

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](http://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](http://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](mailto:info@boomamsterdam.nl)  
[www.boomuitgeversamsterdam.nl](http://www.boomuitgeversamsterdam.nl)

# De kakofonie van Big en small Data: reflecties over digitale etnografie

Roy Gigengack\*

De uitnodiging van Joost Beuving (JB) om een reflectie te schrijven over zijn Big Data-artikel (2018) nam ik graag aan. Het idee dat een kwantitatieve econoom verwantschap zou claimen met etnografie, vond ik fascinerend; dat hij ook nog iets zou hebben met naturalisme, nam ik op de koop toe. Joost was ook zo vriendelijk mij het boek *Everybody Lies* van Stephens-Davidowitz (SD) toe te sturen.

## De voorspelling en de ironie van de verwantschap

Toch meen ik een totaal ander boek gelezen te hebben! De ‘verwantschapsclaim’ die mij was beloofd, kwam ik niet tegen; de woorden etnografie en naturalisme ontbreken in de index, en in de tekst vind je ze evenmin. Nu moet ik toegeven dat het ook mij moeite heeft gekost *Everybody Lies* uit te lezen, want met al zijn flauwe grappen en onbeduidende zijpaden is het boek erbarmelijk slecht geschreven. Desalniettemin vond ik het uiteindelijk lonend me over de irritaties heen te zetten. SD blijkt een zelfverklaarde volgeling van Steven Levitt, de bekende *rogue economist* die mooie boeken (Levitt & Dubner, 2005, 2009) schreef over de ‘verborgen kanten van alles’, zoals het frauderen van onderwijzers en sumoworstelaars, het effect van abortus op misdadigers, of waarom drugsdealers bij hun moeder blijven wonen. Overigens schreef Levitt *Freakonomics* samen met een journalist, Stephan Dubner; vandaar wellicht dat een en ander ook nog eens goed leesbaar was. Naar zijn grote voorbeeld slaagt Stephens-Davidowitz er in om op basis van grote datasets fascinerende inzichten te berde te brengen. Hij meldt bijvoorbeeld dat citaten van prijswinnende economieboeken bijna altijd afkomstig zijn van de eerste vijftig pagina’s, hetgeen hem doet concluderen dat deze boeken nauwelijks worden uitgelezen. Hoe dan ook, *Everybody Lies* blijkt vrij accuraat te hebben voorspeld hoe Joost Beuving het boek zou citeren: de kracht van Big Data illustreert hier zichzelf.

Mijn grootste probleem met het artikel van JB is niet dat hij een verwantschap ziet tussen etnografie en Big Data à la Stephens-Davidowitz, maar dat hij die verwantschapsclaim aan de laatste toeschrijft. Het zou interessanter zijn geweest als JB zelf de verwantschap tussen deze twee tegenovergestelde epistemologieën had gesignaleerd en uitgewerkt. Nu roept hij in eerste instantie toch vooral ironie op. Ga maar na: voor een positivist als SD, met zijn geloof in ‘de waarheid’, kan een derge-

\* Roy Gigengack is Postdoctoral Research Fellow Military Mental Health Care aan de University of Edinburgh. Tevens is hij Senior Assistant Professor aan de afdeling Sociale en Culturele Antropologie van de Vrije Universiteit te Amsterdam. E-mail: r.a.gigengack@vu.nl.

lijke discrepantie tussen tekst en interpretatie enkel de juistheid van *Everybody Lies* bevestigen. Voor een naturalist à la JB zouden de zaken iets ingewikkelder moeten liggen. Hij zou zich allereerst moeten gaan afvragen of deze interpretatie een weergave van het boek zelf is, of de versie ervan die speciaal voor hem is uitgeserveerd, dan wel het resultaat van zijn persoonlijke projectie. Daar de eerste en de tweede mogelijkheid niet opgaan, zouden we te maken hebben met, in Beuvings woorden, ‘een klassiek geval van *Hineininterpretieren*’, oftewel ‘het onkritisch projecteren van eigen vooronderstellingen’.

### **De stromannen en de epistemologie van wat ‘waar’ is**

In zijn betoog geeft JB de argumenten van SD op meerdere plaatsen heel anders weer dan wat er in werkelijkheid staat. Neem het voorbeeld van de meest gezochte termen over zwangerschap in Mexico en de VS. Stephens-Davidowitz (2017, p. 20) heeft het alleen maar over ‘cultural differences I had never even contemplated’. De hele invulling van romantiek en familiewaarden versus praktische consequenties staat niet in het boek. Evenmin is terug te vinden dat SD nooit met een Mexicaanse of Amerikaanse man gesproken zou hebben; dat laatste lijkt me overigens wel heel onwaarschijnlijk. Het zouden kleine slordigheden kunnen zijn, ware het niet dat JB keer op keer de *Methodenstreit* ermee oprakelt, zoals Reinoud Bosch (2018) terecht opmerkt.

Joost Beuving schuwt de stroman-drogreden niet. Onder verwijzing naar de bekende beeldspraak van de naald in de alsmaar uitdijende hooiberg suggereert hij bijvoorbeeld dat SD correlaties verkoopt als causaliteit. Maar dat is niet zo. SD waarschuwt zelfs voor ‘de vloek van dimensionaliteit’, de correlatie met het toeval dat zeker zal plaatsvinden. Als je elke ochtend maar genoeg muntjes opgooit, zul je er altijd eentje hebben die een jaar lang elke dag correct de beursgang voorspelt, en toch is het niet aan te raden al je geld op dat ene muntje te zetten. Een andere stroman is de retorische vraag ‘of we in het digitale verkeer echt wel zo lichtgelovig zijn dat we ons alleen wanen’. In de aanval op SD wordt dit keer het werk van Sherry Turkle als stormram gebruikt. Maar het is opnieuw een vertekening van argumenten. Het is een cruciaal onderdeel van *Everybody Lies* dat de informatie van Facebook, LinkedIn of Twitter zorgvuldig geregisseerd is. SD komt er meerdere keren op terug (bijvoorbeeld in ‘The Truth about Your Facebook Friends’, p. 150-153).

Met zijn formuleringen over ‘digitaal waarheidsserum’ lijkt Stephens-Davidowitz het epistemologisch inderdaad niet helemaal op een rijtje te hebben. Maar je kunt hem toch niet van naïviteit jegens zijn eigen vakgebied beschuldigen. Het is juist zijn punt dat Big Data niet voor zichzelf spreekt, en interpretatie behoeft (zie bijvoorbeeld het hoofdstuk ‘Big Data – Handle With Care’). En als je *Everybody Lies* in zijn geheel gelezen hebt, kun je het toch ook niet opvatten als een bedreiging voor kwalitatief onderzoek (‘want goedkoper, efficiënter etc.’). SD (2017, p. 254-256) houdt immers ook een pleidooi voor het combineren van Big Data met wat hij ‘small data’ noemt. Small data worden verkregen door ‘*an old-fashioned approach: asking people what they think.*’ *Facebook data teams*, voegt SD eraan toe, [employ]

*social psychologists, anthropologists, and sociologists precisely to find what the numbers miss.*

Dan is er nog de epistemologische kritiek op wat ‘waar’ is. Volgens JB staat SD’s idee van waarheid voor verborgen motieven (zoals bij wat mensen nu ‘echt’ vinden). Die verborgen motieven zouden te simpel zijn om de complexe werkelijkheid te verklaren, en daarom weidt JB uit over correlatie en causaliteit. Het probleem hier echter is dat de kracht van Big Data niet zozeer in het verklaren ligt, maar in het voorspellen. Big Data kan wel iets verklaren, waarover zo meteen meer, maar SD zit in wat hij zelf noemt *the prediction business* en niet *the explanation business* (2017, p. 71). Zo kan hij wel voorspellen welke pagina’s zijn lezers zullen citeren, maar hij hoeft niet te weten waarom ze dat gaan doen (hoewel hij lijkt te suggereren dat het leven te kort is om economieboeken uit te lezen). Natuurlijk beargumenteert SD niet dat de racistische zoektermen de verkiezing van Donald Trump verklaren; zijn stelling is enkel dat die zoektermen hielpen de politieke uitkomst correct te voorspellen. Kortom, SD denkt als een typische, Amerikaanse pragmaticus. In zijn epistemologie staat ‘waar’ voor ‘wat werkt in voorspellen’.

### **De vlieg, de spion en de trol**

Het doet mij deugd dat Joost begint te twifelen aan zijn naturalistische paradigma. Zo stelt hij dat in de wereld van etnografisch onderzoek de onzichtbare onderzoeker niet of nauwelijks meer als ideaal wordt gezien. Dat begint toch al te schuren met het idee van *unobtrusiveness* dat zo centraal stond in zijn eerdere werk. Weliswaar haalt JB de ‘vlieg op de muur’ weer van stal, maar schoorvoetend begint hij toch het illusionaire karakter van dat beestje toe te geven. Hij verwijst daarbij naar het debat dat wij eerder hebben gevoerd, waarin ik ageer tegen onder meer dit naturalistische geloofsartikel en de onderliggende *biology envy* (Gigengack, 2017a, 2017b). Een saillant verschil tussen etnografie en Big Data-onderzoek is, zoals JB aangeeft, gelegen in de (lijfelijke) aanwezigheid van de onderzoeker. Een etnograaf kan juist daardoor zijn of haar onderzoek mede vormgeven (zie ook Okely, 2012), en de metafoor van de vlieg op de muur bagatelliseert die vormende aanwezigheid. Maar ook Big Data-onderzoek heeft reflexieve dimensies. Zoals hierboven al is gesignaleerd, legt SD goed uit dat de digitale gebruikersprofielen van bijvoorbeeld Facebook een meervoudig sociaal bewustzijn uitdrukken. Bovendien laat ook hij het ‘onderzoekersbewustzijn’ zien van internettende subjecten, bijvoorbeeld in zijn uiteenzettingen over Pornhub en Google Flu Trends. De laatste hoofdstukken van *Everybody Lies* maken echter duidelijk dat er nog veel meer aan de hand is: Stephens-Davidowitz geeft overtuigende redenen om die onschuldig lijkende vlieg op de muur voor eens en altijd dood te slaan.

Big Data-onderzoeken laten het namelijk niet bij het louter ‘monitoren’ van onwetende personen op het internet: ze *manipuleren* ze. SD laat overtuigend zien dat Big Data-onderzoekers constant bezig zijn met causale experimenten zoals A/B testing. Ze kijken steeds naar hoe mensen reageren op één bepaalde digitale prikkel, waarbij alle overige variabelen constant worden gehouden. Ze meten dan bijvoorbeeld

dat er meer mensen op een blauwe dan een groene knop drukken, meer op een rechthoekige dan een ronde, en meer op de ene tekst dan de andere. Zo vogelen ze uit welke combinatie van kleur, vorm en tekst het meeste succes heeft in het verleiden van mensen om op een knop te drukken. Hele teams van Big Data-onderzoekers zijn dag en nacht bezig uit te zoeken hoe ze mensen ('gebruikers') kunnen leiden naar advertenties, politieke campagnes, games en andere applicaties, en hoe ze de aandacht van de gebruikers zo lang mogelijk kunnen vasthouden. De vraag of internettende personen wel of niet daarvan op de hoogte zijn, doet er eigenlijk niet toe: weliswaar moeten gebruikers meestal nog wel toestemming geven, maar als ze dat niet doen kunnen ze veelal geen gebruik maken van de betreffende applicatie. De grootste leugenaars in *Everybody Lies*, en dat vind ik de belangrijkste boodschap van het boek, blijken de internationale corporaties en regeringen te zijn. Zij laten al die Big Data-onderzoeken uitvoeren, incasseren waar mogelijk de commerciële of politieke winst, en wanneer dat zo te stade komt verdoezelen ze het allemaal ook nog eens.

Het wordt langzaam duidelijk dat we het in het kader van Big Data-onderzoek niet meer over louter 'bewustzijn' kunnen hebben, maar dat we een en ander beter duiden met een term als *digital governmentality*. De echte kracht van Big Data ligt niet zozeer in het bestuderen maar in het aansturen van onwetende en wetende mensen. Daarom representeert 'de onzichtbare onderzoeker' juist voor een Big Data-wetenschapper als SD geen ideaal, maar een bron van zorg. Hij is niet de enige. In april 2018 noemde Arjen Lubach (2018), de cabaretier, Facebook 'het grootste spionageproject uit de geschiedenis van de mensheid'. We weten nu iets meer over waartoe het heeft geleid. Het gaat er niet meer om dat Big Data de verkiezing van Trump correct kon voorspellen. Het werkelijke succes van Big Data-onderzoek is de mate waarin de onderzoekers het resultaat van die verkiezing hebben kunnen *maken* – voor zover we nu weten in ieder geval middels het exploiteren van Facebook-accounts door Cambridge Analytica en de algoritmes van een trollenfabriek van Internet Research Agency.

## Etnografie en Big Data

Joost Beuving stelt de vraag hoe etnografie zich verhoudt tot Big Data, en wat de implicaties daarvan zijn voor de etnografische praktijk. Bij het beantwoorden van die belangrijke vraag lijkt zowel Beuving als Bosch Big Data te reduceren tot methodologie. Beuving benadrukt de methodenstrijd, hoewel hij ook suggereert dat etnografie en Big Data 'werkelijk harmonieus' gecombineerd kunnen worden. Een dergelijk geïntegreerd samenspel lijkt ook Reinoud Bosch voor ogen te hebben, waarbij hij wel aangeeft dat die kwalitatievelevingen zich eerst maar eens moeten gaan bij- en omscholen. Ze gaan allebei wel erg gemakkelijk voorbij aan het vraagstuk van de incompatibiliteit van de betrokken paradigma's; etnografie en Big Data stellen nu eenmaal geheel andere vragen.

Maar, zoals ook SD heeft laten zien, Big Data is meer dan methodologie. Etnografie en Big Data verhouden zich niet alleen tot elkaar als rivaliserende of complemen-

taire methoden en epistemologieën. Zij staan ook tot elkaar als studie en studie-object. Anno 2018 behoort etnografie Big Data te bestuderen. In het domein van science and technology-studies bijvoorbeeld exploreren auteurs als Pink et al. (2016) op creatieve wijze hoe mensen Big Data in hun dagelijkse leven gebruiken (een vroege bron is Horst & Miller, 2012). Voor dit debat zou het relevant zijn te documenteren hoe Big Data-onderzoekers digitale materialiteit creëren en gebruiken. De kracht van etnografie hoeft daarbij niet te worden onderschat. Ook Big Data-onderzoekers hangen immers rond in cafés en kantoren, praten veel en graag over hun werk, en een etnograaf kan daarbij prima participeren en observeren (Seaver, 2017). Bij het ontsluiten van de denk- en ervaringswereld van Big Data-onderzoekers zou enige bekendheid met Big Data-jargon best van pas kunnen komen, zoals Bosch (2018) stelt. Maar dat hoeft niet per se. Een etnograaf doet zich immers graag voor als iemand die van zijn of haar informanten wil leren. Onbekendheid met het jargon heeft daarbij het voordeel dat kennis niet verondersteld hoeft te worden; de onenigheden over een basaal begrip als algoritme kunnen dan expliciet worden gemaakt (Seaver, 2017).

Etnografie kan de bevindingen van Stephens-Davidowitz contextualiseren en theoretiseren in een assemblage van wetenschap, technologie en macht. De paradox van *Everybody Lies* is dat als iedereen met en over Big Data liegt, SD dat zelf ook doet: er moeten momenten zijn waarop middels Big Data toch waarheid wordt gesproken. Leurs en Shepherd (2016) laten bijvoorbeeld zien, door het samenbrengen van Big Data en small data, dat de toenemende dataficering van de informatiestaat vooral negatieve gevolgen heeft voor reeds gemarginaliseerde subjecten, zoals ongedocumenteerde migranten. Maar door het ontrafelen van de digitale mechanismes van uitsluiting en ontmenselijking kan Big Data ook een rol spelen in 'het spreken van waarheid tegen macht' (Foucault, 2001), en dat kan uiteindelijk leiden tot verandering. Tevens zijn er legio voorbeelden van digitaal Do-It-Yourself citizenship (Leurs & Shephard, 2016; Lanzeni, 2016; McShane, Wilson & Meredyth, 2016) waarbij burgers Big Data aandragen om sociale verandering te bewerkstelligen. Kortom, we hebben het over meer dan een methodologie om de werkelijkheid te documenteren. Big Data is een set van sociale praktijken die vorm geeft aan onze eigen geschiedenis – en daarin verschilt ze uiteindelijk niet wezenlijk van etnografie.

## Literatuur

- Beuving, J. (2018). De toekomst van etnografie in het Big Data-tijdperk. *KWALON*, dit nummer.
- Bosch, R. (2018). De *Methodenstreit* in tijden van Big Data. *KWALON*, dit nummer.
- Foucault, M. (2001). *Fearless speech* (edited by Joseph Pearson). Los Angeles: Semiotext(e).
- Gigengack, R. (2017a). Ambachtelijke illusies: Het naturalisme geduid in de post-naturalistische tijd. *KWALON* 64, 22(1), 5-16.
- Gigengack, R. (2017b). Het post-naturalisme, de ambachtsman, en de meisjes op de Dam: Een dupliek aan Joost Beuving en Geert de Vries. *KWALON* 64, 22(1), 21-27.
- Horst, H.A., & Miller, D. (Eds.). (2012). *Digital anthropology*. London: Bloomsbury Academic.

- Lanzeni, D. (2016). Smart global futures: Designing affordable materialities for a better life. In S. Pink, E. Ardèvol, & D. Lanzeni (Eds.), *Digital materialities: Design and anthropology* (pp. 45-60). London: Bloomsbury Academic.
- Leurs, K., & Shepherd, T. (2016). Datafication & discrimination. In M.T. Schäfer & K. van Es (Eds.), *Social research in the age of data: The datafied society* (pp. 211-231). Amsterdam: Amsterdam University Press.
- Levitt, S., & Dubner, S.J. (2005). *Freakonomics: A rogue economist explores the hidden side of everything*. New York: HarperCollins Publishers.
- Levitt, S., & Dubner, S.J. (2009). *SuperFreakonomics: Global cooling, patriotic prostitutes, and why suicide bombers should buy life insurance*. New York: HarperCollins Publishers.
- Lubach, A. (2018). *Zondag met Lubach*. Aflevering 10. Verkregen van <https://www.vpro.nl/zondag-met-lubach/kijk/afleveringen/seizoen-8/aflevering-10.html>
- McShane, I., Wilson, C.K., & Meredyth, D. (2016). Digital interventions in declining regions. In S. Pink, E. Ardèvol, & D. Lanzeni (Eds.), *Digital materialities: Design and anthropology* (pp. 195-212). London: Bloomsbury Academic.
- Okely, J. (2012). *Anthropological practice: Fieldwork and the ethnographic method*. Oxford: Berg.
- Pink, S., Ardèvol, E., & Lanzeni, D. (2016). *Digital materialities: Design and anthropology*. London: Bloomsbury Academic.
- Seaver, N. (2017). Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2), 1-12.
- 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 Publishers.