

WAT BETEKENT ‘ALTRUIÏSME’ ?

Filip Buekens

Evolutionaire verklaringen in de psychologie inspireren soms revisionistische opvattingen over en voorstellen tot diepgaande herziening van onze manifeste drijfveren, motieven en projecten. Ze staan, zo lijkt het althans, in de traditie van het ondermijnende denken. De redenering loopt ruwweg als volgt: wanneer de evolutionaire achtergrond of oorsprong van onze centrale denkgewoonten wordt blootgelegd, lijkt ons geloof in de globale integriteit van die denkgewoonten problematisch te worden. ‘Je gelooft dat maar omdat...’ wordt dan ingevuld met een verklaring voor die overtuiging die haar als het ware loskoppelt van de vertrouwde rechtvaardiging voor die overtuiging. Omdat een evolutionaire verklaring van onze centrale denkgewoonten en attitudes deel uitmaakt van de studie van wat ons in staat stelde om die denkgewoonten te hebben, kunnen we dit beeld generaliseren: wie kennis neemt van de wijze waarop we uitgerust werden met die denkgewoonten, leert iets dat onze gewone rechtvaardigingen voor die denkgewoonten lijkt te ondermijnen.

In 2011 verdedigde ik in het *Tijdschrift voor Filosofie* dat onze manifeste houding tegenover onszelf en anderen een beeld bevat dat ons presenteert als wezens die min of meer gepast interageren met anderen en de wereld, en dat dit een assumptie is op grond waarvan we elkaars attitudes, handelingen en reacties begrijpen, beoordelen en corrigeren. Die assumptie impliceert niet de absurde stelling dat we ons nooit kunnen vergissen en ze laat zelfs open dat hele clusters van attitudes voor revisie vatbaar zijn. Maar we kunnen dit beeld niet *in toto* verwerpen. Tegelijk heeft onze natuurlijke psychologie fysische en neurologische instaatstellende voorwaarden die vatbaar zijn voor evolutionaire verklaringen. Ik gebruikte de term ‘instaatstellende voorwaarden’ om te verwijzen naar alle elementen die kunnen optreden in een *verklaring* van manifestaties van cognitieve en morele attitudes en reacties. Centraal in de studie van die instaatstellende voorwaarden staan de ‘onderliggende’ mechanismen die ten grondslag liggen aan gepaste of correcte manifestaties. Een voorbeeld kan hier helpen. Wanneer ik ervan overtuigd ben dat het buiten regent, hebben perceptuele mechanismen hun werk gedaan. Dat mijn overtuiging waar en gerechtvaardigd is, vereist dat die mechanismen goed gefunctioneerd hebben. Wanneer ze niet of slecht functioneren, manifesteert zich dit: ik vergis mij, of ik kan regen niet meer herkennen, of wat dan ook. Wanneer ik kan rekenen en een som correct oplos, hebben de onderliggende cognitieve en neurologische mechanismen goed werk afgeleverd. Welnu, om te kunnen nagaan waarom perceptie of rekenen in

sommige omstandigheden systematisch misloopt, moeten we op manifest niveau ervan terecht een onderscheid kunnen maken wanneer cognitieve reacties gepast zijn en wanneer ze (systematisch) beginnen afwijken. Op grond hiervan kunnen we dan gaan onderzoeken wat er misloopt wanneer iets misloopt, en beetje zoals we nagaan wat er misloopt wanneer een goed werkende machine vast komt te zitten, wat verondersteld dat we kunnen herkennen wanneer ze goed werk levert. Zulke verklaringen veronderstellen dus de globale integriteit van ons manifeste zelfbeeld, van onze manifeste reacties. De globale integriteit van onze natuurlijke psychologie is dus compatibel met wetenschappelijke en met name evolutionaire verklaringen van haar instaatstellende voorwaarden, maar ze kan in het licht van wetenschappelijke en met name evolutionaire verklaringen niet globaal verworpen worden, ook al kan een evolutionaire verklaring soms wat ontluisterende effecten hebben. De structuur en effecten van zulke ‘ondermijnende verklaringen’ zijn ondertussen grondig bestudeerd.

In een interessante reactie plaatsten Stefaan Blancke, Maarten Boudry, Helen De Cruz, en Johan De Smedt terecht wat kanttekeningen bij mijn nogal breedvoerig gebruik van de notie ‘instaatstellende voorwaarde’, maar mijn kerngedachte – dat de globale integriteit van ons manifeste zelfbeeld op grond van wetenschappelijke kennis van evolutionair te verklaren instaatstellende voorwaarden niet in het gedrang kan komen – leken ze niet te betwisten. Ikzelf ben sindsdien wat genuanceerder gaan denken over mogelijke revisies van ons manifeste beeld van onszelf op grond van wetenschappelijke inzichten. Kennis van spontane neigingen en heuristieken helpt ons denkfouten te vermijden en beter in te schatten hoe wij als sociale dieren ons samenleven wat beter kunnen organiseren en de soms wat te optimistische interpretatie van onszelf te nuanceren. Beter inzicht in de evolutionaire oorsprong van patronen van emotionele reacties zou ons kunnen behoeden voor misplaatste stereotypering, een fenomeen dat evolutionair te verklaren valt vanuit instinctieve identificatie- en exclusierelaties, die zelf attitudes zijn die in de ‘Environment of Evolutionary Adaptation’ (EEA) *fitness* bevorderen. Zelf denk ik niet dat kennis van de evolutionaire psychologie noodzakelijk was om tot dit soort inzichten te komen, maar dat terzijde. Hoever kan en mag zulk revisionisme gaan? Neem onze capaciteit om geldig te redeneren. Waarom zouden we epistemisch vertrouwen hebben in evolutionaire psychologie of evolutietheorie, gegeven het feit dat ze tot stand kwamen dankzij cognitieve vermogens die door evolutie niet geselecteerd werden om theoretische inzichten over hun ontstaansgeschiedenis te produceren, en die inzichten bovendien niets bijdragen tot het overleven van de soort? In elk geval kan amelioratieve psychologie niet beweren dat onze mentale economie *zo* inadequaat zou zijn dat het wetenschappelijke gehalte van de theorieën die ze *zelf* produceert daaronder zou gaan lijden. En ze kan evenmin ontkennen dat de vergissingen en neigingen waarop ze wijst, ons altijd al enigszins inzichtelijk waren. Veel van de bekende fenomenen uit de ‘bias en heuristics’-literatuur waren ons *altijd al vertrouwd* en worden doorgegeven in fabels, parabels en een oude en

wat ondergesneeuwde traditie van moralistische teksten die ons leren hoe we moeten leven. Zo hebben we nu harde empirische bewijzen voor pakweg de *confirmation bias*. Vroeger zagen we bij wijze van spreken alleen de balk in het oog van de ander. En toch wisten we altijd al dat we niet perfect waren, een inzicht dat ons een beetje minder imperfect maakt.

Ik ontleende mijn argument aan Peter Strawsons beeld van onze vertrouwde repertoire van reactieve attitudes (schuld, trots, schaamte, medeleven, nijd, ...) als natuurlijk weefsel waarop toeschrijvingen van verantwoordelijkheid of verschoning gedijen.¹ Cognitieve en reactieve attitudes zijn de geëvolueerde en dus contingente objecten van evaluaties als *waar*, *gepast*, of *adequaat* (en hun tegenhangers: *onwaar*, *ongepast*, *inadequaat*). Strawson wees er bovendien op dat de natuurlijke houding die onszelf, anderen en onze verhoudingen tot elkaar begrijpelijk en evalueerbaar maakt, staat tegenover een 'objectieve' houding (*objective attitude*) die aangewezen is wanneer we geconfronteerd worden met manifeste patronen van irrationeel en abnormaal handelen. Zulke manifeste patronen mogen niet verward worden met concrete voorkomens van ongepaste of inadequate attitudes die doorgaans heel goed herkenbaar en corrigeerbaar zijn binnen onze natuurlijke psychologie en dus niet nopen tot een 'objectieve' houding.² Patronen van irrationele attitudes, ongemotiveerde reacties en pathologieën nopen tot een objectieve houding die iemand van verantwoordelijkheid voor haar handelen vrijpleit, maar zulke deviante patronen *kunnen niet de regel worden*. Het deviante en pathologische is slechts herkenbaar op een achtergrond van *gepaste* of *adequate* morele en cognitieve reacties en daarom kan de objectieve houding niet de regel worden. 'It is self-contradictory that abnormality is the universal condition', schrijft Strawson.³ Het weefsel van attitudes dat ons zelfbeeld en dat van anderen constitueert en waarvan het ontstaan en de kenmerkende patronen evolutionair verklaard kan worden, creëert een beeld van onszelf en van anderen dat die attitudes als globaal *adequate* (*gepaste, ware*) attitudes opvat, een adequaatheid waartoe zelfs de evolutionaire psychologie zich moet verbinden om haar 'ultieme' evolutionaire verklaringen plausibel te maken. De globale integriteit van onze natuurlijke attitudes is het niet-theoretisch vertrekpunt voor wetenschappelijk onderzoek over haar instaatstellende voorwaarden.

Boudry & co wijzen erop dat ik een belangrijk fenomeen dat aanleiding kan geven tot zulk revisionisme, wat onderbelicht heb: het gebruik van intentionele en strategische termen in evolutionaire verklaringen. Ik laat hen even aan het woord:

De verleidelijkheid van dat revisionisme is in de eerste plaats gebaseerd op de genetische drogreden (*genetic fallacy*), een informele denkfout waarbij iemand een oordeel velt over een object of persoon louter op basis van oorsprong of ontstaansgeschiedenis. De belangrijkste bron van deze 'genetische misvatting' omtrent evolutionaire psychologie, die door Buckens onvoldoende wordt belicht, is het feit dat ultieme verklaringen doorgaans geformuleerd worden in

intentionele en strategische termen. Evolutie- biologen spreken over stabiele strategieën, signaalfuncties, genetische zelfzuchtigheid, manipulatie, uitbuiting, enzovoort (...)

Toegepast op menselijke (of zelfs dierlijke) emoties en psychologische disposities, versterkt deze intentionele formulering de wijdverbreide misvatting dat ultieme verklaringen in de plaats kunnen treden van manifeste en dus bewuste belevingen, of dat ze uitdrukking geven aan onze werkelijke, onderliggende drijfveren. Evolutionaire psychologie wordt verkapt 'psychoanalyse' van de menselijke geest, waarbij mensen 'onbewust' hun genen willen verspreiden of hun biologische fitness willen verhogen.

Boudry & co verwerpen terecht deze 'freudianisering' van de evolutionaire psychologie. David Buller heeft het over 'the temptation to view evolutionary psychology as a source of insight into latent motives directed at reproductive success... it is the Freudian legacy of the dynamic unconscious that tempts us... to internalize adaptive goals into the unconscious and then view them as the hidden driving force behind our behavior and reproductive success.' (D.J. Buller, *DeFreuding Evolutionary Psychology*, p. 111).

Hier ben ik het helemaal mee eens en het gebruik van termen en concepten uit ons manifeste beeld in wetenschappelijke modellen en theorieën verdient meer aandacht. De achtergrond van het debat is een vorm van explanatorisch revisionisme: vertrouwde 'proximale' rechtvaardigingen van cognitieve of reactieve attitudes zouden concurreren met 'ultieme' evolutionaire verklaringen. Zo houdt iemand van deze of gene persoon niet zozeer omdat het zijn kind is, maar omdat het ongeveer 50 procent van zijn genen draagt.⁴ Manifest altruïsme kan teruggevoerd worden tot genetisch egoïsme. Evolutiebiologen Richard Dawkins en John Krebs verdedigden dat communicatief handelen dat *prima facie* coöperatief lijkt *in feite* gericht is op en altijd teruggevoerd kan worden naar manipulatie van de ander in functie van genetisch gedetermineerde zelfzuchtige motieven.⁵

Wellicht spelen sciëntistische motieven hier een belangrijke rol, maar die schuif ik gemakshalve even terzijde. Hier wil ik, ter afsluiting van mijn bijdrage, wat dieper ingaan op de vraag wat die *terugvoerbaarheid* precies zou kunnen betekenen en of de termen *egoïsme* en *altruïsme* zoals gebruikt in onze manifeste psychologie wel hetzelfde betekenen wanneer ze in biologische theorieën gebruikt worden. Zijn de voorbeelden van persuasief taalgebruik die Dawkins en Krebs opvoeren – voorbeelden van *manifeste* manipulatie van anderen met taal – inderdaad argumenten voor de stelling dat *genetisch egoïsme* correct is?

Achter dit explanatorisch revisionisme schuilt een belangrijk misverstand dat in de hand gewerkt wordt door een vorm van 'conceptual engineering' die impliciet plaats vindt wanneer je *manifeste* concepten gaat gebruiken in *wetenschappelijke* modellen en theorieën. Ik neem als voorbeeld de term altruïsme. We onderscheiden het manifeste concept van altruïsme (M-altruïsme), en het concept van altruïsme dat in de evolutionaire psychologie gebruikt wordt

(B-altruïsme). Beide concepten worden vaak met hetzelfde *woord* uitgedrukt, wat bijzonder misleidend is, en in de evolutionaire psychologie aanleiding geeft tot misverstanden. In ‘What is evolutionary Altruism?’ schrijft Elliott Sober (filosofie van de biologie én wetenschapsfilosoof) het volgende:

“Sociobiologists sometimes confuse vernacular and evolutionary altruism, as when they argue that people cannot really be altruists in the vernacular sense, on the grounds that evolutionary altruism cannot be a reality” (Sober 1988, p. 459).⁶

‘M-altruïsme’ en ‘B-altruïsme’ behoren tot verschillende *conceptuele kaders en kunnen dus niet dezelfde betekenis hebben. Wanneer een term uit het ene kader binnen het andere kader gebruikt wordt, ontstaan ongerijmdheden.* Ik zal nu kort laten zien dat ‘altruïsme’ binnen het manifeste beeld een andere betekenis heeft dan in het wetenschappelijke beeld. Manifest altruïsme (M-altruïsme) vereist dat je verlangens kunt vormen die gericht zijn op het welzijn van de ander (de **kost** (*cost*) van het realiseren van dat verlangen is groter dan het **voordeel** (*benefit*) voor jezelf). M-altruïsme veronderstelt dat motieven of redenen een rol spelen in de rechtvaardiging van de manifeste intentionele handelingen gericht op het welzijn of welbevinden van de ander. Ik moet dus een M-altruïstisch *motief* hebben gehad om (naar aanleiding van een handeling) als altruïstisch beoordeeld te worden. Toevallig iets verrichten (of nalaten!) dat jou beter maakt of jou voordeel verschaft, valt daar niet onder. Doorgaans wordt aangenomen dat ik altruïstische motieven kan hebben zonder dat mijn motieven tot geslaagd of succesvol altruïstische handelen hebben geleid. B-altruïsme is van toepassing op *elk* organisme – op mieren en mensen. B-altruïsme kan instinctief of adaptief zijn. M-altruïsme is, als manifest fenomeen, gericht op het welzijn van de ander als *keuze die we maken. Zelfs indien dit fenomeen een niet-manifeste biologische verklaring kan krijgen, is de manifeste rechtvaardiging voldoende om het belang, de waardevolheid ervan in te zien.* Een tweede kenmerk van M-altruïsme is dat in onze motieven geen *reproductief voordeel hoeft te liggen.* Stel je bent een begenadigd pianist. Ik help je de piano af te betalen, en mijn handeling brengt je ertoe je hele leven lang piano te spelen, met als gevolg dat je lief en leed vergeet – en je vergeet je voort te planten. Mijn handeling verliest daardoor niets van haar M-altruïstisch karakter, maar ze bevordert zeker niet jouw *fitness.* ‘De ander een reproductief voordeel bezorgen’ behoort niet tot het manifeste kader waarin we samenwerken, elkaars handelingen op elkaar afstemmen, etc. Het is geen gebruikelijk motief om je voort te planten. Het is een concept dat behoort tot het evolutionair-biologische denkkader.)

Ten derde: M-altruïsme is een *absoluut* concept. Het laat geen kwalificaties toe. Stel dat iemand altruïstischer is en twee piano’s schenkt. Maakt het feit dat hij meer deed dan ikzelf, hem meer altruïstisch? Helemaal niet: de kwalificatie van mijn handeling A als altruïstisch is *conceptueel* onafhankelijk van het bestaan van een handeling van B die ook altruïstisch is. Het is niet omdat je het altruïsme van de schatrijke filantroop zelfs niet kunt benaderen dat je geen altruïstische

handelingen kunt stellen. B-altruïsme is een kenmerk van organismen in een populatie waarvoor de evolutionair relevante vraag luidt: wanneer is B-altruïsme fitness-bevorderend binnen een populatie? ‘Hij handelde onzelfzuchtig’ is een commentaar, een kwalificatie, van zijn of haar handelingen in het licht van redenen of motieven. Het is geen algemeen commentaar gebaseerd op een inzicht uit de biologie.

Tenslotte: Motieven voor altruïstisch handelen mogen niet ingegeven zijn door verder liggende zelfzuchtige motieven. De motieven en handelingen mogen niet in een instrumentele relatie staan tot een verder liggend doel dat mij iets zou opleveren. Mijn verlangen is niet-instrumenteel gericht op de ander. (Zuiver altruïsme vereist dat je *afstand* doet van verlangens die op jezelf gericht zijn en daarvoor een prijs betaalt.) Uit het banale feit dat ik mijn verlangens wil realiseren, volgt niet dat het object van die verlangens (hun inhoud) op mijzelf betrekking heeft, op mijn welbevinden, op een voordeel voor mij. Dat we handelen om *onze* verlangens te realiseren is een triviale waarheid (hoe kan ik handelen op grond van jouw verlangens, tenzij ik ze begrijp en overneem, en als motief accepteer voor wat ik doe?). Het is de *inhoud* van een verlangen dat het altruïstisch karakter ervan vastlegt. Zelfs indien geen enkele handeling helemaal zuiver is – we denken altijd wel ergens aan onszelf, we zijn ijdele wezens en plaatsen onszelf graag in goed daglicht – kunnen we dit onderscheid maken. Wanneer we ontdekken dat aan gul gedrag van de buurman een verborgen egoïstisch motief ten grondslag lag, *zien we de ander in een ander daglicht*.

Samengevat: M-altruïsme kenmerkt onze keuzes en handelingen – zo treffen we elkaar aan in de wereld. M-altruïsme vereist een psychologie, een mentale economie, is niet gericht op reproductie en niet comparatief. Het is een uitgesproken **antropocentrisch** concept. B-altruïsme is niet-psychologisch, rechtstreeks gericht op reproductie en comparatief. Binnen het manifeste beeld van onszelf kunnen we de vraag ‘zijn wij altruïstische wezens?’ alleen maar bevestigend beantwoorden. Dat B-altruïsme louter gericht is op fitness-bevordering, maakt niet uit. Je kunt niet aantonen dat mensen altijd M-zelfzuchtig handelen louter omdat onze genen B-zelfzuchtig zijn.

Ik vermoed dat het dit soort spraakverwarringen zijn die evolutionaire psychologie een slechte naam bezorgen, en de meeste evolutionaire psychologen zijn verstandig genoeg om niet in deze valkuilen te trappen. Helaas treffen we in de media lieden aan die deze conceptuele verwarringen in stand houden en er verregaande en vaak niet al te flatterende conclusies over onszelf en ons manifeste beeld van onszelf uit afleiden. Ze gebruiken evolutionaire psychologie om een ondermijnend punt te maken. Ik heb geprobeerd hun denkfout te verhelderen.

Noten:

¹ De reactieve attitudes zijn volgens Flanagan geculturaliseerde versies van proto-deugden zoals sympathie, trouw en moed die overlevingswaarde hebben. Zie O. FLANAGAN, ‘Ethical Expressions: Why Moralists Scowl, Frown and Smile’, in J. HOGDE & G. RADICK,

The Cambridge Companion to Darwin, Cambridge, Cambridge University Press, 2003, pp. 377-398.

² Zie P.F. STRAWSON, 'Freedom and Resentment', in Michael McKenna & Paul Russell (Eds.), *Free Will and Reactive Attitudes. Perspectives on P.F. Strawson's 'Freedom and Resentment'*, Farnham: Ashgate, 2008, pp. 19-37.

³ STRAWSON, *o.c.*, p. 25.

⁴ Zie bijvoorbeeld Anthony WALSH, 'Evolutionary Psychology and Criminal Behavior', in J. BARKOW (ed.), *Missing the Revolution: Darwinism for Social Scientists*, Oxford, University Press, 2006, pp. 225-268.

⁵ Zie bv. J.T. KREBS, en R. DAWKINS, 'Animal Signals: Information or Manipulation?' in J.R. KREBS and N.B. Davies (eds.), *Behavioural Ecology: an Evolutionary Approach*. Oxford, Blackwell Science, 1978, pp. 282-309.

⁶ Een voorbeeld: Trivers 1971.