

Een nudge voor objectiever personeelsbeleid

Vooroordelen leiden bij personeelsbeleid tot suboptimale beslissingen. Door een verandering van de evaluatiemethode valt het effect van vooroordelen bij promotie en wervingsbeslissingen terug te dringen. Door kandidaten gezamenlijk in plaats van individueel te evalueren, kunnen evalueerders de beste kandidaat identificeren en de kwaliteit van hun beslissingen verbeteren.

ALEXANDRA VAN GEEN

Promovendus aan Harvard University en vanaf september 2014 universitair docent aan de Erasmus Universiteit Rotterdam

In een recent experimenteel onderzoek (Moss-Racusin *et al.*, 2012) onder wetenschappers van vooraanstaande universiteiten in de VS moesten zij het profiel van een student evalueren wat betreft een vacature voor laboratoriummanager. De helft van deze wetenschappers kreeg een profiel met een meisjesnaam (Jennifer) onder ogen. De andere helft kreeg een dergelijk profiel gepresenteerd voorzien van een jongensnaam (John). Wetenschappers die het mannelijke profiel evalueerden, noemden deze kandidaat capabeler, waren meer bereid hem aan te nemen, boden hem meer begeleiding aan, en wilden hem een 4000 dollar hoger startsalaris geven dan de wetenschappers die het vrouwelijke – maar verder identieke – profiel evalueerden. Wetenschappers zijn wat dit betreft niet anders dan mensen in andere sectoren: experimentele studies wijzen uit dat ook daar genderdiscriminatie een factor is bij beslissingen over promoties en het aannemen van kandidaten (Riach en Rich, 2002; Neumark *et al.*, 1996).

Vaak is zulke discriminatie helemaal niet bewust of opzettelijk, vandaar dat men kan spreken van ‘impliciete discriminatie’. Onderzoek toont aan dat de meeste mensen, als het bijvoorbeeld gaat om huidskleur, gender, leeftijd en religie, stereotiepe denkbeelden hebben. Mensen zijn zich dikwijls

helemaal niet bewust van die denkbeelden, maar laten zich er – ongemerkt – toch door beïnvloeden (Bertrand *et al.*, 2005). Het is overigens niet zo dat mannen meer geneigd zijn tot genderdiscriminatie dan vrouwen (Moss-Racusin *et al.*, 2012). Mannen en vrouwen zijn het product van dezelfde maatschappij en houden er dus onbewust dezelfde vooroordelen op na.

BELEIDSOPTIES

Genderdiscriminatie is moeilijk goed aan te pakken. Diversiteitstrainingen werken niet goed (Dobbin *et al.*, 2007) en omdat gendervooroordelen bij zowel mannen als vrouwen voorkomen, is het veranderen van de gendersamenstelling van werving-en-selectiecommissies ook niet effectief (Bagues en Esteve-Volart, 2010). Het instellen van quota werkt wel uitstekend om het percentage mensen uit ondervertegenwoordigde groeperingen te vergroten, maar het is een weinig fijnzinnige methode die de keuzevrijheid vermindert en die er bovendien in individuele gevallen toe kan leiden dat de beste kandidaat buiten de boot valt. Zogeheten ‘blinde evaluaties’ zijn wel een succesvolle manier om genderdiscriminatie uit te sluiten. Goldin en Rouse (2000) bestudeerden het effect van audities achter een scherm bij het selecteren van nieuwe musici voor een symfonieorkest, zodat de selectiecommissie de kandidaten niet kon zien. Uit hun onderzoek bleek dat vrouwen 25 procent meer kans hadden te worden aangenomen bij deze blinde audities. Zo konden de selectiecommissies onvooroordeeld dus de beste kandidaat bepalen. Helaas is zo’n scherm in de praktijk bij de meeste wervings- en selectieprocedures niet goed haalbaar.

NUDGE

Bohnet *et al.* (2012) stellen een alternatieve methode voor om genderdiscriminatie aan te pakken, een zogenoemde *nudge* (Thaler en Sunstein, 2009). Een ‘nudge’ (por of duwtje) is een interventie die de vrije keuze van beslisters niet beperkt, maar die – door gebruik te maken van inzichten uit de psycholo-

Overzicht van evaluatiemethoden TABEL 1

Methoden	Evaluatiemethode	Hoge presteerder	Lage presteerder
1	Gezamenlijk	Man	Vrouw
2	Gezamenlijk	Vrouw	Man
3	Individueel	Man	x
4	Individueel	x	Man
5	Individueel	Vrouw	x
6	Individueel	x	Vrouw

gie – mensen helpt de kwaliteit van hun keuzes te verbeteren.

De nudge van Bohnet *et al.* (2012) is een verandering in de evaluatiemethode van afzonderlijke evaluatie (beoordeling van één individuele kandidaat) naar gezamenlijke evaluatie (het onderling vergelijken van meerdere kandidaten tegelijkertijd). Deze methode is geïnspireerd op psychologisch onderzoek dat aantoont dat gezamenlijke vergelijking tot rationelere keuzeprocessen leidt dan afzonderlijke evaluatie. Bijvoorbeeld: bij afzonderlijke evaluatie van een product wordt de merknaam belangrijker gevonden dan producteigenschappen en prijs, terwijl bij gezamenlijke evaluatie het omgekeerde het geval is (Nowlis en Simonson, 1997). Ook willen mensen bij afzonderlijke evaluatie meer betalen voor een kleine portie ijs in een kleine, tot de rand gevulde beker dan voor een veel grotere portie ijs in een grote, maar niet helemaal gevulde beker. Bij een gezamenlijke evaluatie is het tegendeel het geval (Hsee *et al.*, 1999). Bij afzonderlijke evaluatie willen mensen een hoger bedrag doneren voor dierenwelzijn, terwijl ze bij een gezamenlijke evaluatie meer willen doneren voor mensenwelzijn (Kahneman en Ritov, 1994). Bohnet *et al.* (2012) onderzochten of dit ook opgaat voor promotie- en rekruteringsbeslissingen. Bij veel van deze beslissingen worden op dit moment afzonderlijke evaluatiemethodes gebruikt. In een enquête onder senior bedrijfsleiders van grote bedrijven in de VS blijkt dat bij dertig procent van de promotiebeslissingen slechts één kandidaat wordt overwogen (Penn, Schoen & Berland, 2012). Afzonderlijke evaluatie methodes komen ook vaak voor bij rekruteringen; ongeveer de helft van de kandidaten wordt een voor een vergeleken, in de andere helft is een zoekproces met meerdere kandidaten tegelijk overwogen (Van Ommeren en Russo, 2009; Oyer en Schaefer, 2010).

De effectiviteit van de nudge van Bohnet *et al.* (2012) werd onderzocht door middel van experimenten bij het Decision Science Lab van de Harvard Kennedy School.

METHODE VAN ONDERZOEK

Het onderzoek bestaat uit twee fases en het doel ervan is om het effect van de interventie op de selectie van kandidaten te testen. In de eerste fase van het onderzoek vergaren de auteurs de informatie om de te evalueren profielen op te stellen. In deze onderzoeksfase voert een eerste groep deelnemers aan de studie (de ‘werkzoekenden’) in verschillende rondes een taak

uit. Vervolgens worden profielen opgesteld van deze kandidaten, met onder andere de informatie over hun prestaties in de eerste ronde van de taak. In de tweede fase van het onderzoek kan dan – door middel van deze profielen – de evaluatiemethode getest worden op een nieuwe groep proefpersonen (de ‘werkgevers’). Proefpersonen in de tweede fase van het onderzoek (de ‘werkgevers’) moeten dus profielen van werkelijk bestaande personen evalueren, en worden ook betaald al naar gelang de prestatie van de geselecteerde kandidaat in een latere ronde van de taak.

Bij afzonderlijke evaluatie van een product wordt de merknaam belangrijker gevonden dan producteigenschappen en prijs, terwijl bij gezamenlijke evaluatie het omgekeerde het geval is

Eerste fase

In de eerste fase van het onderzoek werd een groep mannelijke en vrouwelijke proefpersonen gevraagd om in het lab verschillende rondes van een rekenopdracht (series van tweecijferige getallen optellen) of een taalopdracht (een woordzoeker) uit te voeren. Uit onderzoek blijkt dat ‘verbaal sterk’ met vrouwen wordt geassocieerd en ‘wiskundig sterk’ met

Resultaten van selectiebeslissing

TABEL 2

	Afzonderlijke evaluatie		Gezamenlijke evaluatie	
	Mannelijke kandidaat	Vrouwelijke kandidaat	Mannelijke kandidaat	Vrouwelijke kandidaat
Wiskunde-opdracht				
Hogere presteerder	0,66 (N=29)	0,44 (N=32)	0,52 (N=31)	0,57 (N=35)
Lagere presteerder	0,65 (N=26)	0,53 (N=30)	0,03 (N=35)	0,06 (N=31)
Taalopdracht				
Hogere presteerder	0,64 (N=22)	0,81 (N=21)	0,52 (N=31)	0,55 (N=29)
Lagere presteerder	0,35 (N=20)	0,50 (N=22)	0,07 (N=29)	0,16 (N=31)

mannen. Een taalopdracht is daarom stereotypisch gezien een vrouwelijke taak, terwijl een rekenopdracht stereotypisch gezien een mannelijke taak is (Perie *et al.*, 2005). De feitelijke prestatieverschillen tussen de genders zijn overigens niet eenduidig en verschillen per land en studiepopulatie. Soms vindt men het verwachte verschil, soms vindt men geen verschil; en recentelijk vindt men in diverse landen zelfs een omkering van het genderverschil, met vrouwen die in rekentaken beter presteren dan mannen (Xie en Shauman, 2005; Guiso *et al.*, 2008). In het onderzoek van Bohnet *et al.* (2012) verschilden mannen en vrouwen niet significant van elkaar in hun gemiddelde score bij de reken- of taalopdracht. Er was ook geen significant verschil in de distributie van de scores. Het was bijvoorbeeld niet zo dat mannen significant extremere scores hadden dan vrouwen. Het was ook niet zo dat mannen sneller leerden dan vrouwen en dat ze hun scores in latere rondes van de opdracht sterker verbeterden. Wel zijn de individuele scores goede voorspellers voor scores in latere rondes.

Van de proefpersonen in de gezamenlijke evaluatieprocedure kiest slechts 8 procent voor een kandidaat met een slechte prestatie in de eerste ronde, tegenover 51 procent van de proefpersonen in de afzonderlijke evaluatieprocedure

Tweede fase

Voor de tweede fase van het onderzoek participeerde er een nieuwe groep proefpersonen. Deze proefpersonen moesten de profielen van de eerste groep evalueren en een kandidaat selecteren. De betaling van de proefpersonen werd bepaald door de prestaties in de tweede ronde van de door hen geselecteerde kandidaat. Zij hadden er dus belang bij om de persoon te selecteren van wie ze de hoogste score verwachtten. Er werden twee verschillende evaluatiemethodes getest in het experiment: een afzonderlijke evaluatiemethode en een gezamenlijke (tabel 1). Bij de afzonderlijke methode moesten proefpersonen het profiel van één kandidaat evalueren. Zij konden kiezen tussen deze kandidaat en een randomselectie uit de groep kandidaten door de onderzoekers, het profiel van deze random geselecteerde kandidaat is dus niet bekend bij de proefpersoon. Zij wisten wel wat de gemiddelde prestatie van alle kandidaten was. Bij de gezamenlijke evaluatiemethode moesten proefpersonen profielen van twee kandidaten evalueren (een mannelijke en een vrouwelijke). In dit geval konden ze een van de kandidaten kiezen, of kiezen voor een

randomselectie van een kandidaat uit de kandidatenpool door de onderzoekers. Zoals in tabel 1 te zien is, zijn er bij elke evaluatiemethode diverse behandelingen. De eerste rij van tabel 1 beschrijft de methode waarbij de proefpersonen profielen van een hoog presterende man en een laag presterende vrouw moeten evalueren. In de tweede rij wordt de methode beschreven waarbij proefpersonen juist een hoog presterende vrouw en een laag presterende man moeten evalueren. De hoog presterende man en de hoog presterende vrouw hebben overigens een identieke score, evenals de laag presterende man en de laag presterende vrouw. Door gender en prestatie te variëren per methode kan het effect van zowel gender als prestatie worden onderzocht.

RESULTATEN

Wat kiezen de proefpersonen in de tweede fase van het onderzoek? Tabel 2 geeft de resultaten weer. In het individuele geval blijkt de gender van de kandidaat een belangrijke drijfveer in hun selectiebeslissing. Voor de taalopdracht zijn proefpersonen veel meer geneigd om een vrouw te selecteren dan een man met dezelfde score in de eerste ronde. Bij de individuele methode met een hoger presterende kandidaat kiest namelijk 81 procent van de evalueerders voor de kandidaat als het een vrouw is (kolom 2 van tabel 2), en maar 64 procent als het een man is met een identieke hoge score (kolom 1 van tabel 2). Hetzelfde patroon is te zien bij de individuele methode met een lager presterende kandidaat: 50 procent kiest voor deze kandidaat als dit een vrouw is, en maar 35 procent als het een man is. In de evaluatie van kandidaten voor de rekenopdracht zijn proefpersonen meer geneigd om een man te selecteren dan een vrouw met dezelfde score in de eerste ronde. Bij de individuele methode met een hoger presterende kandidaat kiest 66 procent van de evalueerders voor de kandidaat als het een man is (kolom 1 van tabel 2), en maar 44 procent als het een vrouw is (kolom 2 van tabel 2). Hetzelfde patroon treedt op bij de individuele methode met een lager presterende kandidaat: 65 procent kiest voor deze kandidaat als dit een man is, en maar 53 procent als het een vrouw is. Dit gendereffect is zelfs zo sterk bij de selectie voor de rekenopdracht dat proefpersonen meer geneigd zijn om een man met een lage score te kiezen dan een vrouw met een hogere score. Prestatie van de kandidaten speelt geen enkele rol in de selectie van de kandidaten in de afzonderlijke evaluatie.

In de evaluatie bij de gezamenlijke vergelijking ziet het er heel anders uit. Voor zowel de wiskundeopdracht als de taalopdracht selecteren mensen de kandidaat met de hoogste score in de eerste ronde, en speelt gender geen significante rol (tabel 2). Door middel van de gezamenlijke evaluatiemethode baseren proefpersonen zich op de relevante indicatoren, en laten zich niet afleiden door vooroordelen en heuristieken. Ze maken daardoor betere keuzes en verkrijgen een hogere opbrengst. Van de proefpersonen in de gezamenlijke evaluatieprocedure kiest slechts 8 procent voor een kandidaat met een slechte prestatie in de eerste ronde, tegenover 51 procent van de proefpersonen in de afzonderlijke evaluatieprocedure.

CONCLUSIE EN BELEIDSIMPLICATIE

Organisaties die willen voorkómen dat rekrutering, promoties en functietoedeling bepaald worden door irrelevante criteria die worden getriggerd door stereotypen en die willen dat

de beste kandidaat de baan krijgt, kunnen gebruikmaken van een nudge – kandidaten niet afzonderlijk maar gezamenlijk evalueren – en de vergelijking gebruiken als basis voor hun beslissingen. Deze nudge is relatief makkelijk in te passen. Als er niet werkelijk meerdere kandidaten voor een functie zijn, is een vergelijking ook vaak mogelijk – bijvoorbeeld door naar profielen te kijken van kandidaten in voorgaande jaren, of door te kijken naar profielen van kandidaten in vergelijkbare organisaties en functies. Voor het succes van de interventie is het overigens wel belangrijk om de diverse criteria, en het bijbehorende gewicht ervan, vooraf te bepalen en constant te houden. Onderzoek wijst namelijk uit dat, wanneer er meerdere criteria zijn waar kandidaten mee kunnen worden vergeleken, proefpersonen meer gewicht geven aan het criterium waar de stereotiep bevoorrechte groep relatief hoog op scoort. (Norton *et al.*, 2004). Men vond bijvoorbeeld dat als een man en een vrouw werden vergeleken voor de functie van manager, en de man meer ervaring heeft maar de vrouw meer opleiding, dat proefpersonen dan aangeven ervaring belangrijker te vinden dan opleiding en daarom voor de man te gaan. Echter, in het experimentele scenario waar de man meer opleiding heeft, maar de vrouw meer ervaring, geven proefpersonen aan dat ze opleiding belangrijker vinden, en dáárom voor de man kiezen.

Hoewel het niet altijd mogelijk zal zijn om beslissingen te bundelen en kandidaten expliciet te vergelijken, toont dit onderzoek aan dat gezamenlijke evaluatie in het personeelsbeleid zowel efficiency als gelijkheid kan vergroten.

LITERATUUR

- Bagues, M. en B. Esteve-Volart (2010) Can gender parity break the glass ceiling? Evidence from a repeated randomized experiment. *Review of Economic Studies*, 77(4), 1301–1328.
- Bertrand, M., D. Chugh en S. Mullainathan (2005) Implicit discrimination. *American Economic Review*, 95(2), 94–98.
- Bohnet, I., A. van Geen en M. Bazerman (2012) When performance trumps gender bias. *HBS Working Paper*, 12(083).
- Dobbin, F., A. Kalev en B. Kelly (2007) Diversity management in corporate America. *Contexts*, 6(4), 21–28.
- Goldin, C. en C. Rouse (2002) Orchestrating impartiality: The impact of blind auditions on female musicians. *American Economic Review*, 90(4), 715–741.
- Guiso, L., F. Monte, P. Sapienza en L. Zingales (2008) Diversity: culture, gender, and math. *Science*, 320(5880), 1164–1165.
- Hsee, C.K., S. Blount, G.F. Loewenstein en M.H. Bazerman (1999) Preference reversals between joint and separate evaluations of options: A review and theoretical analysis. *Psychological Bulletin*, 125(5), 576–590.
- Kahneman, D. en I. Ritov (1994) Determinants of stated willingness to pay for public goods. A study in the headline method. *Journal of Risk and Uncertainty*, 9(1), 5–38.
- Moss-Racusin, C.A., J.F. Dovidio, V.L. Brescoll et al. (2012) Science faculty's subtle gender biases favor male students. *PNAS*, 109(41), 16474–16479.
- Neumark, D., R.J. Bank en K.D. Van Nort (1996) Sex discrimination in restaurant hiring: an audit study. *The Quarterly Journal of Economics*, 113(3), 915–941.
- Norton, M.I., J.A. Vandello en J.M. Darley (2004) Casuistry and social category bias. *Journal of Personality and Social Psychology*, 87(6), 817–831.
- Nosek, B., M. Banaji en A.G. Greenwald (2002) Math = male, me = female, therefore math ≠ me. *Journal of Personality and Social Psychology*, 83(1), 44–59.
- Nowlis, S.M. en I. Simonson (1997) Attribute-task compatibility as a determinant of consumer preference reversals. *Journal of Marketing Research*, 34(2), 205–218.
- Oyer, P. en S. Schaefer (2011) Personnel economics: hiring and incentives. In: Card, D. en O. Ashenfelter (red.) *Handbook of Labor Economics*, vol. 4B. North Holland: Elsevier, 1769–1823.
- Penn, Schoen & Berland Associates, inc. (2012) *The Capstone Project*. Publicatie op msb.georgetown.edu.
- Perie, M., R. Moran en A.D. Lutkus (2005) *NAEP 2004 trends in academic progress: three decades of student performance in reading and mathematics*. Washington, DC: U.S. Department of Education. Office of Educational Research and Improvement. National Center for Education Statistics.
- Riach, P.A. en J. Rich (2002) Field experiments of discrimination in the market place. *The Economic Journal*, 1124, 480–518.
- Thaler, R.H. en C.R. Sunstein (2009) *Nudge. Improving decisions about health, wealth, and happiness*. Londen: Penguin Books.
- Ommeren, J. van, en G. Russo (2014) Firm recruitment behaviour: sequential or non-sequential search? *Oxford Bulletin of Economics and Statistics*, 76(3), 432–455.
- Xie, Y. en A.K. Shauman (2005) *Women in science: career processes and outcomes*. Cambridge, MA: Harvard University Press.