

Grote kwesties

Ter introductie

Volgde u ook met belangstelling de discussie die zich vorig jaar kort voor de zomer in de Vlaamse media ontspon over de plannen voor de hervorming van het secundair onderwijs? Bijna een jaar geleden, inderdaad. En daarom wellicht al ondergesneeuwd en vergeten door de overvloed aan andere kwesties die ons sindsdien bezig hielden. Of de stad van iedereen is, bijvoorbeeld, en waarom (en hoe) mensen achter loketten (best) neutraliteit moeten belichamen. Of Lance ons nu al dan niet echt belazerde en of de media hem te veel (of nog niet genoeg) aan de schandpaal nagelden. Of doping enkel maar in de wielersport voorkomt. Of er iets (en eventueel wat) te doen valt aan de golf van bedrijfssluitingen en ontslagen. Of het woord 'allochtoon' nu al dan niet moet worden geschrapt en waarom dan wel. Of we ongeweten paardenvlees hebben gegeten. Of de UNESCO wel goed begrijpt wat Aalsters carnaval juist betekent. Misschien werkte het bij u anders. Maar ondanks deze maalstroom van belangrijke en tot veel nadenken stemmende kwesties heeft een bepaald aspect van de discussie over de onderwijs-hervorming mij niet meer losgelaten. Niet overheersend en constant, maar toch wel hardnekkig. Een beetje zoals men zich soms moeilijk kan ontdoen van een of andere melodie die men vroeger op de dag of eerder in de week heeft gehoord. Bewitched, bothered and bewildered. Da's het gevoel waar ik sindsdien iets meer dan tevoren mee rondloop.

Laat me maar meteen met de deur in huis vallen. Waar ik sinds vorige zomer een mengeling van fascinatie, bezorgdheid en verwarring aan overhoud, is de vaststelling dat men op basis van het toen gevoerde debat nauwelijks een duidelijk, min of meer eenduidig en door iedereen gedragen oordeel bleek te kunnen vormen over de zin of onzin van het hervormingsvoorstel of van belangrijke onderdelen ervan. Een gemis aan eenstemmigheid

bij wat tegenwoordig zo mooi 'de stakeholders' heet. Te verwachten wellicht, gezien de onvermijdelijke belangentegenstellingen. Maar veel minder verwacht, alleszins op het eerste gezicht, was ook het ontbreken van eenstemmigheid bij de deskundigen. Deels wellicht omdat men elk hervormingsvoorstel vanuit meerdere doelstellingen of objectieven kan beoordelen. Mikt men op een grotere gelijkheid van onderwijsuitkomsten? Willen we globaal bekeken betere schoolprestaties? Moet men opteren voor meer welzijn op school? Beogen we de optimale ontplooiing van de individuele talenten van elk kind? Is een grotere ongelijkheid in uitkomsten aanvaardbaar als iedereen erop vooruit gaat, maar de beteren meer dan de anderen? Best mogelijk dus en zelfs waarschijnlijk dat de deskundigen verschillen inzake de criteria die zij hanteren om hervormingsvoorstellen te beoordelen en dat ze aan de mogelijke doelstellingen een verschillende rangorde of een verschillend gewicht geven. Geen probleem, wat mij betreft. Integendeel, dat deskundigen oordelen vanuit verschillende invalshoeken en geen slachtoffer blijken van een 'pensée unique' vind ik eerder geruststellend, want een belangrijke voorwaarde voor een vruchtbaar maatschappelijk debat en – als ik er dan toch wat grote woorden mag tegenaan gooien – zelfs een garantie voor meer transparantie en een groter democratisch gehalte van de besluitvorming.

Maar toch. De discussie van vorige zomer illustreerde vooral, lijkt mij, dat onenigheid onder deskundigen in dat geval slechts ten dele voortkwam uit een verschillende waardering van de doelstellingen van de hervorming, maar ook veel te maken had met de interpretatie van onderzoeksresultaten. Toegegeven, mogelijk laten sommige deskundigen zich ook bij de interpretatie van onderzoeksresultaten meer dan ze zelf beseffen leiden door hun beleidsmatige of politieke voorkeuren, en iedereen leest zeker niet even zorgvuldig. Niettemin – en al beschouw ik mij

niet als een specialist in de materie – lectuur van het materiaal waarmee de deskundigen elkaar in de media te lijf gingen, maakte mij alleszins duidelijk dat het gebrek aan eenstemmigheid ook berustte op een verschillende lectuur en interpretatie van de onderzoeksresultaten zelf. Alweer begrijpelijk, vind ik. Deskundigen behoren immers tot verschillende vakgebieden en de werktuigen eigen aan hun respectievelijke domeinen zijn niet identiek. Ze hantieren verschillende methoden, gebruiken een ander jargon en lezen daarom mogelijk met verschillende brillen. Zelfs binnen een bepaald onderzoeksdoelmeisn gebruikt men verschillende statistische technieken en kan de informatiebasis erg divers zijn.

In ieder geval en wat de oorzaak ook moge wezen, het gevolg is meestal dat er ook onder deskundigen, zelfs als men alle kennis samenlegt, nog een ruime marge van onzekerheid kan bestaan over de interpretatie van onderzoeksresultaten en dus a fortiori over mogelijke effecten van de voorgestelde hervormingen. En vandaar dus de vraag die in mijn hoofd blijft rondspoken: wat betekent deze vaststelling nu eigenlijk voor de relatie tussen onderzoek en beleid in het algemeen en voor de roep naar meer ‘evidence-based policy’ in het bijzonder? Toegegeven, misschien niet de meest belangrijke aller grote kwesties. Maar zoals Gertrude Stein ooit stelde als aanhef van één van haar *Lectures in America*: “One cannot come back too often to the question what is knowledge and to the answer knowledge is what one knows”. Vandaar.

Over argwaan en dergelijke

Alvorens verder te gaan, moet ik u bekennen dat ik mij al langer ongemakkelijk voel bij het gebruik van de term ‘evidence-based policy’. Ik heb dat wel meer met termen die men overneemt uit een andere taal of invoert uit een andere discipline. ‘Competenties’ bijvoorbeeld. Of ‘capability’.

Het boekje dat mij ongeveer 30 jaar geleden blijvend met deze argwaan opzadelde is *Freud and Man's Soul* van de psychoanalyst en kinderpsychiater Bruno Bettelheim. Film liefhebbers kennen hem wellicht als de man die zichzelf speelde in Zelig van Woody Allen, werkers in de geestelijke gezondheidszorg mogelijk als grondlegger van de University of Chicago's Sonia Shankman Orthogenic

School en de bedenker van een ondertussen uitermate controversiële en gecontesteerde theorie over de oorzaken van autisme, u en ik eventueel als de auteur van *The Uses of Enchantment*, een psychoanalytische interpretatie van sprookjes, *The Informed Heart* en *Surviving*, geschriften gebaseerd op zijn ervaringen in de concentratiekampen, of *A Good Enough Parent*, een Doctor-Spock-achtig boek met als belangrijkste boodschap de geruststelling dat ouders niet perfect moeten zijn; ‘goed genoeg’ voldoet, zegt Bettelheim.

Freud and Man's Soul is bedoeld als een korte, ongeveer honderd kleine bladzijden lange bespreking van de kernbegrippen van de Freudiaanse denkwijze. Inleidend dus, maar tegelijk ook pleitend voor een herinterpretatie van het werk van de Weense grondlegger van de psychoanalyse. Zo'n herinterpretatie is nodig, zegt Bettelheim, omdat de Engelse vertaling van Freuds werk een ware ravage heeft aangericht. De eerste zin van het voorwoord van het boekje kan niet duidelijker. “The English translations of Freud's writings are seriously defective in important respects and have led to erroneous conclusions, not only about Freud the man but also about psychoanalysis. This applies even to the authoritative *Standard Edition of the Complete Psychological Writings of Sigmund Freud*”.

Let wel: ‘defective’; zegt Bettelheim over de vertaling, niet noodzakelijk ‘erroneous’ dus. Al wijst hij ook op passages die gewoonweg fout zijn vertaald. Zijn belangrijkste bezwaar is dat de Engelse vertaling heel dikwijls de elementen uitvlakt, die impliciet verwijzen naar de context waarin Freud schreef, of aanleiding geeft tot een interpretatie die ver staat van wat Freud eigenlijk bedoelde en de psychoanalytische praktijk sterk hinderde. Voor alle duidelijkheid: deze stelling is redelijk boud. Want de vertaling van de Standard Edition is verzorgd door één van Freuds vertrouwelingen, mee gesuperviseerd door Freuds dochter en tot op zekere hoogte door Freud zelf goedgekeurd.

Welke argumenten legt Bettelheim op tafel om zijn stelling te staven?

Een goed voorbeeld van zijn werkwijze betreft het essay *Das Unbehagen in der Kultur*, vertaald als *Civilization and its Discontents*. Over de keuze om

'Kultur' te vertalen als 'Civilization' en niet als 'Culture' kan men redetwisten, zegt Bettelheim, ook al is er in het Duits een duidelijk verschil tussen 'Kultur' en 'Zivilisation' en weten we, onder meer uit zijn brieven, dat Freud de voorkeur gaf aan 'Kultur'. Maar voor 'Unbehagen' geldt dit niet. Het correcte equivalent van 'discontent' is in het Duits 'Unzufriedenheit'. Ontevredenheid dus en niet onbehagen. Een betere vertaling zou dus geweest zijn 'malaise' of 'discomfort'. Voor een ander goed voorbeeld zorgt het woord 'Trieb', vertaald als 'Instinct'. Welnu, zegt Bettelheim, ook Freud gebruikt in zijn geschriften wel degelijk het woord 'Instinkt'. Maar enkel wanneer het gaat over aangeboren instincten van dieren. Waarom gebruikt de vertaler niet 'drive' (drijfveer) of 'impulse'? Dit laatste heeft als bijkomend voordeel dat men er het adjectief 'impulsive' gratis bovenop krijgt. De Franse vertaling gebruikt trouwens 'pulsion'. Maar het voorbeeld dat Bettelheim nog het meest ergert, is de vertaling van 'Seele' (ziel) door 'Mind' (geest). De Engelse taal beschikt met 'soul' over een woord dat voor een perfecte vertaling kan zorgen. Waarom gebruikt men dat dan niet?

Voor Bettelheim wijzen al deze voorbeelden op een vertaling die heel sterk behavioristisch gekleurd is. Ten onrechte, zegt Bettelheim, want "Behaviorism concentrates on what can be seen from the outside, what can be studied objectively by an uninvolved outsider, what can be replicated and assigned numerical values" (p.108). Het meest duidelijke is dit nog door de manier waarop men twee van de centrale begrippen van het Freudiaanse begrippenapparaat – 'Ich' en 'Es' – niet heeft vertaald door dagelijks gebruikte Engelse woorden – 'I' en 'It' – maar woorden uit het Latijn – 'Ego' en 'Id'.

Ik neem aan dat voorgaande duidelijk maakt dat mijn argwaan tegenover woorden die men leent uit een andere taal of een andere discipline niet voortkomt uit een afkeer van vreemde talen, Engels, Frans, Duits, Spaans, Italiaans, Russisch, Arabisch of om het even welke andere taal. Ik vrees niet alleen dat het dikwijls moeilijk is om de betekenis van de term in de originele taal adequaat te begrijpen – het lenen van een term uit een vreemde context suggereert op zich al dat een exacte vertaling moeilijk is, want waarom anders de vreemde term gebruiken – maar vooral ook dat in zo'n geval wat taalkundigen aanduiden als de 'indexicale betekenis' van de term gaat overwegen.

En terzijde nog even een tip. Wie een beter inzicht wil krijgen in de oorsprong en de ontwikkeling van de psychoanalytische beweging, de receptie van het werk van Freud, zijn relaties met de eerste generaties van 'medestanders' en de strijd tussen de verschillende fracties kan ik de lectuur aanbevelen van Georg Makari, *Revolution of the Mind. The Creation of Psychoanalysis*, Harper Collins Publishers, New York, 2010, 613 p.

Wat betekent 'evidence-based'?

Gaat het bij 'evidence-based policy' om meer of om iets anders dan een 'beleid gebaseerd op kennis van zaken'? Zo nee, waarom dan het Nederlandse equivalent niet gebruiken? Zo ja, waarin ligt dan het verschil en waarom vindt men dan geen vertaling die dit verschil duidelijk aangeeft? En suggereert de roep naar 'evidence-based policy' dat beleid vroeger en vandaag doorgaans geen rekening hield of houdt met kennis? Vreemd toch. Dat beleidmakers slechts beschikken over onvolledige kennis, mogelijk zelfs werken met kennis die achteraf fout of onjuist blijkt, en dus in de praktijk wellicht onvermijdelijk uitgaan van een mengelmoes van kennis en gokwerk of belangenverdediging, wie zal dat tegenspreken? Maar strikt genomen gaat het in elk van deze gevallen toch om 'beleid gebaseerd op kennis van zaken'. Niet? In dergelijke gevallen is mijn eerste reactie meestal een zelfkritische. Heb ik de betekenis wel goed begrepen, vraag ik me dan af en probeer soelaas te vinden in een reeks verklarende woordenboeken. 'Based' en 'policy' lijken mij weinig problemen op te leveren, maar wat betekent 'evidence' nu exact? Volgens de Oxford English Dictionary is 'evidence' "the available body of factors or information indicating whether a belief or proposition is true or valid". De Chambers English Dictionary omschrijft het als "means of proving an unknown or disputed fact, support for a belief, indication, information in a law case, testimony, a witness or witnesses collectively". De Collins Cobuild English Language Dictionary zegt over 'evidence': "1.1 anything that you see, experience, read, or are told that causes you to believe that something is true or has really happened; 1.2 the information which is used in a court of law to try to prove something and which is obtained from documents, objects or witnesses; 1.3 the way a person behaves in a particular situation, which shows what they feel or think about it".

Gegeven deze omschrijvingen is 'evidence' misschien inderdaad niet zo gemakkelijk te vertalen. Mogelijkheden zijn 'aanwijzing' of 'indicaties', 'spoor', 'getuigenis', 'bewijs'. Maar geen van deze vertalingen lijkt mij exact weer te geven wat er in het Engels taalidoom mee wordt bedoeld. Een beetje zoals je de betekenis van 'fair' niet adequaat kan vatten zonder te weten dat het ook kan verwijzen naar haarkleur. Misschien is het daarom nog best om 'evidence-based policy' te vertalen als 'een beleid gebaseerd op de beschikbare aanwijzingen'. Of 'aanwijzingen' hetzelfde is als 'met kennis van zaken' laat ik aan uw oordeel over. Belangrijker is evenwel dat het echte probleem met de term niet te remediëren valt via mijn geijkte werkwijze – consulteren van verklarende woordenboeken. Inderdaad, wie even 'googlet', merkt onmiddellijk dat wat men in de literatuur over 'evidence-based policy' onder 'evidence' begrijpt iets heel erg specifiek is en naar een erg strikt omschreven praktijk verwijst.

Ik ga u niet vervelen met een uitvoerige bespreking van wat u zoal op het web vindt, maar beperk me tot een korte situering. Wie dieper wil graven, kan Wikipedia als uitgangspunt nemen. Het item over 'evidence-based policy' bevat voldoende verwijzingen en de nodige links naar de sites van de belangrijke betrokken organisaties. Wie op een gemakkelijke en aangename manier zicht wil krijgen op het wel en wee van 'evidence-based policy' kan best (via de link www.bbc.co.uk/programmes/b01p-hhb9) de vorig jaar op de BBC-radio uitgezonden documentaire 'Bad Evidence' van Ben Goldacre beluisteren. Niet echt neutraal, maar wel voldoende informatief. De multi-taskers onder u hoeven er trouwens geen tijd door te verliezen. Luisteren naar de documentaire kan zonder problemen tijdens het uurtje strijken of joggen.

Verrassend misschien, maar de georganiseerde pleidooien voor 'evidence-based policy' in de zin zoals de term vandaag wordt gebruikt, dateren van de tweede helft van de jaren 1990 en verwijzen expliciet naar het voorbeeld van de zogeheten 'evidence-based medicine'.

Een van de iconische teksten in de ontwikkeling van de beweging voor 'evidence-based medicine' is een veel geciteerd artikel van Iain Chalmers, 'What do I want from health research and researchers when I am a patient?', in de *British Medical Journal* van 1995. De boodschap van het artikel wordt kernachtig uitgedrukt in de titel van een artikel dat Chalmers ongeveer een decennium later schrijft:

"Because professionals sometimes do more harm than good when they intervene in the lives of other people, their policies and practices should be informed by rigorous, transparent, up-to-date evaluations". (Iain Chalmers, Trying to Do More Good than Harm in Policy and Practice: The Role of Rigorous, Transparent, Up-to-date Evaluations. In: *Annals of the American Academy of Political and Social Sciences*, 2003, 589, 22-39)

In het BMJ-artikel gebruikt Chalmers twee persoonlijke anekdotes om zijn punt te maken. De eerste anekdote betreft de ervaringen van zijn echtgenote met een chiropractor, naar wie ze door haar huisarts verwezen was. Na ongeveer vijf behandelingen begon zij te geloven dat dit soort behandelingen haar konden verlossen van schouder- en rugpijn. Natuurlijk was ik opgetogen over het verdwijnen van symptomen waar ze al erg lang last van had, zegt Chalmers. Maar zelf begon ik dat geloof pas een paar jaar later te delen nadat ik een rapport las met een systematisch overzicht van de relevante gecontroleerde experimenten. "For me 'reliable evidence' about the effects of health care will usually mean evidence derived from systematic review of carefully controlled evaluative research" (p.1315). Als medische ingrepen dramatische verbeteringen in iemands toestand teweeg brengen – bijvoorbeeld door het vervangen van een slecht werkende heup door een prothese – dan zijn dergelijke studies niet echt nodig. Maar dergelijke situaties zijn eerder schaars. Meestal zijn de effecten van een ingreep minder dramatisch, zeker als er vormen van suggestie kunnen mee gemoeid zijn. "If these moderate but important effects are to be detected reliably, then systematic reviews of carefully controlled research will be required to produce the kind of evidence that I am likely to believe, and that I would wish those offering me care to take into account" (p.1316).

Illustreer dat de medische wereld niet noodzakelijk dezelfde mening is toegedaan, maar dikwijls ten onrechte koppig blijft aanmodderen, is de functie van de tweede anekdote. Enkele jaren geleden brak ik domweg mijn kuitbeen, zegt Chalmers. Een Amerikaans orthopedisch chirurg zegde me dat hij na het verdwijnen van de zwelling een gipsverband zou aanleggen dat er de volgende zes weken diende aan te blijven. Na thuiskomst, zesendertig uur later, zegde een Britse orthopedische chirurg dat een gipsverband niet nodig was. Een steunverband

dragen was voldoende en het advies was om het been zoveel als mogelijk te gebruiken bij het wandelen. "Because these two orthopaedic surgeons had prescribed very different forms of care, I asked the second one whether I might be entered into a randomized controlled trial to help resolve the contradictory advice that I had received. He explained to me that only doctors who are uncertain whether they are right or wrong collaborate in randomized trials – and he was certain that he was right" (p.1316). Waarvan akte, zegt Chalmers.

Van 'evidence-based medicine' naar 'evidence-based policy'

Het pleidooi van Chalmers en zijn geestesgenoten voor een medische praktijk gebaseerd op systematische overzichten van experimenteel onderzoek waarbij een behandelde groep vergeleken wordt met een groep die een placebo krijgt toegediend, vindt zijn oorsprong in een boek uit 1972. In *Effectiveness and Efficiency* argumenteerde de Britse epidemioloog Archie Cochrane dat "medicine had not organized its knowledge in any systematic, reliable and cumulative way; the result was chaotic, individualistic, often ineffective and sometimes harmful, patterns of care".

De pogingen om in deze situatie enige verandering te brengen, leidde in 1993 in de UK tot de oprichting van een organisatie, de Cochrane Collaboration, met als missie het up-to-date houden van de kennis over gezondheidszorg en gezondheidsbeleid. De twee methodologische principes die men hierbij hanteert zijn "the need for unbiased comparisons of different interventions ('randomised controlled trials') and the importance of collating, and constantly updating, evidence from different studies to arrive at the most reliable estimates of effects". Het belangrijkste instrument dat men hierbij gebruikt zijn de zogeheten 'Cochrane reviews', een volgens een vast stramen geschreven overzicht van de resultaten van onderzoek op basis van 'randomized controlled trials' (RCTs). De 'nursery' van de Cochrane Collaboration was de National Perinatal Epidemiology Unit met als directeur (Sir) Ian Chalmers.

In de tweede helft van de jaren 1990 verspreidt de belangrijkste argumentatie voor een 'evidence-based medicine' zich ook over andere beleidsdomeinen en merkt men in de Angelsaksische wereld

een toenemende interesse voor deze aanpak bij onderzoekers en beleidsmakers. Een belangrijke rol in de missionering rond 'evidence-based policy' wordt gespeeld door EPPI, het Evidence for Policy and Practice Information and Coordination Centre van het Institute of Education, University of London. Voor een gedetailleerde bespreking van de rol van dit instituut verwijs ik naar Ann Oakley, David Gough, Sandy Oliver and James Thomas, The politics of evidence and methodology: lessons from the EPPI-Centre, in: *Evidence & Policy*, 1, 1, 5-31.

1999 lijkt in zeker opzichten een sleuteljaar in de ontwikkeling van de beweging voor 'evidence-based policy'. Zowel Association of Research Centres in the Social Sciences als de School of Public Policy van de University College London, organiseren een belangrijke conferentie over 'evidence-based policy'. Men start de Campbell Collaboration, een zusterorganisatie van de Cochrane Collaboration die zich concentreert op overzichten van de onderzoeken naar de effecten van sociaal en onderwijsbeleid. Daarnaast beslist één van de belangrijkste Britse onderzoeksfinanciers, de Economic and Social Research Council (ESRC), om de beweging naar meer 'evidence-based policy' te ondersteunen door een toelage van 1,3 miljoen pond toe te kennen aan het 'Evidence Network', een centrum vergelijkbaar met de Cochrane Collaboration en de Campbell Collaboration.

Maar wellicht de belangrijkste impuls tot het uitbreiden van het begrip 'evidence-based policy making' naar andere domeinen dan de medische wereld – naar domeinen zoals 'evidence-based policing', 'evidence-based sentencing' of 'evidence-based teaching' om er slechts enkele te noemen – zijn een reeks documenten van de Britse regering ten tijde van New Labour, die een explosie van audits en andere dergelijke praktijken teweeg brachten. Zo stelt een White Paper vanuit het Department of Health in 1997 uitdrukkelijk: "what counts is what works... services and treatment that patients receive across the National Health Service should be based on the best evidence of what does and does not work and what provides best value for money. ... How could anyone claim that services should not be based on what we know to work?". Maar hierop volgen uiteraard een reeks moeilijke vragen: "... what constitutes valid evidence? Who decides? Do certain types of evidence seem to be treated as more legitimate than

others? What happens when evidence is fragmented or even contradictory? How much evidence does there need to be before we can confidently develop and roll out a particular policy?" In 1999 volgen de *Modernizing Government White Paper* en een document met als titel *Professional Policy Making for the Twenty First Century*, waarin men uitdrukkelijk verwijst naar 'evidence-based policy making'. Institutioneel krijgt dit beleidskader vorm door de uitbouw van nieuwe onderzoekseenheden binnen de overheidsadministratie, zoals de Performance and Innovation Unit, de Social Exclusion Unit, of de Centre for Management and Policy Studies. Dat een dergelijke visie op de relatie tussen beleid en onderzoek niet eigen is aan New Labour, maar diep ingebed zit in het wereldbeeld van Britse beleidsmakers bewijst dat de huidige Conservatieve regering in dezelfde richting verder werkt. In juni 2012 organiseert Cameron als onderdeel van het Cabinet office het 'Behavioural and Science Team' – in het dagelijks taalgebruik door insiders, met een verwijzing naar de bestseller van Richard Thaler en Robert Sunstein, ook het 'Nudge Team' genoemd.

Over RTCs als gouden standaard en de 'audit cultuur'?

De roep om meer 'evidence-based policy' bevat, zo leert deze korte historische terugblik, twee belangrijke ingrediënten. Een algemene visie die men simpel gezegd kan samenvatten als "meer en betere kennis leidt tot meer en beter beleid". Of om het te zeggen met de woorden waarmee de Amerikaanse Coalition for Evidence-Based Policy haar doelstelling omschrijft: "Increasing Government Effectiveness Through Rigorous Evidence About What Works" (<http://coalition4evidence.org>). Maar naast deze algemene visie, die impliceert dat men ook meer en meer audits uitvoert om de effecten van beleid te meten, is er natuurlijk ook het tweede ingrediënt: de 'rigorous evidence' waarvan sprake is, kennis verzameld op basis van RCTs of, beter nog, kennis die voortvloeit uit overzichten van een reeks van RCTs naar het model voorgeschreven door de Cochrane-procedures. Twee aspecten dus, waarvan vooral de tweede nu toch al een tijdje onder vuur ligt van critici.

Een eerste punt van discussie betreft het ethisch gehalte van deze aanpak. Omwille van deze nadruk

op experimentele methodes en het gebruik van behandelde versus controle-groepen krijgt deze aanpak inderdaad nogal eens het verwijt dat men riskeert in onethische praktijken verzeild te geraken. Benadeelt men patiënten niet als men hen omwille van het experiment bewust een bepaald medicijn onthoudt waarvan men toch minstens sterk vermoedt dat het een positieve werking heeft? Is het ethisch verantwoord om leerlingen in een school aan verschillen in de vormen van lesgeven bloot stellen als men verwacht dat niet elk van deze vormen even effectief is? Mag men werkzoekenden of uitkeringsgerechtigden in het algemeen aan verschillende condities onderwerpen om uit te vinden of en waarom de effecten van deze verschillende condities verschillen? In elk van deze gevallen maak men puur op basis van het toeval een onderscheid tussen wie men behandelt en wie niet. Mag dat wel? Pleitbezorgers van meer 'evidence-based policy', gebaseerd op RCTs, werpen hier tegen op dat het ethische probleem ook vanuit de andere kant moet worden bekeken. Is het ethisch te verantwoorden tegenover toekomstige generaties, zo zeggen zij, dat men verder gaat met het hanteren van lesmethodes of medische ingrepen waarvan men niet weet welke de effecten zijn of waarvan men op basis van niet-experimenteel onderzoek vermoedt, maar niet echt kan bewijzen dat er betere methodes of ingrepen beschikbaar zijn of beschikbaar gemaakt kunnen worden? Of, zo zeggen ze ook, is het verantwoord en ethisch te rechtvaardigen om door rechtbanken straffen te laten uitspreken zonder te weten of die straffen effectief ook het bedoelde effect hebben?

Een ander heikel punt is gekoppeld aan de vaststelling dat 'evidence-based policy making' meestal gevat zit in een discours dat een afweging maakt tussen kosten en effectiviteit. 'Most evidence-based methodologies fit broadly within a cost-benefit (or at least cost effectiveness) framework, designed to determine an estimated (net) payoff to society. It is a robust framework that provides for explicit recognition of costs and benefits, and requires the policy maker to consider the full range of potential impacts. But it hasn't been all that commonly or well used, even in relatively straightforward tasks such as infrastructure project evaluations.', zegt Gary Banks, voorzitter van de Productivity Commission – een onafhankelijk onderzoeks- en adviesbureau van de Australische regering (Evidence-based

policy making: What is it? How do we get it? ANU Public Lecture Series, presented by ANZSOG, 4 February 2009, Productivity Commission, Canberra). De vraag stelt zich of een dergelijke aanpak het beleid niet noodzakelijk verenigt en andere belangrijke waarden onvermijdelijk buiten het blikveld van het beleid houdt?

Misschien wel het meest belangrijke punt van kritiek, dat opduikt in publicaties van lieden die duidelijk behoren tot de beweging voor 'evidence-based policy making', betreft de beperkte informatiebasis van de aanpak, gepropageerd door de Cochrane Collaboration. Na hun kritische bespreking van deze procedures wijzen Merete Konnerup, Hans Christian Kongsted ('Do Cochrane reviews provide a good model for social science? The role of observational studies in systematic reviews', in: *Evidence & Policy*, 8, 2012, 1, 57-79) erop dat dergelijke overzichten, die nagenoeg uitsluitend gebruik maken van resultaten van gecontroleerde experimenten, gebaseerd zijn op een erg enge waarnemingsbasis. Waarom gaat men ervan uit dat andere onderzoeksmethoden geen betrouwbare en valide kennis opleveren? Zij pleiten er dan ook voor om de waarnemingsbasis voor deze overzichten te verbreden.

Vanuit een zelfde intuïtie pleit Louise Shaxson ('Is your evidence robust enough? Questions for policy makers and practitioners', in: *Evidence & Policy*, 1, 1, 101-111) voor een verruiming van de gehanteerde definitie – "Evidence for policy making is any information that helps to turn a department's strategic priorities and other objectives into something concrete, manageable and achievable." (p.103) – en een uitbreiding van de criteria voor 'robustness'. Naast de meer courante criteria (geloofwaardigheid, veralgemeenbaarheid, betrouwbaarheid en objectiviteit) wijst zij ook op het belang van 'rootedness'. "Rootedness is about understanding the nuance of the evidence, exploring assumptions with an open mind, encouraging others to question the status quo as we see it, and thinking about who uses what evidence for what purpose". Een algemene vraag die toelaat om deze component van 'robustness' te onderzoeken is, zegt Shaxson, "Could the initial question, the evidence, the analytical methods, the conclusion and subsequent action be confirmed, refuted, explained or enriched by information from other sources?" (p.108)

En wat met de externe validiteit?

Het voorgaande lijstje van bedenkingen bij het gebruik van RCTs beperkt zich hoofdzakelijk tot kritiek die zich ook bij de pleitbezorgers van 'evidence-based policy making' zelf begint te ontwikkelen en kan rustig nog worden uitgebreid. Laten RCTs toe om uitspraken te doen over beleid dat pas na langere tijd een effect kan sorteren? En wat ondertussen met prangende beleidsvragen? En kan men er zicht mee krijgen op beleidsmaatregelen die men als structureel en iedereen omvattend kan beschouwen? Ik kan het lijstje nog wel met een aantal andere netelige kwesties aanvullen, maar ik nader het aantal mij toegemeten woordjes en daarom volsta ik met nog één belangrijk probleem aan te kaarten.

In een essay ('Excursions to the Islands of the Happy Few', in: *Living, Thinking, Looking*, Hodder & Stoughton, London, 2012, p.126) waarin ze de relatie tussen kunst, psychoanalyse en neurowetenschap aftast, beschrijft Siri Hustvedt wetenschappelijke kennis als het resultaat van een moeizaam collectief productieproces dat hoe dan ook zorgvuldige interpretatie vereist en waarin 'feiten' omgevormd worden tot 'argumenten' om een bepaalde stelling of uitspraak al dan niet (voorlopig en tot bewijs van het tegendeel) te aanvaarden. "Hard science is a plodding business of findings, re-findings and findings again. It is incremental, often contradictory, and dependent on the creativity of the mind doing the research, a mind that can grasp what the research means at the time. At the end of many science papers there is a section called 'Discussion', where the researchers tell the reader how their study might be understood or how it could be followed up. Results rarely speak for themselves". Hustvedt beschrijft hier op adequate wijze, lijkt me, hoe wetenschappelijke uitspraken idealiter tot stand komen. Deskundigen komen samen en spreken af welke gegevens als bewijsmateriaal mogen worden aangedragen. Ze wegen de eventuele elkaar tegensprekende vaststellingen tegen elkaar af en komen zo tot een weloverwogen en door iedereen gedragen besluit over de houdbaarheid of niet van een of andere hypothese of theorie. Karl Popper heeft deze werkwijze ooit vergeleken met de manier waarop een rechtbank tewerk gaat.

Wat hierin centraal staat, duidt men aan als de 'interne validiteit' van het proces. In het geval van

de toepassing van wetenschappelijke kennis in het beleid gaat het in wezen om iets heel anders. Hier stelt zich niet meer het probleem van de 'interne' maar van de 'externe' validiteit.

In een uitvoerige reeks publicaties en een recent boek (Nancy Cartwright/Jeremy Hardie, *Evidence-Based Policy. A Practical Guide to Doing it Better*, Oxford University Press, 2012) betoogt Nancy Cartwright, wetenschapsfilosofe aan de London School of Economics, dat er zich juist op dat vlak grote problemen stellen. De sterkte van RCTs, zo zegt Cartwright ('Are RCTs the Gold Standard?', in *Biosocieties*, 2, 2007, 11-20) is dat ze een grote interne validiteit opleveren. Gecontroleerde experimenten doen wat de term suggereert. Ze zijn in staat om alle condities die de effecten van verschillen in behandeling binnen het kader van het experiment beïnvloeden te controleren. In wezen maakt dit RCTs tot deductieve methodes. Als aan de veronderstellingen die aan de test ten grondslag liggen wordt voldaan, dan impliceert een positief resultaat dat men de aangepaste causale conclusie mag trekken. Maar eerst en vooral is het zo dat deze deductieve werkwijze niet specifiek is voor RCTs. Integendeel, ze deelt deze karakteristiek met tal van methoden van wetenschappelijk onderzoek. In die zin kan men RCTs zeker niet beschouwen als de gouden standaard waaraan men andere werkwijzen moet afmeten. Men moet er trouwens rekening mee houden, gaat ze verder, dat deze redenering enkel opgaat als aan *alle* veronderstellingen is voldaan. Het experiment of de methode is immers maar zo sterk als de zwakste schakel. "Suppose you have 10 premises, 9 of them almost certain, one dicey. Your conclusion is highly insecure, not 90% probable." (p.14) Maar juist omdat de interne validiteit zo groot is, zijn de resultaten van een dergelijk onderzoek niet of nauwelijks over te dragen naar situaties waarin de context waarbinnen de behandeling plaats vindt, wijzigt.

Een interessante opmerking van Cartwright in dit verband is dat de deductieve methode, zoals meestal gebruikt in de praktijk van 'evidence-based policy making', niet echt overeenkomt met de wijze waarop Karl Popper ze indertijd propageerde. Volgt men de werkwijze voorgesteld door Popper, zo zegt Cartwright, dan formuleert men een hypothese op basis waarvan men een bepaalde uitkomst verwacht. Doet die uitkomst zich niet voor dan vervalt de hypothese. Cartwright noemt deze conclusie

een 'clincher'; er valt een definitief verdict en de kwestie wordt afgesloten. In het geval van een RCTs volgt men een andere redenering. Men formuleert een hypothese op basis waarvan men een bepaalde uitkomst verwacht. Doet de uitkomst zich voor dan gaat men ervan uit dat de waarschijnlijkheid van de hypothese (ceteris paribus) is toegenomen. Cartwright noemt deze conclusie een 'voucher' – van 'to vouch', te vertalen als 'garanderen', 'verzekeren', 'beweren'.

Gaat het te ver om deze opmerking van Cartwright te herformuleren in de zin dat wat hier plaatsvindt, gelijk staat aan de overdracht van een uitspraak uit een 'beschrijvend' of '(voorlopig) verklarend' register naar een register van 'voorschriften'. Of beter nog: de omvorming van een vaststelling naar een voorschrift met causaliteit, verbonden met de wetenschappelijke hypothese, als argument voor het voeren van een bepaald beleid?

En als uitsmijter ...

Een boek dat in mijn wetenschappelijke jeugd voorwerp van veel en heftige discussie uitmaakte, is 'Against Method' van Paul Feyerabend, wetenschapsfilosoof, leerling van Popper en als ik me niet vergis voordien ooit nog regie-assistent van Bertold Brecht. De discussies over dit boek laaiden hoog op, omdat Feyerabend als anarchist het motto 'Anything Goes' huldigde en zo ver ging om na de scheiding van Kerk en Staat ook de Scheiding van Staat en Wetenschap te bepleiten. Waarom, zo chargeerde de man, laat men het parlement ook niet stemmen over het aanvaarden van wetenschappelijke wetmatigheden? Op gezag van enkele wetenschappers bepalen ze immers ons gedrag en wereldbeeld meer dan sommige zaken waar men wél over stemt.

Ik zei het al. Feyerabend chargeerde en provoceerde graag. Maar is het overdreven om mij af te vragen of vertegenwoordigers van het volk beleidsmaatregelen (zoals het terug invoeren van de doodstraf) mogen afwijzen, zelfs als uit onderzoek (op basis van 'evidence-based sentencing') zou blijken dat de verhouding tussen kosten en effectiviteit (van de strafuitvoering) daardoor verbetert?

Walter Van Trier