

Bedankt voor het downloaden van dit artikel. De artikelen uit de (online)tijdschriften van Uitgeverij Boom zijn auteursrechtelijk beschermd. U kunt er natuurlijk uit citeren (voorzien van een bronvermelding) maar voor reproductie in welke vorm dan ook moet toestemming aan de uitgever worden gevraagd.

Boom

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch door fotokopieën, opnamen of enig andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van kopieën uit deze uitgave is toegestaan op grond van artikelen 16h t/m 16m Auteurswet 1912 jo. Besluit van 27 november 2002, Stb 575, dient men de daarvoor wettelijk verschuldigde vergoeding te voldoen aan de Stichting Reprorecht te Hoofddorp (postbus 3060, 2130 KB, www.reprorecht.nl) of contact op te nemen met de uitgever voor het treffen van een rechtstreekse regeling in de zin van art. 16l, vijfde lid, Auteurswet 1912.

Voor het overnemen van gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (artikel 16, Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten, postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any way whatsoever without the written permission of the publisher.

info@boomamsterdam.nl
www.boomuitgeversamsterdam.nl

ESSAY, DEBAT EN DIALOOG

Deze rubriek is een forum voor debat over kwesties die betrekking hebben op kwalitatieve methodologie. Wie een idee heeft voor een thema of een reactie wil leveren op een eerder verschenen bijdrage, vragen wij contact op te nemen met Fred Wester: f.wester@maw.ru.nl.

De toekomst van etnografie in het Big Data-tijdperk*

Joost Beuving**

Inleiding

Big Data klopt aan de poort van sociaalwetenschappelijk onderzoek en dat roept in kringen van sociale denkers over het algemeen heftige reacties voor en tegen op. ‘Big Data’, in de rest van dit essay afgekort tot BD, is een populaire maar weinig precieze term die onder meer verwijst naar het gebruik van grote elektronische databestanden in het internettijdperk (zie Mayer-Schönberger & Cukie, 2013).

Voorstanders wijzen erop dat BD een belangrijke rol speelt in het herstellen van het maatschappelijke prestige van de sociale wetenschappen. Dat is hard nodig want na diverse emancipatiegolven die hebben geleid tot mondige burgers, alsmede de positieve wetenschap ondermijnende hype van het postmodernisme, is bij het grote publiek en de politiek de gedachte steeds meer gaan leven dat sociale wetenschap ook maar een mening is (Pepi, 2013). Tegenstanders wijzen er daarentegen op dat de groeiende populariteit van BD in sociaalwetenschappelijk onderzoek gepaard gaat met talloze problemen rondom privacy – de recente maatschappelijke ophef in Nederland over de zogeheten Sleepwet is slechts een voorbeeld – en vragen zich af of het moreel wel juist is dat er zoveel BD in handen van grote bedrijven is: er lijkt sprake van een heus digitaal-zakelijk complex waarin de belangen van gewone burgers steeds meer ondergeschikt raken aan het maken van bedrijfs-winst (O’Neil, 2016).

Afgemeten aan deze levendige ideologische discussies bestaat er beduidend minder

* Dit essay is ontstaan uit een lange professionele samenwerking met Geert de Vries en Jan Kees van Donge. Beiden wil ik bedanken voor hun steun en suggesties. Verder bedank ik mijn studenten in Nijmegen voor hun commentaar op delen van dit essay, meestal bediscussieerd in de coulissen van academisch onderwijs – de vrijplaats van de academie waar vaak de interessantste ideeën ontstaan. Ten slotte bedank ik Roy Gigengack en Reinoud Bosch voor hun kritische reflectie op dit essay.

** Dr. Joost Beuving is universitair docent bij Antropologie aan de Radboud Universiteit te Nijmegen en redacteur van KWALON. E-mail: j.beuving@maw.ru.nl.

aandacht voor de kennisaanspraken van BD: hoe verstaat deze nieuwe ster aan het methodische firmament zich tot andere sociaalwetenschappelijke strategieën zoals bijvoorbeeld enquêtes en etnografisch veldwerk? Dit gebrek aan aandacht valt des te meer op wanneer je bedenkt dat BD recent een aantal klinkende successen heeft geboekt. Neem de Amerikaanse verkiezingen bijvoorbeeld. Analyses van Google Trends-data waarmee in Google gebruikte zoektermen bestudeerd kunnen worden, lieten al vroeg in het verkiezingsproces zien dat de online wind sterk pro-Trump waaide. Dit terwijl peiling na peiling suggereerde dat de Democraten op winst stonden en Clinton zich over Trump geen zorgen hoefde te maken.¹ Ook rond de Brexit was er een succes. Bestudering van Facebookprofielen gekoppeld aan een psychometrisch analysemodel wezen op een ruime voorsprong van het 'Leave'-kamp, die aan de aandacht van veel waarnemers ontsnapte (Graslegger & Krogerus, 2017). In de categorie kleine succesjes zijn er geautomatiseerde analyses van auto's op Google Street View, die snellere en goedkopere demografische informatie opleveren dan met gangbare methoden (enquêtes) gerealiseerd kan worden (Gebu et al., 2017). BD is kortom meer dan een leuke hobby voor computernerds waar serieuze sociale wetenschappers zich niet te veel van hoeven aan te trekken (Burrows & Savage, 2014).

De kennisaanspraken van BD bespreek ik in dit essay aan de hand van Seth Stephens-Davidowitz' recent gepubliceerde, veelgeprezen, boek *Everybody Lies: Big Data, New Data, and What the Internet Can Tell Us About Who We Really Are* (2017). Dit is een belangrijke stem in de discussie omdat de auteur zich nadrukkelijk als sociaalwetenschapper afficheert, hij zich scherp kritisch over enquêtes uitlaat als een sociale constructie, en verwantschap met etnografie suggereert.² Dat laatste lijkt nogal merkwaardig aangezien BD en etnografie meestal als diametraal tegenovergestelde methodische bloedgroepen worden gezien: kwantitatief-positivistisch-grootschalig versus kwalitatief-interpretatief-kleinschalig (Ford, 2014). Stephens-Davidowitz suggereert echter dat er maar liefst drie overeenkomsten met etnografie bestaan: (a) het naturalistische ideaal van onopvallende aanwezigheid (*unobtrusiveness*), (b) de bestudering van kleine populaties die gewoonlijk buiten beeld blijven (*small populations*), en (c) het zichtbaar maken van verborgen motieven (*hidden motives*). Echter, wat volgens hem overeenkomsten lijken, blijken bij nadere beschouwing diepgaande misverstanden over centrale begrippen te betreffen. De BD-revolutie maakt etnografie dus geenszins overbodig (want goedkoper, efficiënter et cetera), maar roept wel prikkelende vragen op over een nadere methodische bepaling. Die werk ik tot besluit van dit essay kort uit aan de hand van de zogenoemde symfonische benadering die de Britse sociologen Susan Halford en Mike Savage voorstellen.

Onopvallende aanwezigheid versus de onzichtbare onderzoeker

BD gebruikt voor onderzoek digitale sporen die we achterlaten op het internet, bijvoorbeeld tijdens het googelen naar informatie, het bijhouden van een Facebook- of LinkedIn-profiel en het versturen van tweets. BD-wetenschappers verzamelen deze sporen, coderen en interpreteren de gegevens die ze bevatten, en doen

er sociaalwetenschappelijke uitspraken over. Belangrijk is dat deze gegevens niet tot stand komen speciaal voor een onderzoeksproject. In de woorden van BD-onderzoeker Eytan Adar: *'Big data collection is a secondary artifact of the information production and consumption process of people living their ordinary digital lives'* (Adar, 2015, p. 771). Er wordt dus geen onderzoekssituatie geschapen, zo is de gedachte, en er vindt geen verstoring plaats van alledaagse handelingen door de aanwezigheid van een onderzoeker. Een andere BD-onderzoeker, Michael Kosinski, gaat zelfs nog een stap verder en beweert: *'[Big data] can be used as a powerful data-recording tool because it stores (...) actual behaviour expressed in a natural environment'* (Kosinski et al., 2015, p. 543).

Deze statements suggereren een naturalistische ambitie die we met etnografie associëren: *'studying people in everyday circumstances by ordinary means'* (Beuving & De Vries, 2015, p. 15). Maar dan natuurlijk zonder het gedoe van reflexieve introspectie. Etnografische onderzoekers moeten zich immer in allerlei epistemologische bochten wringen om in het reine te komen met de vraag of wat ze observeren de samenleving zelf is, of de versie ervan die speciaal voor hem of haar wordt uitgeserveerd, of het resultaat van persoonlijke projectie. Zo niet de BD-onderzoeker. De internettende informant blijkt een onwetend persoon die zich er niet van bewust is dat hij of zij wordt gemonitord: hij of zij waant zich alleen en de onderzoeker blijft dan ook onzichtbaar. Het grote voordeel, zo stelt Stephens-Davidowitz: *'People will admit more if they are alone than if others are in the room with them'* (2017, p. 108). Probleem van reflexiviteit dus opgelost. Of toch niet?

Het is allereerst maar de vraag of we in het digitale verkeer echt wel zo lichtgelovig zijn dat we ons alleen wanen. De invloedrijke Amerikaanse internetpsychologe Sherry Turkle, bijvoorbeeld, denkt van niet. Zij beschouwt het internet als een Goffmaniaans theater waarin rollen, coulissen en publiek evenzeer het alledaagse leven vormgeven als in de offline wereld. Op basis daarvan argumenteert ze dat digitale gebruikersprofielen (Facebook, LinkedIn) onderdeel vormen van een zelfpresentatie. Om effectief en geloofwaardig te zijn is er een publiek nodig, en dat veronderstelt sociaal zelfbewustzijn. Dus wat er in eerste instantie uitziet als neutrale ego-data die op afstand, dat wil zeggen zonder aanwezige onderzoeker, en met grote snelheid door een computeralgoritme kunnen worden geogst voor sociaalwetenschappelijk onderzoek, is in werkelijkheid meestal een zorgvuldig geregiseerde *performance* waarin de negatieve randjes zorgvuldig worden weggemasseerd (Turkle, 2011). Daarnaast worden op datzelfde internet BD-onderzoeksresultaten natuurlijk gewoon gedeeld (het invloedrijke *Wired Magazine* is bijvoorbeeld erg populair), en valt dus te verwachten dat dit wel degelijk tot een soortement 'onderzoekersbewustzijn' leidt, dat weer gevolgen heeft voor wat er precies gedeeld/gezocht wordt op het internet.

Ook deel van dit misverstand is de gedachte dat de onzichtbare onderzoeker als een ideaal wordt gezien. Dat doet weliswaar opgeld in BD-kringen, maar niet of nauwelijks in de wereld van etnografisch onderzoek. 'Onderzoek' verwijst daar eerder naar een reflexieve praktijk waarin de onderzoeker consequent het effect van zijn of haar aanwezigheid kritisch betreft op het opgebouwde beeld van de onderzochte samenleving. Reflexiviteit is dus geen probleem dat even 'opgelost' kan wor-

den met onzichtbaarheid, maar gaat over sociale relaties in het veld die serieuze verkenning behoeven. Sociale relaties spelen zich uiteraard ook steeds meer op het internet af, waarvoor de reflexieve uitdaging misschien nog wel groter is dan die we al kennen van meer traditioneel, 'offline' onderzoek. A priori standpunten zoals we onder BD-onderzoekers aantreffen, zijn daarbij weinig behulpzaam. Soms is het onderzoeker-effect minimaal en valt de etnograaf als een *'fly on the wall'* nauwelijks op in het leven van alledag (Winkler, 1990). En soms, zoals Roy Gigengack terecht opmerkt in diens bijdrage aan een recente KWALON-discussie over naturalistisch onderzoek, wil de etnografisch onderzoeker weliswaar graag onzichtbaar zijn, maar steekt de door hem of haar onderzochte groep daar een stokje voor – en wordt de onderzoeker onbedoeld middelpunt van belangstelling (Gigengack, 2017, p. 9; zie ook Beuving, 2017).

Statistische en theoretische populaties

Tweede misverstand: de betekenis van de term populatie. Methodisch gesproken heeft dat te maken met de kwestie van representativiteit, een term waarin nogal eens het positivistische ideaal van de natuurwetenschappen doorklinkt: met een aselechte steekproef verworven onderzoeksresultaten zijn representatief voor de hele onderzoekspopulatie. BD daarentegen heeft geen probleem van statistische representativiteit. De informatiezee waaruit ze vist, omvat veelal de gehele onderzoekspopulatie: denk aan alle Facebook-profielen, of de volledige datastroom van zoekgegevens in een bepaald tijdvak; 'N=all' (Borgman, 2014). Er hoeft dus geen steekproef te worden getrokken. Belangrijk aanvullend voordeel is dat op die deelpopulaties kan worden ingezoomd die vanwege hun geringe omgang anders buiten beeld blijven. Bij steekproeftrekking doet omvang er immers toe: een kleine deelpopulatie heeft statistisch minder kans om in een aselechte procedure getrokken te worden (O'Neil, 2016). Maar dit betekent niet dat kleine deelpopulaties niet interessant zijn. Redenerend als statisticus: wat er buiten de puntenwolk gebeurt, kan onverwacht licht werpen op wat er daarbinnen plaatsvindt.

Het volgende voorbeeld (overgenomen uit Stephens-Davidowitz) van de Amerikaanse econoom Raj Chetty en diens medewerkers illustreert dat. Zij kregen toegang tot de database van de *Internal Revenue Service* (IRS), de BD van de Amerikaanse belastingdienst, waarmee ze inkomensgegevens van alle Amerikanen konden bestuderen.³ Deze gegevens verbonden ze aan sterftcijfers en zo kon de statistische relatie tussen inkomen en levensverwachting worden bestudeerd. Het is uit de socio-economische literatuur genoegzaam bekend dat arme mensen (veel) jonger sterven dan rijke mensen (kwaliteit van voedsel en medische zorg speelt daarbij een belangrijke rol) (Wilkinson & Pickett, 2009), maar daar waar het voor rijke Amerikanen vrijwel niet uitmaakt waar ze wonen, blijkt dat voor de armen een heel ander verhaal: arme Amerikanen worden op sommige plekken veel ouder dan elders. Chetty c.s. (2016) zoomden in op specifieke locaties en vonden dat de relatie tussen arm en rijk een grote voorspellende waarde heeft: de aanwezigheid van rijke Amerikanen draagt bij aan een hogere levensverwachting van arme landgenoten.

Inzoomen op kleine deelpopulaties is ook een kenmerk van etnografie. Dit type onderzoek staat of valt immers met het aanknopen van een persoonlijke relatie met een onderzoeksgroep, en dat kost nu eenmaal tijd, en dat zet een natuurlijke rem op de aantallen informanten. Maar er is een belangrijk verschil met BD dat voor verwarring zorgt. Vanuit etnografisch gezichtspunt is een populatie helemaal geen statistische, maar een theoretische eenheid; het statistische ‘sample’ is, gezien door etnografische ogen, een kunstmatige constructie die geen, of maar een beperkte, weerklank heeft in het alledaagse leven. Sociale eenheden zoals relaties, groepen en instituties worden derhalve uitsluitend beschouwd als deel van een populatie voor zover ze behulpzaam zijn in het ontwikkelen van abstract-theoretische gedachten over het mechanisme of proces dat in deze populatie werkzaam is. Doorredenerend op Chetty c.s.: de etnografisch onderzoeker zal met name geïnteresseerd zijn in een verklaring van het aanwezigheidsmechanisme dat voor arme Amerikanen levensverlengend werkt, en zoekt dan ook doelbewust naar die sociale situaties waarin dit direct kan worden geobserveerd. Daarbij volgt hij of zij dus het principe van *purposeful sampling* (Glaser & Strauss, 2012) om zo bij te dragen aan theorievorming over de relatie tussen arm en rijk – overigens zonder te claimen dat de onderzochte situaties statistisch representatief zijn voor alle arm-rijk-naburenschappen in de gehele VS.

Verstehen gerelativeerd?

Een derde misverstand betreft de status van motieven voor gedrag: gedachten, neigingen, voorkeuren, verlangens et cetera die ons handelen sturen. Veel BD-onderzoekers lijken een psychologisch model aan te hangen waarin mensen direct toegang hebben tot hun eigen motieven. Deze zouden in het alledaagse leven echter onderdrukt worden door de aanwezigheid van andere mensen, en die aanwezigheid wordt psychisch geïnternaliseerd via schaamte en dergelijke. De anonimiteit van het internet verlost ons dan van sociale druk zodat motieven volledig tot expressie kunnen komen; deze tonen zich vervolgens in zoekpatronen. In het vocabulaire van Stephens-Davidowitz: *‘In the pre-digital age, people hid their embarrassing thoughts from other people. In the digital age, they still hide them from other people, but not from the internet and in particular sites such as Google and PornHub, which protect their anonymity. These sites function as a sort of digital truth serum’* (2017, p. 53-54). Hier zijn een paar problemen. Allereerst, vanuit etnografisch perspectief is de gedachte dat motieven in het individu zijn gelokaliseerd maar in de alledaagse praktijk verborgen blijven, tamelijk naïef. Zo is het maar de vraag in hoeverre motieven een stabiele eigenschap van het individu zijn. Doet dat geen afbreuk aan de in kwalitatief-sociaalwetenschappelijke kringen breed gedragen gedachte dat motieven een uitdrukking van sociale krachten zijn? Neem de volgende quote uit een interview dat de Amerikaanse etnografe Alice Goffman gaf aan de New York Times (2016) naar aanleiding van haar recente boek *On the Run*, waarin ze het dagelijkse leven beschrijft van de zwarte onderklasse in de VS: *‘I don’t think that [we] have real preferences, just desires that emerge in social interactions.’* Dit raakt, denk ik, een belangrijke kern: motieven ontstaan in sociale interactie en om deze te kunnen dui-

den, is inzicht in sociaal gedrag dus essentieel. Dat zulk gedrag op internet mogelijk speciale eisen stelt aan zowel de onderzoeksopzet als aan de vaardigheden van de onderzoeker lijkt evident, maar verandert weinig aan de geldigheid van het onderliggende sociale principe.

Daarnaast is er de vraag of motieven wel direct sociaalwetenschappelijk geduid kunnen worden, dat wil zeggen zonder tussenkomst van de betrokken personen/informanten. In BD-onderzoek wordt gemakkelijk de analytische sprong van beschrijving van gedrag naar verklaring ervan gemaakt, waarbij de tussenliggende stap van *Verstehende* interpretatie simpelweg wordt overgeslagen (Boyt & Crawford, 2012). Voor de BD-analist zijn digitale sporen dus direct verklaarbaar als sociale feiten – dit lijkt de noodzaak tot *Verstehen* te nuanceren – terwijl een etnografisch onderzoeker juist centraal maakt hoe informanten over hun handelen zelf nadenken: welke betekenissen geven ze aan hun digitaal gedrag, en in welke sociale praktijken resulteert dat? Een voorbeeld van Stephens-Davidowitz zelf laat zien hoe problematisch de directe interpretatie kan zijn. Hij bestudeerde met Google Trends zoektermen van mannen met zwangere vrouwen en zag dat in Mexico de meest gezochte term over zwangerschap omvat ‘kooswoordjes voor mijn zwangere vrouw’ terwijl dat in de VS is ‘mijn vrouw is zwanger en wat moet ik nu’ (Stephens-Davidowitz, 2017, p. 20). Dit wijst volgens hem op een cultureel verschil tussen Mexico en de VS: de romantische, op familiewaarden gerichte Mexicaanse man voor wie een kind een verrijking is versus de moderne Amerikaanse man die tobt over de praktische consequenties van een zwangerschap. Dit verschil poneert Stephens-Davidowitz overigens zonder ooit met een Mexicaanse of Amerikaanse man gesproken te hebben. Vanuit etnografisch gezichtspunt lijkt hij daarmee een klassiek geval van *Hineininterpretieren* te demonstreren: het onkritisch projecteren van eigen vooronderstellingen op onderzoeksgegevens.

Discussie: BD en de relatie met etnografie

Nu we de misverstanden in de etnografisch getinte kennisaanspraken van BD wat scherper in beeld hebben, rest er nog wel een serieuze kwestie om te behandelen: zijn de vermeende successen van sociaalwetenschappelijke BD-studies belangrijk voor etnografisch onderzoek en tot welke nadere bepaling van de etnografische praktijk leidt dat eventueel?

Om de BD-successen te bespreken is Stephens-Davidowitz’ werk andermaal relevant. Diens boek opent met de tweede verkiezingswinst van Obama, die veel minder overtuigend was dan diens eerste, meer eclatante, succes. Hij stelt dat dit het gevolg was van opvlammend racisme in de VS gedurende het afgelopen decennium, en dat die door de meeste politieke waarnemers erg werd onderschat: ze bezagen de wereld vanuit een liberale bubbel in Washington die weinig oog had voor het groeiende racisme van gewone Amerikanen, zo is zijn verklaring. Stephens-Davidowitz beweert verder dat hij dit wel zag aankomen. Zijn Google Trends-data registreerden namelijk het opvlammen van racisme door een sterke toename van met racisme geassocieerde zoektermen in de maanden voor de verkiezingen. In het ar-

tikel waarin hij hierover rapporteert (en dat aan zijn boek ten grondslag ligt), meldt Stephens-Davidowitz hoe deze zoektermen een kijkje in de ziel van gewone Amerikanen geeft: *'Racially charged search rate is a significant, negative predictor of Obama's (...) vote shares'* (Stephens-Davidowitz, 2014, p. 27).

Deze uitspraak is belangrijk want hij suggereert causaliteit: een sterke troef die *pro* directe interpretatie pleit en daarmee het belang van *Verstehen* relativeert. Immers, als met BD een betrouwbare voorspelling gemaakt kan worden op basis van een herhaalbaar oorzaak-gevolg-mechanisme zoals in een experiment, wat is dan nog de toegevoegde waarde van het perspectief van binnenuit: de *Verstehende* interpretatie?⁴ In het artikel wordt inderdaad op causaliteit aangestuurd, maar Stephens-Davidowitz doet dat op een wat omfloerste manier die suggereert dat hij er misschien zelf ook over twijfelt: *'this paper's methodology (...) allows for a more robust test of a causal effect of racial animus, relative to other papers in the literature'* (p. 27), en *'there is strong evidence for a causal interpretation'* (p. 32), en *'our findings further support a causal interpretation'* (p. 33). Dat is weliswaar handig taalgebruik maar nog niet hetzelfde als hard bewijs van een oorzaak-gevolgrelatie. Zelf ben ik te weinig statistisch onderlegd om de claim op waarde te schatten. Daarom heb ik deze kwestie voorgelegd aan drie collega's, allen onderzoekers met een serieuze statistische trackrecord, en twee ervan blijken niet overtuigd van de causaliteitsclaim; de derde aarzelt.

Deze uitkomst is niet zo verrassend. Zoals door anderen al is opgemerkt, het is in de echte sociale wereld lastig om causaliteit te bewijzen. De sociaalkritische statisticus Nassim Taleb, bijvoorbeeld, fulmineert in diens recente boek *Antifragile* dan ook tegen de herhaalde pogingen van onderzoekers om als causaliteit te verkopen wat eigenlijk correlaties zijn – een veel zachtere maat voor samenhang. Het is iets dat elke eerstejaars student leert: in jaren dat er veel ooievaars rondvliegen, ligt het geboortecijfer bij mensen doorgaans ook hoger, maar dat betekent niet dat baby's door de ooievaar gebracht worden; er is correlatie maar geen causaliteit. Taleb waarschuwt verder tegen een belangrijk gevaar inherent in BD-onderzoek: hoe groter de onderzoekspopulatie, en hoe groter het aantal onderzochte variabelen daarin, des te hoger de kans dat er correlaties gevonden worden – ook al blijkt er van een werkelijk oorzaak-gevolgrelatie geen sprake te zijn (Taleb, 2012, p. 417). In een interview liet hij zich dan ook ontvallen: *'I am not saying here that there is no information in big data. There is plenty of information. The problem – the central issue – is that the needle comes in an increasingly larger haystack'* (Taleb, 2013).

Mijn vermoeden is dan ook dat de relatie tussen racistische zoektermen op het internet en de politieke wind rondom verkiezingen wat ingewikkelder is dan Stephens-Davidowitz stelt: er is weliswaar correlatie en dat is op zich best interessant, maar dat betekent nog niet dat de geconstateerde toename in zoektermen de politieke uitkomst verklaart. Het suggereert eigenlijk dat Amerikanen zich dus niet zo gemakkelijk in de ziel laten kijken via hun zoekgedrag op Google. Het BD-verkiezingsonderzoek in de VS toont vooral aan dat mensen allerlei gevoelens hebben, vaak tegenstrijdig en onfris – ook racistische – maar niet dat juist de op het internet geventileerde versies ervan meer 'waar' zijn dan andere. Ze bestaan naast andere die ook echt en waar zijn en hoe die twee tezamen onze leefwereld

vormen waar vanuit we de samenleving bekijken, en waarin ons handelen gevormd wordt, is zeker geen *a priori* uitgemaakte zaak maar vereist een begrip van binnen-uit – dus pro *Verstehend* interpreteren.

Naar een methodologische symfonie?

De discussie die Stephens-Davidowitz aangaat, staat in een ontluikend debat waarin kennisaanspraken over de samenleving op het spel staan. Omdat hij zich uitgesproken als sociaalwetenschapper positioneert, biedt zijn werk een uitgelezen kans om de positie van etnografie (en daarmee op naturalistische leest geschoeide kwalitatieve sociale wetenschappen in bredere zin) vis à vis BD aan te scherpen. Stephens-Davidowitz neemt daarin een radicaal standpunt in: BD levert met etnografie vergelijkbare kennis op over de motieven van gedrag, kan daarbij inzoomen op kleine deelpopulaties die anders over het hoofd worden gezien, en verstoort daarbij natuurlijk gedrag niet. Hierop doorredenerend: BD kan etnografisch onderzoek mettertijd vervangen, want: goedkoper, efficiënter et cetera. De enige serieuze beperking daarbij lijkt vooralsnog de wereldwijde onderklasse, Paul Colliers 'Bottom Billion', die geen/nog geen toegang tot internet heeft (2007). Een kwestie van tijd voordat ook die aangesloten zijn op de digitale stromen, en de BD-droom werkelijkheid kan worden. Maar dit radicale standpunt stoelt op belangrijke epistemologische misvattingen. BD is weliswaar uitstekend in staat om digitale stromen in kaart te brengen; een serieuze rivaal voor etnografie is ze niet.

Een minder radicaal standpunt ziet er vooralsnog beloftevoller uit, onder meer verwoord door de Britse sociologen Susan Halford en Mike Savage (2017). Zij werken een pleidooi uit voor wat ze een symfonische sociale wetenschap noemen, en waarin etnografie en BD naast elkaar bestaan – maar wel met een duidelijk gedefinieerde arbeidsverdeling. Ze onderscheiden in hun bespreking de bekende trits data-methoden-theorie, die ik tot besluit van dit essay kort behandel. Allereerst beschouwen ze als een grote winst van BD om 'data' in een nieuw licht te beschouwen: persoonlijke profielen op internet, registerdata, mobiele-telefoonverkeer, tweets, et cetera. Maar in plaats van het naïeve realisme dat de houding van BD naar data kenmerkt, stellen Halford en Savage een kritisch-pragmatische benadering voor. Centraal daarin is de gedachte dat in elke informatiebron bias zit, en deze moet worden verkend. Tweets bijvoorbeeld oververtegenwoordigen het hoogopgeleide, meer politiek uitgesproken smaldeel van de bevolking en zijn dus niet zomaar representatief (een stille meerderheid van lager opgeleiden roert zich minder op Twitter). Datzelfde geldt voor registerdata, denk aan hoe sociale minima zorg mijndend zijn en dus ondervertegenwoordigd blijven in de registers van huisartsen.

Op het terrein van methoden propageren Halford en Savage pluralisme: het naast en door elkaar heen gebruiken van diverse (kwalitatieve en kwantitatieve) strategieën voor dataverzameling. Dit is niet hetzelfde als het aanbevelen van methodisch eclecticisme: effectief pluralisme is probleem-gedreven. In de onderzoekspraktijk komt dat neer op een voortdurende wisselwerking tussen structuur en betekenis: BD is sterk in de (statistische) beschrijving van sociaal gedrag (Durk-

heimiaanse ‘sociale feiten’), die het vertrekpunt vormt voor een *Verstehende* interpretatie in etnografisch veldwerk; het omgekeerde proces is natuurlijk ook goed denkbaar. Zoals we in Beuving & De Vries (2015, p. 31-33) al stelden: sociale structuur kan niet worden begrepen zonder terug te grijpen op hoe mensen deze begrijpen; omgekeerd is een interpretatie van betekenis zinloos zonder dit te plaatsen in het kader van structurele (on)mogelijkheden die mensen tot hun beschikking hebben. Het voorbeeld hierboven beschreven van Ray Chetty c.s. illustreert dit voor de behandeling van belastingregisterdata in relatie tot vragen over armoede in de VS. Wat betreft theorie, ten slotte, wijzen Halford en Savage op het gevaar om dataverzameling, waar BD in uitblinkt, te verwarren met theorievorming: de praktijk die de Amerikaanse socioloog Wright-Mills (1959) treffend duidt als *abstracted empiricism*: het verzamelen van feiten en feitjes zonder deze te verbinden aan een bredere betekenis of toepassing. Hieruit volgt dat niet alles dat geobserveerd kan worden ook relevante data zijn, en data spreken niet voor zichzelf: waarnemingen worden tot leven gebracht in relatie tot zich ontvouwende abstracte gedachten over de onderzochte sociale groep/samenleving (Swedberg, 2016). We voegen aan dit procesmatige idee van sociale theorie toe dat dit niet over de hoofden van de samenleving gedaan kan worden; om zin van onzin in de interpretatie van digitale patronen te scheiden zijn de perspectieven van mensen zelf een noodzakelijk vertrekpunt. Alleen dan klinken de stemmen van BD en etnografie in een werkelijk *harmonieuze* symfonie.

Noten

- 1 Peilingen van de Nederlandse econoom Arie Kapteyn, werkzaam aan de University of Southern California, vormen een belangrijke uitzondering: aan de hand van een anoniem internetpanel dat hij langdurig volgde, voorspelde hij Trumps winst, zie <https://cesr.usc.edu/election>.
- 2 Stephens-Davidowitz is beslist niet de enige, noch de eerste, die deze verwantschap suggereert, zie bijvoorbeeld het werk van John Curran (2016) en Heather Ford (2014) voor vergelijkbare statements. Echter, Stephens-Davidowitz is wel een van de meer uitgesproken stemmen en zijn boek biedt een uitgewerkt raamwerk voor een nadere bepaling.
- 3 Weliswaar speelt internet bij deze registerdata een beperkte rol maar dit wordt in de regel desondanks tot de categorie van BD gerekend. Het gaat hierbij nadrukkelijk om een heuristisch voorbeeld dat het behandelde principe illustreert.
- 4 De Nederlandse antropoloog André Köbben wierp een halve eeuw geleden dezelfde vraag al eens op in een baanbrekend (maar weinig geciteerd) artikel (Köbben, 1952) waarin hij de relatie tussen antropologie en de destijds snel opkomende, nieuwe methode van sociale statistiek kritisch bevroeg. Zijn conclusie: *‘[The statistical method] ... notes certain phenomena in a society and isolates them from their cultural context, thereby conveying an entirely different impression from the one they make when studied within this context’* (Köbben, 1952, p. 133).

Literatuur

- Adar, E. (2015). The two cultures and big data research. *I/S: A Journal of Law and Policy for the Information Society*, 10(3), 765-781.
- Beuving, J. (2017). The anthropologist as jester, anthropology as jest? *Social Anthropology*, 25(3), 353-363.
- Beuving, J., & Vries, G. de. (2015). *Doing qualitative research: The craft of naturalistic inquiry*. Amsterdam: Amsterdam University Press.
- Borgman, C. (2014). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: The MIT Press.
- Boyt, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Community & Society*, 15, 662-679.
- Burrows, R., & Savage, M. (2014). After the crisis? Big data and the methodological challenges of empirical sociology. *Big Data & Society*, April-June, 1-6.
- Chetty, R. et al. (2016). The association between income and life expectancy in the United States. *Journal of the American Medical Association*, 135, 16.
- Collier, P. (2007). *The bottom billion: Why poor countries are failing and what can be done about it*. Oxford: Oxford University Press.
- Curran, J. (2016). Big data or 'big ethnographic data'? Positioning big data within the ethnographic space. *Ethnographic Praxis in Industry Conference*, 62-73.
- Ford, H. (2014). Big data and small: Collaborations between ethnographers and data scientists. *Big Data & Society*, July-September, 1-3.
- Gebru, T. et al. (2017). Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. *Proceedings of the National Academy of Sciences of the United States of America*, <https://doi.org/10.1073/pnas.1700035114>.
- Gigengack, R. (2017). Ambachtelijke illusies: Het naturalisme geduid in het post-naturalistisch tijdperk. *KWALON* 64, 22(1), 5-16.
- Glaser, B., & Strauss, A. (2012 [1967]). *The discovery of grounded theory: Strategies for qualitative research*. New Brunswick/London: Aldine Publishers.
- Goffman, A. (2015). *On the run: Fugitive life in an American city*. Chicago: Chicago University Press.
- Graslegger, H., & Krogerus, M. (2017). *The data that turned the world upside down*. Retrieved from https://motherboard.vice.com/en_us/article/mg9vvn/how-our-likes-helped-trump-win
- Halford, S., & Savage, M. (2017). Speaking sociologically with Big Data: Symphonic social science and the future for big data research. *Sociology*, 51(6), 1132-1148.
- Köbben, A. (1952). New ways of presenting an old idea: The statistical method in social anthropology. *Man/JRAI*, 82(2), 129-146.
- Kosinski, M., Matz, S., Gosling, S., Popov, V., & Stillwell, D. (2015). Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. *American Psychologist*, 70(6), 543-556.
- Mayer-Schönberger, V., & Cukie, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: John Murray.
- O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. London: Penguin.
- Pepi, M. (2013). The Postmodernity of Big Data. *The New Inquiry*, December 30.
- Stephens-Davidowitz, S. (2014). The cost of racial animus on a black candidate: Evidence using Google Search data. *Journal of Public Economics*, 118, 26-40.

- Stephens-Davidowitz, S. (2017). *Everybody lies: Big data, new data, and what the internet can tell us about who we really are*. New York: HarperCollins.
- Swedberg, R. (2016). Before theory comes theorizing or how to make social science more interesting. *The British Journal of Sociology*, 67(1), 5-22.
- Taleb, N. (2012). *Anti-fragile: Things that gain from disorder*. London: Penguin.
- Taleb, N. (2013). *Beware the big errors of 'big data'*. Retrieved from <https://www.wired.com/2013/02/big-data-means-big-errors-people/>
- Turkle, S. (2011). *Alone, together: Why we expect more from technology and less from ourselves*. New York: Basic Books.
- Wilkinson, R., & Pickett, K. (2009). *The spirit level: Why more equal societies almost always do better*. New York: Allen Publishers.
- Winkler, J. (1990). The fly on the wall of the inner sanctum: Observing company directors at work. In G. Moyser & M. Wagstaffe (Eds.), *Research methods for elite studies* (p. 129-146). London: Allen & Unwin.
- Wright-Mills, C. (1959). *The sociological imagination*. Oxford: Oxford University Press.