tot grote fouten in de besluitvorming.

Neem bijvoorbeeld de ‘Myers-Briggs Type Indicator’,³ een persoonlijkheidstest die volgens de uitgever ervan bij ruim vier op de vijf Fortune 500-bedrijven een rol speelt in de besluitvorming op personeelsgebied. De Myers-Briggs-test bestaat uit 93 vragen met steeds twee mogelijke antwoorden. Op basis van de gekozen antwoorden wordt degene die de test neemt ingedeeld in één van de zestien persoonlijkheids categorieën die worden onderscheiden in de Myers-Briggs-methodiek. Het probleem is dat deze vragen een ingewikkelde en continue beoordeling vereisen. Neem bijvoorbeeld de volgende vraag: ‘Ga je eerder af op feiten of op je intuïtie?’ De meesten van ons zouden waarschijnlijk antwoorden: ‘Nou, dat hangt ervan af.’ Maar dat antwoord kun je in de test niet geven. Je moet dus het ene of het andere kamp kiezen en een keuze maken die je misschien niet opnieuw zou maken als je de test nog een keer deed. De antwoorden op de vragen worden samengevat en je krijgt het etiket ‘extravert’ in plaats van ‘introvert’, of ‘beoordelaar’ in plaats van ‘waarnemer’ opgeplakt. Deze categorisering is eenvoudigweg niet geldig. De test is ook niet nuttig; het persoonlijkheidstype voorspelt geen resultaten zoals werksucces en tevredenheid.

Waarom is Myers-Briggs dan zo populair? Omdat categorisch denken krachtige illusies genereert. Maar

er kleven vier gevaarlijke risico’s aan: je kunt de leden van een categorie als uniform gaan beschouwen en hen behandelen alsof ze meer op elkaar lijken dan ze in werkelijkheid doen; je kunt de verschillen tussen de leden van verschillende categorieën overdrijven; je kunt bepaalde categorieën discrimineren of voortrekken; en je kunt ‘verstarren’ in je manier van kijken, doordat je de categorische structuur die je hebt opgelegd behandelt alsof die statisch is.

UNIFORME BLIK

Wanneer je dingen indeelt in categorieën denk je in termen van prototypen. Daarmee riskeer je echter dat je al snel de veelheid aan variaties vergeet die er bestaan binnen een categorie die je hebt afgebakend.

De mythe van de doelklant

In zijn boek *The End of Average* vertelt Todd Rose dat een krant in Cleveland in 1945 een wedstrijd uitschreef om de anatomisch prototypische vrouw te vinden. Kort daarvoor waren in een onderzoek de gemiddelde waarden bepaald voor diverse anatomische metingen; de redactie van de krant gebruikte die waarden om haar prototype te definiëren. In totaal stuurden 3.864 vrouwen hun metingen in. Hoeveel van hen zaten voor alle aspecten dicht bij het gemiddelde, denkt u?

Dat was er niet één. Mensen verschillen op zoveel manieren van elkaar dat het hoogst onwaarschijnlijk is dat één enkele persoon op alle aspecten in de buurt komt van het gemiddelde.

Zo is het ook met klanten. Bedenk wat er gebeurt in segmenteringsonderzoek, een van de meest gebruikte instrumenten van marketingafdelingen. Bij segmentering verdeel je je klanten in categorieën en bepaal je vervolgens je doelklanten – dat wil zeggen, de categorie die speciale aandacht en strategische focus verdient.

In een segmenteringsonderzoek wordt klanten om te beginnen meestal gevraagd naar hun gedrag, wensen en demografische kenmerken. Een clustering-algoritme verdeelt de respondenten vervolgens in groepen op basis van overeenkomsten in de antwoorden. Dit soort analyse levert zelden sterk gedifferentieerde

categorieën op. Maar in plaats van serieus te onderzoeken of de clusters valide zijn, voeren de marketeers gewoon de volgende stappen van het segmenteringsproces uit: ze bepalen gemiddelde waarden, stellen profielen op en formuleren typen klanten.

Zo ontstaan ‘ruimtewagenmoeders’ en meer van dat soort categorieën. Nadat een enquête is uitgevoerd, benoemt iemand bij Marketing een interessant uitziend cluster waarin bijvoorbeeld zestig procent van de respondenten vrouw is, met een gemiddelde leeftijd van begin veertig, en met gemiddeld 2,75 kinderen. Kijkend naar deze gemiddelden drijf je dan gemakkelijk af van de gegevens en ga je denken in termen van een prototypische klant die precies die eigenschappen heeft: de ruimtewagenmoeder.

Dergelijke etiketten maken ons blind voor de variatie die er binnen categorieën bestaat. Een voorbeeld. In een onderzoek dat in 2011 werd uitgevoerd kregen de deelnemers twee versies te zien van een figuur bestaande uit negen silhouetten van vrouwen (Froni & Rothbart, 2011). De ene versie bestond uit enkel de naast elkaar geplaatste silhouetten (versie ‘A’ in figuur 1). In de andere versie waren de silhouetten uitgezet op een schaal die de Body Mass Index (BMI) voorstelde, met daaronder de etiketten ‘anorexisch’, ‘normaal’ en ‘obees’ (versie ‘B’ in figuur 1).

De deelnemers werd voor iedere versie gevraagd de beelden op verschillende aspecten te beoordelen. De vrouwen in de figuur met de etiketten bleken anders te worden beoordeeld dan in de figuur zonder etiketten, hoewel het om precies dezelfde silhouetten ging. De deelnemers gaven bijvoorbeeld aan dat de persoonlijkheid en levensstijl van vrouw nr. 7 meer leek op die van vrouw nr. 9 toen die allebei als ‘obees’ werden bestempeld. Evenzo werden de silhouetten nr. 4 en nr. 6 meer aan elkaar gelijkgesteld toen ze beide als ‘normaal’ werden bestempeld.

Net als bij soorten lichamen zijn de segmenten waarmee de meeste bedrijven werken niet zo duidelijk als ze lijken. De klanten in één bepaald segment gedragen zich onderling vaak heel verschillend. Om de effecten van een uniforme blik te vermijden, zouden analisten en managers zich kunnen afvragen hoe groot de kans is dat twee klanten uit verschillende clusters meer op

FIGUUR 1. ETIKETTEN MAKEN BLIND VOOR VARIATIE BINNEN CATEGORIEËN – VOORBEELD



Bron: Ontleend aan Froni & Rothbart, 'Category Boundaries and Category Labels', Social Cognition (Guildford Press, 2011).

elkaar lijken dan twee klanten uit hetzelfde cluster. Hoe groot is bijvoorbeeld de kans dat het favoriete kledingmerk van een ruimtewagenmoeder meer lijkt op dat van een hippiemoeder dan op dat van een andere ruimtewagenmoeder? Die kans ligt vaak dichterbij de vijftig procent dan bij nul procent.

Het screeningseffect

Een uniforme blik kan ook misleidend werken bij werving en selectie. Neem een manager die verantwoordelijk is voor de werving in zijn bedrijf. Hij heeft onlangs een vacature geplaatst waarop twintig mensen hebben gereageerd. De manager voert een eerste screening uit, rangschikt de kandidaten op basis van hun technische vaardigheden en nodigt de vijf hoogstgeplaatste kandidaten uit voor een gesprek.

Hoewel de technische vaardigheden van de vijf kandidaten sterk uiteenlopen, wordt de uiteindelijke beslissing van de manager daar niet sterk door beïnvloed. Wanneer de kandidaten eenmaal op basis van hun technische kunnen zijn geselecteerd, worden degenen die dóór zijn naar de volgende ronde van de procedure als gelijkwaardig beschouwd op dat punt. Aldus beïnvloed door categorisch denken neemt de manager

zijn uiteindelijke beslissing vooral op basis van de zachte vaardigheden die de kandidaten tonen in de gesprekken: hoe aardig ze zijn, hoe effectief ze communiceren enzovoort. Dat zijn natuurlijk allemaal belangrijke aspecten, maar voor veel functies is het vooral belangrijk om de allerbeste technische vaardigheden in huis te halen. Door het screeningseffect raakt dat aspect echter juist buiten beeld.

Afwijkingen bij beleggingen

De uniforme blik komt ook voor in de financiële markten. Beleggers rangschikken hun activa grofweg naar omvang (small-cap- of large-cap-aandelen), sector (bijvoorbeeld energie of gezondheidszorg), geografie enzovoort. Deze classificaties maken het mogelijk om het grote aantal beschikbare beleggingsmogelijkheden te doorgronden. Dat is ook belangrijk. Maar diezelfde classificaties leiden er wel toe dat beleggers hun kapitaal inefficiënt toewijzen in termen van risico en rendement. Tijdens de internetzeepbel van eind jaren negentig staken beleggers bijvoorbeeld veel geld, en vrijwel meteen, in bedrijven die een 'dot-com'-naam hadden aangenomen, ook al was er verder niets aan die bedrijven veranderd. Die fout heeft veel beleggers veel geld gekost (Cooper e.a., 2001).

Nog een voorbeeld: wanneer het aandeel van een onderneming wordt opgenomen in de S&P 500, tendeert de koers ervan in de richting van de aandelenkoersen van andere ondernemingen in die index, ook al is er eigenlijk niets aan die onderneming en het aandeel veranderd.

OVERDRIJVING

Categorisch denken zet ertoe aan om verschillen tussen categorieën te overdrijven. Dat kan ertoe leiden dat mensen uit andere groepen worden gezien als stereotypisch, dat er in de besluitvorming willekeurige drempelwaarden worden gekozen, en dat er foute conclusies worden getrokken.

Groepsdynamiek

Overdrijving kan ernstige gevolgen hebben wanneer dat beïnvloedt hoe je denkt over de leden van sociale of politieke groepen. Uit onderzoek is gebleken dat mensen verbonden aan tegenovergestelde politieke partijen geneigd zijn om elkaars standpunten extremer in te schatten dan ze zijn.

Wie geeft er volgens u meer om sociale gelijkheid: progressieve of conservatieve partijen? U zegt de progressieve partijen? Dat klopt: gemiddeld vinden progressieve politici sociale gelijkheid belangrijker dan conservatieve politici. Sommige conservatieven geven

echter meer om sociale gelijkheid dan sommige progressieven. Stel dat we twee willekeurige mensen op straat nemen: eerst iemand die conservatief stemt, en dan iemand die progressief stemt. Hoe groot is de kans dat de eerste persoon sociale gelijkheid belangrijker vindt dan de tweede? Veel dichter bij de vijftig procent dan je zou denken.

Gemiddelden verhullen de overlap tussen groepen, waardoor de waargenomen verschillen worden versterkt. Ondanks het gemiddelde in dit geval geven veel conservatieven meer om sociale gelijkheid dan veel progressieven.

Als je een *liberal* (progressief denkend iemand) bent in de Verenigde Staten, ga je er waarschijnlijk vanuit dat alle conservatieven tegen abortus, wapenbeheersing en het sociale vangnet zijn. Ben je conservatief, dan denk je waarschijnlijk dat alle *liberals* open grenzen en universele gezondheidszorg van de overheid willen. In werkelijkheid is er een breed spectrum aan ideologieën en standpunten.

Overdrijving als gevolg van categorisch denken is vooral zorgwekkend in deze tijd van big data en profilering van klanten. Van Facebook, bijvoorbeeld, is bekend dat het politieke etiketten toekent aan zijn gebruikers op basis van hun surfgeschiedenis ('gematigd', 'conservatief', of 'progressief' (*liberal*)) (Merrill, 2016), en die informatie deelt met adverteerders. Adverteerders denken dan misschien dat de mensen in de categorieën gebruikers van Facebook sterker van elkaar verschillen dan in werkelijkheid het geval is. Dat

VROEGER HOEFDE
CATEGORISCH DENKEN
EEN BEDRIJF NIET
FATAAL TE WORDEN,
MAAR DAT VERANDERT
NU DOOR DE
DATA REVOLUTIE

kan de werkelijke verschillen juist vergroten, doordat het adverteerders stimuleert om uitdrukkelijk op maat gemaakte boodschappen te sturen naar elke groep. Dat is wat er in 2016 lijkt te zijn gebeurd tijdens de campagnes voor het Brexit-referendum in het Verenigd Koninkrijk en voor de presidentsverkiezingen in de Verenigde Staten: Facebook stuurde ‘conservatieven’ en ‘progressieven’ (*liberals*) duizenden berichten waarmee verdeeldheid werd gezaaid (Singer, 2018).

Veel bedrijven worstelen intern met een vergelijkbare overdrijvingsdynamiek. Goede bedrijfsresultaten vereisen vaak synergie tussen de verschillende bedrijfsonderdelen. Categorisch denken kan er echter toe leiden dat het vermogen van teams uit verschillende delen van de onderneming om met elkaar samen te werken zwaar wordt onderschat. Als het topmanagement bijvoorbeeld denkt dat de datawetenschappers in de organisatie weliswaar veel technische expertise hebben maar niet goed begrijpen hoe de onderneming in elkaar zit, en dat de marketingmanagers wel veel weten van marketing maar geen raad weten met grote hoeveelheden aan gegevens, dan zal men niet snel een team samenstellen met experts uit die beide domeinen. Dat verklaart onder andere waarom zoveel analytische initiatieven mislukken.

Besluitvorming

Overdrijving beïnvloedt de besluitvorming ook op subtielere manieren. Neem bijvoorbeeld coaches in de Amerikaanse basketbalcompetitie NBA. De kans dat die hun opstelling aanpassen na een nipte nederlaag (100-101) is zeventien procent groter dan wanneer hun team nipt heeft gewonnen (100-99), ook al bedraagt het verschil in de scores van het andere team slechts twee punten (Lefgren e.a., 2015). Bij een nederlaag met 100-106 in plaats van 100-108 blijken coaches echter veel minder geneigd om de opstelling te veranderen, ook al bedraagt het verschil nog steeds maar twee punten. Een nederlaag voelt kwalitatief anders aan dan een overwinning omdat uitslagen van sportwedstrijden niet worden beschouwd in termen van een continuüm.

Wanneer je een beslissing neemt op basis van een afbakening in een continuüm, zul je kleine verschillen waarschijnlijk extra aanzetten. Tijdens de financiële crisis van 2008 greep de Belgische overheid in bij Fortis om de bank te redden en kreeg daardoor vervolgens een belang van miljoenen aandelen in BNP Paribas. Volgens de Belgische krant *De Standaard* besloot de regering eind januari 2018, toen de koers van het aandeel BNP Paribas iets meer dan 67 euro bedroeg, om haar aandelen te verkopen zodra de koers tot 68 euro was gestegen. Dat laatste gebeurde echter niet: de koers daalde scherp en de aandelen zijn nu slechts 44 euro waard.

Niemand in de Belgische regering had kunnen voorspellen dat de aandelenkoers zo sterk zou dalen. De regering beging echter de fout om van de verkoop van haar aandelen een ‘alles of niets’-kwestie te maken. Een betere aanpak zou zijn geweest om een deel van het pakket tegen één prijs te verkopen, een ander deel tegen een tweede prijs enzovoort.

VEEL BEDRIJVEN MISSEN DE BENODIGDE KENNIS VOOR GROOTSCHALIGE GEGEVENSANALYSE, MAAR DAT IS OP TE LOSSEN

Statistische significantie

Met de toenemende invloed van de gedragseconomie en de gegevenswetenschap vertrouwen bedrijven steeds meer op A/B-testen om de effectiviteit van keuzen in te schatten. Eén van de redenen is dat A/B-testen eenvoudig te implementeren en te analyseren zijn: je maakt twee versies van de wereld die op één element na identiek zijn aan elkaar; je laat één groep deelnemers versie A ervaren, en een andere groep versie B; vervolgens meet je of het gedrag van de groepen wezenlijk verschilt. Nu zal er door het toeval altijd een verschil optreden tussen de beide groepen, zelfs al heeft de manipulatie (het verschil tussen de beide versies) geen enkel effect gehad. Om te bepalen of het verschil groot genoeg is om te kunnen worden toegeschreven aan de manipulatie, wordt dan een statistische test losgelaten op de proef. De uitkomst van de test is de kans dat je een verschil van een dergelijke grootte zou hebben waargenomen als de manipulatie geen effect had gehad. Deze kans staat bekend als de p-waarde. Naarmate deze p-waarde dichterbij nul ligt, kun je met meer vertrouwen concluderen dat een

eventueel verschil kan worden toegeschreven aan het aspect dat je hebt gemanipuleerd, en niet alleen aan het toeval. Maar hoe dicht bij nul is dicht genoeg?

In 1925 besloot Sir Ronald Fisher, een Britse statisticus en geneticus, willekeurig dat .05 een handige drempelwaarde was. Fisher had net zo goed .03 kunnen kiezen. Hij adviseerde ook dat de drempel voor de p-waarde afhankelijk moest worden gesteld van de specifieke kenmerken van het betreffende onderzoek. Van dat laatste advies werd echter nauwelijks notie genomen, en in de decennia die volgden namen complete wetenschappelijke disciplines blindelings de waarde .05 over als de magische grens die signaal en ruis van elkaar scheidt; dit is ook de norm geworden in de dagelijkse gang van zaken in het bedrijfsleven.

Dat is een probleem. Wanneer een A/B-test een p-waarde van .04 oplevert, kan een interventie worden gepleegd. Maar is de uitkomst .06, dan wordt wellicht besloten om niets te doen – ook al heeft het verschil tussen $p=.04$ en $p=.06$ op zich geen betekenis. Maar het wordt nog erger: veel onderzoekers kijken geregeld naar binnenkomende gegevens om de statistische significantie ervan te testen en houden op met het verzamelen van gegevens zodra ze een p-waarde onder de .05 zien. Dat vergroot de kans dat men tot de conclusie komt dat een interventie zin heeft terwijl dat in feite niet het geval is. In een recent onderzoek (Berman e.a., 2018) is gekeken naar de werkwijze van onderzoekers die een populair online-platform gebruikten om A/B-tests uit te voeren. De meesten van hen bleken aan dergelijke ‘p-hacking’ te doen, en het aantal foutieve bevindingen steeg daardoor van 33 naar 42 procent.

DISCRIMINATIE

Als je eenmaal een categorische structuur hebt ingevoerd, ontwikkel je de neiging om de voorkeur te geven aan bepaalde categorieën ten opzichte van andere categorieën. Tekortschietende aandacht voor andere categorieën kan echter schadelijk zijn.

DE SEGMENTEN WAAR DE MEESTE BEDRIJVEN MEE WERKEN ZIJN NIET ZO DUIDELIJK ALS ZE LIJKEN

Te scherpe ‘targeting’

Laten we ons eens verplaatsen in de rol van een directeur digitale marketing van een webwinkel voor interieurartikelen met unieke en creatieve ontwerpen. We voeren een segmenteringsonderzoek uit en identificeren een klantsegment met de volgende kenmerken: mannelijke professionals in de leeftijd van 18 tot 34 jaar met creatieve banen in de mode, de marketing of de media, en met een middelmatig besteedbaar inkomen. We kunnen tienduizend dollar besteden aan digitale advertenties en overwegen drie strategieën: (1) Geen ‘targeting’. De advertentie wordt met een gelijke mate van waarschijnlijkheid verstuurd naar alle Facebook-gebruikers en kost 40 cent per klik. (2) Volledig ‘getarget’. De advertentie gaat uitsluitend naar leden van de doelgroep en kost 60 cent per klik. (3) Gedeeltelijke ‘targeting’. In dit geval investeren we de helft van ons budget in marketing aan het doelsegment en de andere helft in massamarketing; in dit geval betalen we 48 cent per klik.

Wat is de beste keuze? Waarschijnlijk B of C want dan kunnen we onze doelgroep scherper afbakenen. Toch?

Fout. De beste optie is waarschijnlijk A, de breedste markbenadering. Waarom? Omdat een brede ‘targeting’ vaak een hogere ROI oplevert dan een scherper afgebakende ‘targeting’. Uit onderzoek is gebleken dat online-advertenties de kans op een aankoop met slechts een fractie van een procent verhogen (Gordon e.a., 2018). Als de kans dat iemand je product koopt zonder een advertentie te hebben gezien 0,10 procent bedraagt, dan stijgt die kans door het zien van een advertentie tot 0,13 procent. Het positieve effect van de advertentie kan een beetje groter zijn voor de doelgroep, maar in veel gevallen zal dat niet opwegen tegen de extra kosten per klik. Marketeers raken echter geobsedeerd door de klanten in hun doelgroep en negeren de waarde die andere klanten kunnen genereren.

Facebook heeft zich ingespannen om adverteerders het belang van bereik ten opzichte van smalle ‘targeting’ duidelijk te maken. Het haalt het geval aan van

een biermerk dat traditioneel was gericht op mannen. Toen het merk verhuisde naar digitale mediaplatforms kon het zijn ‘targeting’ scherper afbakenen. Dat leek een goede zaak, maar het bereik van de campagnes werd er sterk door beperkt en de resultaten verslechterden. Na enig onderzoek beseftte de brouwer dat een aanzienlijk deel van zijn omzet werd gegenereerd door vrouwen. Toen de ‘targeting’ en creatieve boodschap werden verbreed, leverde dat onmiddellijk positieve resultaten op.

Net Promoter Score

Discriminatie kan de interpretatie van gegevens verstoren. Wanneer wij college geven in gegevensanalyse, vragen we onze studenten vaak of ze hebben gehoord van de Net Promoter Score (NPS) en of hun bedrijven die maatstaf op de een of andere manier gebruiken. Meestal steken de meesten in de zaal dan hun hand omhoog, en met reden. Nadat Frederick Reichheld het concept had geïntroduceerd in de *Harvard Business Review* (Reichheld, 2003) werd het al snel een van de belangrijkste succescriteria in het bedrijfsleven, en dat is het nog steeds.

Wat is de NPS en hoe werkt die? Je vraagt aan je klanten (of medewerkers) om op een schaal van 0 tot en met 10 aan te geven hoe groot de kans is dat zij jouw bedrijf zullen aanbevelen bij familie of vrienden. Een score van 0 betekent ‘volstrekt onwaarschijnlijk’ en 10 staat voor ‘zeer waarschijnlijk’. Nadat ze de vraag hebben beantwoord, worden de klanten verdeeld in drie categorieën, namelijk: *detractors* (scores 0 tot en met 6), *passives* (scores 7 en 8) respectievelijk *promoters* (scores 9 en 10). De totale NPS wordt bepaald door het percentage klanten in elke categorie te berekenen en vervolgens het percentage *detractors* af te trekken van het percentage *promoters*. Wanneer 60 procent van de klanten *promoters* zijn en 10 procent *detractors*, dan is je NPS 50 procent.

Er zijn goede redenen om met NPS te werken. Het is een duidelijke en gemakkelijk te begrijpen maatstaf. Ook helpt het de neiging tot overdrijven (*amplification bias*) te voorkomen die eigen is aan categorisch denken.

Reichheld zelf sprak in het voornoemde artikel van ‘*grade inflation*’, de vertekening die optreedt in uitkomsten van traditionele enquêtes over klanttevredenheid doordat iedereen die ook maar een fractie hoger dan ‘neutraal’ scoort meteen belandt in de bak ‘tevreden’.

Dat is dus een nuttig effect van de NPS. Tegelijkertijd vertoont de NPS-systematiek zelf de neiging tot overdrijving waartegen het bedrijven juist zou moeten wapenen. Klanten die bijvoorbeeld een 6 scoren, zitten veel dichterbij een 7 dan bij een 0. Toch worden ze op één hoop gegooid met de *detractors* en niet ingedeeld bij de *passives*. Met andere woorden: kleine verschillen tussen verschillende categorieën hebben een grote invloed op de totaalscore, maar vergelijkbare of nog grotere verschillen binnen één enkele categorie niet.

De NPS heeft nog een ander probleem gerelateerd aan categorisch denken: de score houdt geen rekening met het aantal *passives*. Laten we eens twee extreme onderzoeksuitkomsten nemen: één bedrijf heeft nul procent *detractors* en nul procent *promoters*, en een ander bedrijf heeft vijftig procent *detractors* en vijftig procent *promoters*. De NPS is voor beide bedrijven hetzelfde, maar hun klantenbestand is duidelijk zeer verschillend en moet op verschillende manieren worden benaderd.

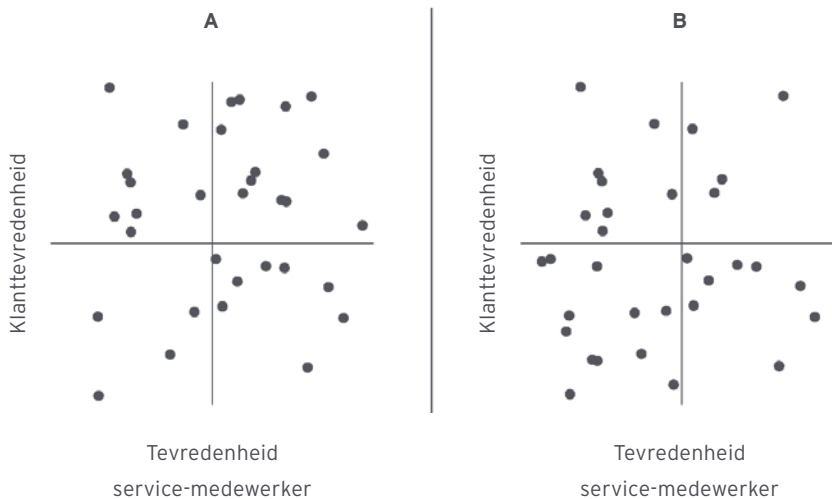
Eenzijdige interpretatie van correlaties

Categorisch denken kan ook leiden tot een vertekening in de interpretatie van gegevens. Een voorbeeld: de manager van een serviceafdeling denkt dat de tevredenheid van haar medewerkers de klanttevredenheid beïnvloedt en laat daar onderzoek naar doen. Een paar weken later krijgt ze van de analyse-afdeling van Personeelszaken de gegevens toegestuurd, weergegeven in een scatter-diagram (beeld A in figuur 2).

Hoe zal deze manager de samenhang tussen medewerker- respectievelijk klanttevredenheid inschatten? De meeste mensen zouden in dit diagram een matige samenhang zien. Maar wat nu als het onderzoek een andere uitkomst had opgeleverd, zoals weergegeven in beeld B van figuur 2? Hoe zou de samenhang tus-

MARKETEERS HEBBEN
DE NEIGING OM
HUN AANDACHT TE
CONCENTREREN OP
HUN DOELKLANTEN EN
DE REST TE NEGEREN

FIGUUR 2. WEL OF GEEN CORRELATIE?



Bron: auteurs.

sen medewerker- respectievelijk klanttevredenheid in dat geval worden ingeschat? De meeste mensen zien in beeld B een veel zwakkere of helemaal geen samenhang. Maar in werkelijkheid is de samenhang van 2B ongeveer gelijk aan die van 2A. De scatter-diagrammen zijn identiek, behalve acht datapunten die van het kwadrant rechtsboven in het eerste beeld (2A) zijn verplaatst naar het kwadrant linksonder in het tweede (2B).

Waarom zien mensen dan een sterkere samenhang in de eerste grafiek (2A)? Omdat mensen geneigd zijn de voorkeur te geven aan het kwadrant rechtsboven. In de eerste grafiek (2A) ziet men veel tevreden medewerkers en tevreden klanten, dus concludeert men dat de correlatie vrij sterk is. In de tweede grafiek (2B) ziet men weinig tevreden medewerkers met tevreden klanten en concludeert dat de correlatie zwakker is. Hier kunnen we uit leren dat als we niet evenveel aandacht besteden aan alle categorieën, we relaties tussen variabelen minder goed waarnemen.

VERSTARRING

Categorieën leiden tot een vaststaand wereldbeeld. Ze geven ons het gevoel dat dit is hoe de dingen zijn, en

niet hoe iemand heeft *gekozen* de (blik op de) wereld te *organiseren*. John Maynard Keynes bracht dat ooit heel mooi onder woorden toen hij schreef: 'Het is niet zozeer moeilijk om nieuwe ideeën te bedenken als wel om te ontsnappen aan de oude.'

In de jaren vijftig werd de Amerikaanse markt voor fietsen gedomineerd door de Schwinn Bicycle Company. Schwinn richtte zich op de markt voor kinderen en bouwde zware, verchroomde fietsen met grote wielen waarmee kinderen door hun buurt konden fietsen. Maar in de jaren vijftig tot zeventig veranderde de markt sterk. Veel volwassenen begonnen te fietsen als sport en zochten lichtere fietsen waarmee betere prestaties konden worden behaald. Schwinn paste zich niet aan en de Amerikaanse consument koos steeds vaker voor Europese en Japanse fietsen; Schwinn belandde in een langdurige en pijnlijke, en uiteindelijk fatale, neergang. De decennialange succesvolle verkoop van fietsen aan kinderen had de kijk van het bedrijf op de markt doen verstarren, met als gevolg dat men de drastische verandering in de vraag niet door had.

Innovatie

Bij innovatie gaat het erom af te rekenen met de neiging tot categorisch denken. Veel bedrijven proberen

hun bedrijfsvoering efficiënter te maken met behulp van categorisering. Ze wijzen taken toe aan mensen, mensen aan afdelingen enzovoort. Dergelijke disciplinaire grenzen dienen een doel maar gaan ook ten koste van dat doel. De uitdagingen in de markt van morgen passen niet netjes binnen de afbakeningen die ooit werden gecreëerd met het oog op de uitdagingen van gisteren. Denken in termen van bestaande categorieën kan ook vertragend werken op de ontwikkeling van nieuwe kennis omdat mensen dan worden belemmerd om elementen op nieuwe manieren met elkaar te combineren.

Illustratief in dit verband is wat onderzoekers aan de universiteit van Toronto in 2016 ontdekten toen ze ongeveer tweehonderd mensen vroegen om met legosteentjes een buitenaards wezen te bouwen. De deelnemers werden verdeeld in drie groepen. Eén groep kreeg een assortiment steentjes dat in groepen was onderverdeeld, een tweede groep kreeg een willekeurig samengesteld assortiment steentjes, en een derde groep moest het werk van de beide groepen beoordelen op creativiteit. De derde groep wees de tweede groep, die met het willekeurig samengestelde assortiment legosteentjes, aan als het meest creatief.

Versterkte categorieën kunnen ook nog op een andere manier een belemmering vormen voor innovatie. Ze kunnen ons namelijk belemmeren ons voor te stellen hoe objecten (of ideeën) op een ongebruikelijke manier zouden kunnen worden gebruikt. Dit is het probleem van *functionele fixatie*. Wat doe je wanneer iemand jou een schroef en een Engelse sleutel geeft en vraagt die schroef in de wand te bevestigen? Dan klem je die schroef misschien in de Engelse sleutel en probeert hem vervolgens borend in de muur te schroeven – wat behalve een hoop frustratie waarschijnlijk niet veel resultaat oplevert. De meest effectieve aanpak – de Engelse sleutel gebruiken als een hamer om de schroef in de wand te slaan – komt misschien niet eens bij je op.

HOE BEPERKEN WE DE RISICO'S VAN CATEGORISCH DENKEN?

Hoe kun je nu, als weldenkend leider, de schade vermijden die wordt veroorzaakt door categorisch denken? Daarvoor hebben wij een proces bedacht dat uit vier stappen bestaat:

1. Vergroot het bewustzijn omtrent categorisch denken

We denken allemaal categorisch, en dat heeft een reden. Iedereen die betrokken is bij besluitvorming moet zich echter bewust zijn van de risico's van categorisch denken: het zet aan tot verleidelijke simplificaties en vertekeningen, wekt de indruk dat de problematiek gemakkelijk te doorgronden is, en werkt onzichtbare vooroordelen in de hand. De bedrijven die deze valkuilen het beste weten te vermijden, zijn de bedrijven die hun werknemers helpen om zich meer op hun gemak te voelen met onzekerheid, nuance en complexiteit. Is een indeling in categorieën geldig? En is die bovendien ook nuttig? Die vragen horen een vast bestanddeel te zijn van de mantra van de besluitvorming.

HET POSITIEVE
EFFECT VAN
DE ADVERTENTIE KAN
EEN BEETJE GROTER
ZIJN VOOR
DE DOELGROEP,
MAAR DAT WEEGT
VAAK NIET OP TEGEN
DE EXTRA KOSTEN
PER KLIK

2. Ontwikkel het vermogen om aanhoudend gegevens te analyseren

Om de fouten in de besluitvorming veroorzaakt door categorisch denken te vermijden, is een goede continue gegevensanalyse van essentieel belang. Veel ondernemingen missen echter de benodigde kennis daarvoor. Voor de segmentering, bijvoorbeeld, besteden ze de *analytics* uit aan gespecialiseerde bedrijven – maar vervolgens interpreteren ze de ingekochte informatie verkeerd. Dit kan betrekkelijk eenvoudig worden opgelost. Met een beetje training kunnen beproefde maatstaven voor het evalueren van de validiteit van een bepaald segment worden toegepast. Elk bedrijf dat in het kader van zijn marketingonderzoek of strategische planning de markt segmenteert, zou dergelijke maatstaven moeten gebruiken en dergelijke trainingen moeten doorlopen. Ze vormen een

gouden kans voor slimme organisaties om interne expertise te ontwikkelen en concurrentievoordeel op te bouwen.

3. *Kijk kritisch naar de criteria voor besluitvorming*

Veel bedrijven besluiten dat ze pas in actie komen zodra ze een arbitraire drempel op een continuüm hebben overschreden. Dit heeft twee nadelen.

Om te beginnen vergroot het je risico. Stel dat een bedrijf aan de hand van een marktonderzoek te weten wil komen of een nieuw product kans van slagen heeft. Misschien wordt het groene licht gegeven voor de marktintroductie als uit een grootschalig onderzoek blijkt dat de reacties in de markt een vooraf bepaalde drempel halen, of als de uitkomsten van een experiment een p-waarde opleveren die kleiner is dan het magische getal .05. Omdat echter het verschil tussen een toevallige treffer of misser van de drempelwaarde minuscuul is, wordt die drempelwaarde misschien alleen maar gehaald als gevolg van een willekeurige variatie in de steekproef of doordat er een bepaalde kleine nadruk zit in de gekozen methode voor het verzamelen van de gegevens. Een klein en volstrekt betekenisloos verschil kan dan leiden tot een drastisch andere beslissing – en misschien wel de verkeerde. Dat zagen we, zoals gezegd, in het voorbeeld van de Belgische regering en het gekozen koersniveau voor de verkoop van de aandelen dat niet werd bereikt. In een dergelijke situatie is een gefaseerde aanpak veel verstandiger. De Belgen hadden zich bij hun desinvestering kunnen laten leiden door het gewicht van de feiten in plaats van vast te houden aan dat ene koersdoel.

Ten tweede kan een arbitraire drempelwaarde belemmerend werken voor het lerend vermogen van de organisatie. Denk aan een bedrijf dat zich voorneemt om veranderingen door te voeren in de organisatie wanneer een bepaalde omzetdoelstelling niet wordt gehaald. Als dat doel op een haar na wordt gemist, denkt het management misschien dat er inderdaad iets mis is in de organisatie en treft maatregelen. Wordt het doel echter nipt gehaald, dan denkt men

misschien dat de organisatie in orde is en laat alles bij het oude. Maar de uitkomst is in beide gevallen vrijwel identiek.

Om dergelijke problemen te vermijden, is het een goed idee om in de hele organisatie de criteria voor besluitvorming kritisch onder de loep te nemen. Men zal waarschijnlijk verstandig staan van het aantal beslissingen dat wordt genomen op basis van *go/no-go*-criteria. Soms is dat onvermijdelijk. Maar meestal bestaan er alternatieven, en die bieden een andere mogelijkheid om concurrentievoordeel te behalen.

4. *Plan regelmatig bijeenkomsten tegen ‘verstarring’*

Zelfs wanneer de drie bovengenoemde stappen worden gezet, blijft het risico op verstarring bestaan. Om dit te voorkomen is het verstandig om geregeld bijeenkomsten te beleggen waarin wordt gebrainstormd over de meest fundamentele overtuigingen in de onderneming. Kijk daarbij kritisch naar wat er in de branche gebeurt.

Is de bestaande kijk op de markt nog steeds relevant? Veranderen de behoeften en wensen van de klant?

Een manier om te innoveren is om na te denken over de afzonderlijke componenten die deel uitma-

ken van bestaande categorieën en daar nieuwe functies voor te bedenken. Bijvoorbeeld: auto's vervoeren mensen van A naar B en postmedewerkers vervoeren post van A naar B – nietwaar?

Ja, dat is wel zo, maar als je zo denkt, zie je waarschijnlijk interessante mogelijkheden over het hoofd. Amazon beseftte dat. Toen men daar begon na te denken over de functie van de auto, realiseerde men zich dat auto's ook konden worden gebruikt om pakketten aan te bezorgen. Daarop ging het bedrijf in de Verenigde Staten bezorgen aan de kofferbakken van auto's van de leden van Amazon Prime. Een ander voorbeeld vinden we in Nederland. Daar dacht PostNL na over de functie van haar postbestellers. Het bedrijf beseftte dat die tijdens hun ronde regelmatig foto's zouden kunnen maken van onkruid om aan de hand daarvan de effectiviteit van bestrijdingsmiddelen af te meten. Dit was een waardevolle nieuwe functie waar men met categorisch denken nooit op zou zijn gekomen.

DE MENSELIJKE GEEST IS EEN CATEGORISERINGS- MACHINE

CONCLUSIE

Categorieën helpen ons de wereld te begrijpen en onze ideeën te delen met anderen. Wij mensen werken er echter zoveel mee dat we vaak categorieën zien die er niet zijn. Dat vertekent onze kijk op de wereld en ondermijnt onze besluitvorming. Vroeger hoefden dergelijke fouten een bedrijf niet fataal te worden. Door de aanzwellende datarevolutie wordt het nu echter steeds belangrijker om te leren hoe we de consequenties van categorisch denken kunnen ondervangen.

Noten

1. <https://www.youtube.com/watch?v=imesoWCHG3c>.
2. <https://www.youtube.com/watch?v=AskAQwOBvhc>.
3. <https://www.themyersbriggs.com/en-US/Products-and-Services/> Myers-Briggs.

Literatuur

Berman, R., L. Pekelis, A. Scott & Ch. Van den Bulte (2018). p-Hacking and False Discovery in A/B Testing. Marketing Research Institute, 2018-2020 Research Priorities Working Paper Competition Winner, 18-130-10.

Cooper, M.J., O. Dimitrov & P.R. Rau (2001). A Rose.com by Any Other Name. *The Journal Of Finance*, Vol. LVI, No. 6, December.

Foroni, F. & M. Rothbart (2011). Category Boundaries and Category Labels: When Does a Category Name Influence the Perceived Similarity of Category Members? *Social Cognition*, 29/5, pp. 547-576.

Gordon, B.R., F. Zettermeyer, N. Bhargava & D. Chapsky (2018). A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook. Kellogg School of Management, Northwestern University, April 12. (https://www.kellogg.northwestern.edu/faculty/gordon_b/files/fb-comparison.pdf)

Lefgren, L., B. Platt & J. Price (2015). Sticking with What (Barely) Worked: A Test of Outcome Bias. *Management Science*, 61/5, May, pp. 931-1196.

Merrill, J.B. (2016). See How Facebook Labels You. *The New York Times*, 23 augustus.

Reichheld, F.F. (2003). The One Number You Need to Grow. *Harvard Business Review*, 81/12, December.

Singer, N. (2018). Tech Giants Now Share Details on Political Ads. What Does That Mean For You? *The New York Times*, 1 september.

Over de auteurs

B. de Langhe is universitair hoofddocent marketing aan ESADE Business School (Ramon Llull-universiteit) in Barcelona. Ph. Fernbach is hoogleraar marketing aan de Leeds School of Business van de universi-

teit van Colorado in Boulder en medeauteur van *The Knowledge Illusion: Why We Never Think Alone* (Riverhead Books, 2017).

Translated and reprinted by permission of Harvard Business Review. This article was originally published under the English title 'The Dangers of Categorical Thinking', by Bart de Langhe and Philip Fernbach, in the September-October 2019 Issue of the Harvard Business Review, vol. 97, issue 9-10. © 2019 Harvard Business School Publishing Corp. (Distributed by The New York Times Syndicate); all rights reserved. This translation, © 2019 Harvard Business School Publishing Corp. (Distributed by The New York Times Syndicate). Vertaling: Raymond Gijsen.