

DRIE MODELLEN VAN HET MENSELIJK HANDELEN

Je hebt soms van die kinderen die het eigenlijk allemaal overkomt. Zomaar ineens hebben ze slaande ruzie, staan ze met modderlaarzen in de woonkamer of doen ze een plas in hun broek. Ook voor hun gelden de geveugelde woorden van Miss Anscombe, "Ik doe wat gebeurt", maar het lijkt er soms verdacht veel op dat er in hun gedrag vooral sprake is van gebeuren, en nauwelijks van doen.

In dit boek wordt een overzicht gegeven van hoe er in de hedendaagse Anglo-Amerikaanse filosofie nagedacht wordt over het onderscheid tussen dat wat je doet en dat wat je overkomt. Daarbij worden drie modellen uitgewerkt: (1) handelingen worden verricht door actoren; (2) handelingen zijn gebeurtenissen; (3) handelingen zijn sociale fenomenen. In de bespreking van het eerste model ligt de nadruk op de zelfstandigheid van de actor en op de structuur en de functie van praktische redeneringen. In de bespreking van het tweede model wordt tegen de achtergrond van naturalisme en determinisme nader ingegaan op het probleem van mentale veroorzaking. En in de bespreking van het derde model wordt handelen vergeleken met toneelspelen, waarbij er aandacht wordt besteed aan de problematiek van het regelvolgen, en aan de problematische relatie tussen het sociale en het normatieve.

Het boek is een echte inleiding: in een prettig leesbare stijl wordt een toch tamelijk complex vakgebied met veel voorbeelden toegankelijk gemaakt. Het boek wakkert daarbij een filosofische houding aan: de lezer gaat begrijpen waarom de vragen belangrijker zijn dan de antwoorden.

ISBN 90-429-0712-6



9 789042 907126

Drie modellen van het menselijk handelen

JAN BRANSEN



PEETERS

Jan Bransen Drie modellen van het menselijk handelen

Drie modellen van het menselijk handelen

Drie modellen van het menselijk handelen

JAN BRANSEN

Reeks: Tertium Datur

1. C. LANSINK, Vrijheid en Ironie. Kierkegaards ethiek van de zelfwording, 1997, XXIV-309 p.
2. J. BRANSEN, Drie modellen van het menselijk handelen, 1999, X-130 p.

ISBN 90-429-0712-6

D. 1999/0602/1

© 1999- Uitgeverij Peeters, Bondgenotenlaan 153, B-3000 Leuven

Foto omslag: Ron Hendriks

*voor mijn kinderen Stijn, Lotte en Camiel
die mij iedere dag weer laten zien
hoe talrijk en gevarieerd de verschillen en verbanden zijn
tussen wat je doet en wat je overkomt*

Voorwoord

Dit boekje is eigenlijk per ongeluk ontstaan. Twee Vlaamse filosofen hebben daarbij een hoofdrol gespeeld. Filip Buekens heeft mij in eerste instantie weten te overtuigen van het belang van een Nederlandstalig overzicht van de Anglo-Amerikaanse filosofie van het handelen, en zijn geflatteerde opmerking dat ik de enige ben die een dergelijk overzicht zou kunnen schrijven, kon ik natuurlijk niet naast me neerleggen. Nu het uiteindelijk zo ver is, ben ik dan ook blij dat ik ingegaan ben op zijn verzoek deze tekst te schrijven. Het is vervolgens Stefaan Cuypers geweest die mij ervan overtuigde dat het de moeite waard zou zijn deze tekst in boekvorm uit te laten geven. Aan hem dank ik de belangstelling van Uitgeverij Peeters voor dit overzicht.

Naast Filip Buekens en Stefaan Cuypers hebben ook Patrick Delaere, Wim de Muijnck en Eva van Ooijen het manuscript gelezen en becommentarieerd. Ik wil hen allen daarvoor graag bedanken.

Tenslotte: de eerste paragraaf van hoofdstuk 1, “Over zelfstandigheid”, verscheen oorspronkelijk onder de titel “Acterschap en zelfstandigheid” in het *Tijdschrift voor Filosofie*, 60 (4), 1998, pp. 723-739. Ik ben blij dat de redactie van het *Tijdschrift voor Filosofie* mij toestemming heeft gegeven dit artikel hier opnieuw te publiceren.

Jan Bransen
Utrecht, december 1998

INHOUDSOPGAVE

Voorwoord	VI
Inleiding: Wat gebeurt er als iemand handelt?	1
Wijsgerige antropologie en de filosofie van het handelen	2
Drie handelingsmodellen	8

HOOFDSTUK 1

Handelingen worden verricht door actoren	13
1. <i>Over zelfstandigheid</i>	13
In kunnen grijpen	13
Het hiërarchische model van zelfstandigheid	15
Drie problemen	19
Waarde-oordelen	21
Coherentie	25
De actor-identificerende handeling	28
Hoe belangrijk is <i>zelfstandigheid</i> ?	31
2. <i>Over delibereren, motiveren en rechtvaardigen</i>	32
Praktische rationaliteit	32
Rationele principes en opvattingen	35
Pro-attitudes	39
Intenties	43
Motiveren en rechtvaardigen	48

HOOFDSTUK 2

Handelingen zijn gebeurtenissen	57
1. <i>Naturalisme en Determinisme</i>	57
Methodologisch naturalisme	57
Ontologisch naturalisme	58

Methodologisch determinisme	60
Ontologisch determinisme	62
2. <i>Het probleem van mentale veroorzaking</i>	64
Het probleem: twee oorzaken voor één effect	64
Eén effect? Dan één oorzaak!	66
Twee oorzaken? Dan twee effecten!	70
Twee oorzaken? Dat is één te veel!	74
Oorzaken en verklaringen zijn twee.	76
Verklaringen en oorzaken 'zijn' één.	81
Conclusie	83

HOOFDSTUK 3

Handelingen zijn sociale fenomenen	85
1. <i>Het sui generis karakter van het sociale</i>	85
Het dramaturgisch model	87
Over regels en sociale conventies	90
Over rolconflicten en rolafstand	96
2. <i>Individualisme, holisme, en het normatieve</i>	101
Inleiding: Succes en sociaal succes	101
Over regelvolgen en regelvolgers	103
Verklaringen in de sociale wetenschappen	107
Nawoord	113
Literatuursuggesties	117

Inleiding

Wat gebeurt er als iemand handelt?

Je hebt soms van die kinderen die het eigenlijk allemaal overkomt. Zomaar ineens hebben ze slaande ruzie, staan ze met modderlaarzen in de woonkamer of doen ze een plas in hun broek. Ook voor hen gelden de gevleugelde woorden van Miss Anscombe, "Ik doe wat gebeurt", maar het lijkt er soms verdacht veel op dat er in hun gedrag vooral sprake is van gebeuren, en nauwelijks van doen.

Dit onderscheid tussen wat je doet, en wat er met je, door je, of in je gebeurt kennen we allemaal. Nagenoeg iedereen wordt op een gegeven moment zindelijk, en de meesten van ons leren nog veel meer dan alleen hun willekeurige spieren beheersen: we denken na bij wat we doen, of *voor* we iets doen, en dat helpt. Je leert er bij stil te staan dat je eerst je laarzen uit doet voor je binnenkomt, en je leert op het juiste moment te zwijgen, omdat je liever geen ruzie krijgt. Nog weer later leer je hoe je wiskundeproblemen op kunt lossen; je krijgt de fijne motoriek onder controle zodat je tandarts kunt worden; en je leert moeilijke beslissingen te nemen als de eerste kalverliefde overgaat, of als je vergaderingen moet gaan voorzitten. Hoewel we nagenoeg allemaal leren handelen, weet eigenlijk niemand precies hoe we dat doen, hoe we gebeurtenissen in ons lichaam onder controle kunnen krijgen. En we weten eigenlijk ook niet wat dat precies betekent: je gedrag onder controle krijgen.

In de hedendaagse analytische filosofie wordt er nog steeds veelvuldig gepubliceerd over deze kwestie, ondanks het feit dat we al de beschikking hebben over vele bibliotheken vol vaak al eeuwenoude boeken over het menselijk handelen. Dat er steeds weer nieuwe publicaties op het terrein van de filosofie van het handelen nodig zijn, heeft vooral te maken met de eigenaardige plaats van filosofische inzichten in het alledaagse leven: het zijn vooronder-

stellingen die zowel aan het begin als aan het eind staan van een intellectuele reflectie op de dagelijkse praktijk. Filosofische inzichten kunnen antwoorden opleveren maar ook vragen oproepen. Filosofische inzichten zijn aannames die aanvankelijk veelbelovend klinken en een uitroep tekenen lijken te verdienen. Het zijn uitgangspunten waarmee enthousiasme gewekt kan worden voor een model waarmee we (delen van) ons leven eindelijk in kaart lijken te kunnen brengen. Het inzicht, bijvoorbeeld, dat een mens niet uit een lichaam en een geest bestaat, maar uit iets dat zowel lichamelijk als geestelijk is, zou een enorme stimulans kunnen betekenen voor een visie op ziekte en gezondheid waarmee we recht zouden kunnen doen aan al die psychosomatische klachten die de huidige geneeskunde zo in verlegenheid brengen. Maar diezelfde inzichten verliezen hun stelligheid, en lijken alleen nog maar een vraagteken te verdienen, als het model zijn voltooiing nadert en we gaan merken wat er allemaal *niet* mee in kaart gebracht kan worden. Dan kunnen we echter al vele tientallen jaren verder zijn, of zelfs enkele eeuwen, en dan blijkt het alledaagse leven zelf deels de gestalte van het model aangenomen te hebben, zoals dat nu min of meer het geval is met ons dagelijkse leven dat zo grondig gevormd is door het Cartesiaanse lichaam-geest dualisme. Meer dan ooit lijkt er dan behoefte aan een nieuw model, een nieuw filosofisch inzicht, dat niet in een absolute zin nieuw hoeft te zijn, maar ruimte creëert voor dat wat intellectueel mogelijk moet zijn, ook al lijkt het praktisch onmogelijk geworden.

Wijsgerige antropologie en de filosofie van het handelen

Wie zich verdiept in het handelingsbegrip, zal al gauw ontdekken dat hij een uiterst vruchtbare ingang heeft gevonden in de wijsgerige antropologie. Denken over het menselijk bestaan kan eigenlijk nergens beter beginnen dan bij het denken over het menselijk handelen. Dat komt doordat het handelingsbegrip ons dwingt ons opnieuw te bezinnen op een aantal begripsmatige onderscheidingen die in de geschiedenis van het denken over het menselijk

bestaan een problematische hoofdrol spelen: (1) mens versus dier; (2) lichaam versus geest; en (3) individu versus gemeenschap. Ik heb mij in deze inleiding gepermitteerd een aantal tamelijk algemene opmerkingen te maken over de verhouding tussen het handelingsbegrip en ieder van deze onderscheidingen, zodat men een idee krijgt van de wijze waarop de filosofie van het handelen bij zou kunnen dragen aan een beter begrip van het menselijk bestaan in ruimere zin.

1. Mens-dier. De wijsgerige antropologie is als zelfstandige discipline pas ontstaan in de eerste helft van deze eeuw. De naam werd ontleend aan wat in die tijd een vakwetenschap aan het worden was: de antropologie, of menskunde. Als vakwetenschap ging de antropologie uit van de aanname dat mensen dieren zijn en dus, net als andere dieren, bestudeerd kunnen worden op empirisch wetenschappelijke wijze. Antropologie zou daarom een wetenschap kunnen zijn die vooral verwant is aan de biologie en de ethologie. Op grote schaal gingen antropologen aan het werk, en zij deden allerlei wetenswaardige ontdekkingen. Zo bleek, bijvoorbeeld, dat mensen veel minder gespecialiseerd zijn dan andere dieren, dat ze over veel minder instinctieve reacties beschikken dan gemiddeld welk ander dier, dat ze veel – echt veel – langer voor hun voortbestaan afhankelijk zijn van de zorg van hun ouders, en dat ze fenotypisch opvallend gekenmerkt worden door de historische en culturele context waarin ze leven.

Filosofen, als vanouds geïnteresseerd in het menselijk zelfbegrip, reageerden op de resultaten van deze empirische wetenschap door ze te proberen te duiden. Zij hadden niet de neiging zich te verzetten tegen de aanname dat mensen dieren waren, maar meenden wel dat een wijsgerige interpretatie van de empirische resultaten nodig was. Een dergelijke interpretatie zou, zo beweerden zij, laten zien dat er een diepe waarheid schuilt in het aloude idee dat mensen weliswaar dieren zijn, omdat we immers ook gewoon schepsels zijn, maar tegelijkertijd ook veel meer dan dieren, omdat ons bestaan zich, heel anders dan dat van de overige dieren, wezenlijk in een ‘morele ruimte’ voltrekt. Ook de empirische

antropologie, zo stelden de wijsgerig antropologen, zou slechts bewijsmateriaal aandragen ter ondersteuning van wat met een oud beeld inzichtelijk gemaakt kan worden: mensen zijn half dier, half God.

Het gaat er mij nu niet om de plausibiliteit van deze positie te onderzoeken. Ik wil alleen maar laten zien hoe een begrip van het menselijk handelen nieuw licht kan werpen op dit oude idee, zodanig dat we preciezer, en zonder beeldspraak, onder woorden kunnen brengen wat de strekking is van dit idee. Want de beeldspraak, dat begrijpt iedereen, leidt tot allerhande associaties die als generalisaties niet plausibel kunnen zijn. Er zijn vast mensen die in hun dagelijkse bestaan voortdurend een strijd voelen tussen hun goddelijke en hun dierlijke deel, maar als, zoals in de rest van dit boek verdedigd zal worden, dit nooit zomaar een onmiddellijke en onbetwifelbare ervaring kan zijn, maar altijd al een kwestie is van meer of minder geslaagde modellering, dan is het nog maar de vraag of we geen betere modellen kunnen ontwikkelen om verslag te kunnen doen van onze interne conflicten.¹

Welnu, zoals in de rest van dit boek duidelijk zal worden, is een wezenlijk kenmerk van onze handelingen gelegen in het feit dat we ze kunnen (of zelfs moeten) beschrijven met behulp van een normatief vocabulaire: wie handelt doet *A* in plaats van *B*, omdat hij *liever A* doet *dan B*. En natuurlijk kan 'liever...dan' nog van alles betekenen, maar het impliceert minimaal dat wie handelt minstens twee opties heeft, die, in termen van een nog nader te bepalen criterium, *gerangschikt* kunnen worden. Door het normatieve karakter van het menselijk handelen te onderzoeken kan de filosofie van het handelen ons daarom onmiddellijk helpen bij het verhelderen van wat wijsgerig antropologen proberen te begrijpen in hun poging het onderscheid tussen mens en dier te bepalen. En dan gaat het er eigenlijk helemaal niet om dat de filosofie van het

¹ Zoals in de volgende hoofdstukken zal blijken beschikken zeker twee van de drie modellen die ik zal bespreken over goede middelen voor het conceptualiseren van onze interne conflicten. Zie, bijvoorbeeld, de discussies over het hiërarchische model van zelfstandigheid (pp. 15-31), over motiveren en rechtvaardigen (pp. 48-54) en over rolafstand en rolconflicten (pp. 96-101).

handelen ons helpt aan een heldere articulatie van het onderscheid tussen mens en dier. Eerder is het zo dat de filosofie van het handelen ons kan helpen aan een helder inzicht in wat het nu precies betekent dat het menselijk bestaan zich wezenlijk voltrekt in een 'morele ruimte', door, bijvoorbeeld, uiteen te zetten waarin de zelfstandigheid van mensen bestaat, of door duidelijk te maken wat het verschil is tussen het begrijpen van het deliberatieproces als enerzijds een kwestie van motiveren en anderzijds een kwestie van rechtvaardigen, of door duidelijk te maken wat het verschil is tussen het volgen van een regel en het vertonen van regelgeleid gedrag.

We kunnen dit, samenvattend en vooruitlopend op de volgende hoofdstukken, ook zo zeggen: door ons te verdiepen in de filosofie van het handelen, kunnen we ontdekken dat het ons in de wijsgerige antropologie (anders dan bijvoorbeeld in de ethiek van vegetariërs) niet zo zeer gaat om het verschil tussen mens en dier, maar om een inzicht in hoe het menselijk bestaan, gekenmerkt als het is door normativiteit, begrepen zou kunnen worden als een natuurlijk (d.i. tot de natuur behorend) verschijnsel.

2. *Lichaam-geest*. Sinds het verschijnen van Gilbert Ryle's *The Concept of Mind* in 1949 staat het 'lichaam/geest'-probleem bijzonder in de belangstelling in die delen van de analytische wijsbegeerte die in het Nederlands aangeduid zouden worden met de term 'wijsgerige antropologie'. Ryle probeert in zijn boek te laten zien dat het Cartesiaanse dualisme, dat stelt dat mensen uit een geest en een lichaam bestaan die op problematische wijze met elkaar interacteren, gebaseerd is op een denkfout. Per ongeluk heeft de geschiedenis van de moderne wijsbegeerte ons opgezadeld met een onverantwoorde reïficatie, waardoor alle termen die slaan op de mentale dimensie van ons bestaan lijken te verwijzen naar een geestelijke substantie, de geest, die als een spook door ons lichaam waart. Vanwege dit dualisme schiet ons zelfbegrip, en daarom moet het ons in de wijsgerige antropologie toch te doen zijn, helaas schromelijk te kort.

Het logisch behaviorisme dat Ryle in zijn boek als alternatief

voor het Cartesiaanse dualisme voorstelt, heeft slechts een korte bloeitijd gekend, waarna het serieuze concurrentie heeft gekregen van fysicalistische en functionalistische theorieën. Hoezeer deze theorieën onderling ook mogen verschillen en ieder hun verdiensten en tekortkomingen hebben, wat ze met elkaar delen is de wijze waarop zij menen dat het ‘lichaam/geest’-probleem geconcentreerd moet worden. Enerzijds, zo is de idee, *moet* het mentale *wel* in onze lichamelijkheid bestaan, maar anderzijds *kan* het mentale *niet* begrepen worden in termen van onze lichamelijkheid.

Ook hier kunnen we heil verwachten van een omtrekkende beweging via de filosofie van het handelen. Immers, hoe we ook wensen te spreken over onze handelingen, we zullen er nooit aan ontkomen om, sprekend over dit menselijk handelen, een beroep te doen op zowel het mentale als het fysische (c.q. lichamelijke) vocabulaire. Je kunt niet over een handeling spreken door alleen maar te spreken over lichamelijke bewegingen. Ergens zullen onze wensen en overtuigingen, onze intenties en deliberaties hun intrede moeten doen in een spreken dat ook maar enigszins plausibel de indruk kan wekken dat het een spreken over menselijk handelen is. En net zo min zul je over een handeling kunnen spreken door het alleen maar te hebben over onze wensen, opvattingen, motieven, redenen, overwegingen, en intenties. Natuurlijk zullen we die moeten noemen, maar ergens zullen lichamelijke bewegingen hun intrede moeten doen in dit spreken als het over meer wil kunnen gaan dan slechts over ons denken, namelijk over ons *handelen*.

Vooruitlopend op met name hoofdstuk 2 (“Handelingen zijn gebeurtenissen”) kan daarom gesteld worden dat een gedegen kennis van de filosofie van het handelen ons zou kunnen bevrijden van de intellectuele onrust waarmee het ‘lichaam/geest’-probleem ons al weer zo lang lastig valt. Dit wil niet zeggen dat we het ‘lichaam/geest’-probleem zullen kunnen oplossen met behulp van een adequaat handelingsmodel, maar het wil wel zeggen dat als we ons verdiepen in de filosofie van het handelen, we ons daarmee in één keer ook verdiepen in het probleem dat ten grondslag ligt aan ons verlangen te begrijpen hoe lichaam en geest, als ze zo op zichzelf beschouwd zouden kunnen bestaan, zouden moeten samenwerken.

3. *Individu-gemeenschap*. Een centrale kwestie in de wijsgerige antropologie, en een kwestie waarmee ze raakt aan de politieke en sociale filosofie waarmee ze, zeker in de Frans-Duitse traditie, nauw verbonden is, is de kwestie hoe we moeten denken over de manier waarop menselijke individuen zich tot elkaar en de gemeenschap verhouden. Er zijn verschillende manieren om de kwestie aan de orde te stellen. Gevraagd kan bijvoorbeeld worden of, en zo ja in hoeverre, individuen zelfstandigheid bezitten tegenover de gemeenschap waar ze lid van zijn. Maar er kan ook, bijvoorbeeld, gevraagd worden of er eigenschappen zijn die een individu alleen maar kan hebben dankzij het feit dat ze lid is van een gemeenschap. Ongetwijfeld heeft de huidige gemeenschap, met haar hoogwaardige informatietechnologie, een grote rol gespeeld in de totstandkoming van mijn vurige verlangen naar een multimedia computer. Maar nu ik eenmaal dit verlangen heb, en het expliciet onderschrijf als een verlangen dat werkelijk *van mij* is, zou het nochtans best zo kunnen zijn dat het aan mij, en slechts aan mij is, dit verlangen te betrekken in mijn overwegingen die er toe zouden kunnen leiden dat ik een beroep doe op de gemeenschappelijke middelen van mijn werkgever voor de aanschaf van zo'n fantastisch apparaat. Op een niet eenvoudig te doorgronden manier is de kwestie blijkbaar steeds in hoeverre mensen zijn wat ze zijn als gevolg van de gemeenschap waar ze deel van uitmaken, en in hoeverre deze sociale gesitueerdheid van mensen al dan niet een aantasting van hun individuele authenticiteit is.

Net als met betrekking tot de twee voorgaande kwesties is mijn suggestie hier niet dat we deze – precies deze – kwestie zullen kunnen oplossen door ons te beroepen op resultaten uit de filosofie van het handelen. Mijn suggestie is echter wel dat we in termen van de filosofie van het handelen deze kwestie op een veelbelovende manier kunnen herformuleren. Ook hier is dan mijn voornaamste overweging, en ik loop hiermee vooral vooruit op hoofdstuk 3 (“Handelingen zijn sociale fenomenen”), dat het niet goed mogelijk is over handelingen te spreken zonder beide begrippen, individu *én* gemeenschap, in ons vocabulaire te betrekken. Handelingen zullen niet als zodanig begrepen kunnen worden

zonder te spreken over de normatieve achtergrond waartegen ze begrijpelijk gevonden kunnen worden. En we zullen niet over zo'n normatieve achtergrond zullen kunnen spreken zonder een beroep te doen op de gemeenschappelijke toegankelijkheid van de regels die kenmerkend zijn voor deze achtergrond. We kunnen echter niet alleen maar spreken over de sociale begrijpelijkheid, c.q. redelijkheid, van bepaalde handelingstypen, zonder het te hebben over de individuen die een concrete instantiatie van zo'n type realiseren, als we daadwerkelijk over het menselijk handelen willen spreken. In die zin is het bij uitstek in de filosofie van het handelen dat we de kwestie van de relatie tussen individu en gemeenschap aan de orde kunnen stellen.

In de rest van dit boek zal ik niet meer expliciet spreken over deze drie traditioneel wijsgerig antropologische thema's. Mijn onderwerp is immers de filosofie van het handelen. Maar tussen de regels door zal de goede verstaander heel veel materiaal aantreffen dat rechtstreeks relevant is voor een wijsgerige reflectie op de thematiek die van oudsher geconcipieerd wordt in termen van de verschillen en verbanden tussen (1) mens en dier, (2) lichaam en geest, en (3) individu en gemeenschap.

Drie handelingsmodellen

Ik heb in het voorgaande al een aantal malen laten doorschemeren dat ik in dit boek een overzicht zal geven van drie handelingsmodellen. Daarbij zal ik telkens ook duidelijk maken hoe deze drie modellen in de hedendaagse analytische filosofie om elkaar heen draaien en elkaar voortdurend aantrekken en afstoten. De modellen laten zich goed introduceren aan de hand van een analyse van de boeiende vraag waaraan David Velleman in 1992 een artikel wijdde: "Wat gebeurt er als iemand handelt?" De vraag is interessant om een aantal redenen.

1. In de eerste plaats is het veelbetekenend dat Velleman in zijn vraag de woorden 'gebeurt' en 'handelt' vergezeld laat gaan van de woordjes 'wat' en 'iemand'. We zouden immers raar opgekeken

hebben als Velleman ons bijvoorbeeld deze vraag voorgelegd zou hebben: “Wie gebeurt er als iets handelt?” Met ‘wie’ en ‘iemand’ wordt verwezen naar personen, en *dat* zijn de veroorzakers van handelingen; met ‘wat’ en ‘iets’ wordt verwezen naar anonieme grootheden, brokken realiteit die lange oorzaak-gevolg ketens vormen en naar believen in grotere of kleinere stukken opgedeeld kunnen worden langs lijnen die tamelijk willekeurig zijn. Zoals het zinloos lijkt te zeggen dat iets handelt, lijkt het betekenisloos te beweren dat iemand gebeurt.

Handelingen worden, zoveel is duidelijk, voortgebracht door personen, of misschien, als we een iets liberalere instelling hebben, door *actoren* – mensen, dieren of artefacten met een ‘zelf’, een organisatieprincipe, of functionele eenheid. Hoe het nu precies mogelijk is, of wat het nu precies betekent, dat actoren handelingen voortbrengen is een kwestie waaraan in de hedendaagse analytische filosofie zeer veel publicaties worden gewijd – vaak (maar dit is helemaal niet noodzakelijk) in het kader van het probleem van de vrije wil. In *Hoofdstuk 1*, “Handelingen worden verricht door actoren”, zal ik een overzicht geven van de huidige stand van zaken betreffende deze kwestie.

2. In de tweede plaats verbindt Vellemans vraag, net als Anscombe’s stelling, handelingen met gebeurtenissen. Dat is in de hedendaagse analytische filosofie heel vanzelfsprekend. Daar zijn een aantal inhoudelijke redenen voor, die te maken hebben met de verschillen en verbanden tussen wat je doet en wat je overkomt, maar er is in ieder geval ook een historische verklaring voor. In 1963 publiceerde Donald Davidson een zeer druk becommentarieerd artikel, “Actions, Reasons, and Causes”, waarin hij betoogde dat handelingen een bepaalde klasse gebeurtenissen vormen, namelijk de klasse van gebeurtenissen die door redenen veroorzaakt worden. Het is de discussie die Davidson met dit artikel ontketend heeft die centraal zal staan in *Hoofdstuk 2*, “Handelingen zijn gebeurtenissen”.

3. Tenslotte gaat er van alles schuil achter dat kleine woordje ‘als’ waarmee Velleman in zijn vraag een verbinding legt tussen enerzijds de handelende actor en anderzijds het anonieme gebeuren.

Het woordje ‘als’ heeft een veelbetekenende temporele connotatie; het verwijst naar dat ene moment waarop iets gebeurt en, *tegelijkertijd*, iemand handelt. Met dat ene woordje lijkt het alsof het mogelijk is de individuatie van momenten te gebruiken om te beweren dat handelingen gebeuren, en actoren bijdragen aan het verloop der gebeurtenissen. Wat daarbij zo interessant is, is dat de individuatie van een moment altijd een tijdsverloop vooronderstelt, een geschiedenis met behulp waarvan de opeenvolging, of volgorde, van momenten vastgesteld kan worden.

Van de geschiedenis hebben we echter twee tamelijk verschillende interpretaties. Enerzijds begrijpen we het als een natuurlijk proces, een fysisch, causaal en mechanistisch gebeuren dat met de *Big Bang* begon en evolutie heet. Anderzijds begrijpen we de geschiedenis echter als een met betekenissen geladen biografie van het leven op aarde, een narratief geïnformeerd verloop der gebeurtenissen, onze *historie* die een niet te traceren begin heeft gekend in de prehistorie, en een toekomst heeft die wij niet kunnen overzien. En wie met dit in het achterhoofd naar Vellemans vraagt, voelt de aantrekkingskracht van beide interpretaties van het temporele karakter van handelingen. De eerste interpretatie versterkt het model dat in *Hoofdstuk 2* centraal zal staan. Dit is het model dat voortbouwt op de aanname dat handelingen een bepaalde klasse van gebeurtenissen zijn; handelingen moeten op de één of andere manier opgenomen zijn in dat grote natuurlijke, causale, wellicht wetmatig te beschrijven proces.

Maar de tweede interpretatie van de geschiedenis introduceert een nieuw model. In dit derde model wordt een handeling primair, of uitsluitend, gezien als een bijdrage aan een groot sociaal en cultureel verhaal; een *praktijk* waarin een grote rol wordt gespeeld door selectieve aandacht en een opbouw rondom grote thema's. Wat er is gebeurd, en staat te gebeuren, als iemand handelt, is iets dat alleen begrepen, en beschreven, kan worden vanuit een historisch en sociologisch perspectief. Handelingen zijn in dit model niet primair het resultaat van de activiteit van een rationele en autonome actor, en ook niet vooral gebeurtenissen, maar essentieel *sociale* grootheden. In *Hoofdstuk 3*, “Handelingen zijn sociale

fenomenen”, zal ik een overzicht geven van de actuele discussie over dit derde model.

Mijn bespreking van de diverse handelingsmodellen in de nu volgende drie hoofdstukken zal niet historisch van karakter zijn: ik zal me niet vooral bezighouden met het uiteenzetten van wie nu precies wanneer wat heeft beweerd. Waar het mij om zal gaan, is het articuleren van een aantal modellen, waarbij ik gebruik maak van het conceptuele instrumentarium van de hedendaagse analytische wijsbegeerte. Natuurlijk zal ik daarbij namen noemen, en publicaties. Maar geheel in overeenstemming met de analytische instelling zal ik primair de *conceptuele*, en niet de historische, ontwikkeling van een drietal filosofische aannames betreffende de aard van handelingen in kaart trachten te brengen.

Zeer uitgebreide bronverwijzingen ontbreken overigens natuurlijk niet in dit overzicht. Ik heb ze, voorzien van korte opmerkingen, en gestructureerd in overeenstemming met de opbouw van dit boek, in de *Literatuursuggesties* opgenomen.

Handelingen worden verricht door actoren

1. Over zelfstandigheid

In kunnen grijpen

Wil het mogelijk zijn dat wij handelingen verrichten, dat onze aanwezigheid als actoren effect heeft op het verloop der gebeurtenissen, dan lijkt het erop dat er sprake moet zijn van *vrijheid*. Het is met het oog op het probleem van de vrije wil dat er in de analytische filosofie uitgebreid is gediscussieerd over de kansen voor een model van het menselijk handelen waarin de these centraal staat dat handelingen bestaan omdat ze verricht worden door actoren. De these lijkt in eerste instantie twee implicaties te hebben. Als actoren verantwoordelijk zijn voor het bestaan van handelingen, dan moeten ze enerzijds spontaan kunnen zijn – ze moeten iets kunnen bewerkstelligen waartoe ze door de omstandigheden niet genoodzaakt zijn – en anderzijds in zichzelf voldoende grond kunnen vinden voor het verrichten van de handeling die ze feitelijk (besluiten te) verrichten. Dat ik een boterham met pindakaas eet, is een handeling die door mij verricht wordt, als ik er niet toe genoodzaakt was door de omstandigheden (ik had ook niets kunnen eten, of een boterham met hagelslag, of een croissantje brie), en er zelfstandig toe kon besluiten (het is op grond van mijn wens eens lekker te genieten van een boterham met pindakaas dat deze handeling verricht wordt).

De vanzelfsprekende juistheid van deze constatering wordt echter tamelijk problematisch als we precies onder woorden proberen te brengen wat het zou kunnen betekenen dat actoren *zelfstandig*, of autonoom, handelingen verrichten. Dit is op een heel aardige manier in beeld gebracht door Thomas Nagel (1976) die zich afvraagt of er sprake zou kunnen zijn van 'moreel geluk'. De vanzelfsprekende aanname hierbij is dat er in het morele domein

geen sprake van geluk behoort te zijn, omdat alleen dat morele goed- of afkeuring verdient, wat ook inderdaad door een actor zelfstandig gedaan wordt. Morele goed- of afkeuring behoort niemand toevalligerwijze 'in de schoot' te vallen, zo is de suggestie. Niettemin spelen toevalsfactoren een grote rol in onze gewoonten anderen vanuit moreel perspectief te beoordelen. Wie toevallig nooit ter plaatse is als er een kind dreigt te verdrinken, zal nooit als deugdzaame held door het leven kunnen gaan. Wie een akelig karakter heeft, krijgt ook nog eens de pech gestraft te worden voor de nare dingen die zij doet, terwijl een blijmoedige vrouw, met enorm veel energie en een heel groot hart, ook nog eens het geluk heeft al die goedkeuring te oogsten voor gedrag dat haar nauwelijks moeite kost. We kunnen mensen niet verantwoordelijk houden voor de omstandigheden waarin ze verkeren, maar, zo betoogt Nagel, bij nader inzien zou je haast gaan denken dat de actor zelf ook gewoon maar een onderdeel is van de omstandigheden. Wie alle toevalsfactoren geduldig en zorgvuldig in ogenschouw neemt, kan haast niet anders dan concluderen dat het domein dat *onder controle* van de actor valt ineen krimpt tot een oneindig klein punt.

Als we de these willen verdedigen dat actoren zelfstandig handelingen verrichten, dan lijkt het erop dat we begrijpelijk zullen moeten kunnen maken wat er bedoeld wordt met de these dat er een domein is waarover de actor *controle* uitoefent, een domein waarin slechts gebeurt wat door de actor wordt *veroorzaakt*. Als we ons probleem echter zo opvatten, en daarmee het spoor trachten te volgen dat Roderick Chisholm in 1964 in "Human Freedom and the Self" heeft uitgezet, dan zullen we het erg moeilijk krijgen, omdat we de term 'domein' niet letterlijk zullen kunnen opvatten, maar ook niet goed weten hoe we hem dan wel moeten opvatten. Het punt zal duidelijk zijn: de biologische of staatkundige term 'territorium' kan goed gebruikt worden om uit te leggen wat bedoeld wordt met het uitoefenen van controle. Iemand die binnen een territorium de macht heeft, kan die macht uitoefenen, en dat wil dan zeggen dat binnen de *ruimtelijke* grenzen van dat territorium alleen dat zal gebeuren wat veroorzaakt is door (de uitdrukkelijke toestemming van) deze machtshebber.

Een dergelijke *ruimtelijke* interpretatie is echter onbruikbaar als we het over het uitoefenen van onze vrije wil hebben. Want waar zouden de grenzen van dat domein, waarbinnen we onze autonomie kunnen realiseren, zich bevinden? Kinderen die leren praten, ontdekken dat het domein waarin zij het ‘voor het zeggen hebben’ zich tot ver buiten de grenzen van hun lichaam uitstrekt. Psychiatrische patiënten, daarentegen, ontdekken juist weer dat ze zelfs binnen hun eigen lichaam niets, of niet veel, of niet genoeg, te zeggen hebben. En Chisholms suggestie dat de hersenen een dergelijk domein zouden zijn, creëert vooral problemen, omdat de hersenen natuurlijk ook, en bij uitstek, een fysisch domein vormen waarin, zoals we nu weten, allerhande neurotransmitters heel veel te ‘zeggen’ hebben.

Het hiërarchische model van zelfstandigheid

Een grote vooruitgang op dit terrein werd mogelijk gemaakt door Harry Frankfurt, die in 1971 het zeer invloedrijke artikel “Freedom of the Will and the Concept of a Person” publiceerde. Met dit artikel heeft Frankfurt de discussie losgemaakt van de idee dat het begrip ‘controle’ (dat nauw gelieerd is aan het begrip ‘veroorzaking’) een hoofdrol zou moeten spelen in analyses van de these dat actoren zelfstandig handelingen verrichten. We hebben met betrekking tot het probleem van autonomie volgens Frankfurt geen behoefte aan een analyse van het begrip ‘controle’, maar aan een analyse van de structuur van iemands *eigen* wil. De werkelijk relevante vragen hebben te maken met het begrijpen van de these dat een handeling *van iemand* is, door iemands eigen wil is gewild.

Wanneer is gedrag dat door mijn lichaam wordt vertoond een handeling *van mij*? Als ik met mijn handen omhoog in de loop van een revolver kijk en een boef mij twee handelingsalternatieven aanbiedt, “Je geld of je leven!”, dan heb ik niets aan een analyse van het begrip ‘controle’. Want de boef zal werkelijk, vanuit een dergelijk filosofisch perspectief, gelijk hebben als hij zou stellen dat ik *zelf* mijn geld gegeven heb, omdat ik op het juiste moment

mijn controle heb uitgeoefend in dat niet nader aan te wijzen domein waarin alleen *ik* het voor het zeggen heb. Niettemin weten we allemaal dat een slachtoffer van een overvaller heeft gehandeld ondanks zichzelf, waarmee we bedoelen dat zo'n slachtoffer niet zelfstandig, niet uit *eigen* wil heeft gedaan wat hij deed. Hij stond, zo zouden we kunnen zeggen, geenszins achter zijn daad. En met deze formulering treffen we precies de kern van Frankfurts betoog.

In eerste instantie is het fundamentele begrip in Frankfurt's theorie van autonomie het begrip van de 'wilstoestand van de tweede-orde'. Frankfurt verdedigt een zgn. *hiërarchisch* model van zelfstandigheid, een model waarin er wilstoestanden, of wensen, op verschillende niveaus voorkomen. Ik kan bijvoorbeeld enorm veel zin hebben in het opeten van een heel pak speculaasjes. Maar ik kan ook, en zelfs tegelijkertijd, de wil hebben om me te beheersen en me niet te laten (ver)leiden door mijn verlangen naar al die speculaasjes. Het verlangen naar speculaasjes is een wilstoestand van de eerste orde, omdat hij gericht is op een bepaalde (on)gewenste stand van zaken in de wereld. Het verlangen om mijn gedrag niet te laten sturen door deze wilstoestand van de eerste orde, is zelf een wilstoestand van de tweede orde. Hij is van de tweede orde, omdat zijn intentionele object (dat is de stand van zaken waar hij als wilstoestand op gericht is), niet een toestand in of van de wereld is, maar een toestand van mijn eigen wil. Als ik mij wil beheersen, dan gaat mijn wil *over* mijn eigen wil, of, preciezer gezegd, dan heb ik een wilstoestand van de tweede orde (namelijk de wil me te beheersen) die als intentioneel object een wilstoestand van de eerste orde heeft (namelijk mijn verlangen naar al die speculaasjes), en ten aanzien van dit intentioneel object wenst dat hij niet effectief is.

Voor alle duidelijkheid zal ik nog een omgekeerd voorbeeld geven. Stel je houdt niet van spruitjes, maar je bent ergens op bezoek en je krijgt een bord vol van die akelige groene bolletjes voorgeschoteld. Het is niet ondenkbaar dat je naast je afkeer van spruitjes (wat een negatieve wilstoestand van de eerste orde is) ook nog graag zou willen dat je spruitjes lekker vond. Want was dat

het geval, dan zou je een heel plezierige maaltijd kunnen hebben. Deze wens om spruitjes lekker te willen vinden, is een wilstoestand van de tweede orde, omdat hij niet over spruitjes gaat, maar over je eigen wilstoestanden ten opzichte van spruitjes. Vond je spruitjes maar lekker! Kon je je eigen wil maar veranderen! Dat is wat je op zo'n moment eigenlijk wil.

Frankfurt wijst op deze interne structuur van de wil, op het vermogen van mensen zelf-reflectief te zijn, op het vermogen zich te bekommeren om de wenselijkheid van wensen. En Frankfurt betoogt dat als we willen begrijpen dat mensen handelen in overeenstemming met hun eigen wil, als we willen begrijpen dat mensen zelfstandig handelingen verrichten, we dit moeten begrijpen als handelen *gemotiveerd* door effectieve *eerste-orde* wilstoestanden die *gesteund* worden door wilstoestanden van de *tweede-orde*. De eigen wil van mensen bestaat in de wilstoestanden die gaan over de eigen wilstoestanden. Iemand die dol is op speculaasjes, maar zich tevens wenst te beheersen, verricht handelingen, d.w.z. handelt zelfstandig, als hij zich gedraagt conform de wens zich te beheersen, dat is conform zijn tweede-orde wilstoestand. Iemand die dol is op speculaasjes, en zich wenst te beheersen, maar daar niet in slaagt en zich laat leiden door wensen die hij eigenlijk zelf niet wil hebben..., zo iemand handelt niet autonoom, of, zoals Frankfurt het ook wel zegt, niet uit vrije wil. Een verslaafde verricht geen handelingen, maar laat van alles met zich gebeuren, zoals het slachtoffer van een overvaller niet zelfstandig, uit vrije wil, zijn geld geeft, maar helaas moet toezien dat dit met hem gebeurt.

Het hiërarchisch model van zelfstandigheid is in eerste instantie een veelbelovend model. Het laat zien dat een handelingsfilosofie veel aandacht moet besteden aan *deliberatie*, aan het praktisch redeneren, het afwegen van pro's en contra's, het overwegen van verschillende opties, het oog hebben voor verschillende, vaak tegengestelde motieven. Het model laat ook zien hoe men inhoud kan geven aan de these dat actoren handelingen verrichten zonder te verzanden in metafysisch onhandelbare problemen betreffende de kwestie van het uitoefenen van controle. *Hoe* iemand erin

slaagt iets te doen is vanuit het perspectief van dit model niet iets dat verder geanalyseerd hoeft te worden. Wat er toe doet is *dat* iemand er in slaagt iets te doen in overeenstemming met zijn wil. Dit is argumentatief een belangrijke manoeuvre.

De eerste stap werd al in 1965 voorbereid door Arthur Danto die betoogde dat er 'basic actions' bestaan, handelingen waarvan we weten hoe we ze moeten doen, zonder dat we verder iets kunnen zeggen over *hoe* het komt dat we dat weten. Dergelijke handelingen verrichten we gewoon, en de vraag naar *hoe* we dat dan doen is zelfs een beetje raar. Ik weet *hoe* ik van mijn kiespijn af kan komen. Door naar de tandarts te gaan. Ik weet *hoe* ik naar de tandarts kan gaan. Door een afspraak te maken. Ik weet *hoe* ik een afspraak kan maken. Door te bellen. Ik weet *hoe* ik moet bellen. Door de hoorn van de haak te nemen en het nummer te toetsen. Ik weet *hoe* ik de hoorn van de haak kan nemen. Door hem beet te pakken en op te tillen. Ik weet zelfs nog *hoe* ik de hoorn beet kan pakken. Door mijn hand er naar toe te bewegen en een grijpbeweging te maken. Maar dan houdt het op! Ik weet niet *hoe* ik mijn hand naar de hoorn kan bewegen. Dat kan ik gewoon! Het bewegen van mijn hand is een 'basic action'. Ik weet zelfs niet wat voor soort antwoord ik zou moeten geven op de vraag *hoe* het toch komt dat ik mijn hand kan bewegen. Door mijn spieren te bewegen? Maar *dat* is nooit iets wat ik probeer te *doen*. Dat is wat gebeurt als ik iets anders probeer te doen: namelijk, mijn hand bewegen. Zoals een klein kind ook leert zijn plas op te houden, en niet, daar als het ware aan vooraf gaand, zijn willekeurige kringspier beheersen.

De hele opzet te zoeken naar het domein waarbinnen de actor zelfstandig controle kan uitoefenen, is vanuit het model van Frankfurt misleidend. Daar valt verder geen conceptuele verheldering te verwachten. En dat is ook helemaal niet erg. Want als we de these willen verdedigen dat actoren zelfstandig handelingen verrichten, dan gaat het ons er om te kunnen begrijpen wat het betekent dat mensen alleen maar die 'basic actions' verrichten *die in overeenstemming zijn met hun eigen vrije wil*. En om dit te kunnen begrijpen hebben we in eerste instantie niet meer nodig dan

een adequaat model van een actors eigen vrije wil, van zijn zelfstandigheid. Dit model meent Frankfurt te hebben gevonden met behulp van het idee van de hiërarchische relaties tussen wilstoestanden van de eerste- en de tweede-orde.

Drie problemen

Uit de vele publicaties die er over deze kwestie sinds 1971 zijn verschenen blijkt echter duidelijk dat het inmiddels gemeengoed is dat er enkele zeer serieuze problemen aan Frankfurts model kleven. Op de eerste plaats is er de dreiging van de oneindige regressus. Waarom zou mijn eigen wil nu precies huizen in mijn wilstoestanden van de *tweede* orde? Waarom zou ik me niet soms, meestal, of altijd moeten afvragen of ik nu wel zo gelukkig ben met deze tweede-orde wilstoestanden? Zou het niet mogelijk zijn dat ik op een oneigenlijke wijze aan die tweede-orde wilstoestand ben gekomen die behelst dat ik niet toe wil geven aan mijn verlangen naar een heel pak speculaasjes? Dergelijke vragen liggen vooral voor de hand in de huidige cultuur (zeker als we in Nederland boven de grote rivieren blijven), een cultuur waarin we al zo vaak tot de ontdekking zijn gekomen dat het helemaal niet nodig is gebukt te gaan onder gewetensvolle Calvinistische trekjes. De suggestie van het regressus-probleem is natuurlijk overigens niet dat we dan pas werkelijk onze eigen wil ontdekken in de wilstoestand van de derde-orde die gaat over ons (oneigenlijke) verlangen ons te beheersen. Nee, de suggestie van het *regressus*-probleem is nu juist dat, zodra je de willekeur van de tweede-orde inziet, je niet anders kunt dan concluderen dat er in principe geen enkele goede reden is om op welk niveau dan ook te stoppen en te denken dat je op dat niveau werkelijk de inhoud van je eigen wil zou kunnen aantreffen.

Een parallel probleem is wat John Christman (1989) het 'ab-initio'-probleem heeft genoemd. Als we op ieder hoger niveau geen andere toestanden aan zullen treffen dan *wilstoestanden*, waarom zeggen we dan niet dat onze eigen wil huist in iedere wilstoestand die we hebben, onafhankelijk van welke orde dan ook.

Waarom zou onze wil niet intern verscheurd kunnen zijn? Waarom moeten we zulke drastische stappen nemen, en beweren dat onze *eigen* wil in de hogere-orde toestanden bestaat, daarmee beweerend dat sommige wilstoestanden van de eerste-orde vreemde indringers zijn in onze motivationele huishouding? We weten toch allemaal dat er wel iets zit in psychoanalytische beweringen dat men zich niet zo maar ongestraft kan dissociëren van ongewenste wilstoestanden? De kwestie die hier aan de orde wordt gesteld, is niet dat we maar moeten accepteren dat we in motivationeel opzicht geen eenheid zijn, maar wel dat, als er geen andere optie is dan onze *eigen* wil te identificeren met sommige van onze eigen wilstoestanden, er geen enkele goede reden lijkt te zijn om te denken dat de wilstoestanden op een hoger niveau meer aanspraak zouden kunnen laten gelden dan die op een lager niveau.

Samengevat komen beide problemen eigenlijk neer op het betwijfelen van de autoriteit van de tweede-orde wilstoestanden met het oog op het bepalen van een actors *eigen* wil. Of zoals Gerald Dworkin (1981) zich afvraagt, daarmee respectievelijk het *ab initio*-en het *regressus*-probleem kernachtig karakteriserend: “Waarom *beginnen* aan het tweede niveau? Maar ook, waarom *stoppen* op het tweede niveau?”

Het lijkt erop alsof Frankfurt, maar ook andere verdedigers van het hiërarchische model, onterecht gebruik maken van een bepaalde ambiguïteit in de idee van een hogere-orde mentale toestand. Een bepaalde mentale toestand kan hoger zijn dan een andere omdat de eerste de tweede als zijn intentioneel object heeft. Dit is de ‘meta-’ connotatie. Maar er is ook de ‘autoritatieve’ connotatie: een bepaalde mentale toestand kan hoger zijn dan een andere omdat hij meer gewicht met betrekking tot de inhoud van de wil van de actor in de schaal kan leggen. Nu is het echter één ding om te beweren dat bepaalde wilstoestanden hoger zijn in de ‘meta-’ zin van het woord, maar een heel ander ding om te beweren dat bepaalde wilstoestanden hoger zijn in de ‘autoritatieve’ zin van het woord. Zodra men deze connotaties verwart, verzeilt men hetzij in het *regressus*-probleem (door steeds weer te ontdekken dat de ogenschijnlijk ‘autoritatieve’ hogere-orde wilstoestand slechts in de

‘meta’-zin van het woord hoger is), hetzij in het *ab initio*-probleem (door al bij voorbaat te stellen dat het ogenschijnlijk ‘autoritatieve’ karakter slechts, en onterecht, opgeroepen wordt door het misbruik van de term ‘hoger’ – een misbruik dat mogelijk wordt als men de ‘meta’-zin van het woord versluiert). Zodra men echter beseft dat deze connotaties uiteen gehouden moeten worden, ontstaat een nieuw probleem, maar ook een nieuwe oplossingsstrategie.

Deze redenering ligt ten grondslag aan de naam die John Christman gegeven heeft aan nog een derde probleem met het hiërarchische model van zelfstandigheid: het ‘onvolledigheids’-probleem. Als het ‘meta-’ karakter van bepaalde hogere orde mentale toestanden op zichzelf beschouwd nog niet voldoende is voor de stelling dat een dergelijke hogere orde toestand meer gewicht in de schaal legt bij het bepalen van de inhoud van een actors eigen wil, dan impliceert dit dat er een extra stap nodig is voor het toekennen van autoriteit aan deze hogere orde toestand. In de literatuur zijn verschillende oplossingen voor dit probleem voorgesteld, waarvan ik er een drietal kort zal karakteriseren.

Waarde-oordelen

De eerste oplossing van het ‘onvolledigheids’-probleem is voorbereid door Gary Watson (1975), en wordt gekenmerkt door het opgeven van de centrale Humeaanse vooronderstelling in Frankfurts hiërarchische model. Deze vooronderstelling is dat er maar één type mentale toestand is dat een rol speelt in het motiveren van een actor *qua actor*: *wil*toestanden. Watson betoogt dat het karakteristieke verschil in autoriteit tussen de mentale toestanden die *niet* en de mentale toestanden die *wel* bepalend zijn voor de inhoud van iemands eigen wil, nooit uitgelegd kan worden door een verschil in niveau te maken tussen *wil*toestanden. Waar het volgens Watson op neer komt, is in te zien dat er verschillende typen mentale toestanden zijn die allemaal een rol spelen in de zelfreflectie waarmee een actor de inhoud van zijn eigen wil tracht te bepalen. Naast de *wil*toestanden, die traditioneel de passies genoemd worden, treffen we ook de waarde-oordelen aan, die

binnen een Platoons denkkader begrepen kunnen worden als voortgebracht door de *ratio*.

Hume, en daarmee Frankfurt, heeft ongelijk als hij meent dat de *ratio* slechts de slaaf is van de passies. Volgens Watson moeten we waarde-oordelen beschouwen als onafhankelijk voorkomend naast de wilstoestanden, en moeten we zelfs tot de conclusie komen dat onze eigen wil in belangrijker mate gearticuleerd wordt door onze waarde-oordelen dan door onze wilstoestanden. Het verschil in autoriteit tussen verschillende motivationele toestanden zit volgens Watson niet in een verschil in de positie in de hiërarchische opbouw van onze wilstoestanden, maar in een verschil in de *bron* van de desbetreffende motivationele toestanden. En voor zover gesproken kan worden van een actor die handelingen verricht, moet er niet zo zeer gekeken worden naar het zogenaamde 'motivationale systeem' (waarin allerlei zaken als driften, emoties, wensen, verlangens, en waarde-oordelen bijdragen aan het 'aansturen' van ons lichaam), maar naar het zogenaamde 'evaluerende systeem'. In dit systeem gaat de persoon zelf als het ware op in deliberaties en komt tot waarde-oordelen die, vervolgens, in iemands 'motivationale systeem' de eigen wil van de actor vertegenwoordigen en de diverse krachten die in het 'motivationale systeem' aanwezig zijn trachten te mobiliseren voor de goede zaak – dat wil zeggen, voor gedrag dat in overeenstemming is met deze waarde-oordelen.

Er is een subtiel argument van Charles Taylor (1985) voor de these dat waarde-oordelen inderdaad *naast* wilstoestanden moeten bestaan, en er niet toe herleid kunnen worden. Alleen zo immers, betoogt Taylor, zullen we kunnen begrijpen dat het mogelijk is een zinnig oordeel te vellen over de (on)wenselijkheid van onze wensen, iets wat wij voortdurend doen wanneer we ons bezighouden met het bepalen van de inhoud van onze eigen wil. Als ik zin heb in een heel pak speculaasjes, en me afvraag of het gewenst is dat ik daar zin in heb, dan vraag ik me in feite af of mijn verlangen naar speculaasjes gepast is; ik vraag me dan af of de sterkte van mijn verlangen naar speculaasjes in overeenstemming is met de waarde van die speculaasjes. Maar om zinvol te kunnen praten

over dit op elkaar afgestemd zijn van de sterkte van de wens en de waarde van het gewenste object, moet er natuurlijk een onafhankelijke toegang zijn tot het vaststellen van de waarde van dat object. Dit betekent niet vanzelf dat we hiermee een argument hebben voor de these dat waarden werkelijk objectief, onafhankelijk van onze oordelen, zijn. Het is best mogelijk dat we met onze waarde-oordelen bepaalde waarden in een object projecteren. Het enige dat door dit argument bewerkstelligd wordt, is dat we de waarde van een object niet kunnen afleiden uit de sterkte van ons verlangen naar dat object. En natuurlijk weten we dit allemaal, want ook al maakt liefde blind, toch zullen we altijd blijven beweren dat we naar onze geliefde verlangen omdat ze lief *is*, en altijd blijven ontkennen dat ze lief is omdat we naar haar verlangen.

De strekking van deze Watson/Taylor oplossing voor het 'onvolledigheids'-probleem, is dat we opnieuw stuiten op een variant van het probleem dat we eerder al tegenkwamen in termen van het afbakenen van het domein waarbinnen de actor controle kan uitoefenen. Immers, de suggestie die we hier onderzoeken gaat uit van de these dat er niet zo zeer sprake is van een hiërarchische ordening van de verschillende motivationele toestanden die zich in de psyche van een actor voordoen, maar dat er sprake is van een hiërarchische ordening tussen verschillende subsystemen die allemaal een bijdrage leveren aan het bepalen van de inhoud van de eigen wil van een actor. En de stelling is dat één van de aanwezige subsystemen de werkelijke vertegenwoordiger is van de actor zelf.

Dit idee van de intern gespleten actor is niet nieuw. We treffen het inderdaad al aan in de metafoor van Plato die de ziel in drieën deelt en stelt dat de ratio als een koetsier is die zijn twee paarden, de goede wil en de (kwade) passie, in toom moet zien te houden. En we treffen het ook aan in Freuds onderscheid tussen Id, Ego, en Super-Ego.

Maar de bewering dat we tussen de verschillende, in de actor aanwezige subsystemen er eentje zullen kunnen vinden die met recht de werkelijke vertegenwoordiger van de actor zelf genoemd

zou kunnen worden, is een stelling die zeer moeilijk te verdedigen is. Natuurlijk gaat het ook hier weer niet om het letterlijk aan kunnen wijzen van een empirisch subsysteem, alsof er ergens in het lichaam van de actor een speciaal orgaan zou zijn dat functioneert als werkelijke vertegenwoordiger van de actors eigen *wil*. Maar als dat niet zo is, op grond waarvan zouden we dan categorisch kunnen beweren dat het altijd de waarde-oordelen zijn, en nooit de driften, emoties of verlangens, die het meeste recht van spreken hebben? Om het bij één bezwaar te laten, dat wat mij betreft de kracht heeft van een 'knock-down' argument, parafraseer ik graag een voorbeeld van Marilyn Friedman (1986): "Hoe zelfstandig handelt een vrouw die oprecht en weloverwogen haar aspiraties op een carrière opgeeft als ze trouwt, die bewust besluit een schoonmaakbaantje te accepteren om haar man zijn proefschrift te laten voltooien, en het waarde-oordeel velt dat ze, alles welbeschouwd, het best met werken op kan houden als de kinderen komen?"

Natuurlijk wordt het filosofische probleem dat hier aan de orde is, niet minder groot als we recalcitrant zouden beweren dat dit allemaal de schuld is van onze overgerationaliseerde cultuur, en dat we juist altijd ons *eigen gevoel* moeten volgen, of moeten leren luisteren naar de stem van ons lichaam. Iedere bewering dat het één of andere subsysteem de werkelijke vertegenwoordiger is van de actors eigen wil, blijft immers iets willekeurigs hebben. Het lijkt erop alsof de in deze paragraaf gesuggereerde oplossing, dat we naast wilstoestanden waarde-oordelen nodig hebben om te kunnen rechtvaardigen dat sommige mentale toestanden een grotere rol behoren te spelen in het bepalen van de eigen wil, het 'onvolledigheids'-probleem niet echt oplost, maar slechts een stap opschuift. Het model blijft immers onvolledig: ook het onderscheiden van diverse subsystemen legt nog niet onmiddellijk meer autoriteit bij deze of gene bepaling van de wil.

Er is overigens een tweede interpretatie mogelijk van de aandacht die door Watson, Taylor, en andere geestverwanten gevestigd wordt op de cruciale rol van waarde-oordelen. Deze tweede interpretatie

legt de nadruk niet op een voorondersteld subsysteem dat verantwoordelijk is voor de productie van waarde-oordelen, maar legt de nadruk op het proces van het articuleren, vormen of vellen van waarde-oordelen. Dit proces, dat we deliberatie of praktisch redeneren noemen, is, volgens deze interpretatie, bij uitstek de zaak waarin zich de zelfstandigheid van de actor toont. De bewering is hier dat het niet zo heel belangrijk is een verhaal te vertellen dat duidelijk maakt dat een bepaalde handeling door mij *en dat wil zeggen niet door jou* wordt verricht, maar om duidelijk te maken waarin de activiteit van een actor verschilt van de activiteit van, bijvoorbeeld, een planeet, een amoebe, een benzeenmolecule, een thermometer of een beuk. Door de nadruk te leggen op deliberatie, als dat wat ten grondslag ligt aan het verrichten van handelingen, verschuift de aandacht van het onderscheid tussen de ene actor en de andere naar het onderscheid tussen enerzijds de redelijkheid van actoren en anderzijds de anonieme wetmatigheid van fysische processen. Op deze verschuiving zal ik nog uitgebreid terugkomen.¹ Hier volstaat het op te merken dat deze verschuiving in ieder geval niet het probleem van de zelfstandigheid van de actor oplost, maar het terzijde schuift als niet van fundamenteel belang voor het begrijpen van de these dat handelingen verricht worden door actoren.

Coherentie

Een tweede oplossing voor het 'onvolledigheids'-probleem wordt door verschillende mensen met verschillende oriëntaties verdedigd. Het sleutelwoord is steeds 'coherentie': een bepaalde motivationele toestand komt de autoriteit toe de beslissende stem te zijn in de bepaling van de eigen wil van een actor als die toestand de *samenhang* of *coherentie* in de motivationele huishouding van de actor vergroot. Deze oplossing is in eerste instantie te zien als een reactie op de problemen die er zijn met de eerste oplossing. Als de introductie van verschillende subsystemen fenomenologisch

¹ Zie de discussie in de volgende paragraaf van dit hoofdstuk, "Over delibereren, motiveren en rechtvaardigen", pp. 32-55, en de discussie in hoofdstuk 2, "Handelingen zijn gebeurtenissen".

correct is, maar het aanwijzen van het een of andere subsysteem als de categorisch geautoriseerde vertegenwoordiger van de actor zelf altijd willekeurig blijft, dan kan de coherentie-oplossing goed werk doen. De idee is dat alle motivationele toestanden in principe gerechtvaardigd kunnen beweren dat ze de beslissende stem articuleren. Geen enkele motivationele toestand komt meer autoriteit toe vanwege zijn afkomst, om het zo te zeggen. Soms zal het een bepaalde passie kunnen zijn die tegenover een overdreven gerationaliseerd waarde-oordeel het meeste gewicht in de schaal legt, maar ook zou het soms een bepaalde overtuiging kunnen zijn, of een intentie, of een emotie, enz. Inspiratie voor deze oplossing kan gevonden worden in de idee van het zogenaamd reflectief equilibrium dat dankzij het werk van John Rawls (1972) ingang heeft gevonden in het debat over de juiste structuur van normatieve argumentatie. Binnen dit debat stellen verdedigers van het reflectief equilibrium dat een moreel oordeel niet altijd afgeleid kan worden uit principes, maar dat zo'n oordeel altijd het resultaat is van een evenwicht dat na langdurige deliberatie resulteert uit een confrontatie van ethische principes, morele intuïties, moreel relevante achtergrondinformatie, precedentes, etc. Overeenkomstig stellen verdedigers van de coherentie-oplossing van het 'onvolledigheids'-probleem dat het niet gepast is categorisch meer gewicht toe te kennen aan bijvoorbeeld de waarde-oordelen, maar dat het meeste gewicht toekomt aan die motivationele toestand die de samenhang in de psyche van de actor, in diens handelen of diens biografie, het meest behartigt.

Het grote probleem voor deze oplossing is dat ze niet goed kan uitleggen waarom er zo'n groot belang wordt gehecht aan coherentie. Waarom zou 'coherentie' de ultieme waarde zijn (dat wat Charles Taylor (1989) een 'hypergood' noemt) waaraan motivationele toestanden hun belang ontlenen?

De verdedigers van deze oplossing zullen hun positie overigens niet in termen van coherentie *als 'hypergood'* karakteriseren. Coherentie is voor hun niet een doel op zich, maar is een teken van het feit dat de actor *één* eigen wil heeft. Zij nemen daarbij aan dat we deze eenheid van de wil van de actor nodig hebben om de these

dat handelingen verricht worden door actoren begrijpelijk, en vervolgens verdedigbaar, te maken. Dit lijkt een gerechtvaardigde aanname die zelfs ten grondslag lijkt te liggen aan de hele discussie: zoals ik immers al in de *Inleiding* schreef impliceert de idee van een actor een 'zelf', en dat wordt doorgaans beschouwd als een organisatieprincipe of functionele eenheid. De suggestie van de verdedigers van deze oplossing is dat we niet kunnen, maar ook niet hoeven, zoeken naar deze 'zelf' *achter* iemands handelingen. Deze 'zelf' toont zich in de handelingen en in de daaraan ten grondslag liggende motieven als de rode draad die eenheid geeft aan de biografie van een actor.

Hoe juist dit ook mag klinken, toch is het geen argument voor de these dat het 'onvolledigheids'-probleem opgelost kan worden door het naar voren schuiven van zo'n formele waarde als 'coherentie'. De moeilijkheid is hier vergelijkbaar met de fundamentele moeilijkheid voor een procedurele ethiek. Als we willen weten wat het beste is om te doen, en we hanteren als uitgangspunt de opvatting dat de beste optie *die* is die door een bepaalde procedure aangewezen wordt, dan wordt het voor ons onmogelijk de procedure op te starten tenzij we ongefundeerd beginnen met een aanname van wat we voorafgaand aan de procedure denken dat de uitkomst zal zijn, of zal behoren te zijn. Het resultaat is dan natuurlijk dat de uitkomst van de procedure beslissend bepaald wordt door deze ongefundeerde voor-oordelen (zoals Hans-Georg Gadamer (1960) deze veelzeggend en instemmend noemt). Maar als deze voor-oordelen zo'n grote rol spelen, blijkt de functie van de procedure niet meer dan slechts een heuristische indicatie achteraf te kunnen zijn, en blijkt de procedure zelf helemaal geen rol van betekenis te kunnen spelen in het bepalen van wat het beste is om te doen.

Hetzelfde geldt voor de rol van 'coherentie' in het oplossen van motivationele conflicten. Pas tegen de achtergrond van een voorhanden hoeveelheid motivationele toestanden met een voorgegeven gewicht kan coherentie een rol gaan spelen. Maar ze kan nooit de rol gaan spelen die bestaat in het herzien van het gewicht van deze of gene motivationele toestand, terwijl dat nu juist is wat

moet gebeuren in echte motivationele conflicten. Want die bestaan bij de gratie van een evenwicht tussen de factoren *voor* en de factoren *tegen* een bepaalde handeling. Het blootleggen van de echte eigen wil van de actor kan in zulke situaties nooit geschieden door 'coherentie' een ultiem gewicht te geven. Immers, als er een evenwicht bestaat tussen de voors en de tegens, dan helpt de eis van coherentie ons niet bij het maken van de keuze voor het elimineren van hetzij de voors hetzij de tegens. Bovendien impliceert de eis van coherentie dat we conservatisme moeten beschouwen als een grondtrek van ons actorschap: ik kan het gemakkelijkst de coherentie van mijn motieven waarborgen door nooit iets anders te willen dan wat ik eerder al eens wilde.

De actor-identificerende handeling

De derde oplossing van het 'onvolledigheids'-probleem komt van Frankfurt zelf. Deze oplossing is beter dan de vorige, omdat zij zich niet beweegt op het niveau van de symptomen, maar op het niveau van de zaak zelf. Frankfurts suggestie is dat een tweede-orde wilstoestand de eigen wil van de actor zelf kan bepalen, doordat de actor zich door middel van een specifieke mentale handeling *identificeert met* een bepaalde eerste-orde wilstoestand. Met zo'n 'actor-identificerende handeling' formeert de actor als het ware een tweede-orde wilstoestand die een bepaalde eerste-orde wilstoestand aanwijst. Daarmee zegt de actor als het ware dat het deze specifieke eerste-orde wilstoestand behoort te zijn die bepaalt wat de inhoud is van de eigen wil van de actor. Geconfronteerd met de botsing tussen bijvoorbeeld de wens naar het strand te gaan en de wens dit hoofdstuk uit te lezen, kan een actor zich door middel van een 'actor-identificerende handeling' identificeren met één van deze wensen en bijvoorbeeld besluiten dat het *zijn* wens is naar het strand te gaan (en, bijgevolg, niet echt *zijn* wens dit hoofdstuk uit te lezen). We moeten dit niet begrijpen alsof er op ieder willekeurig moment in de psyche van een actor allerlei tweede-orde wilstoestanden voorhanden zijn die als het ware stiltes liggen te wachten tot ze als werkelijke vertegenwoordiger

van de actor zelf gemobiliseerd worden om de juiste eerste-orde wilstoestand aan te wijzen. Nee, voorhanden zijn alleen maar allerlei *eerste-orde* wilstoestanden, die natuurlijk niet stilletjes liggen te wachten, maar voortdurend in de weer zijn hun energie over te dragen op de bewegende onderdelen van het lichaam van de actor (om het maar zo, *al te plastisch*, te zeggen). Een patstelling tussen deze eerste-orde wilstoestanden dwingt de actor tot deliberatie (je wipt bijvoorbeeld steeds op je stoel heen en weer, dan weer haast op weg naar het strand, dan weer bijna gegrepen door deze tekst, tot je besluit dat je nu toch echt een beslissing moet nemen), en dit proces van deliberatie komt tot een einde door middel van een 'actor-identificerende handeling'. Zo'n handeling bestaat uit het formeren van een tweede-orde wilstoestand die namens de actor als scheidsrechter optreedt (om Frankfurts eigen vergelijking te gebruiken) en het interne conflict beslecht in het voordeel van één bepaalde eerste-orde wilstoestand (je gaat naar het strand).

Natuurlijk staat het de andere eerste-orde wilstoestanden vrij de autoriteit van deze tweede-orde wilstoestand in twijfel te trekken (ook al moeten we dit natuurlijk helemaal niet op zo'n antropomorficerende wijze begrijpen). Als er zo een nieuw intern conflict rijst, kan er behoefte ontstaan aan een derde-orde wilstoestand, die wederom als scheidsrechter de zaak in het voordeel van één van de strijdende partijen beslecht. En dit kan zo doorgaan, maar als het telkens dezelfde partij is die de verliezende partij is, dan zal, zo betoogt Frankfurt, deze partij de lust tot oppositie vergaan. Dit impliceert geen willekeur, maar slechts gezond verstand. En als we even door de antropomorficerende metafoor heen kijken, zullen we de zinnigheid van deze redenering inzien.

Een aardige vergelijking van Frankfurt kan hierbij helpen. Als we een rekensom maken, en ons afvragen of de uitkomst goed is, kunnen we de som nog eens maken. Is het resultaat hetzelfde, dan kunnen we dit resultaat natuurlijk *weer* controleren, door de som een derde keer te maken. Maar op een gegeven moment zal ons gezond verstand zegevieren, en zullen we inzien dat nog weer een controle niet meer uit zal maken. We hebben het blijkbaar

gewoon bij het juiste eind. (Het is natuurlijk ook mogelijk dat we systematisch dezelfde fout maken, maar ook daar komen we niet achter door het nogmaals opnieuw te doen. Ook de fout zal zich dan weer opnieuw voordoen.) Eenzelfde proces zien we in de deliberatie, en waar het feitelijk op neer komt, is dat de deliberatie eindigt in een 'actor-identificerende handeling', in een besluit, in de formatie van een hogere-orde wilstoestand waarmee de actor uitspreekt dat hij zich van harte identificeert met een specifieke eerste-orde wilstoestand.

Een moeilijkheid met deze oplossing van het 'onvolledigheids'-probleem is dat niet duidelijk is hoe we een *decisionistische* interpretatie van de 'actor-identificerende handeling' kunnen voorkomen. Of, om het in termen van de metafoor te zeggen: op grond van welke overwegingen, redenen, motieven, waarde-oordelen of andere criteria-genererende zaken zal de 'scheidsrechter' een intern conflict tussen bepaalde wilstoestanden van de eerste orde kunnen beslechten? Op grond waarvan komt de actor tot de formatie van *deze* tweede-orde wilstoestand met zijn voorkeur voor deze ene eerste-orde wilstoestand? Waarom komt hij niet tot de formatie van een tegenovergestelde tweede-orde toestand die een heel andere eerste-orde wilstoestand prefereert? Het lijkt erop alsof de *ratio* achter deze of gene tweede-orde wilstoestand noodzakelijkerwijs in dit model moet ontbreken, alsof we slechts achteraf in staat zijn te constateren dat het nu eenmaal zo blijkt te zijn dat de actor zich identificeert met *deze* en niet met *die* wilstoestand als was het een vorm van blinde willekeur. In zijn jongste publicaties werkt Frankfurt aan dit probleem waarbij hij een plaats probeert toe te kennen aan de overtuiging dat personen, om personen te zijn, niet kunnen zonder een *contingent welbepaalde wil*. Het feit dat iemand een persoon is, brengt met zich mee dat hij wezenlijk gekenmerkt wordt door bepaalde 'wils-noodzakelijkheden' die nu eenmaal een welbepaalde hiërarchie in zijn eigen wil realiseren. Dat deze noodzakelijkheden gegeven zijn, ervaren we, volgens Frankfurt, niet als dwang. De noodzaak die zich bijvoorbeeld toont in het beroep dat geliefden op ons kunnen doen (denk aan een ziek kind dat, huilend in de nacht, om aandacht vraagt), ervaren

we niet als een externe macht die ons stuurt, maar als de essentie van een onontkoombare reeks gebeurtenissen waarmee wij ons van harte en volledig identificeren. We willen niet anders dan voor ons kind zorgen, als het ons 's nachts nodig heeft.

Natuurlijk heeft Frankfurt met dit voorbeeld op een suggestieve manier gelijk. Toch blijven we hier nog met twee problemen zitten. Want hoe zullen we kunnen begrijpen dat een 'wils-noodzakelijkheid' conceptueel, en niet alleen maar terminologisch of fenomenologisch, verschilt van een fysische noodzakelijkheid? En hoe zullen we kunnen begrijpen dat enerzijds 'wils-noodzakelijkheden' *wel* voldoende grond geven voor de these dat actoren zelfstandig zijn, maar anderzijds fysische noodzakelijkheden *niet* voldoende grond geven voor de these dat bepaalde complexe systemen zelfstandig zijn?

Hoe belangrijk is *zelfstandigheid*?

Naast de twee bovengenoemde problemen is er nog een derde probleem dat in het licht van een verdediging van de aanname dat handelingen verricht worden door actoren zeer gewichtig is: als we de zelfstandigheid van een actor nodig hebben om te begrijpen wat een handeling is, maar de idee van een 'actor-identificerende handeling' nodig hebben om te begrijpen wat die zelfstandigheid behelst, zijn we dan niet op een onacceptabele wijze in een cirkel verstrikt geraakt? Of anders gezegd, het is toch *question-begging* om de notie van een 'actor-identificerende handeling' te gebruiken in een analyse van het begrip 'handeling'?

Een voor de hand liggende reactie is nu te beweren dat het hier niet om een vicieuze cirkel gaat, omdat we niet op zoek zijn naar primitieve noties, maar werken aan een noodzakelijkerwijs holistische articulatie van de conceptuele verbanden tussen een heel cluster van begrippen. Deze reactie is zeker plausibel, maar ze geeft de aanname op die in dit hoofdstuk centraal staat. Ik zal haar in *Hoofdstuk 3* bespreken, als ik het niet meer heb over de aanname dat we handelingen beter begrijpen als we actoren beter begrijpen, maar het hebben over het *sui generis* karakter van het

begrippenkader met behulp waarvan we zinvol kunnen spreken over handelingen als sociale fenomenen.

Een tweede reactie op de problemen waarmee de analyse van de zelfstandigheid van de actor ons opgezadeld heeft, is te beweren dat niet de zelfstandigheid maar de *rationaliteit* van een actor een cruciale rol speelt in de verdediging van de aanname dat handelingen verricht worden door actoren. Het gaat er niet om dat actoren op de één of andere manier behept zijn met zelfstandigheid, waardoor we hun gedrag kunnen begrijpen als het product van hun eigen spontaniteit, hun vermogen initiatief te tonen, maar het gaat er om in te zien dat actoren onder een ander regime vallen dan planeten, amoeben, benzeenmoleculen, thermometers en beuken. Actoren bewegen zich in de 'space of reasons'; voor hen gelden niet alleen, of niet vooral, de wetten van de natuur, maar meer dan dat: voor actoren gelden de wetten van rationele normativiteit. Het cruciale verschil tussen actoren die handelingen verrichten, en entiteiten die gebeurtenissen ondergaan, zit hem in het vermogen van actoren om te delibereren, om op grond van goede redenen en met het oog op goede redenen dit dan wel dat te doen. Zelfstandigheid is belangrijk, zo zouden we kunnen zeggen, maar het is niet een primitieve notie. Zelfstandigheid is zelf het resultaat van ons vermogen goede redenen te waarderen, van ons vermogen praktisch te redeneren. Dat is waar Frankfurts analyse ons uiteindelijk op wijst. Een verdediging van de aanname dat actoren handelingen verrichten, zal daarom gebouwd moeten worden op een analyse van ons vermogen praktisch te redeneren.

2. *Over delibereren, motiveren en rechtvaardigen*

Praktische rationaliteit

De gedachte dat onze rationaliteit een centrale rol speelt in het verklaren van handelingen gaat minstens terug tot het werk van Aristoteles. Diens analyse van het praktische syllogisme speelt nog steeds een rol van betekenis in de hedendaagse analytische filosofie

van het handelen, al wordt daarin minstens een even grote rol gespeeld door de opvattingen van Hume, en eigenlijk ook door die van Kant.

Een aantal termen hebben in deze lange traditie een min of meer stabiele positie veroverd. Op de eerste plaats hebben handelingen, *qua* handelingen, een intentioneel karakter. Een handeling vindt 'opzettelijk' plaats; zij kan alleen maar als handeling beschreven worden door te spreken over de intentie van de actor. Hoe handelingen en intenties zich verder precies tot elkaar verhouden is iets waarover van mening verschild kan worden. Op de tweede plaats is de intentie, of misschien wel de handeling zelf, het resultaat of de conclusie van een praktische redenering. In een dergelijke redenering wordt een hoofdrol gespeeld door een pro-attitude van de actor, dat wil zeggen door een mentale toestand, doorgaans een wens, die de actor in een bepaalde situatie een doel geeft. Ook een hoofdrol in een praktische redenering wordt gespeeld door een concrete opvatting of reeks van opvattingen van de actor. Deze concrete opvattingen bemiddelen als het ware tussen de situatie waarin de actor zich bevindt en de situatie waarin de actor zich zou willen bevinden, door informatie over mogelijke middelen ter beschikking te stellen aan de actor. Tenslotte wordt er in een praktische redenering een grote rol gespeeld door rationele principes die de normativiteit die inherent is aan iedere redenering zichtbaar maken. De functie van deze principes bestaat daarin goede gevolgtrekkingen te onderscheiden van slechte, en, bijgevolg, een rationele ordening van handelingsalternatieven mogelijk te maken. Het resultaat van dit alles is dat een handeling gezien kan worden als een rationele respons op een praktisch probleem.

Laat ik een voorbeeld geven. Stel ik ben een vrachtwagenchauffeur, onderweg van Nederland naar Italië. Met een gezellig deuntje op de radio bevind ik me ergens in Oostenrijk op de linker rijbaan, halverwege een inhaalmanoeuvre die vele kilometers in beslag neemt. Plotseling zie ik in grote letters boven de weg staan dat verkeer voor de Brennerpas over 600 meter rechtsaf moet slaan. Blijkbaar heb ik, luid meeneuriënd, een eerder verkeersbord

over het hoofd gezien. In paniek realiseer ik me dat ik die Brennerpas op moet. Wat nu te doen?

De presentatie van het voorbeeld, vanuit het perspectief van de eerste persoon, brengt met zich mee dat we de noodzaak om te delibereren niet kunnen missen. Ze stelt zich overweldigend. Ik kan niet anders dan delibereren, d.w.z. praktisch redenerend de intentie vormen waarmee ik zal doen wat ik zal besluiten te doen. Maar ook vanuit het derde persoons perspectief van een mede-weggebruiker, een verkeersagent, of een wetenschapper die een onderzoek in stelt naar de ware toedracht (zoals dat heet) zal wat plaatsvindt beschouwd worden vanuit de vooronderstelling dat het hier om *intentioneel* gedrag gaat. Iemand (ik!) doet iets op grond van een praktische redenering waarin een grote rol gespeeld wordt door een pro-attitude (ik moet naar Italië!), een reeks opvattingen (mijn 'diesel' bevindt zich op de linker inhaalstrook, de vrachtwagen rechts naast mij rijdt bijna even hard als ik, ik heb nog 600 meter, ik weet niet zeker of er iets vlak achter me zit, het lijkt er op alsof er ruimte is achter de vrachtwagen die ik aan het inhalen ben, etc.), en rationele principes (men kan niet naar rechts als men aan het inhalen is, vrachtwagens kunnen niet zowel recht-door als rechtsaf, etc.). De overwegingen die welhaast met de snelheid van het licht door me heen gaan, en die gestructureerd zouden kunnen worden in de vorm van een reeks premissen, zullen tot één of andere conclusie leiden. Het is goed denkbaar dat ik, omwille van de veiligheid, besluit mijn inhaalmanoeuvre te voltooien, dan de eerstvolgende afslag te nemen om in tegengestelde richting terug te keren naar de afslag die ik nu helaas mis.

De these die ik in deze paragraaf nader wil bekijken luidt dat iedere handeling gereconstrueerd kan worden als het resultaat van een praktische redenering. Deze stelling geeft een bepaalde interpretatie van de aanname dat handelingen verricht worden door actoren, doordat ze handelingen beschouwt als het resultaat van een proces dat praktisch redeneren heet en zich afspeelt in actoren. Niet de zelfstandigheid van actoren is cruciaal, volgens deze interpretatie, maar het feit dat actoren *redelijke* wezens zijn. In wat

volgt zal ik eerst wat dieper ingaan op de vier kernbegrippen (principes, opvattingen, pro-attitudes, intenties) die een rol spelen in het praktisch redeneren, om vervolgens de cruciale kwestie van de functie van praktische redeneringen in een filosofie van het handelen nader onder de loep te nemen.

Rationele principes en opvattingen

Er zijn overeenkomsten en verschillen tussen dat wat ik rationele principes noem en dat wat ik concrete opvattingen noem. De cruciale overeenkomst is dat de mentale toestanden van een actor waarin opvattingen en principes gerepresenteerd worden een zogenaamde 'mind-to-world direction of fit' hebben. Daarmee wordt bedoeld dat het hier om mentale toestanden met een propositionele inhoud gaat, met betrekking waartoe de actor die deze mentale toestanden heeft, een overeenstemming zoekt met de wereld, en wel zodanig dat de actor geneigd zal zijn de inhoud van deze mentale toestanden aan te passen aan de werkelijke stand van zaken in de wereld. Wie een opvatting heeft, zal geneigd zijn de inhoud van deze opvatting af te stemmen op, en aan te passen aan de stand van zaken waarover deze opvatting gaat. Mijn opvatting dat ik nog 600 meter tot aan de afslag heb, is een opvatting die als opvatting getrouwheid aan de actuele stand van zaken vooronderstelt. Daarin verschillen deze toestanden van de pro-attitudes die een zogenaamde 'world-to-mind direction of fit' hebben. Dat wil zeggen: wensen, bijvoorbeeld, zijn mentale toestanden met een propositionele inhoud met betrekking waartoe de actor die deze toestanden heeft, een overeenstemming zoekt met de wereld, maar dan zodanig dat de actor geneigd zal zijn de wereld aan te passen aan de inhoud van deze mentale toestanden. Als ik de wens heb de afslag naar de Brennerpas te nemen, zal ik, al handelend, proberen de wereld zo te veranderen dat zij in overeenstemming is met de inhoud van mijn wens: ik zal dan dus inderdaad proberen die afslag daadwerkelijk te nemen.

Het fundamentele verschil tussen rationele principes en concrete opvattingen is een verschil in algemeenheid. Dit verschil

werd al onderkend door Aristoteles, volgens wie een praktisch syllogisme opgebouwd is uit een zogenaamde 'major' die een universeel karakter heeft ("Alles wat zoet is moet geproefd worden") en een zogenaamde 'minor' die een particulier karakter heeft ("Dit snoepje is zoet"). Juist de combinatie van zo'n algemene en zo'n concrete overtuiging bracht volgens Aristoteles op zuiver theoretische gronden een aanzet tot een concrete handeling met zich mee ("Proef dit snoepje").

Hoewel een Aristoteliaanse 'major' in het geval van het praktisch redeneren een duidelijk prescriptief karakter heeft, is dit niet een *conditio sine qua non* voor wat we rationele principes kunnen noemen. In de vorige paragraaf gebruikte ik als voorbeeld van een rationeel principe de opvatting dat vrachtwagens niet tegelijkertijd rechtdoor en rechtsaf kunnen. Dit is slechts impliciet prescriptief, op grond van de aanname dat je niet behoort te doen wat je niet *kunt* doen. In beide gevallen kan er over rationele principes gesproken worden, omdat er in beide gevallen sprake is van een algemeen geldende regel. De regel hoeft niet expliciet prescriptief te zijn, omdat het niet alleen over gedragsregels gaat. Maar de regel zegt wel dat *Y* geldt in ieder geval dat *X*, waarbij de verbinding van *X* naar *Y* een verbinding is die gebaseerd is op de betekenis van *Y* en *X*, en niet op een of andere fysische (causale) relatie die tussen *X* en *Y* bestaat.

Net als de concrete opvattingen kunnen de rationele principes alleen een rol spelen in de deliberaties van een actor als de actor de normatieve geldingskracht van deze principes op de één of andere manier kent. Het is gebruikelijk dit kennen op te vatten in termen van het kunnen toeschrijven aan de actor van een propositionele attitude met als inhoud een verwijzing naar, of representatie van het principe in kwestie. Zo beschouwd zijn de rationele principes gewoon opvattingen, net als dat wat ik de concrete opvattingen heb genoemd. In principe is hier niets op tegen; het sluit aan bij de Humeaanse visie dat redenen combinaties zijn van opvattingen en pro-attitudes.

Rationele principes en concrete opvattingen hebben echter heel verschillende functies.

De functie van de concrete opvattingen is het operationeel maken van de huidige situatie van de actor op het niveau van het praktisch redeneren. De actor is een belichaamd, gesitueerd wezen; hij bevindt zich ergens op een bepaalde manier. De informatie over zijn huidige positie in de context waarin hij zich bevindt is voor de actor die praktisch wil delibereren van groot belang.

De functie van de rationele principes is een andere: ze geven informatie over de betekenisvolle interne structuur van de context waarin de actor zich bevindt – niet alleen de spatio-temporele context, maar ook allerlei andere rationele contexten. Denk hierbij aan de structurering van bijvoorbeeld de alledaagse wereld van vrachtwagenbestuurders, waarin er sprake is van een groot aantal meer of minder duidelijk zichtbare ordeningen: de technologische, economische, sociale, esthetische, morele, emotionele, etc. Dat men niet naar rechts kan als men aan het inhalen is, is zowel meer als minder dan een regel die onze spatio-temporele context gestalte geeft, zoals het ook zowel meer als minder is dan een regel die onze morele of technologische context gestalte geeft.

In dit verband kunnen we, als ik me een uitstapje mag permitteren, een, zij het wat kleine, plaats geven aan het enorme debat over (interculturele, of universele) rationaliteit en relativisme. Dit debat, dat overigens in eerste instantie in de wetenschapsfilosofie gevoerd wordt, gaat over de vraag of er uiteindelijk slechts één fundamentele ordening relevant is voor al het menselijk handelen; een ordening waartoe al die andere, meer oppervlakkige ordeningen gereduceerd kunnen worden; een ordening die gearticuleerd kan worden in een beperkt aantal universele rationele principes die altijd een hoofdrol behoren te spelen in iedere praktische redenering die ten grondslag ligt aan zinvol menselijk handelen.

Vanuit het perspectief van een handelingsfilosoof is de importantie die nogal eens aan dit debat kleeft schromelijk overdreven. Natuurlijk is het belangrijk dat wetenschappers erg kritisch zijn over het als geldig accepteren van een of ander rationeel principe, en natuurlijk is het begrijpelijk dat wetenschappers heel veel belang hechten aan het integreren van alle principes die ze willen

of wel moeten accepteren. Immers, aangezien deze principes informatie geven over de betekenisvolle interne structuur van de context waarin wij ons bevinden, is het van groot belang als we zouden begrijpen hoe de diverse principes zich tot elkaar verhouden. Maar het is helemaal niet verstandig een simpele eenheid van de wetenschap als onontkoombaar axioma te introduceren. En het is ook helemaal niet nodig vergaande relativistische conclusies te trekken uit het feit dat we in allerlei verschillende omstandigheden te maken hebben met allerlei onvergelykbare (incommensurabele, is het technische woord) rationele principes. Natuurlijk roept die incommensurabiliteit vragen op; natuurlijk weten we soms niet welk principe de voorkeur verdient; en natuurlijk zijn er overtuigende argumenten voor de these dat het accepteren van sommige principes vooraf gaat aan de mogelijkheid om überhaupt te kunnen redeneren. Maar dat wil nog helemaal niet zeggen dat er verschillende rationaliteiten zouden zijn. Wat zou dat moeten betekenen? En het wil natuurlijk ook nog helemaal niet zeggen dat er zoveel werkelijkheden zijn als dat er onvergelykbare clusters rationele principes zijn. Want, nogmaals, wat zou dat moeten betekenen?

Ik ben hier niet in de gelegenheid me in dit debat te begeven, maar beide partijen lijken me te snel geneigd te accepteren dat er samenhangende clusters rationele principes zijn die met betrekking tot een bepaald domein omvattend, volledig, en exclusief zijn. Dit lijkt me, maar dat is dan hier echt het laatste woord daarover, een onacceptabele extrapolatie uit een te gemakkelijke interpretatie van Wittgensteins ideeën over taalspelen en levensvormen. De alledaagse praktijk speelt zich eigenlijk nooit af in gesloten domeinen. De context waarin we ons bevinden is er altijd één die bol staat van conceptuele 'gaten' en 'incongruenties', van het impliciet, en in de achtergrond, door elkaar heen lopen van allerlei verschillende 'levensvormen'.

De menigvuldigheid en incommensurabiliteit van rationele principes belemmert overigens geen enkele actor in het praktisch redeneren. Wil er sprake kunnen zijn van praktisch redeneren dan is het immers niet nodig dat de kennis die de actor van zijn

omgeving heeft objectief en juist is. Er hoeft, om het met een voorbeeld te zeggen, helemaal geen sprake te zijn van een context die objectief gestalte heeft gekregen in termen van de regel dat vrachtwagens kunnen vliegen om mij als bestuurder *praktisch redenerend* te laten besluiten dat ik vol gas moet geven en tegelijkertijd ‘kort-kort-lang’ moet toeteren, als ik onterecht de opvatting heb dat vrachtwagens gaan vliegen als hun bestuurder zo gas geeft en toetert. Het zijn niet de feiten en de werkelijk geïnstitutionaliseerde regels die een rol spelen in het praktisch redeneren, maar de (subjectieve) representatie van die feiten en die regels in de concrete en algemene opvattingen van de individuele actor.

Natuurlijk is het wel zo dat *correct* praktisch redeneren (en succesvol handelen) altijd ook vraagt om het kritisch doorlichten van de opvattingen en pro-attitudes van de actor. Ik kom hier nog op terug als ik in het kader van het onderscheid tussen motiveren en rechtvaardigen het onderscheid zal bespreken tussen een praktische redenering en een *goede* praktische redenering.² Wat hier slechts besproken is, is dat handelingen het resultaat van praktische redeneringen kunnen zijn, ook als deze redeneringen feitelijk ondeugdelijk zijn omdat ze uitgevoerd worden door actoren die niet beschikken over correcte opvattingen inzake hun concrete situatie, of dit gebrek aan kennis nu de coördinaten van hun positie betreft of de regels die inhoud geven aan deze coördinaten. Daarbij is voorondersteld dat het succes van een handeling niet op een simpele, rechtstreekse wijze gekoppeld is aan de kwaliteit van de er aan ten grondslag liggende redenering. Mijn vrachtwagen zal heus de lucht niet invliegen als ik vol gas geef en ‘kort-kort-lang’ toeter.

Pro-attitudes

Praktische redeneringen en de handelingen die daaruit voortvloeien spelen zich af tegen de achtergrond van het voorhanden zijn van een doel of verlangen van de actor. Deze pro-attitudes

² Zie pp. 50-54.

hebben de juiste 'world-to-mind direction of fit' om actoren tot handelingen aan te zetten. De vrachtwagenchauffeur zal al die kilometers niet hebben kunnen rijden, en zal zich niet genoodzaakt hebben gezien te delibereren over wat nu te doen, als hij al die tijd al niet op de achtergrond de motiverende kracht had ervaren van de pro-attitude dat hij naar Italië wilde gaan.

Pro-attitudes, zoals ze genoemd worden, spelen niet alleen in een praktische redenering een grote rol, maar ook voorafgaand aan praktische redeneringen. Het is vanwege de gegeven wens naar Italië te gaan dat de constatering dat men over 600 meter moet afslaan voor de Brennerpas mij dwingt tot deliberatie. Dat is één van de redenen voor een Humeaanse visie op de (praktische) rede: dat ze nooit meer kan zijn dan een slaaf van de passies, dat ze nooit greep kan krijgen op de verlangens die stomweg *al gegeven* zijn, en in dienst waarvan een deliberator op zoek zal gaan naar het beste middel voor het bereiken van het door deze passies gestelde doel.

In de hedendaagse analytische filosofie van het handelen is er aandacht voor drie kwesties betreffende de status en functie van de pro-attitudes in het delibereren.

De eerste kwestie betreft de al dan niet unieke status van de pro-attitudes als motiverende toestanden. Er wordt dan doorgaans gesproken over het debat tussen internalisme en externalisme in de motivatietheorie. (Let op: deze termen worden ook gebruikt in de *philosophy of mind*, maar betekenen daar iets heel anders.) De internalist beweert dat het hebben van een opvatting op zichzelf al een goede reden kan genereren die motiverende kracht heeft. De externalist beweert dat het hebben van een opvatting op zichzelf nooit een goede reden genereert, maar altijd alleen maar voor zover de actor naast deze opvatting een corresponderende pro-attitude heeft. De externalist beweert dat *alleen* de pro-attitudes een inherente motiverende kracht hebben. Daarin zijn zij uniek. En de externalist wordt zo genoemd, omdat hij beweert dat het overtuigen van een ander door het geven van alle goede redenen voor het doen van X nooit voldoende kan zijn zonder een beroep te doen op nog een extra, externe factor: namelijk het aanwezig zijn

van een pro-attitude bij de toehoorder. Dat wil zeggen: als men mij vertelt dat ik naar Italië moet, dat de afslag naar de Brennerpas al over 600 meter is, dat ik mijn inhaalmanoeuvre dus maar snel moet afbreken, enz. enz., dan stelt de externalist dat zelfs als ik de ander geloof en ik alle opvattingen heb die mij zouden kunnen doen besluiten te doen wat het beste is, ik nog helemaal niet de neiging zal hebben het beste te doen, tenzij ik de pro-attitude heb, of krijg, dat ik naar Italië *wil* gaan, of het beste *wil* doen. De internalisme-externalisme kwestie is vooral belangrijk voor het moreel redeneren, omdat we juist in de morele sfeer soms graag zouden willen dat we anderen zouden kunnen motiveren door ze voldoende te informeren. En meer dan dat, we hebben er eigenlijk niet alleen belang bij te denken dat informatie ons kan motiveren, maar we hebben vanuit een voor de hand liggende interpretatie van waardeoordelen vaak ook een soort fenomenologisch geïnspireerde overtuiging *dat* kennis van wat het beste is om te doen, ons ook inderdaad motiveert dat te gaan doen. Dat wil zeggen, een of ander internalisme (dat opvattingen op zichzelf goede redenen kunnen zijn) lijkt plausibel, terwijl tegelijkertijd een of ander externalisme (dat kennis van het goede nooit noodzakelijkerwijs mensen aanzet tot het willen doen van het goede) even plausibel lijkt.

De tweede kwestie betreft het al dan niet redelijke karakter van de pro-attitudes. Men zou kunnen denken dat deze kwestie geïntroduceerd zou kunnen worden door de vraag te stellen of gevoel en verstand al dan niet drastisch van elkaar verschillen en nooit in termen van elkaar begrepen kunnen worden. De suggestie zou dan kunnen zijn dat onze wensen voortkomen uit ons gevoel en niet kritisch doorgelicht zouden kunnen worden. De tegenstelling tussen gevoel en verstand is echter vooral een ondoorzichtige die gelukkig niet veel vaste voet aan de grond heeft kunnen krijgen in de analytische filosofie. De vraag is daar vooral of Hume gelijk heeft met zijn these dat de pro-attitudes *non-rationeel* zijn, omdat ze altijd als voorwaarden vooraf gaan aan deliberaties en zelf nooit onderworpen kunnen worden aan een kritische doorlichting. Zoals ik de vraag formuleer, blijkt al dat er sprake is van minstens twee beweringen waarvan niet onmiddellijk duidelijk is dat ze

onlosmakelijk met elkaar verbonden zijn. Het is, met andere woorden, één zaak om te beweren dat de pro-attitudes altijd als voorwaarden vooraf gaan aan deliberaties, en een heel andere zaak om te beweren dat de pro-attitudes zelf nooit onderworpen kunnen worden aan een kritische doorlichting. De eerste bewering kan een tamelijk onschuldige zijn: een actor kan alleen delibereren als hij bepaalde wensen heeft, waaronder de wens door middel van deliberatie te komen tot inzicht in wat het beste is om te doen. Deze wens, om slechts te willen doen waar men het meest reden toe heeft, is volgens David Velleman (1989) precies dat waarin een actor *actor is*. Daarom betoogt Velleman ook dat de these dat handelingen verricht worden door actoren niets meer en niets minder behelst dan de these dat handelingen het resultaat zijn van praktische deliberaties. Immers, een actor is volgens Velleman *als actor* niets meer en niets minder dan de wens dat te doen waar men het meest reden toe heeft. De eerste bewering introduceert daarmee in ieder geval één wens die noodzakelijk vooraf gaat aan het praktisch redeneren, maar aangezien dit de wens is te delibereren, komt het wellicht een beetje vreemd over dit een non-rationele wens te noemen, ook al is duidelijk dat deze wens zelf niet door middel van een praktische redenering kritisch doorgelicht kan worden. Voor zover de eerste bewering ook impliceert dat andere wensen aan het praktisch redeneren vooraf gaan, wil dit vervolgens *niet* zeggen dat deze wensen daarom ook niet kritisch doorgelicht kunnen worden. Er lijken minstens twee kritische oordelen over deze verdere wensen mogelijk: (1) het is rationeel slechts die toestanden te wensen die gekenmerkt worden door, en voor zover ze gekenmerkt worden door, gewenste eigenschappen; en (2) het is rationeel slechts die eigenschappen te wensen die het waard zijn gewenst te worden. Natuurlijk kan het eerste kritische oordeel herleid worden tot het tweede, en natuurlijk roept het tweede kritische oordeel om een niet onproblematisch waardenrealisme, maar ook al is het moeilijk, toch lijkt het voor zich te spreken dat men zijn wensen kritisch doorlicht, of althans kritisch *kan* doorlichten. We zagen dit al in de vorige paragraaf waar dit doorlichten betrekking had op de vraag of een bepaalde wens wenselijk is in het licht van

de *eigen* wil van de actor.³ En we zien het nu weer met betrekking tot de vraag of een bepaalde wens wenselijk is in het licht van wat rationeel, c.q. waardevol is. Aan het eind van deze paragraaf zal ik nogmaals even op deze kwestie terugkomen.⁴

De derde kwestie betreft de mate waarin de propositionele inhoud van een wens een expliciet en gearticuleerd karakter heeft. In mindere mate speelt deze kwestie ook voor de opvattingen en rationele principes. Ik zeg 'in mindere mate' omdat het op een bepaalde manier onproblematisch lijkt dat men de opvattingen en rationele principes die men een rol laat spelen in deliberaties, voor zover ze die rol spelen, door middel van reflexieve tweede-orde mentale toestanden introduceert alsof ze expliciete, gearticuleerde proposities vertegenwoordigen. Men kan immers alleen maar delibereren door het op elkaar betrekken van proposities. Maar het is niet zo heel duidelijk of alle pro-attitudes op deze zelfde manier geïntroduceerd kunnen worden als behept met een expliciete, gearticuleerde propositionele inhoud. Wat in deze kwestie aan de orde wordt gesteld is het wellicht over-generationaliseerde karakter van wat mensen beweegt en het miskennen van zogenaamde sub-intentionele processen. Wat ter discussie wordt gesteld is in feite de these dat *iedere* handeling beschouwd kan worden als het resultaat van een praktische redenering. De moeilijkheid van deze these is dat, als men haar accepteert, men een grote klasse van gebeurtenissen (handelingen?) creëert die zich weliswaar voltrekken door middel van menselijk gedrag, maar niet verklaard of begrepen kunnen worden in termen van het praktisch redeneren. Men kan hierbij denken aan irrationeel gedrag als wilswakte en zelfbedrog. Ik kom hier aan het eind van deze paragraaf nog even op terug.⁵

Intenties

Intenties spelen een grote rol in de hedendaagse filosofie van het handelen. Ze spelen die rol vooral omdat het al sinds Anscombe

³ Zie de bespreking onder het kopje "Waarde-oordelen", pp. 21-24.

⁴ Zie de bespreking onder het kopje "Motiveren en rechtvaardigen", pp. 52-54.

⁵ Onder het kopje "Motiveren en rechtvaardigen", pp. 48-55.

gemeengoed is te denken dat handelingen alleen maar als handelingen begrepen kunnen worden door het te hebben over de intentie van de actor. Anders gezegd: handelingen zijn slechts wat ze zijn *onder een bepaalde beschrijving*, nl. de beschrijving waarin een hoofdrol wordt gespeeld door de intentie van de actor. Natuurlijk levert dit het probleem op dat het kan lijken alsof alleen de actor het voor het zeggen heeft als we ons zorgen maken om wat de juiste beschrijving van een handeling is. Immers, als iedere beschrijving een verwijzing naar de intentie van de actor bevat, en als het zo is dat alleen de actor op een geprivilegieerde wijze toegang heeft tot zijn intenties, dan kan alleen de actor de juistheid van een beschrijving vaststellen. Maar let wel: dit probleem dreigt alleen als we de geprivilegieerde toegang van een actor tot zijn intenties heel strikt opvatten, en er lijken geen doorslaggevende argumenten te zijn voor zo'n strikte interpretatie van de autoriteit van een actor. Maar laat ik eerst eens een voorbeeld geven om deze eerste rol van intenties in de filosofie van het handelen een beetje tot de verbeelding (!) te laten spreken.

Stel ik zet 's ochtends vroeg altijd thee. Dat wil zeggen, ik vul de waterkoker en schakel de stroom in. Maar stel dat ik die ochtend de waterkoker niet zie staan waar hij altijd staat, en dat ik hem ook na enig rondkijken nog niet zie. Ik besluit de fluitketel te gebruiken, vul hem met water, zet hem op het fornuis, en steek het vuur aan. Stel nu, dat door een wonderbaarlijke samenloop van omstandigheden het opendraaien van de gaspit op die kwade dag een ongeluk veroorzaakt in een nabijgelegen gascentrale. (Stel even dat deskundigen de technische mogelijkheid hiervan kunnen uitleggen.) De honderd meter hoge steekvlam die het effect hiervan is, wordt ergens in de ruimte opgemerkt door een vreemd buitenaards volkje dat juist in een ruimteschip ons zonnestelsel passeert. Deze vreemde wezens communiceren met elkaar door gebruik te maken van hun vermogen temperatuurswisselingen te veroorzaken. In hun taal staat de steekvlam die ik veroorzaakt heb voor een zeer hartelijke welkomstgroet. Het vervolg laat zich raden: de *Temperaturi*, zoals wij ze zullen noemen, landen op aarde en slagen er in enkele decennia in onze aarde te koloniseren

en ons allemaal aan het werk te zetten in de productie van een soort gasaanstekers.

Wat heb ik gedaan? Dat zal ik, en niet alleen ik, me geregeld afvragen. Heb ik thee gezet, die ochtend? Heb ik het gas open gedraaid? Heb ik een gascentrale laten ontploffen? Heb ik een nieuw tijdperk in de geschiedenis van de mensheid ingeluid? Heb ik de Temperaturi uitgenodigd? Hoeveel handelingen heb ik eigenlijk verricht?

Deze vragen zijn variaties op het volgende tweetal: (1) Wat is er gebeurd? (2) Welke handeling heb ik verricht? Het is niet duidelijk of er vanuit een neutraal, metafysisch standpunt een groot verschil is tussen deze twee vragen, hoewel er, ook vanuit dit standpunt, allerlei theoretische complicaties kunnen bestaan die te maken hebben met het integreren van alle mogelijke correcte beschrijvingen van wat er gebeurde. Maar vanuit het standpunt van mijzelf, mijn verwanten en alle andere *betrokkenen* – d.w.z. vanuit het perspectief van bij het gebeuren betrokken *actoren* – is er een zeer groot, cruciaal verschil tussen beide vragen. En dit verschil heeft te maken met de rol van dat woordje ‘ik’ in de tweede vraag, dat een actor introduceert die verantwoordelijk gesteld kan worden, omdat een actor verantwoordelijkheid draagt en behoort te nemen voor zijn handelingen. En het is met betrekking tot die verantwoordelijkheid dat op dit punt in mijn betoog gewezen kan worden op de cruciale rol van de intenties van de actor, die het resultaat zijn van een proces van deliberatie dat zich heeft afgespeeld in de actor.

Natuurlijk heb ik de Temperaturi *niet* uitgenodigd; dat is een onbedoeld effect van wat ik duidelijk *wel* heb gedaan: thee zetten. En natuurlijk heb ik de gascentrale niet laten ontploffen; ook dat was een onbedoeld effect van wat ik *wel* heb gedaan: het gas open gedraaid. Verschillende soorten intenties en hun onderlinge relaties komen hier in beeld. Sommige intenties zijn helemaal niet het resultaat van een proces van deliberatie. De intentie om thee te zetten, bijvoorbeeld, is vooral het resultaat van gewoontevorming. ’s Ochtends vroeg word ik nu eenmaal gewoon in die geestestoestand wakker. De intentie om het gas open te draaien is ech-

ter wel een intentie die ik vanwege een praktische redenering heb. Omdat ik thee wilde zetten, maar de waterkoker niet kon vinden, besloot ik de fluitketel te gebruiken, en dat bracht met zich mee dat ik het gas open moest draaien. Verdergaande onderscheidingen kunnen gemaakt worden. Sommige intenties gaan als voldoende aanleiding aan een handeling vooraf, terwijl anderen zich precies realiseren *in* de zich voltrekkende handeling. De intentie thee te willen zetten met de fluitketel is een voorafgaande intentie, die het resultaat is van een expliciete praktische redenering. De intentie het gas open te draaien is daarentegen een intentie die ik doorgaans niet eerst voorafgaand aan de handeling hoeft te vormen, maar die zich toont in de handeling. Dit is zeker het geval als het een ‘basic action’ betreft, d.w.z. een handeling waarvan ik weet *dat* ik die kan verrichten zonder dat ik weet *hoe* ik dat doe, anders dan door het gewoon te doen – zoals bijvoorbeeld in de handeling die bestaat uit het maken van een draaibeweging met de hand waarbij de hand een gasknop beet heeft.⁶ Het belang van dit laatste onderscheid is niet dat sommige intenties zich slechts zouden *kunnen* tonen in het verrichten van de handeling, maar dat er vaak situaties zijn waarin de intenties zich nu eenmaal slechts tonen in de handeling, omdat het handelingen betreft die zo routineus zijn dat de actor er niet expliciet over na hoeft te denken.

Het zou kunnen dat er nog andere intenties in het spel zijn, de intenties die ik al dan niet opzettelijk verborgen tracht te houden. Was het immers wel zo dat ik slechts thee wilde zetten? Zou het niet kunnen dat ik op een geniale manier een uitweg heb gevonden voor mijn boosheid op mijn buurman, die directeur is van de gascentrale? Stel dat ik wist dat het opendraaien van het gas dit rampzalige gevolg voor de gascentrale zou hebben (dit te weten zou al genoeg zijn, ook zonder zelf betrokken geweest te zijn bij het knoeien aan de gasleidingen). Stel dat ik zelf de waterkoker heb zoekgemaakt, om het gebruik van het gas op ongecompliceerde

⁶ Zie ook de discussies onder de kopjes “Het hiërarchisch model van zelfstandigheid” en “Motiveren en rechtvaardigen”, resp. p. 18 en pp. 48-50.

wijze te rechtvaardigen. Dat zou mij een heel andere intentie hebben gegeven. Nog steeds zou ik de Temperaturi niet hebben uitgenodigd (dat blijft een onbedoeld effect – het is *te* bizar ook daarachter nog een complot te zoeken), maar ik zou *wel* de gascentrale hebben laten ontploffen, omdat ik deed wat ik deed op grond van deze intentie.

Een manier om vast te kunnen stellen welke de intentie is geweest die ten grondslag heeft gelegen aan het plaatsvinden van een bepaalde gebeurtenis (en deze dus te kunnen beschrijven als een bepaalde handeling), is de zogenaamde ‘counterfactual test’. Stel dat ik beweer dat een bepaald effect onbedoeld was. Dan moet het zo zijn dat ik het had voorkomen als ik het had kunnen voorzien, en als ik in omstandigheden had verkeerd waarin ik dat wat ik *wel* wilde doen zou hebben *kunnen* doen zonder dit effect te veroorzaken. Als de waterkoker op zijn plaats had gestaan, dan zou ik die gebruikt hebben als ik had kunnen voorzien dat het opendraaien van het gas desastreuze gevolgen zou hebben voor de gascentrale, en als het mij inderdaad niet ging om het vernietigen van de gascentrale. En ook: als er niet geknoeid was met de gasleidingen, en het mij juist *wel* ging om het vernietigen van de gascentrale, dan had ik gezocht naar manieren om mijn doel te bereiken (het produceren en plaatsen van een bom, bijvoorbeeld).

Deze bespreking van het kenbaar zijn van de werkelijk aan de handeling ten grondslag liggende intentie, brengt mij bij de belangrijke, gerelateerde kwestie van wat een gepaste houding is tegenover voorziene maar onbedoelde effecten van wat wij doen. Deze kwestie is belangrijk, omdat de vraag naar de werkelijk beslissende intentie altijd een vraag is die gesteld wordt vanuit het perspectief van de bij de gebeurtenis betrokken actoren, en daarom altijd de kwestie van verantwoordelijkheid introduceert. Als ik onbedoelde effecten voorzie (het gebruik van de fluitketel brengt bijvoorbeeld met zich mee dat ik deze verhit waardoor de vetkrijttekening die mijn zoontje er op maakte, zal smelten) – als ik onbedoelde effecten voorzie, maar geen mogelijkheid zie ze te vermijden, zal ik op gepaste wijze moeten duidelijk maken dat ik de verantwoordelijkheid neem voor deze effecten. Ook al verricht

ik niet de handeling ‘het verwijderen van een vetkrijttekening van de fluitketel’, toch zal ik mijn verantwoordelijkheid voor dit effect moeten thematiseren door haar te nemen, als ik weet dat wat ik *wel* doe, ‘het zetten van thee met water dat in de fluitketel wordt opgewarmd’, dit voorziene maar onbedoelde effect heeft. Hoeveel verantwoordelijkheid hier genomen *behoort* te worden, en voor hoeveel van de voorziene maar onbedoelde effecten, is een subtiel kwestie waar ik in *Hoofdstuk 3* nog wat meer over te zeggen zal hebben. Eveneens geldt dit voor de effecten die ik niet voorzag, maar had kunnen voorzien, en wellicht had moeten voorzien, als ik attenter was geweest – wat ik wellicht had moeten zijn.

Hiermee introduceer ik al het thema dat ik in de volgende, laatste paragraaf van dit hoofdstuk aan de orde wil stellen, en dat een grote rol zal spelen in *Hoofdstuk 3*, waarin ik de these zal bespreken dat handelingen sociale fenomenen zijn. Dit is het thema van de onontkoombaar *normatieve* context waarin we over handelingen als handelingen kunnen spreken.

Motiveren en rechtvaardigen

Nu we een idee hebben gekregen van de begrippen die een hoofdrol spelen in een praktische redenering, wordt het tijd ons eens wat nauwkeuriger af te vragen wat we nu eigenlijk precies onder een deliberatie verstaan. Wat *is* een praktische redenering? Wat *doet* een praktische redenering?

Tot nu toe heb ik over deliberaties gesproken alsof het rede-neerprocessen zijn die zich afspelen in de actor. Ik heb hierbij dit drietal woorden ‘zich afspelen in’ gebruikt, omdat ik wilde ontsnappen aan het verwijt dat ik pas over deliberaties zou kunnen spreken als ik eerst een succesvolle rekenschap zou hebben kunnen geven van de zelfstandigheid van een actor. Dat verwijt zou ik immers onmiddellijk over me afgeroepen hebben, als ik over deliberaties had gesproken als bedoelde ik er een bepaald soort handelingen mee, redeneringen die door actoren *uitgevoerd* kunnen worden. Natuurlijk zou ik dit over deliberaties willen zeggen, maar dan wel slechts in die zin dat een deliberatie een ‘*basic action*’

is, een 'handeling' waarvan de actor weet *dat* hij die kan verrichten zonder te weten *hoe* hij dat doet, anders dan door het gewoon te doen. In die zin lijkt een deliberatie op een elementaire, simpele lichaamsbeweging, zoals het opheffen van een arm, waarvan ik weet *dat* ik dat kan, maar niet weet *hoe* ik dat kan, of doe, anders dan door het gewoon te doen.

Een dergelijk primitieve notie van 'deliberatie' hebben we nodig als we met behulp van dit begrip willen kunnen uitleggen wat handelingen zijn. Zo'n uitleg kan immers niet, op straffe van een kwalijke circulariteit, beginnen met de introductie van het begrip 'deliberatie' als we, om dat begrip uit te kunnen leggen, weer een beroep zouden moeten doen op het begrip 'handeling'. Dit meende ik te kunnen ondervangen door over deliberaties te spreken als redeneerprocessen die *zich afspelen in* de actor. Dit wil overigens helemaal niet zeggen, zoals ik in deze paragraaf zal aantonen, dat we hier te maken hebben met een ongepaste reductie van het rationele naar het mechanische niveau. Dit zou slechts dan zo zijn, als de regels die een deliberatie reguleren zelf slechts strikt mechanische regels zouden zijn, en delibereren slechts een kwestie is van 'dom' of 'blind' motiveren, en niet een kwestie van rechtvaardigen.

Over dit onderscheid tussen motiveren en rechtvaardigen zal ik zo meer zeggen, maar als opstapje daartoe eerst nog even dit. Het woordje 'hoe' in de opmerking "dat ik niet weet *hoe* ik mijn arm op hef, anders dan door het gewoon te doen" is dubbelzinnig. Het kan enerzijds slaan op het *middel* dat ik aanwend om het opheffen van mijn arm te realiseren; het kan anderzijds echter slaan op de *wijze waarop* ik mijn arm ophef, dat wil zeggen, op de 'kwaliteit' van mijn handeling als een opheffing van mijn arm. De notie van een 'basic action' slaat op een handeling die niet verricht kan worden door het aanwenden van een middel daartoe. In die zin is er geen antwoord op de vraag naar het 'hoe'. Dit wil echter niet zeggen dat een 'basic action' een handeling is waarbij ik geen invloed zou kunnen uitoefenen op de wijze waarop ik haar verricht. Ik kan mijn arm op allerlei manieren opheffen: langzaam, voorzichtig, wild, beredeneerd, uit de losse pols, geconcentreerd, etc. Hoewel

ik dus niet weet *hoe* ik het voor elkaar krijg, kan ik een 'basic action' op allerlei manieren verrichten en daarbij trachten de regels te volgen die van toepassing zijn op deze handeling van mij in deze situatie. Het maakt niet veel uit hoe ik mijn arm hef als ik daarmee tijdens een vergadering te kennen geef dat ik instem met een bepaalde maatregel (hoewel ik in de wijze waarop ik mijn arm hef allerlei commentaar kan geven op mijn eigen stemgedrag). Als ik echter tijdens een reddingsoperatie boven op een ladder staand een klein kind op een balkon tracht te tillen, maakt het wel degelijk uit *hoe* ik mijn armen hef, en daarmee te kennen geef me bewust te zijn van wat er in deze situatie van me gevraagd wordt. Dit geldt nu precies ook voor deliberaties begrepen als 'basic actions'. In de wijze waarop ik delibereer, toon ik mijn aandacht en respect voor de normatieve geldingskracht van de regels die ik relevant acht in de situatie waarin ik me bevind.

Deze aandacht voor de wijze waarop een actor een 'basic action', en dus ook een deliberatie, kan verrichten is belangrijk omdat het ons wijst op een punt waarop de rationaliteit van de actor een onherleidbaar aspect is van het proces van deliberatie. Rationaliteit komt niet alleen maar in beeld doordat er überhaupt sprake is van een delibererende actor, maar vooral doordat er in de wijze waarop de actor delibereert ruimte is voor de relevantie van regels als regels die het waard zijn gevolgd te worden. En pas dat introduceert de normatieve dimensie die een essentieel aspect is van deliberatie of praktisch redeneren. Met andere woorden, cruciaal voor een deliberatie is niet het naakte feit dat het een proces is dat zich in de actor afspeelt (en waarin een hoofdrol wordt gespeeld door rationele principes, opvattingen, pro-attitudes en intenties), maar dat het een proces is waarin de rol van deze principes, opvattingen, pro-attitudes en intenties nu precies daarin is gelegen dat *ze regels* vertegenwoordigen die de actor, al delibererend, tracht te volgen omdat ze normatieve geldingskracht hebben.

Daarmee komen we op het verschil tussen de motiverende en de rechtvaardigende functie van een deliberatie. Stel, om het standaard voorbeeld eens te gebruiken, dat je naar de keuken loopt om een biertje uit de koelkast te halen. Gevraagd naar het waarom

zeg je dat je dorst hebt (je pro-attitude) en dat je weet dat er altijd bier in de koelkast ligt (je opvatting). De reden die je geeft (pro-attitude + opvatting) verklaart jouw gedrag omdat ze ons informeert over je motivering, over het interne proces dat vooraf ging aan je handeling en resulteerde in de intentie die zich in je handeling toont. Ons te informeren over je motieven is echter niet hetzelfde als je gedrag te rechtvaardigen. Zelfs als we je antwoord gehoord, begrepen, en als verklaring geaccepteerd hebben, kunnen we nog steeds vragen waarom je deed wat je hebt gedaan. Nu vragen we niet meer naar een verklaring, maar naar een rechtvaardiging. Was het gepast dorst te hebben (stel dat er al 25 lege bierflesjes rondom je heen staan)? Was het gepast te denken dat bier de dorst lest? Was het gepast te denken dat er bier in de koelkast ligt (gegeven al die lege flesjes)? Etc.

Het onderscheid tussen motiveren en rechtvaardigen is belangrijk, maar ook problematisch. Het is belangrijk, omdat een praktische redenering zich kan voltrekken zonder dat de normativiteit die er in aanwezig is als zodanig gethematiserd wordt. Er is voor een praktische redenering niet meer nodig dan een interne afstemming van opvattingen, pro-attitudes, en intenties, waarbij een strikt subjectieve semantiek op geen enkele manier het redeneren nadelig beïnvloedt. Zoals ik hierboven⁷ al schreef wordt men geenszins gehinderd in het correct redeneren door het hebben van volstrekt onzinnige (onhoudbare) opvattingen over bijvoorbeeld het laten vliegen van vrachtwagens. Daarbij moet men zich blijven realiseren dat redeneren iets anders is dan handelen op grond van een redenering. Men kan correct *redeneren* over hoe te handelen als men vrachtwagens wil laten vliegen; men komt daarentegen onvermijdelijk terecht in allerlei problemen als men *tracht te handelen* op grond van een overigens correcte redenering waarin onhoudbare opvattingen een hoofdrol spelen. Ook is het zo dat heel veel van ons gedrag verklaard kan worden door te wijzen op de motiverende kracht van pro-attitudes die eigenlijk volstrekt niet gepast zijn, waarbij ‘gepast’ een woord is dat een heel spec-

⁷ Onder het kopje “Rationele principes en opvattingen”, p. 39.

trum kan bestrijken – variërend van een heel neutraal ‘niet in overeenstemming met de aard van het gewenste object’ (zoals, bijvoorbeeld, het geval is bij een ongelukkige supporter die nog steeds de wedstrijden van zijn echt waardeloos spelende en altijd verliezende club blijft bezoeken) tot een heel zwaar ‘moreel verwerpelijk’ (zoals, bijvoorbeeld, het geval is bij een brute moordenaar die iemand om zeep helpt omdat de kleur van haar lippenstift hem niet aanstaat).

In die zin kan het er op lijken dat we voor een adequaat verslag van wat zich werkelijk voordoet, niet meer hoeven (kunnen? behoren?) te vertellen dan welke daadwerkelijk de elementen zijn geweest die een rol gespeeld hebben in het proces van deliberatie (begrepen als een kwestie van *motivatie*) dat feitelijk ten grondslag heeft gelegen aan het doen plaatsvinden van een bepaalde handeling. In die zin kan het er met andere woorden op lijken alsof een verslag van de feiten niets te maken heeft met de waarden die in het spel zijn, omdat de enige rol die weggelegd is voor deze waarden een causale rol is waarin er geen sprake hoeft te zijn van het door wie dan ook op normatief relevante wijze *onderschrijven* van deze waarden. Net zo min lijkt er sprake te moeten zijn van opvattingen die *waar* zijn; al het werk kan gedaan worden door wat de delibererende actor voor waar houdt. Deliberatie lijkt zo een proces dat voldoende heeft aan ‘narrow content’⁸, en bijgevolg begrepen kan worden als een volstrekt interne aangelegenheid.

Het onderscheid tussen motiveren en rechtvaardigen is, ik zei het al, niet alleen belangrijk, maar ook problematisch. Dit is zo, omdat het hele begrip van motiveren slechts denkbaar wordt tegen de achtergrond van voor vanzelfsprekend aangenomen rationele principes, principes die informatie geven over de betekenisvolle interne structuur van de context waarin de actor zich bevindt.

⁸ Dit is een term uit de *philosophy of mind*, die slaat op de inhoud van mentale toestanden als een betekenisinhoud die volstrekt geïdentificeerd kan worden op basis van slechts deze mentale toestanden zelf, zonder de relaties die deze mentale toestanden hebben met de natuurlijke, sociale, en morele omgeving waarin de delibererende actor zich bevindt.

Zodra men denkt dat een bepaalde redenering een handeling *motiveert*, of dat het citeren van bepaalde opvattingen en pro-attitudes een handeling *verklaart*, vooronderstelt men een achtergrond van rationele principes die op geen enkele manier normatief neutraal is. Verklaringen parasiteren in die zin op rechtvaardigingen. Iets is slechts dan een verklaring, is slechts dan een verslag van een gegeven motivatie, als ze mogelijk gemaakt wordt door een specifieke, zij het impliciete, rechtvaardiging van het perspectief dat deze deliberatie een motivatie laat zijn. Laat ik dit uitleggen.

Stel dat Karel de kamer uit gaat en met een biertje terugkomt, en stel dat ik de mededeling dat hij dorst heeft als causale verklaring accepteer. Dit betekent dat kennis nemen van Karels dorst mijn behoefte aan een verklaring (aan begrip van dit gebeuren) bevredigt. Dit is slechts mogelijk tegen een achtergrond die zowel in ontologisch als normatief opzicht zeer rijk is. Er is van alles dat ik moet weten, en waarvan ik moet aannemen dat Karel dit ook weet. En dan heb ik het niet alleen maar over kennis betreffende de dorstlessende eigenschap van bier, of over het ontbreken van dorstlessende zaken in de kamer, of over de nabijheid van de keuken en de koelkast waarin het bier staat. Maar ik heb het ook over allerlei normatief significante relaties tussen Karel, mij en het huis waarin we ons bevinden. Immers, Karel heeft mij niet gevraagd of ik ook iets wil drinken, en hij biedt mij ook nu nog niets aan. Blijkbaar gaat hij er, en ogenschijnlijk terecht, van uit dat ik in dit huis gewoon wel voor mijzelf kan zorgen. En ik ga daar blijkbaar ook van uit, gegeven het feit dat ik uit zo weinig informatie een bevredigende verklaring voor zijn gedrag kan destilleren. Meer dan dat, blijkbaar voelt Karel zich ook niet geroepen zich te verantwoorden voor zijn bezoek aan de keuken, en voor het pakken van een biertje uit de koelkast. Ogenschijnlijk kan hij in dit huis ook prima voor zichzelf zorgen. In de Nederlandse context kan dit haast niet anders dan betekenen dat wij beiden in dit huis *wonen*, en, daarom, beiden de vrijheid genieten die we tonen in onze motivaties en verklaringen. (In de Verenigde Staten gaat dit heel anders, zoals ik ooit tot mijn grote verbazing merkte, toen ik zag

hoe iemand die voor het eerst bij een ander thuis op bezoek was, daar de koelkast opende en iets te drinken pakte *alsof het de gewoonste zaak van de wereld was!*)

Het feit dat Karels dorst het motief was voor wat hij deed blijkt bij nader inzien zeer veelzeggend. Het feit dat slechts de introductie van deze dorst epistemologisch beschouwd een voldoende verklaring levert, en, bijgevolg, ontologisch beschouwd een proces van deliberatie kan laten culminerend in de vorming van een effectieve intentie een biertje in de keuken te gaan halen, maakt duidelijk dat Karel niet alleen maar de benodigde 'narrow content' paraat had. Nee, veel meer dan dat alleen, laat dit simpele feit zien dat Karel en ik gesitueerde rationele actoren zijn, voor wie 'dorst' een motief is in een context die normatief verregaand gestructureerd is.

De consequenties hiervan zijn aanzienlijk. Enerzijds suggereert de problematische relatie tussen delibereren als motiveren en delibereren als rechtvaardigen dat het wellicht de moeite loont eens grondig stil te staan bij het probleem van mentale veroorzaking. Dit is het centrale probleem binnen het handelingsmodel dat gebaseerd is op de aanname dat handelingen gebeurtenissen zijn. Want als handelingen gebeurtenissen zijn, en bijgevolg van nabij veroorzaakt worden door processen die zich in het lichaam van de actor afspelen (delibereren begrepen als motiveren), hoe is het dan mogelijk in dit model een plaats te geven aan rationele overwegingen die hun betekenis volstrekt ontleen aan de normatieve structuur van de handelingscontext (delibereren begrepen als rechtvaardigen). Deze kwestie zal ik bespreken in *Hoofdstuk 2*, waarin ik de verdiensten en beperkingen zal bespreken van de aanname dat handelingen gebeurtenissen zijn. Deze aanname zal ons een nieuwe kijk kunnen geven op de in dit hoofdstuk besproken problematiek, omdat, mocht er uit die hoek een oplossing komen voor het probleem van mentale veroorzaking, we geen last meer hoeven hebben van de moeizame verhouding tussen motiveren en rechtvaardigen. Dat zou dan slechts een probleem hoeven zijn voor hen die vertrekken vanuit de aanname dat handelingen verricht worden door actoren.

Anderzijds suggereert de onduidelijke, maar prominente plaats die normativiteit in een correct begrip van deliberatie inneemt, dat het wellicht de moeite loont eens grondig stil te staan bij het probleem van de al dan niet sociale oorsprong van de geldingskracht van regels. Immers, als handelingen verricht worden door actoren, en als dat betekent dat handelingen de resultaten zijn van deliberaties, en als deliberaties processen zijn waarin een hoofdrol gespeeld wordt door de rechtvaardigende normativiteit die zich toont in het volgen van de juiste regels, dan zou men kunnen denken dat we handelingen misschien vooral moeten begrijpen in termen van het volgen van regels, en dan zou men wellicht kunnen denken dat handelingen, meer dan iets anders, sociale fenomenen zijn (tenminste als zou blijken dat regels *als regels* sociale aangelegenheden zijn). Ook deze aanname, dat handelingen sociale fenomenen zijn, kan, op dit punt aangekomen, een verdienstelijk alternatief lijken te zijn voor de aanname dat handelingen verricht worden door actoren. Ik zal haar in *Hoofdstuk 3* bespreken.

Laten we nu echter eerst eens wat beter kijken naar die andere aanname, dat handelingen gebeurtenissen zijn – een aanname die zo voor de hand ligt in de Angelsaksische cultuur waarin filosofen vooral in de wetenschap een handlanger zien.

Handelingen zijn gebeurtenissen

1. Naturalisme en Determinisme

In een tijd waarin de natuurwetenschappen een hoge vlucht hebben genomen, en in een cultuur waarin men geneigd is bij 'kennis' en 'inzicht' onmiddellijk te denken aan *wetenschappelijke* kennis en *wetenschappelijk* inzicht, is het niet zo verwonderlijk dat filosofen een groot belang hechten aan het ontwikkelen van een begrip van handelen dat in het verlengde ligt van het natuurwetenschappelijk plausible begrip 'gebeurtenis'. In dit laatste begrip liggen twee vooronderstellingen opgesloten: naturalisme en determinisme. Beiden kunnen beschouwd worden als een methodologische, maar ook als een ontologische these.

Methodologisch naturalisme

De methodologische these van het naturalisme zegt dat alle fenomenen die het bestuderen waard zijn, op één-en-dezelfde manier benaderd, onderzocht, en in kaart gebracht kunnen worden; en wel op de manier die al met veel succes gebruikt wordt in de natuurwetenschappen. Vanzelfsprekend is daarmee nog niet precies één wetenschappelijke methode aangeduid, gegeven het feit dat er helemaal geen consensus is in de hedendaagse wetenschapsfilosofie. Maar de strekking van het methodologisch naturalisme is duidelijk: als we menselijk handelen zouden willen kennen en begrijpen, dan moeten we het behandelen als alle andere natuurlijke fenomenen: als (1) empirisch te benaderen zaken, die we in kaart kunnen brengen door middel van (2) het opstellen van hypothesen, die we (3) testen in experimenten op grond waarvan we (4) generalisaties formuleren in termen die (5) in een transparant begrippenkader hecht met elkaar samenhangen, en weer (6) aanleiding geven tot nieuwe hypothesen.

Ontologisch naturalisme

De ontologische these van het naturalisme sluit op het methodologisch naturalisme aan. De these fundeert het succes van de wetenschappelijke methode door te stellen dat het tot de natuur van alle fenomenen behoort zich zo te gedragen dat ze in herhaalbare, objectieve experimenten aanleiding geven tot wetmatige generalisaties. Deze wetmatigheden kunnen geformuleerd worden met behulp van slechts een beperkt aantal begrippen; begrippen betreffende de soort entiteiten waaruit de natuur is opgebouwd en de soort relaties die tussen deze entiteiten bestaan. Als het menselijk handelen begrijpelijk gemaakt kan worden, dan is dat mogelijk, aldus de these van het ontologisch naturalisme, omdat het tot de aard van het menselijk handelen behoort zich (1) natuurlijk te gedragen, daar het (2) gerealiseerd wordt door natuurlijke entiteiten (menselijke wezens die weliswaar buitengewoon complex zijn maar slechts die eigenschappen bezitten die ze kunnen bezitten dankzij het feit dat ze bestaan uit de elementaire bouwstenen van de natuur) die zich (3) op een natuurlijke (hoogstwaarschijnlijk een causale) wijze verhouden tot hun omgeving.

Een paar opmerkingen zijn hier op hun plaats. Ten eerste kunnen we, dankzij de gebruikte als-dan structuur, een uitweg bieden aan hen die zich al te ongemakkelijk voelen bij de wetenschappelijke benadering van het menselijk handelen. De uitweg is overigens niet zo heel aantrekkelijk, omdat ze neerkomt op het ontkennen van de stelling dat het menselijk handelen uiteindelijk volstrekt begrijpelijk gemaakt zou kunnen worden. Deze uitweg wil ik, als iemand die zich wil bezighouden met de analytische filosofie van het handelen, in het achterhoofd houden voor het uiterste geval dat..., en daarmee bedoel ik ook slechts en inderdaad voor het *uiterste* geval dat we werkelijk geen andere optie meer hebben dan te beweren dat het niet tot de aard van het menselijk handelen behoort zich natuurlijk (d.i. begrijpelijk) te gedragen. Let er hierbij op dat deze optie in algemene termen is gesteld. De uitweg behelst niet dat sommige handelingen onbegrijpelijk zijn, maar dat alle handelingen wezenlijk onbegrijpelijk zijn.

Ten tweede zien we dat de ogenschijnlijk strenge rigiditeit van het methodologisch naturalisme gesteund wordt door een aanzienlijk soepelere ontologische these. Want het is natuurlijk niet zo heel schokkend, en nog niet zo heel veelzeggend, te beweren dat mensen natuurlijke entiteiten zijn die natuurlijke relaties hebben met hun omgeving. Want wat is 'natuurlijk', in deze context? Peter Strawson, die daar in 1985 een heel aardig en toegankelijk boekje over schreef, onderscheidt een zwak en een sterk naturalisme. De sterke of *reductionistische* variant beweert dat met 'natuurlijk' slechts dat bedoeld kan worden dat past binnen het fysisch-biologische begrippenkader. Natuurlijke entiteiten zijn volgens dit begrippenkader nooit iets anders dan (complexe) conglomeraten van moleculen, en natuurlijke relaties zijn nooit iets anders dan wetmatig-causale (hoewel we op het terrein van de biologie behoefte lijken te hebben aan functionele relaties die een eigen niet-causale wetmatigheid lijken te bezitten). De zwakke of *non-reductionistische* variant beweert dat 'natuurlijk' in eerste instantie moet slaan op dat wat met betrekking tot een bepaald fenomeen vanzelfsprekend of normaal is, op dat wat 'als vanzelf' hoort bij de manier waarop fenomenen zich in onze alledaagse wereld presenteren.

Dit is een wel heel tolerante interpretatie van het naturalisme, maar de strekking ervan is belangrijk: het naturalisme stelt dat er geen enkele goede reden is om te vertrekken vanuit de aanname dat wat zich 'als vanzelf' aan ons voordoet niet geïntegreerd zou kunnen worden in een systematisch samenhangend begrippenkader. Deze aanname loslaten moeten we slechts in het uiterste geval doen, als we niet anders kunnen dan concluderen dat het menselijk handelen volstrekt onbegrijpelijk is. Vóór het zover komt, hebben we echter alle kans te speculeren over hoe te begrijpen dat zowel atomen als mensen *natuurlijke* entiteiten zijn, en zowel causale als intentionele relaties *natuurlijk*. En in deze speculaties, om ze maar even zo te noemen, hoeven we ons helemaal niet uitsluitend te laten leiden door de nauwe blik van de succesvolle fysicus. Niemand, immers, heeft ons beloofd dat filosofie gemakkelijk zou zijn.

Methodologisch determinisme

De methodologische these van het determinisme zegt dat alle fenomenen verklaard kunnen worden door een beroep te doen op causaliteit. Een verklaring is een belangrijk intellectueel instrument dat ons veel wetenswaardigs kan vertellen over een fenomeen waar we graag meer van zouden willen weten. Een verklaring kan alleen gegeven worden als we op de een of andere manier al geconfronteerd werden met het te verklaren fenomeen. Zo'n confrontatie kan resulteren in een beschrijving van dat fenomeen, maar een dergelijke beschrijving hoeft ons niet veel te leren over het fenomeen. Het kan juist ook een uitdrukking zijn van ons onbegrip over het fenomeen in kwestie. Als ik bijvoorbeeld rustig de krant aan het lezen ben en Emma van 5 slaat ineens de krant uit mijn hand, boos roepend dat ik een 'stommerd' ben, dan zou ik heel precies kunnen beschrijven wat mij zo-even overkomen is, maar dat zou nog helemaal niet betekenen dat ik ook begrijp wat mij overkomen is. Pas als ik het gedrag van Emma kan verklaren, zou ik kunnen vertellen dat ik inderdaad begrijp wat mij overkomen is.

De formuleringen die ik hier nu gebruik kunnen wellicht de indruk wekken dat ik het onderscheid tussen 'verklaren' en 'begrijpen' niet respecteert. Dit onderscheid hangt samen met het onderscheid tussen de *Natur-* en de *Geisteswissenschaften* dat begin deze eeuw werd geïntroduceerd, en dat vaak wordt geassocieerd met een onderscheid tussen twee methoden. De suggestie is dat verklaringen ten grondslag liggen aan natuurwetenschappelijk inzicht, en begrip ten grondslag ligt aan sociaal-wetenschappelijk inzicht. Natuurlijk is de suggestie tevens dat een ontologische these dit onderscheid rechtvaardigt: sommige fenomenen, zo luidt de suggestie, zijn natuurlijk van aard en sommige sociaal-cultureel.

Het zal duidelijk zijn dat in de huidige context, waarin ik me bezighoud met het uitwerken van de aanname dat handelingen gebeurtenissen zijn, dit ontologische onderscheid tussen natuurlijke en sociaal-culturele fenomenen niet zomaar geaccepteerd kan worden. Handelingen, zo is immers de aanname in dit verband,

zijn natuurlijke gebeurtenissen, en het onderscheid tussen verklaring en begrip kan dan ook niet als onoverkomelijk beschouwd worden. Nu hoeft dit ook niet, en daarom heb ik me het losse gebruik van begrip en verklaring in bovenstaand voorbeeld gepermitteerd. Waar het in de *Erklären-Verstehen* controversie feitelijk om gaat, is niet om een onderscheid tussen verklaring en begrip, maar om een onderscheid tussen causale en intentionele verklaringen. De eerste soort verklaringen introduceert gebeurtenissen die aan het te verklaren fenomeen vooraf gingen, en laat zien dat deze voorafgaande gebeurtenissen het fenomeen veroorzaakt hebben. De tweede soort verklaringen introduceert redenen, motieven, overtuigingen en dergelijke, en laat zien dat het te verklaren fenomeen begrijpelijk is in het licht van deze intentionele mentale toestanden. De eerste soort verklaringen moet een beroep doen op bepaalde opvattingen over causaliteit; de tweede soort moet een beroep doen op bepaalde opvattingen over begrijpelijkheid. Ik kan het gedrag van Emma *causaal* verklaren door te wijzen op een eerdere gebeurtenis (bijvoorbeeld dat ik beloofd had iets te drinken voor haar in te schenken) en een plausibel verhaal te vertellen over de causale relaties die er bestaan tussen mijn belofte, mijn gedrag en Emma's gedrag. En ik kan het gedrag van Emma *intentioneel* verklaren door te wijzen op haar verwachting ten aanzien van mijn beloofde gedrag, haar frustratie vanwege mijn actuele gedrag, en haar leeftijd waarbij bepaalde gedragspatronen horen.

Nu is de these van het methodologisch determinisme precies deze dat intentionele verklaringen in feite causale verklaringen zijn. *Verstehen* is volgens het methodologisch determinisme, met andere woorden, een vorm van *Erklären*. Als ik bepaalde handelingen wil verklaren, en daarbij wijs op de relevantie van bepaalde frustraties, overtuigingen, verlangens, en soortgelijke mentale toestanden, dan ken ik deze mentale toestanden in feite een bepaalde causale rol toe. Waar ik volgens de these van het methodologisch determinisme niet aan ontkomen kan, als ik fenomenen wil verklaren, is ze te behandelen als gebeurtenissen en het als oorzaak aanwijzen van andere gebeurtenissen. De these van het methodologisch determinisme zegt dat alle verklaringen een causaal karakter

hebben, en als we bepaald gedrag kunnen verklaren door een beroep te doen op bepaalde mentale toestanden, dan moet dat betekenen dat deze toestanden zodanig beschreven kunnen worden dat we inzien dat ze in een causale relatie staan tot het te verklaren gedrag. Daarmee is natuurlijk nog niet gezegd wat voor opvattingen we over causaliteit zouden moeten accepteren.

Ontologisch determinisme

Net als in het geval van het naturalisme werkt de ontologische these van het determinisme als het ware als een verklaring voor het succes van het methodologische determinisme. De ontologische these stelt dat de werkelijkheid zo in elkaar zit dat inderdaad ieder fenomeen een gebeurtenis is die in een zeer complex causaal netwerk opgenomen is en zich altijd als oorzaak en gevolg verhoudt tot allerlei andere in dit netwerk opgenomen gebeurtenissen. Omdat de werkelijkheid causaal georganiseerd is, is het mogelijk fenomenen te verklaren door te wijzen op hun oorzaken.

Er zijn verschillende opvattingen over causaliteit. In de hedendaagse analytische filosofie wordt doorgaans gewerkt met een oorspronkelijk Humeaanse opvatting, waarbij men echter vergeten is dat deze opvatting zelf in eerste instantie niet als een ontologische opvatting bedoeld was, of, voor zover ze dat wel was, beschouwd werd als een opvatting die geen wetmatigheden impliceerde, maar slechts regelmatige gewoonten. Laat me dit toelichten.

Voor David Hume was een causale relatie een relatie tussen twee gebeurtenissen die onafhankelijk van elkaar geïdentificeerd kunnen worden, en die zich zo tot elkaar verhouden dat steeds als de eerste gebeurtenis zich voordoet, de tweede gebeurtenis er op volgt, zodanig dat wij niet anders kunnen dan denken dat hier sprake is van noodzakelijke opeenvolging. Het eerste deel van de these is nodig om duidelijk te maken dat het hier om een empirische relatie gaat, en niet om een conceptuele. De dood van Xantippe veroorzaakt het weduwnaarschap van Socrates niet, maar impliceert het op conceptuele wijze. Het drinken van de gifbeker impliceert daarentegen de dood van Socrates niet op conceptuele

wijze, maar veroorzaakt hem. Het tweede deel van de these geeft de karakteristieke temporele asymmetrie weer van causale relaties. Bepaalde gebeurtenissen met kenmerk X worden altijd gevolgd door gebeurtenissen met kenmerk Y. En het derde deel keert zich eigenlijk tegen de Aristoteliaanse idee dat oorzaken invloed zouden uitoefenen. Te denken dat de opeenvolging noodzakelijk zou zijn, kan volgens Hume geen empirisch feit zijn, en daarom kan het niet meer zijn dan een ongepaste extrapolatie die kenmerkend is voor de mate waarin ons verstand afhankelijk is van gewoonten.

De ontologische these van het determinisme gaat veel verder dan wat voor Hume eigenlijk een epistemologische these was, door de eerste twee delen van Hume's opvatting over te nemen, maar ten opzichte van het derde deel voor gerechtvaardigd te houden dat de opeenvolging die tussen X-gebeurtenissen en Y-gebeurtenissen waargenomen kan worden, alleen dan constant zou kunnen zijn als het inderdaad zo is dat de werkelijkheid aan wetten van oorzaak en gevolg gehoorzaamt. Wat vooral belangrijk is aan deze op Hume's opvattingen gebaseerde *nomologische* causaliteitstheze, is dat er in de these geen plaats is voor wat bij Aristoteles de 'werkoorzaak' werd genoemd. De hedendaagse ontologische these van het determinisme heeft het niet over oorzaken die invloed uitoefenen, maar heeft het over wetmatigheden waaraan natuurlijke fenomenen gehoorzamen, of, veel beter, want niet antropomorficerend gezegd, over wetmatigheden die in het verloop van de gebeurtenissen ontdekt kunnen worden.

Deze paragraaf samenvattend kan gesteld worden dat naturalisme en determinisme leiden tot een visie op de werkelijkheid als één grote verzameling verschijnselen die allemaal op wetmatige wijze met elkaar verbonden zijn. Dat een bepaalde gebeurtenis een handeling is, verandert niets aan de status van deze gebeurtenis voor zover die gekarakteriseerd kan worden als een verschijnsel dat op wetmatige wijze gerelateerd is aan alle andere verschijnselen. Dit betekent natuurlijk helemaal niet dat alle gebeurtenissen handelingen zouden zijn. Slechts sommige gebeurtenissen zijn ook handelingen, namelijk die gebeurtenissen die veroorzaakt worden

door mentale toestanden als intenties, redenen, verlangens en opvattingen. Dat deze bewering niet onmiddellijk onprobleematisch is, is het onderwerp van de volgende paragraaf.

2. Het probleem van mentale veroorzaking

Het probleem: twee oorzaken voor één effect

Zodra we beweren dat handelingen gebeurtenissen zijn, zullen we al snel geconfronteerd worden met de meest actuele variant van het lichaam-geest probleem: het probleem van mentale veroorzaking. Als we immers beweren dat een handeling een gebeurtenis is, dan lijken we daarmee te zeggen dat sommige gebeurtenissen, namelijk handelingen, veroorzaakt worden door twee ogenschijnlijk zeer ongelijksoortige zaken.

Stel bijvoorbeeld dat ik, na lang beraad, besluit mijn vakantie door te brengen op Terschelling. En stel dat ik, als gevolg van mijn besluit, een reservering maak voor een camping. Ik pak de telefoon en toets een nummer. Wat is de oorzaak van deze handeling? Enerzijds is mijn beslissing de vakantie op Terschelling door te brengen natuurlijk de oorzaak voor het toetsen van dit bepaalde telefoonnummer. Maar anderzijds lijkt er een zuiver fysische oorzaak voor deze handeling te moeten zijn: bepaalde neurofysiologische gebeurtenissen in mijn zenuwstelsel hebben de beweging van mijn handen veroorzaakt. Het is de ongelijksoortigheid van beide oorzaken die het probleem van mentale veroorzaking introduceert, omdat we door deze ongelijksoortigheid geconfronteerd worden met de vraag hoe beide oorzaken zich tot elkaar verhouden. En gegeven de zeer plausibele aanname dat het fysisch universum causaal gesloten is – dat wil zeggen, de aanname dat een fysische gebeurtenis *altijd* een fysische oorzaak heeft – wordt dit vooral de vraag hoe we de mentale oorzaak (mijn beslissing de vakantie op Terschelling door te brengen) een plaats kunnen geven in een fysisch overigens principieel begrijpelijke opeenvolging van gebeurtenissen.

Hier zijn drie manieren het probleem aan te scherpen. (1) We kunnen niet denken dat de beslissing en de neurofysiologische gebeurtenis gezamenlijk de oorzaak voor mijn telefoontje zijn. Het is immers absurd te denken dat ik na het nemen van de beslissing nog afhankelijk ben van het al dan niet optreden van de juiste gebeurtenis in mijn zenuwstelsel, zoals het absurd is te denken dat deze neurofysiologische gebeurtenis mijn handen pas aan het werk kan zetten als ook nog eens, onafhankelijk daarvan, door mij de beslissing genomen wordt te bellen. (2) We kunnen ook niet denken dat de beslissing en de neurofysiologische gebeurtenis ieder op zich de oorzaak van mijn bel-gedrag zijn. Immers, het is absurd te denken dat de juiste gebeurtenis in mijn zenuwstelsel mij het juiste nummer laat toetsen in de afwezigheid van iedere intentie dat nummer te bellen, zoals het absurd is te denken dat mijn beslissing te gaan bellen mij het nummer laat toetsen in de afwezigheid van enige gepaste gebeurtenis in mijn zenuwstelsel. (3) En tenslotte kunnen we niet denken dat de beslissing en de neurofysiologische gebeurtenis een keten vormen en in die opeenvolgende zin gezamenlijk voor mijn gedrag zorgen. Dit kunnen we niet denken, omdat we daarmee zouden beweren dat het fysisch universum niet causaal gesloten zou zijn. Het zou impliceren dat gebeurtenissen in onze hersenen niet veroorzaakt worden door andere fysische gebeurtenissen in onze hersenen, maar door het nemen van beslissingen. Het zou mentale oorzaken niet een begrijpelijke plaats in het fysisch universum geven, maar een raadselachtige, mysterieuze plaats tussen fysische gebeurtenissen. Daarmee zou het niet alleen iets mysterieus maken van het mentale, maar vooral ook iets mysterieus maken van het fysische, dat blijkbaar niet langer meer gedefinieerd kan worden als het domein dat beschreven kan worden in termen die de fysica ons daartoe aanreikt, maar waarin we rekening moeten houden met fysisch intrinsiek onbegrijpelijke relaties.

Er zijn minstens vijf manieren het probleem van mentale veroorzaking, dat ik nu scherp gesteld heb, op te lossen. Ik zal ze in deze paragraaf kort de revue laten passeren. Daarbij gaat het me er

niet om recht te doen aan de vaak zeer technische en gedetailleerde uiteenzettingen die verdedigers van de diverse posities ontwikkeld hebben, en evenmin gaat het me om een gepaste en genuanceerde evaluatie van de kracht en de zwakte van de posities. Het hedendaagse debat over het probleem van mentale veroorzaking is een zeer specialistische aangelegenheid geworden die zich niet leent, en zich ook niet hoeft te lenen, voor een inleidend overzicht. Waar het mij hier slechts om gaat is om een schets van de structuur van de diverse opties. Hoe verhouden zij zich tot elkaar? Welke mogelijkheden boren zij aan? Welke risico's durven zij nemen? Op welke aannames durven zij te bouwen, en welke vinden zij te problematisch? Dat zijn de vragen die mij hier zullen interesseren.

Eén effect? Dan één oorzaak!

Een eerste simpele manier om te ontkomen aan het probleem van mentale veroorzaking is door te stellen dat er helemaal geen twee ongelijksoortige oorzaken voor menselijk gedrag zijn. Er is maar één enkele, echte oorzaak, die, vanwege de eigenaardige structuur van ons bestaan als ervarende wezens, echter op twee radicaal verschillende manieren beschreven kan worden. Deze oplossing suggereert dat er sprake is van een *identiteitsrelatie* tussen mentale en fysische oorzaken. Mijn beslissing de telefoon te pakken om een camping op Terschelling te reserveren *is* de neurofysiologische gebeurtenis die de beweging van mijn handen veroorzaakt. Dat wil zeggen, een bepaalde beslissing is te karakteriseren in termen van de soort beslissing die het is, d.w.z. in termen van haar inhoud (het is bijvoorbeeld de beslissing die gekenmerkt wordt doordat ze over een vakantie gaat, een vakantie van mij, in een tent, op Terschelling, op een camping die tijdig gereserveerd moet worden, wat met de telefoon kan, etc. etc.). Dit betekent dat er een specifieke gebeurtenis is die als een zus-en-zo beslissing gekarakteriseerd kan worden. En dit, zo stellen verdedigers van deze optie, is geheel verenigbaar met de aanname dat deze zelfde gebeurtenis ook fysisch gekarakteriseerd kan worden, d.w.z. op deze wereld

voor deze, menselijke denkers (of wellicht zelfs alleen maar voor mij, nu) in termen van de specifieke neurofysiologische configuratie van mijn hersenen en zenuwstelsel. Als mentale toestand is deze toestand een beslissing, een specifieke instantiatie van een zus-en-zo beslissing. En als fysische toestand is precies deze zelfde toestand een neurofysiologische gebeurtenis.

Het gaat hier om een zogenaamd *token* fysicalisme. Het gaat immers om een concreet geval (deze ene beslissing nu de telefoon te pakken) dat een *token* is van wat in feite een bepaald *type* mentale gebeurtenis is, een type waarvan ook heel andere *tokens* voorkomen (denk maar eens aan die eerste keer dat je de beslissing neemt je geliefde te bellen voor een afspraakje). De idee van het *token* fysicalisme is dat iedere concrete mentale gebeurtenis een instantiatie is van bepaalde mentale eigenschappen, en dat deze ene specifieke gebeurtenis (en dat geldt voor ieder *token* van dit *type* op zich) altijd tegelijkertijd ook een instantiatie is van een fysische eigenschap. Daarbij hoeft het niet zo te zijn dat ieder *token* van dit *type* mentale toestand fysisch beschouwd ook altijd een *token* is van *eenzelfde type* fysische toestand. Te besluiten je geliefde voor het eerst te bellen en te besluiten een camping op Terschelling te bellen zijn te beschouwen als mentale toestanden van één en hetzelfde *type*, terwijl tegelijkertijd zinvol beweerd kan worden dat ze neurofysiologisch tot heel verschillende *typen* behoren. De enige stelling van het *token* fysicalisme is dat er geen *tokens* van mentale toestanden of gebeurtenissen voorkomen die niet tegelijkertijd ook een instantiatie zijn van een fysische eigenschap.

Deze oplossing, die bijvoorbeeld door Donald Davidson verdedigd wordt, suggereert dat we zonder tegenspraak (1) kunnen aannemen dat causale relaties een wetmatig (nomologisch) karakter hebben, (2) kunnen aannemen dat de relaties tussen mentale toestanden en gedrag niet wetmatig van karakter zijn, en (3) kunnen aannemen dat de relaties tussen mentale toestanden en gedrag wel een causaal karakter hebben. Dit kan zonder tegenspraak, omdat deze oplossing een vierde aanname introduceert, die aanleiding geeft tot een boeiend onderzoeksprogramma. Deze vierde aanname

is (4) de *token identiteitsthese*: een bepaalde neurofysiologische gebeurtenis in de hersenen heeft natuurwetmatige relaties met bepaalde lichaamsbewegingen en *is* tegelijkertijd een gebeurtenis die met behulp van mentale termen gekarakteriseerd kan worden. Beschreven met behulp van deze mentale terminologie kan van de gebeurtenis niet gezegd worden dat ze wetmatige relaties onderhoudt met gedrag, maar qua gebeurtenis gaat het om een extensioneel te individueren entiteit die onder een fysische beschrijving natuurlijk wel degelijk een instantiatie is van een wet.

Het onderzoeksprogramma waartoe deze identiteitsthese aanleiding geeft, betreft vooral het conceptueel doorgronden van waar deze identiteit tussen enerzijds mentale en anderzijds fysische entiteiten nu precies op neer komt. En hoewel er veel gepubliceerd wordt over emergentie-, constitutie- en superveniëntie-relaties ziet het er voorlopig nog niet naar uit dat we echt meer gaan begrijpen van de identiteitsthese.

Het grootste probleem voor het token fysicisme schuilt in het eigenaardige gebrek aan verband dat ze impliceert tussen oorzaken en verklaringen. Men zou toch denken dat causale verklaringen hun verklarende kracht ontleenen aan de wetmatige causale relaties waar ze een beroep op doen. Maar dit vanzelfsprekende inzicht moet de verdediger van het token fysicisme opgeven. Immers, enerzijds is er, op fysisch niveau, sprake van een wetmatige causale relatie tussen oorzaak en effect, tussen de neurofysiologische configuratie in mijn hersenen en het bewegen van mijn hand die het nummer op Terschelling draait. Anderzijds is er echter, op mentaal niveau, sprake van een niet-wetmatige relatie tussen mijn beslissing en mijn telefoneergedrag. En gegeven de aanname dat causale relaties een nomologisch karakter hebben, impliceert dit dat de oorzaak *als oorzaak* alleen herkenbaar is op het fysische niveau waar immers wetmatige relaties gelden. In mijn alledaagse verklaring, daarentegen, heb ik het alleen maar over deze oorzaak in termen van mijn beslissing nu eens in Nederland, op Terschelling, de vakantie door te brengen. Hoe komt het dan dat deze verklaring werkt, als ze geen melding maakt van de (fysische) eigenschap die het mogelijk maakt dat de oorzaak een *oorzaak* is

van het te verklaren gedrag? Hoe komt het dat men begrijpt dat mijn gedrag veroorzaakt wordt door mijn beslissing, als het in feite niet deze beslissing is, tenminste niet onder de beschrijving van een beslissing te zijn, die zich op gepaste wijze causaal verhoudt tot mijn gedrag?

Dit probleem wordt meestal gekarakteriseerd door te spreken over al dan niet causaal relevante eigenschappen. Als een rode biljartbal een witte raakt, dan is de beweging van de witte te verklaren door te verwijzen naar de kinetische energie van de rode bal, en niet door te verwijzen naar zijn rode kleur. De kinetische energie is de causaal relevante eigenschap. Maar als dit onderscheid tussen eigenschappen die wel en eigenschappen die niet causaal relevant zijn algemeen toepasbaar is, dan zou een van de consequenties van het token fysiscalisme kunnen zijn dat de mentale eigenschappen van een gebeurtenis causaal niet relevant genoemd moeten worden. De relatie tussen mijn beslissing en mijn telefoongedrag zou dan overeenkomen met de relatie tussen de kleur van de rode biljartbal en de beweging van de witte. De verklaring van mijn telefoontje in termen van de beslissing die ik nam, zou dan net zo zinloos zijn als de verklaring van de beweging van de witte biljartbal door te wijzen op de kleur van de rode. Daarmee zou het token fysiscalisme epifenomenalisme impliceren: immers, de mentale eigenschappen van de gebeurtenis in mijn hersenen (dat het een beslissing was over een camping op Terschelling die nu gebeld moest worden in verband met een reservering) zouden niet de causaal relevante eigenschappen kunnen zijn. De rol van het mentale zou daarmee op het terrein van oorzaak en effect feitelijk nihil zijn: we zouden slechts vertrouwd zijn met de schijn van de causale effectiviteit van het mentale.

Als we dit niet willen accepteren, zullen we meer moeten gaan begrijpen van de precieze relatie tussen oorzaken en verklaringen. Over die relatie zal ik verderop nog het één en ander te zeggen hebben.¹

¹ Namelijk op pp. 76-83, onder de kopjes "Oorzaken en verklaringen zijn twee" en "Verklaringen en oorzaken 'zijn' één".

Twee oorzaken? Dan twee effecten!

Een tweede simpele manier om te ontkomen aan het probleem van mentale veroorzaking, althans in de vorm waarin ik dat in het begin van deze paragraaf geïntroduceerd heb, is door te stellen dat er helemaal geen sprake is van één effect dat veroorzaakt wordt door zowel mijn neurofysiologie als mijn beslissingen. Er is sprake van twee verschillende effecten, bij iedere oorzaak één.

Voor een goed begrip van deze optie is het van belang dat we haar onderscheiden van een vergelijkbare, maar veel radicalere reactie op het probleem van mentale veroorzaking. Deze radicale reactie betwijfelt de juistheid van de centrale stelling van het model dat ik in deze paragraaf bespreek. De radicale reactie stelt dat gedrag een ambigue term is die zowel 'lichaamsbeweging' als 'handeling' kan betekenen. Vanwege deze ambiguïteit, zo stelt deze radicale reactie, kan men onterecht gaan denken dat handelingen gebeurtenissen (in, van of met het lichaam) zijn. En dan krijg je inderdaad een probleem met mentale veroorzaking, omdat je dan twee ongelijksoortige entiteiten (handelingen en lichaamsbewegingen) systematisch met elkaar begint te verwarren in een zoektocht naar één oorzaak voor twee ongelijksoortige effecten. Volgens deze radicale reactie is het probleem van mentale veroorzaking een symptoom van de fundamentele onjuistheid van het naturalistische model dat ik hier aan het bespreken ben. De neurofysiologen kunnen ongestoord praten over de fysische oorzaken van *lichaamsbewegingen*, want deze lichaamsbewegingen worden veroorzaakt door de neurofysiologische configuratie van de hersenen van de actor. Maar neurofysiologen hebben niets zinnigs te zeggen over de mentale (en sociale, historische, enz.) oorzaken van *handelingen*, want handelingen hebben helemaal geen fysische oorzaken, maar psychologische, sociale, etc. Aanhangers van deze radicale reactie zijn meestal aanhangers van het derde model ("handelingen zijn sociale fenomenen") dat ik in *Hoofdstuk 3* zal bespreken.

De optie die ik hier wil bespreken lijkt op de radicale reactie omdat ook zij stelt dat het probleem van mentale veroorzaking

haar onoplosbaarheid te wijten heeft aan een cruciale ambiguïteit in de formulering van het probleem. Maar deze optie lijkt helemaal niet op de radicale reactie voor zover ze de aanname dat handelingen gebeurtenissen zijn volstrekt onderschrijft. De “twee effecten”-oplossing, van bijvoorbeeld Fred Dretske (1988), wijst op het onderscheid tussen lichaamsbewegingen enerzijds en gedrag anderzijds, en zegt dat het verschil tussen beide is dat lichaamsbewegingen enkelvoudige fenomenen zijn, en gedrag een samengesteld fenomeen bestaande uit een lichaamsbeweging *plus* een interne oorzaak. Op dit onderscheid wijzend komen we dan tot het stellen van twee verwante, maar zeer verschillende vragen: (1) Wat is de oorzaak O van deze lichaamsbeweging L? en (2) Wat is de oorzaak M van het feit dat O L veroorzaakt? Om het met een voorbeeld te zeggen: (1) Wat is de oorzaak van de beweging van mijn handen die een telefoonhoorn oppakken en diverse nummertoeetsen indrukken? En, gegeven dat een bepaalde neurofysiologische configuratie in mijn hersenen de oorzaak van deze bewegingen is, krijgen we de vraag die voor het probleem van mentale veroorzaking cruciaal is: (2) Wat is de oorzaak van het feit dat deze neurofysiologische configuratie functioneert als oorzaak van het oppakken van de telefoonhoorn en het intoetsen van bepaalde nummers?

Het is deze tweede vraag die van belang is voor het probleem van mentale veroorzaking. Met de eerste vraag kan de neurofysiologie uitstekend uit de voeten, maar de tweede vraag is eigenlijk een vraag naar het waarom: waarom is de neurofysiologische gebeurtenis O de oorzaak van lichaamsbeweging L? Deze vraag is van een andere orde; ze vraagt niet naar de oorzaak van het feit dat gebeurtenis O zich voordoet (ook die vraag kan de neurofysiologie in principe immers wel beantwoorden), maar ze vraagt naar een verklaring voor het gerealiseerd zijn van een causale relatie tussen een bepaalde hersentoestand O en een bepaalde lichaamsbeweging L. En het beantwoorden van deze vraag impliceert, volgens Dretske, het aantonen van de causale relevantie van de semantische, intentionele, mentale eigenschappen van de neurofysiologische gebeurtenis O.

Dretske's antwoord op deze tweede vraag maakt didactisch gebruik van het gedrag van een thermostaat, en conceptueel gebruik van (1) het onderscheid tussen de actuele werkoorzaak ('triggering cause') en de structurerende oorzaak ('structuring cause'), en (2) de idee van een indicatiefunctie ('indicator function'). Een thermostaat is een schakelaar die 'aan' en 'uit' kan staan afhankelijk van de omgevingstemperatuur. De relatie tussen temperatuur en de stand van de thermostaat is eenvoudig uit te leggen met behulp van natuurkundige termen. Een antwoord op de vraag naar de oorzaak (of het waarom) van de relatie tussen de stand van de thermostaat en het aan- of afslaan van de verwarmingsketel kan echter niet, of niet inzichtelijk genoeg, beantwoord worden met behulp van natuurkundige termen. Als de thermostaat vanwege een daling van de omgevingstemperatuur 'aan' gaat, zal de ketel aanslaan, omdat de verwarmingsmonteur het apparaat zo ontworpen heeft dat er met het oog op het verwarmen van de kamer optimaal gebruik gemaakt wordt van de mogelijkheid de indicatiefunctie van de thermostaat te gebruiken als actuele werkoorzaak. De structurerende oorzaak is gelegen in het ontwerp van de verwarmingseenheid. In dit ontwerp krijgt de thermostaat een bepaalde causale rol. Veranderingen in zijn interne structuur worden een actuele werkoorzaak voor het aan- en afslaan van de verwarmingsketel. En deze causale rol krijgt de thermostaat vanwege zijn indicerende eigenschappen (d.i. vanwege het feit dat hij een indicatiefunctie heeft, omdat veranderingen in zijn interne structuur systematisch afhankelijk zijn van veranderingen in de omgevingstemperatuur).

De vertaling van thermostaat naar mentale toestanden ligt volgens Dretske voor de hand. Hersentoestanden krijgen soms een bepaalde causale rol (ze veroorzaken bepaalde lichaamsbewegingen) vanwege hun indicatiefunctie. In dat geval kan een hersentoestand de actuele werkoorzaak zijn vanwege haar indicerende eigenschappen, d.w.z. vanwege haar intentionele, mentale eigenschappen. Vragen we waarom deze lichaamsbewegingen plaatsvinden, dan wijzen we de desbetreffende hersentoestand aan. Vragen we waarom deze hersentoestand de actuele werkoorzaak is, dan wijzen we op de

intentionele eigenschappen van deze hersentoestand, die, in het organisme in zijn omgeving, de structurerende oorzaak zijn. Dat wil zeggen dat we dan wijzen op de mentale toestand inclusief de hele ontwerpgeschiedenis die geleid heeft tot het aanwenden van deze intentionele, indicerende eigenschappen voor het veroorzaken van bijpassend gedrag. Deze ontwerpgeschiedenis kan, volgens Dretske, analoog aan het evolutieproces begrepen worden als een geschiedenis waarin geen behoefte is aan een, in principe buiten deze geschiedenis staande, ontwerper. Mijn natuurlijke aanleg (zelf het resultaat van miljoenen jaren) en mijn jarenlange opvoeding heeft er niet alleen voor gezorgd dat ik zindelijk ben geworden (=het aanwenden van sommige indicerende prikkels voor het veroorzaken van bijpassend, gewenst gedrag), dat ik mijn laarzen uitdoe als ik binnenkom, mijn mond houd als ik ruzie verwacht, maar ook (en dit is volgens Dretske alleen maar ingewikkelder, niet anders) dat ik de telefoon pak en een nummer op Terschelling draai als ik de beslissing heb genomen een camping te reserveren voor de aanstaande zomervakantie.

De oplossing voor het probleem van mentale veroorzaking waar dit betoog van Dretske aanleiding toe geeft, is te zeggen dat de causale rol van het mentale een structurerende is. Mentale toestanden zoals beslissingen verklaren iets anders dan fysische toestanden. Ze verklaren waarom de fysische toestanden hun specifieke werkcausaliteit bezitten. Ze verklaren, door middel van hun betekenis (en daarmee door middel van een zeer korte verwijzing naar een zeer lang leerproces) hoe de indicerende eigenschappen van hersentoestanden causaal aangewend kunnen worden.

Er zijn echter twee problemen met deze oplossing. Enerzijds is de oplossing circulair als ze een verklaring wil geven voor hoe het mogelijk is dat betekenis een causale rol kan hebben. Immers, alleen door betekenis al te introduceren in de indicerende eigenschappen wordt het mogelijk te zeggen dat het specifieke van structurerende oorzaken is dat ze vanwege de betekenis van hun indicerende eigenschappen sommige hersentoestanden aanwenden als oorzaak voor het gepaste gedrag. Anderzijds is niet duidelijk

waarom de oplossing toch niet reductionistisch blijkt te zijn.² Immers, het is niet duidelijk waarom er uiteindelijk niet een volstrekt neurofysiologisch antwoord gegeven kan worden op de vraag naar de structurerende oorzaak: de gehele evolutionaire en ontwikkelingspsychologische geschiedenis die geleid heeft tot het aanleggen van een causale relatie tussen hersentoestand O en lichaamsbeweging L vraagt toch, binnen het kader van het model dat handelingen gebeurtenissen zijn, om een door en door naturalistische (= fysicalistische) verklaring waarin er geen specifiek eigensoortige rol is voor betekenissen.

Wellicht dat we dan ook het verst komen, binnen het kader van dit model, met een reductionistische aanpak van het probleem van mentale veroorzaking.

Twee oorzaken? Dat is één te veel!

Reductionisme is er in soorten en maten, maar wat ze allemaal met elkaar gemeen hebben is de these dat alles wat er verteld moet, hoeft of kan worden over een zaak verteld kan worden zonder gebruik te maken van de terminologie die gebruikt werd bij het introduceren van de zaak als een zaak die om nadere opheldering vraagt. Met betrekking tot mentale veroorzaking komt de reductionistische these hierop neer dat het mogelijk is alles te vertellen over bijvoorbeeld mijn beslissing een camping op Terschelling te bellen zonder de terminologie te gebruiken die de aandacht vestigt op deze beslissing als een beslissing – zoals ‘beslissing’, maar ook ‘opvatting’, ‘wens’ en ‘intentie’, en vervolgens ook alle woorden waarmee de inhoud van deze opvattingen, wensen, en beslissing gekarakteriseerd worden: dus ‘camping’, ‘tent’, ‘vakantie’, ‘Nederland’, ‘Terschelling’, ‘telefoon’, ‘druktoetsen’, etc. etc.

Als het de reductionist lukken zou alles te vertellen over mijn beslissing zonder deze hele ‘volkpsychologische’ woordenschat te hoeven gebruiken, door bijvoorbeeld alleen maar termen te gebruiken die ontleend zijn aan de neurofysiologie, de biologie, de

² Twee reductionistische oplossingen zal ik hierna bespreken onder het kopje “Twee oorzaken? Dat is één te veel!”

chemie, en de natuurkunde, dan zou de reductionist opgelucht kunnen mededelen dat er geen probleem van mentale veroorzaking is, omdat er helemaal geen sprake is van twee ongelijksoortige oorzaken, maar slechts van één. In die zin lijkt de reductionistische oplossing op de token fysicalistische, hoewel deze laatste over het algemeen juist zijn eigen karakter ontleent aan het feit dat het een niet-reductionistische oplossing is. Het verschil tussen deze beide is dan dit: beide oplossingen beweren dat er maar één oorzaak is, maar de identiteitstheorie beweert dat we iets belangrijks missen als we alleen de neurofysiologische woordenschat hanteren (namelijk de handzame en succesvolle volkpsychologische verklaring in termen van opvattingen, wensen, intenties en beslissingen), terwijl de reductionist beweert dat we niets *kunnen* missen, omdat met behulp van de neurofysiologische terminologie alles wat er over de oorzaak van een handeling te zeggen valt ook inderdaad gezegd wordt.

Er zijn twee reductionistische hoofdstromingen. De één, ontwikkeld en verdedigd door o.a. J.J.C. Smart (1963) en David Armstrong (1966), wil de volkpsychologische verklaringen, of in ieder geval de schijn van hun succes, behouden door ze als een herleidbare deelverzameling op te nemen in de, vanuit wetenschappelijk oogpunt, betere neurofysiologische woordenschat. Ze wil dat doen door het opstellen van zogenaamde 'bridge laws', wetmatigheden die gelden tussen enerzijds volkpsychologische wetmatigheden zoals die binnen bepaalde randvoorwaarden kunnen gelden, en anderzijds de onderliggende, meer fundamentele fysische wetmatigheden die binnen de randvoorwaarden van dit universum gelden voor het gehele domein van het menselijk handelen. Het grote probleem voor deze psycho-fysische 'bridge laws' is het relationele, externalistische karakter van mentale inhouden in de brede zin ("wide content") dat het nauwelijks mogelijk maakt congruentie te vinden tussen volkpsychologische wetmatigheden en fysische wetmatigheden.³

³ Dit is een kwestie die in de *philosophy of mind* druk bediscussieerd wordt. De kwestie is min of meer deze: als mentale toestanden slechts geïndividueerd kunnen worden door ook te verwijzen naar de relaties die deze toestanden hebben

De tweede reductionistische hoofdstroom, verdedigd door o.a. Stephen Stich en Paul & Patricia Churchland, is de eliminativistische, die beweert dat de schijn van het succes van volkspychologische verklaringen helemaal niet verdedigd kan worden, omdat volkspychologie in belangrijke mate, of zelfs uitsluitend, bestaat uit onjuiste en onhoudbare generalisaties. Deze stroming houdt de volkspychologische terminologie niet voor een pre-theoretisch instrument met behulp waarvan menselijk handelen in eerste instantie noodzakelijkerwijs geïntroduceerd moet worden als een zaak die om nadere opheldering vraagt. Nee, volgens de eliminativisten is de volkspychologie een slechte theorie die ons vanaf het eerste moment misleidt doordat ze de natuurlijke fenomenen in een onbruikbaar keurslijf dwingt. Pas als we de gehele volkspychologische woordenschat verwerpen, kan er de ruimte ontstaan om met de wetenschappelijk verantwoorde categorieën van de neurofysiologie een begin te maken met het inventariseren van de verschijnselen waarmee we van doen hebben. Het grote probleem voor het eliminativisme is dat ze als het om de volkspychologie gaat, nog niet eens een vage indicatie kan geven van hoe we die woordenschat van ons af zouden *kunnen* werpen. Het lijkt er op alsof deze woordenschat helemaal geen slechte theorie is, maar een onmisbaar onderdeel van de essentie van menselijke praktijken.

Oorzaken en verklaringen zijn twee.

We hebben tot nu toe drie manieren gezien waarop we zouden kunnen proberen te begrijpen wat het betekent dat handelingen

tot de natuurlijke en sociale omgeving van de persoon wiens toestanden het zijn, dan zullen volkspychologische wetmatigheden een zo sterk holistisch karakter hebben dat we haast niet kunnen beginnen met te gaan begrijpen hoe we zulke wetmatigheden zouden moeten kunnen gaan relateren aan neurofysiologische wetmatigheden, die, zoals alle natuurwetenschappelijke wetmatigheden, zo'n sterk atomistisch karakter hebben. Zie voor meer over deze kwestie bijvoorbeeld de inleidende werken: Paul Churchland, *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*. Cambridge (Mass.): MIT Press, 1984, en Kim Sterelny, *The Representational Theory of Mind. An Introduction*, Oxford, Blackwell Publishers, 1990.

gebeurtenissen zijn, namelijk die gebeurtenissen die veroorzaakt worden door een bijzonder soort oorzaken: mentale toestanden. En de moeilijkheden die we tegen gekomen zijn, zijn karakteristiek voor de specifieke taak die we ons gesteld hebben: aannemelijk te maken dat een bepaalde species tot een bepaalde genus behoort, maar er niet mee samenvalt, juist omdat ze een species is, en niet het genus. Vergelijk dit met de poging aan te tonen dat mensen dieren zijn – weliswaar bijzondere dieren, omdat ze redelijk zijn, maar niettemin toch dieren. Of denk aan een kwestie die veel onbenulliger is: de poging aan te tonen dat vorken bestek zijn – weliswaar bijzonder bestek, omdat ze tanden hebben, maar niettemin toch bestek. Zo'n onderneming zal licht struikelen over dezelfde problemen: óf er komt te veel nadruk op het dierlijke van mensen waarna we ons zorgen gaan maken over hoe dan toch recht te doen aan het redelijke, óf er komt te veel nadruk op de tanden van vorken waarna we ons gaan afvragen hoe nog recht te doen aan het gemeenschappelijke van bestek, óf we zetten zo stevig in dat we de tanden van vorken definiëren als kenmerk van bestek en het redelijke als een vorm van het dierlijke, bijgevolg haast alleen nog maar dieren en bestek overhoudend. Het is de vertrouwdheid met beide categorieën, met de species en het genus, die we nodig hebben om de verhouding tussen beide te doorzien, maar het lijkt erop alsof we die vertrouwdheid met beiden niet in één omvattend inzicht onder woorden kunnen brengen.

Stel nu dat dit komt doordat de oorsprong van de vertrouwdheid met de beide categorieën heel verschillend is. Stel dat de oorsprong van onze vertrouwdheid met een species steeds gelegen is in de praktijk van ons alledaagse handelen, en dat de oorsprong van de vertrouwdheid met de genus steeds gelegen is in de theoretische reflectie op deze praktijk. Ik pretendeer hiermee niet te zeggen dat dit universeel geldt voor alle species en genus waarmee we vertrouwd zijn, maar val voor de verleiding dit hier en nu te beweren over de begrippen 'vork', 'bestek', 'mens', 'dier', 'mentaal', en 'fysisch' voor zover we ons bekommeren om hun onderlinge species-genus verhouding.

Ik ben bijvoorbeeld zes jaar oud en mag (!) de tafel dekken en zoek de vorken, de lepels en de messen. Want die heb ik nodig, en ik weet al aan welke kant van het bord ik de vorken leggen moet. Alleen waar vind ik ze? En als mijn moeder dan zegt: "In de bestekla" dan kijk ik daar, want ook die la is voor mij een vertrouwde plek. Maar als mijn moeder dan een educatieve bui heeft, en mij vraagt waarom de vorken in de bestekla liggen, dan moet ik even afstand nemen van mijn bezigheden en de theoretisch gewichtige klank van het begrip 'bestek' tot me door laten dringen. Ik ben ineens met iets anders bezig; ik verlaat, om dit dan zo te zeggen, even de vanzelfsprekende contouren van het alledaagse leven en verdiep me voor een ogenblik in de zeer prille rudimenten van theoretische klassificaties.

Dat is ook het geval als we proberen te begrijpen dat handelingen gebeurtenissen zijn, die veroorzaakt worden door mentale toestanden. We weten allemaal en op een heel manifeste wijze wat handelingen zijn. We zijn allemaal zindelijk geworden, en we hebben allemaal nog veel meer geleerd dan alleen onze willekeurige spieren beheersen: we denken na bij wat we doen, of vóór we iets doen, en dat helpt. We hebben allemaal, stuk voor stuk, geleerd ons handelen te begrijpen in het licht van onze opvattingen, wensen, intenties en beslissingen, in het licht van onze redenen, die we zelf weer hebben leren kennen als de zaken waar we naar verwijzen om ons handelen te verklaren (en te rechtvaardigen). In onze ervaring is veroorzaken lange tijd eerst en vooral een kwestie van *intentioneel* veroorzaken geweest.

Maar als we dan gevraagd worden dit handelen te begrijpen als geïntegreerd in een naturalistische en deterministische wereld, dan lijken we ineens geconfronteerd met een heel ander, veel strenger begrip van causaliteit en lijken we onze alledaagse verklaringen, die tamelijk los zijn, maar zeer adequaat voor wie, net als wij, grotendeels op routine hun weg door het leven vinden, niet in te kunnen passen in de systematische verklaringspatronen die we als wetenschapper hopen op te kunnen stellen.

Een uitweg is echter denkbaar, en aanwezig in het werk van bijvoorbeeld Frank Jackson & Philip Pettit en Cynthia & Graham

Macdonald. Deze uitweg is niet-reductionistisch, gaat echter wel uit van de vooronderstelling dat handelingen fysische gebeurtenissen zijn, maar heeft geen last van het verwijt van epifenomenalisme (zoals ik dat hierboven⁴ formuleerde). De uitweg grijpt aan bij de token fysicalistische identiteitstheze, maar probeert het maximale te halen uit de bewering dat een eigenschap die causaal relevant maar niet causaal effectief is, wel degelijk effectief kan zijn *als verklarende eigenschap*. Cruciaal voor deze oplossing is de constatering dat verklaringen er in soorten en maten zijn, waarmee deze oplossing ook nog eens het beste probeert te maken van de soort van oplossing die door Dretske werd voorgesteld.⁵

In deze oplossing speelt *superveniëntie* een grote rol.⁶ De idee is dat één en dezelfde gebeurtenis tegelijkertijd een instantiatie kan zijn van zowel hogere-orde als lagere-orde eigenschappen. Van de hogere-orde eigenschappen wordt gezegd dat ze superveniëren op (en dat betekent ongeveer 'mogelijk gemaakt worden door afhankelijk te zijn van') de lagere-orde, of subveniërende, eigenschappen. In letterlijke zin zijn vervolgens, volgens aanhangers van deze oplossing, alleen die eigenschappen causaal effectief die zich bevinden op het laagste, fysische niveau. Dit is zo, omdat er alleen op dit laagste niveau sprake is van de juiste wetmatige relaties die we nodig hebben om van echte causaliteit te kunnen spreken. Dat de gebeurtenis (die bijvoorbeeld tegelijkertijd zowel een instantiatie is van mijn beslissing naar Terschelling te bellen als een instantiatie van een bepaalde neurofysiologische eigenschap) de oorzaak is van mijn telefoontje naar de camping, is, in strikte zin, dan ook alleen maar het geval vanwege het feit dat deze gebeurtenis een instantiatie is van bepaalde neurofysiologische eigenschappen.

⁴ Zie p. 69.

⁵ Zie hierboven, pp. 70-74, onder het kopje "Twee effecten? Dan twee oorzaken!"

⁶ Dit is een technische term die met name in de hedendaagse *philosophy of mind* nog al eens een rol van gewicht toegedicht wordt. Een uitstekend overzichts-artikel over de kracht en de zwakte van dit begrip is Terence Horgan, "From Supervenience to Superdupervenience: Meeting the Demands of a Material World", in *Mind*, Vol. 102, 1993.

Hoewel dit betekent dat de hogere-orde eigenschappen, waarvan deze gebeurtenis ook een instantiatie is (bijvoorbeeld de hogere-orde eigenschap “een beslissing zijn om naar Terschelling te bellen”) niet causaal effectief zijn, wil dit nog helemaal niet zeggen dat deze hogere-orde eigenschappen niet causaal relevant zijn. Dat kunnen ze zijn, als we een verwijzing naar hun causale relevantie nodig hebben om bepaalde hogere-orde eigenschappen van het veroorzaakte gedrag te kunnen verklaren. En dergelijke verwijzingen hebben we volop nodig. De redelijkheid, bijvoorbeeld, of de efficiëntie, van een belletje naar Terschelling als je daar op een camping de zomervakantie door wil brengen, is een eigenschap van het veroorzaakte effect die we niet kunnen verklaren met behulp van de neurofysiologische verklaring. En ook al is het zo dat het effect in letterlijke zin slechts veroorzaakt kan zijn, volgens aanhangers van deze optie, doordat de gebeurtenis die een instantiatie is van de hogere-orde eigenschap van een verstandige beslissing te zijn, tegelijkertijd ook een instantiatie is van een neurofysiologische eigenschap die het echte causale werk doet, toch is het tevens zo dat de redelijkheid van het effect alleen verklaard kan worden, en ook echt causaal verklaard wordt, door te verwijzen naar bepaalde causaal relevante hogere-orde eigenschappen van de veroorzakende gebeurtenis.

Wetenschap vraagt om wetmatigheden, en die kunnen we vinden op het laagste, fysische niveau. Maar verklaringen bieden niet altijd alleen maar wetmatigheden. Omdat verklaringen altijd gerefereerd zijn aan specifieke kennisbelangen, kunnen we soms verklaringen nodig hebben die iets anders bieden, die inzichtelijkheid bieden in een hogere orde patroon. Wat werkt in een verklaring kan een causaal effectieve eigenschap zijn, maar in tal van omstandigheden, in feite in haast alle alledaagse omstandigheden, zijn het de causaal relevante hogere-orde eigenschappen die in verklaringen hun werk doen (ook al kunnen ze dat werk in deze wereld slechts doen dankzij het feit dat gebeurtenissen die instantiaties zijn van deze hogere-orde eigenschappen gelukkigerwijs tevens instantiaties zijn van de fysische eigenschappen die het benodigde causaal effectieve werk kunnen doen).

De conclusie van deze paragraaf kan zijn dat als we niet vergeten om welke verklaring we verlegen zitten, we niet vast hoeven te lopen in het weliswaar zeer lastige, maar niet onoplosbare probleem van mentale veroorzaking.

Verklaringen en oorzaken 'zijn' één.

De laatste oplossing voor het probleem van mentale veroorzaking dat ik hier wil bespreken, is een verreichende oplossing, omdat ze, zonder onverenigbaar te zijn met de centrale aanname dat handelingen gebeurtenissen zijn, wel een tamelijk krachtige bom legt onder de motivatie voor deze aanname. De oplossing kan begrepen worden als liggend in het verlengde van de vorige oplossing, voor zover ze accepteert dat er verschillende niveaus zijn waarop we verschillende eigenschappen (van één en dezelfde entiteit?) kunnen situeren, ieder met hun eigen, kenmerkende relaties, en ieder met hun eigen causale relevantie. Maar anders dan de oplossing die ik in de vorige paragraaf besprak, onttrekt de hier voorgestelde oplossing het fysische niveau de voorrang als het over causale effectiviteit gaat.

De oplossing kan op verschillende manieren geformuleerd worden. Eén manier is te zeggen dat het verschil tussen causaal relevant en causaal effectief niet hard te maken is. Een andere is te zeggen dat causale effectiviteit een term is die *te veel* wil maken van de uiteindelijk tamelijk ondoorzichtige notie 'causal power', en *te weinig* rekening houdt met het Humeaanse inzicht dat causaliteit slechts met (regelmatige, of desnoods noodzakelijke c.q. wetmatige) opeenvolging heeft te maken. Nog een andere is te zeggen dat causale effectiviteit niet meer kan zijn dan verklarende kracht, zodat er op ieder niveau waarop er sprake is van plausible causale verklaringen probleemloos gesproken kan worden van causale effectiviteit. Al deze manieren komen op de één of andere manier neer op de bewering dat er zoveel oorzaken zijn, als dat er verklaringen zijn. Vandaar de titel die ik dit stukje heb gegeven, een titel die ook nog eens het verschil onderstreept tussen deze positie en de positie waarin echte causaliteit alleen kan bestaan op

het laagste, fysische niveau zelfs al kunnen causaal relevante eigenschappen voorkomen op ieder niveau.

Een goede manier om deze positie inzichtelijk te maken, en ik baseer me hier verder op het werk van Lynne Rudder Baker (1995), is te wijzen op de centrale rol die de zogenaamde 'counterfactual test' binnen deze positie speelt. Met deze test kan op een zeer praktische wijze, en met een zeer beperkte theoretische fundering, vastgesteld worden of een bepaalde eigenschap verklarende kracht heeft, en, *daarom*, beschouwd kan worden als een oorzaak (in een dus zeer alledaagse betekenis van het woord). De hele pointe van de positie is dat de alledaagse betekenis van 'oorzaak' goed genoeg is voor iedere belangstelling die we, ook wetenschappelijk, zouden kunnen hebben in het fenomeen van mentale veroorzaking. Dat geeft de positie ook zijn verreichende, dramatische karakter. De stelling komt namelijk in feite neer op het volstrekt trivialisieren van de aanname dat handelingen gebeurtenissen zijn. Natuurlijk zijn handelingen gebeurtenissen, maar 'gebeurtenis' is een heel alledaagse term waarover de natuurwetenschap niet meer te zeggen heeft dan iedereen die wel eens iets heeft meegemaakt. Een gebeurtenis is een voorval dat een oorzaak heeft in deze ongecompliceerde zin dat het een voorval is waarvoor we een verklaring kunnen geven – een verklaring waarvan een 'counterfactual' deel uitmaakt: *als* de oorzaak O zich niet had voorgedaan in overigens gelijkblijvende omstandigheden, *dan* zou het voorval V zich ook niet voorgedaan hebben. Vandaar de 'counterfactual test'.

Een voorbeeld. Te verwijzen naar mijn beslissing de camping op Terschelling te bellen in verband met het reserveren van een plekje voor komende zomer, is het geven van een verklaring voor mijn telefoneergedrag, en inderdaad een causale verklaring, als ze de volgende 'counterfactual' ondersteunt: als ik, in overigens gelijkblijvende omstandigheden, deze beslissing niet had genomen, dan zou ik de hoorn niet opgepakt hebben en niet dat bepaalde nummer hebben getoetst. Vandaar dat mijn beslissing een oorzaak is van mijn gedrag, in de alledaagse betekenis van het woord waarin we ook zeggen dat iemands huwelijk de oorzaak is van diens indeling in een andere belastingtariefgroep, en disfunc-

tionerende software de oorzaak is van de beurskrach die Wall Street in 1987 trof. In al deze gevallen is het zinloos te beweren dat het echte causale werk zich op het laagste, fysische niveau afspeelt, zoals het in al deze gevallen zinloos is te spreken over gebeurtenissen als voorvallen waarover bij uitstek de natuurwetenschap het laatste woord heeft. Natuurlijk waren dit gebeurtenissen, maar dat hoeft ons niet te verleiden tot een heilloos onderzoeksprogramma waarin naturalisme een sterk scientistisch en reductionistisch karakter heeft, en determinisme voorrang voor een streng wetenschappelijk gedefinieerde notie van causaliteit impliceert.

Conclusie

Wat moeten we nu denken van dit tweede handelingsmodel dat ik hier besproken heb? Is het een correct, maar triviaal model? Wijst het ons een doodlopende weg? Heeft het ons geholpen aan een beter inzicht in de relatie tussen motiveren en rechtvaardigen dan we hebben kunnen krijgen in termen van het handelingsmodel dat ik in *Hoofdstuk 1* heb besproken? Zijn we de these dat handelingen gebeurtenissen zijn, omdat redenen oorzaken zijn, beter gaan begrijpen? Wellicht is deze laatste vraag de vraag die ons het snelst aan een globale evaluatie helpt. Ik heb me beziggehouden met het probleem van mentale veroorzaking, met een analyse van de specifieke causaliteit die lijkt te horen bij menselijk gedrag. En hoewel we veel te weten gekomen zijn over causaliteit en over de relatie tussen mentale en fysische eigenschappen, heb ik het eigenlijk helemaal niet meer gehad over redenen. Pas aan het eind, waar ik ingegaan ben op de relatie tussen oorzaken en verklaringen, zijn we weer in de buurt gekomen van wat mentale toestanden als opvattingen en wensen redenen maakt, namelijk dat ze in verklaringen zouden kunnen functioneren als rationalisaties van ons gedrag. Dat was het oorspronkelijke punt van Davidsons 1963 artikel, dat de inspiratiebron is geweest voor de hele discussie die ik in dit hoofdstuk besproken heb.

Maar veel hebben we blijkbaar niet kunnen leren over dat wat redenen nu specifiek tot *redenen* maakt. Blijkbaar helpt het niet

voor een beter begrip van redenen je te buigen over de vraag of redenen oorzaken kunnen zijn. Blijkbaar krijg je niet snel een beter inzicht in rationalisaties als je ze begrijpt als motivaties die bestudeerd kunnen worden als natuurlijke processen die zich afspelen in actoren. Dat is overigens niet zo heel gek: je gaat ook niet beter begrijpen wat een vork nu precies een vork maakt, als je je buigt over de vraag of een vork tot het bestek behoort. Daarom is het ook belangrijk je steeds bewust te blijven van de uitgangspunten van een bepaalde onderzoeksstrategie. En daarom is het ook belangrijk te beseffen dat je niet al te snel moet denken dat je een omvattend model kunt ontwikkelen van een zo complexe zaak als het menselijk handelen door je te concentreren op slechts één aanname, dat handelingen gebeurtenissen zijn.

Dat wil natuurlijk helemaal niet zeggen dat de onderneming tot nu toe zinloos is geweest. Eén van de echte grote intellectuele problemen van deze tijd is de verhouding te begrijpen tussen enerzijds het succes van de natuurwetenschap met hun nadruk op naturalisme en determinisme, en anderzijds de aard van het menselijk bestaan waarin rationaliteit, normativiteit en zinvolheid de boventoon voeren. Ook als het zo zou zijn dat we door middel van een verkenning van het derde handelingsmodel, waarin handelingen beschouwd worden als sociale fenomenen, meer zouden gaan begrijpen van de normativiteit die kenmerkend is voor het menselijk bestaan, zal dit ons niet definitief weg kunnen houden van de vragen die in dit hoofdstuk centraal hebben gestaan. Maar of het werkelijk zo is dat de aanname dat handelingen sociale fenomenen zijn ons helpt meer te begrijpen van de normativiteit die kenmerkend is voor handelingen begrepen als gebeurtenissen die gerechtvaardigd kunnen worden...? Dat moet nog blijken, en daarover gaat het volgende hoofdstuk.

Handelingen zijn sociale fenomenen

1. Het sui generis karakter van het sociale

Aan het eind van *Hoofdstuk 1* heb ik de verhouding tussen motiveren en rechtvaardigen geïntroduceerd als een mogelijk aanknopingspunt voor wie meer wil begrijpen van handelingen.¹ Ik heb daarbij gesuggereerd dat we twee kanten op zouden kunnen gaan. Eén optie was het serieus nemen van de aanname dat handelingen gebeurtenissen zijn en het verkennen van het in het licht van die aanname acute probleem van mentale veroorzaking. De achterliggende gedachte hierbij was dat we zo wellicht een adequate rekenschap zouden kunnen geven van ‘motiveren’ als een natuurlijk proces dat begrepen kan worden als een vorm van rationaliseren. Deze optie is in het vorige hoofdstuk aan de orde gekomen, en ze heeft nog niet geleid tot een probleemloos begrip van het menselijk handelen. De andere optie was het serieus nemen van de aanname dat handelingen sociale fenomenen zijn en het verkennen van de in het licht van deze aanname prominente rol van regels. De achterliggende gedachte hierbij is dat we zo wellicht een adequate rekenschap zullen kunnen geven van ‘rechtvaardigen’ als een intrinsiek sociaal proces en daarmee van handelingen als entiteiten die, vanwege hun normatieve component, een volstrekt eigen karakter hebben. Deze optie zal in dit derde hoofdstuk aan de orde komen.

We zouden mijn organisatie van de diverse discussies in de hedendaagse analytische handelingsfilosofie ook zo kunnen karakteriseren: handelingen hebben een mentale dimensie, een dimensie die ik in kaart heb gebracht in *Hoofdstuk 1* door me te concentreren op de consequenties van de aanname dat handelingen verricht worden door actoren. Daarnaast hebben handelingen ook een

¹ Zie pp. 54-55.

fysieke dimensie, een dimensie die ik in kaart heb gebracht in *Hoofdstuk 2* door me te concentreren op de consequenties van de aanname dat handelingen gebeurtenissen zijn. Beide dimensies kunnen, zo zou een voorlopige conclusie kunnen luiden, niet echt overtuigend gepresenteerd worden als de fundamentele dimensie die het ons mogelijk maakt handelingen te begrijpen als, precies, *handelingen*.

Aanhangers van de derde aanname, dat handelingen sociale fenomenen zijn, menen op dit punt aangekomen dat we niet de conclusie hoeven trekken dat iedere dimensie even belangrijk is. Zij zullen kunnen beweren dat het meest kenmerkende van handelingen is dat ze, naast een fysieke en een mentale, ook nog een normatieve dimensie hebben, en dat het precies die fundamentele dimensie is die het ons eindelijk mogelijk maakt handelingen als handelingen te begrijpen. Hun stelling zou kunnen luiden dat handelingen succesvol kunnen zijn of kunnen mislukken; dat ze niet alleen maar onder natuurwetmatigheden vallen, en ook niet alleen maar onder de controle van zelfstandige, delibererende actoren, maar ook onder regels. Deze regels hebben een heel eigen karakter; ze stellen eisen aan actoren, en ze hebben geldingskracht onafhankelijk van de instemming van actoren, zonder dat dit betekent dat ze een natuurwetmatige status hebben. Ze hebben, zo stellen de aanhangers van de derde aanname, een sociale status.

In dit derde hoofdstuk zal ik de consequenties van deze aanname bespreken. Daarbij zal ik mij, zoals steeds, vooral bezighouden met het weergeven van de actuele stand van zaken. Er doet zich daarbij echter de volgende moeilijkheid voor. Vanwege de tamelijk strenge grenzen die in de analytische filosofie bestaan tussen disciplines als ethiek, politieke theorie, sociale filosofie, taal filosofie, metafysica en de filosofie van het handelen, wordt het thema van het sociale karakter van de normativiteit van het handelen eigenlijk haast nergens expliciet als zodanig aan de orde gesteld. De discussies die ik meen te moeten bespreken, moet ik daarom enigszins zelf construeren. Dit brengt met zich mee dat dit hoofdstuk behalve een 'weergevend' karakter ook een 'explorerend' karakter heeft. Voor beginnende lezers zal dit niet erg zijn; vanuit

hun perspectief is het onderscheid tussen weergave en verkenning nauwelijks te maken. Voor lezers die vertrouwd zijn met de analytische filosofie van het handelen, zal mijn bespreking een aantal verrassende elementen bevatten. Ik hoop echter dat zij de ratio achter de verbanden die ik leg zullen inzien, en hier en daar zelfs het gevoel zullen hebben dat een aantal kwesties door mijn discussie als vanzelf op hun plaats vallen.

Het dramaturgisch model

Er is een sterke tendens, met name in de analytische filosofie, het sociale te begrijpen als een domein dat geen eigen zelfstandigheid bezit, maar een hogere orde aangelegenheid is die herleid kan worden tot het mentale of zelfs tot het fysische niveau. Op deze tendens zal ik om didactische redenen een beroep doen bij de uitleg van het dramaturgische model, een model dat uitdrukking geeft aan de tegenovergestelde these dat juist het sociale domein van fundamenteel belang is, en een onherleidbare zelfstandigheid heeft waartoe zelfs, althans volgens verdedigers van de sterke interpretatie, de mentale en fysische componenten van het menselijk handelen herleid kunnen worden. Door in mijn uitleg van de centrale begrippen van dit model ('rol' en 'regel') reductionistische interpretaties te weerstaan, verwacht ik dat de pointe van het model, dat het sociale een *sui generis* karakter heeft, sterker tot uitdrukking komt.

Het dramaturgisch model heeft zijn oorsprong in het werk van de Amerikaanse socioloog Erving Goffman (1959), en kan vooral, maar eigenlijk ook slechts, bogen op een glansrijke carrière in de theoretische sociologie. Dat ik dit model hier, in navolging van Martin Hollis (1977) en Rom Harré (1979), als uitgangspunt gebruik voor een bespreking van de consequenties van de filosofische aanname dat handelingen sociale fenomenen zijn, komt door het volgende. Handelingen hebben een betekenis-component die we zouden kunnen begrijpen als een symbolische component, wat, heel globaal, zoiets wil zeggen als dat een handeling een fenomeen is dat de aandacht wil vestigen op iets dat het niet zelf is.

Dit brengt, vanuit het perspectief van hen die normativiteit begrijpen als gefundeerd in sociale relaties, twee zaken met zich mee. Blijkbaar is er altijd sprake van interacterende betrokkenen die over middelen beschikken elkaars aandacht op iets te vestigen, als er sprake is van symbolen. Dit is een inzicht dat ook in de taal-filosofie wordt onderschreven. Dit inzicht krijgt in het dramaturgisch model een heel concrete gestalte: de interacterende betrokkenen delen een 'script' met elkaar: ze spelen allen hun rol in hetzelfde toneelstuk omdat ze de regels kennen die de normatieve structuur van het scenario bepalen. Tevens is er, als er sprake is van symbolen, blijkbaar altijd sprake van een verdubbeling: dat wat zichtbaar is, *staat voor* dat waarop de aandacht gevestigd wordt. Deze verdubbeling heeft een respectabele staat van dienst in de geschiedenis van de filosofie die teruggaat tot Plato's allegorie van de grot. Ook deze verdubbeling krijgt een heel concrete gestalte in het dramaturgisch model: handelingen verwijzen naar achterliggende rollen. De twee genoemde elementen, dat handelingen vanwege hun betekenis-component verwijzen naar rollen en dat deze rollen uitdrukkingsvormen zijn van een gemeenschappelijk gedeelde verzameling regels, maken het dramaturgisch model tot een goed uitgangspunt voor een filosofische bespreking van de consequenties van de aanname dat handelingen sociale fenomenen zijn. In het centrale begrip 'rol' komt op een mooie manier de samenhang tussen regels en relaties, tussen het normatieve en het sociale, tot uiting. Een 'rol' is immers te definiëren als de verzameling regels die horen bij een positie.

Wat is een positie? Men zou kunnen denken dat het neerkomt op het bezetten van een locatie in een of ander domein. Maar dan moeten we ons wel realiseren dat het om veel meer moet gaan dan om het simpelweg lichamelijk bezetten van een locatie in het spatio-temporele domein. Een rots kan op deze strikt fysische wijze een positie hebben en als gevolg daarvan zal een boom die in de schaduw van deze rots groeit een bepaalde vorm kunnen krijgen. Dit geldt echter niet voor de positie van de koningin op een schaakbord. Dat is geen strikt fysische waarbij alleen maar causale wetmatigheden betrokken zijn. En het geldt evenmin voor de

sociale positie die ik als vader heb, of die John Travolta in *Saturday Night Fever* heeft. Deze posities zijn alleen maar te beschrijven als locaties in een domein als we begrijpen dat het hier om een sociaal domein gaat dat gedefinieerd wordt met behulp van sociale regels.

Men zou omgekeerd kunnen denken dat het hebben van een positie een kwestie is van het beschikken over macht, over invloed, over het kunnen uitoefenen van controle. Maar ook dat zou een oneigenlijke reductie zijn van wat er aan de orde is. De macht van de acteur die *King Lear* speelt is niet groter dan zijn rol hem toestaat; de controle van een arts over zijn patiënten is minder een kwestie van de vermogens van de arts dan van de regels die er in het medische circuit gelden. Een positie is niet een locatie die een actor zichzelf kan toe-eigenen; het is een product van regels die van sociale origine zijn.

Wat zijn dat voor regels? Hoe horen de regels bij de positie? Wederom, weer natuurlijk niet op de strikt fysische wijze waarop natuurwetmatigheden horen bij de rots die zonlicht wegneemt. En ook niet op de strikt zelfstandige wijze waarop een dictator tracht te bewerkstelligen dat zijn wil wet is. Immers, het moeilijke van de wil van een dictator is dat de onderdanen niet kunnen anticiperen op wat de wet zal voorschrijven. En mochten ze dat wel kunnen, omdat de wil van de dictator verregaand rationeel is, dan moeten ze ook in staat zijn hem te corrigeren als hij een fout maakt, wat hij dan natuurlijk niet kan toegeven, daar hij dan geen dictator zou blijken te zijn, maar een slaaf van zijn eigen redelijkheid. Dit is vanzelfsprekend niet bedoeld als een *knock-down* argument tegen de redelijkheid van dictators, maar het is wel bedoeld als een argument tegen de reductie van het sociale karakter van regels tot de zelfstandigheid van actoren.

Als er zoiets bestaat als een 'rol', en als we handelingen kunnen begrijpen als uitdrukking gevend aan een rol, dan kunnen we handelingen nog steeds begrijpen als gebeurtenissen, en ook als voortgebracht door actoren, maar we kunnen niet langer denken dat het rol-karakter van de handeling te herleiden is tot een natuurwetmatigheid die eigen is aan gebeurtenissen, of tot een

zelfstandigheid die eigen is aan actoren. Wie een rol heeft, zal regels moeten volgen; regels die niet natuurlijk zijn in de zin van causaal-wetmatig, en die men niet zelf heeft opgesteld, maar die tot articulatie zijn gekomen dankzij de sociale relaties waaraan men deelneemt. In de volgende paragraaf zal ik dieper ingaan op dit eigenaardige, sociale karakter van regels.

Maar eerst nog even deze belangrijke terzijde. Een consequentie van het dramaturgisch model die in verschillende publicaties door Hollis is uitgewerkt, is dat personen zelf, als actoren, delen in de verdubbeling die kenmerkend is voor de verhouding tussen handeling en rol. Personen zijn een samenstel van enerzijds het personage dat dankzij de rol gestalte krijgt, en anderzijds de acteur die de rol speelt. De nadruk te leggen op het personage is misleidend, maar begrijpelijk voor wie van mening is dat handelingen gebeurtenissen zijn; de nadruk te leggen op de acteur is eveneens misleidend, maar begrijpelijk voor wie van mening is dat handelingen verricht worden door actoren. Aanhangers van de these dat handelingen sociale fenomenen zijn bestrijden beide, naar hun mening, reductionistische interpretaties: het sociale is volgens hen een volstrekt zelfstandige dimensie waarin het pas mogelijk wordt voor actoren om als actoren te verschijnen en waarin het pas mogelijk wordt voor gebeurtenissen om handelingen te zijn. Ik denk dat deze consequentie een vruchtbaar effect zou kunnen hebben op de discussies over zelfstandigheid en delibereren die ik in *Hoofdstuk 1* besproken heb. Het zou aanleiding kunnen geven zelfstandigheid te begrijpen in termen van de interne dialectiek tussen personage en acteur, en deze dialectiek vervolgens weer in termen van een proces van deliberatie waarbij zelfkennis één van de voornaamste drijfveren is. Wie deze suggestie uitwerkt zal veel kunnen hebben aan het latere werk van Charles Taylor.

Over regels en sociale conventies

In de wereld van het toneel is het script dat waarover de verschillende betrokkenen gezamenlijk beschikken om elkaars aandacht

op iets te vestigen. In de alledaagse wereld die door het dramaturgisch model in kaart gebracht wordt, is dit script nog het best te beschouwen als een verzameling regels waarin de normatieve structuur van het decor en het scenario vastgelegd is. Wie een postkantoor betreedt, kan in geen enkel script letterlijk lezen wat hij precies moet zeggen en doen, maar zal wel slechts dan herkend worden als iemand die betekenisvol handelt, als hij de regels volgt die gelden voor bezoekers van een postkantoor.

Het is Ludwig Wittgenstein geweest die de enorme filosofische belangstelling voor (het volgen van) regels ontketend heeft, door te constateren dat regels normatieve bepalingen zijn die in een oneindig aantal gevallen het onderscheid aangeven tussen de goede en de verkeerde manier om met het oog op dit geval 'door te gaan'. Het standaardvoorbeeld is het opeenvolgend noemen van getallen. Na '1, 2, 3, 4, 5...' ligt het voor de hand '6' te noemen, gebruikmakend van de '+1'-regel. Zo is het ook mogelijk dat de goede toneelspeler na vijf botte opmerkingen in het script gelezen te hebben de 'regel' herkend heeft (een generalisatie over het karakter van het personage) en op eigen kracht een gepaste zesde opmerking 'in de geest van het personage' weet te maken.

Wittgenstein heeft echter ook het grote probleem van het regelvolgen voor het eerst geformuleerd: wie zegt mij dat er na '5' niet '7' komt als de regel niet '+1' is, maar de iets ingewikkelder regel 'tot 5 '+1', na 5 '+2'? Of, wie zegt mij dat het personage tegen iedereen bot is, en niet bijvoorbeeld alleen maar tegen zijn schoonzusje aan wier adres de eerste vijf botte opmerkingen werden gemaakt? Saul Kripke (1982) volgend, kan het probleem van twee verschillende kanten benaderd worden, maar in beide gevallen draait het om de onmogelijkheid het oneindige te pakken te krijgen met eindige middelen.

De ene benadering wijst op het gegeven dat in iedere situatie oneindig veel regels van toepassing zijn die in principe oneindig veel verschillende opties om 'door te gaan' als de juiste aanwijzen. Wellicht is de botterik steeds na vijf keer één keer aardig, behalve iedere vijftigste keer, tenzij op dinsdag, behalve voor twaalf uur, tenzij tegen vrienden op hun verjaardag, behalve op zondagmiddag,

tenzij in het buitenland, etc. etc. Dit geeft geen natuurlijk idee van de complexiteit van een karakter, maar dat onze karakters complex zijn, zelfs ondoorgrondelijk complex, en dat daarin een grote rol gespeeld wordt door de omstandigheden waarin we ons bevinden (tijd, plaats, gezelschap) staat buiten kijf. Wat voor de botterik vanzelfsprekend is om te doen, wat de regel die hij al die tijd al (impliciet) volgt in *deze* situatie van hem vraagt, moet daarom altijd wel weer een verrassing voor ons zijn. We kunnen immers nooit weten welke van de oneindige verzameling regels die compatibel is met al zijn handelingen tot nu toe, de regel is die hij daadwerkelijk blijkt te volgen. In iedere situatie is er, kortom, een oneindige verzameling regels die ieder op zich gelden kan in deze situatie, en ieder op zich een heel andere normatieve structurering van de situatie vastlegt.

De andere benadering wijst op het gegeven dat iedere regel op oneindig veel situaties van toepassing is, en aangezien wij ons geen voorstelling kunnen maken van oneindig veel situaties zijn wij eenvoudigweg volstrekt niet in staat ook maar de simpelste regel te begrijpen en te volgen. De botterik die een voor ons ondoorgrondelijke regel lijkt te volgen, zit met een regel die ook voor hem ondoorgrondelijk is. Hij volgt daarom, zo moeten we wel concluderen, helemaal geen regel, maar doet maar wat, wellicht achteraf wanhopig pogend zijn eigen gedrag te rationaliseren door te suggereren dat er een regel is die hij volgt. Maar als we hem zouden vragen deze regel te noemen dan zou hij ongenadig door de mand vallen. De regel die hij immers zegt te volgen moet, als regel, van toepassing zijn op oneindig veel situaties, en is bijgevolg helemaal niet uitspreekbaar. Hij is dat niet in extensionele zin: dat zou neerkomen op het uitspreken van een oneindige verzameling paren waarbij het eerste element een karakterisering van een denkbare situatie is en het andere element een karakterisering van de juiste optie in die situatie. En hij is dat ook niet in intensionele zin: dat zou neerkomen op het uitspreken van een functie die aan iedere denkbare situatie een juiste optie toekent.

Het zou echter kunnen lijken alsof er in het geval van gezelschapsspelen wel degelijk sprake is van dergelijke functies, van

regels in intensionele zin. Immers, als we naar het schaken kijken, dan zou het kunnen lijken alsof de regels die het gedrag van de stukken bepalen precies zulk soort functies zijn. Voor ieder veld op het schaakbord vertellen deze regels waar het stuk naartoe verplaatst mag worden. Hoewel dit waar is, en we dus o.a. dankzij de gezelschapsspelen een idee kunnen hebben van het begrip 'regel', is dit alleen maar waar gegeven voldoende vooronderstelde achtergrondkennis die zegt wanneer een bepaalde situatie telt als een situatie van een schaakstuk op een bepaald veld, en gegeven de geconstrueerde eindigheid van het aantal betekenisvolle situaties (een stuk kan slechts op één van de 64 verschillende velden staan). Deze achtergrondkennis kan nooit volstrekt gearticuleerd opgenomen worden in de regels. Niettemin is ze op een beslissende wijze relevant: de situatie waarin er een witte loper op c3 staat is op oneindig veel verschillende manieren realiseerbaar (een houten stuk op een bord, een symbool op een diagram, de notatie Lc3, een bepaalde binaire notatie, een mens op een bepaalde plek op het speelveld van een camping, etc. etc.) Op deze wijze kunnen de spelregels van gezelschapsspelen de indruk wekken regels in intensionele zin te zijn, door het kunstmatig verplaatsen van de oneindigheidscomplicatie naar de achtergrondkennis.

Een andere kunstmatige oplossing van deze problematiek treffen we aan in de wijze waarop we zouden kunnen denken dat het script van een klassiek toneelstuk de regels in extensionele zin bevat die de toneelspelers in hun spel volgen. Immers, de afbakening van het toneelstuk door middel van het podium, het decor en het openen en sluiten van de gordijnen kunnen de indruk wekken dat er sprake is van een eindige verzameling paren waarbij het eerste element een bepaald moment is in een bepaald bedrijf, en het tweede element de handeling die het desbetreffende personage op dat moment moet verrichten. Ook hier is de oneindigheid kunstmatig buitengesteld, iets wat we niet in stand kunnen houden op het moment dat we ons gaan voorstellen hoe Gijsbrecht van Amstel, of Kniertje, of Kilo zich zouden moeten gedragen als zij nu onze kamer binnen zouden komen. Of andersom, als wij ons voorstellen dat ons eigen leven als een toneelstuk is, het script

onaf is, het podium onbegrensd, en niemand weet waar de gordijnen hangen.

De bovenstaande argumentatie lijkt ons te willen overtuigen van twee zeer verstrekkende conclusies: (1) regels bestaan niet; (2) regels volgen is onmogelijk. In de volgende paragraaf² zal ik nader ingaan op de tweede conclusie. In de rest van deze paragraaf zal ik nog wat zeggen over de eerste conclusie.

Het zou dodelijk zijn voor het dramaturgisch model als we zouden moeten concluderen dat regels helemaal niet bestaan. Dat zou immers niet alleen betekenen dat het onacceptabel is te stellen dat handelingssituaties normatief gestructureerd zijn, maar ook dat actoren niet gezamenlijk kunnen beschikken over de middelen die ze nodig hebben om elkaars aandacht op iets te vestigen. En dat zou weer betekenen, als we zouden vasthouden aan de aanname dat handelingen sociale fenomenen zijn, dat handelingen helemaal niet bestaan. In mijn karakterisering van de ernst van de problematiek zijn echter kort twee oneigenlijke, kunstmatige manieren om de problematiek buiten de deur te houden de revue gepasseerd. De eerste manier, die ik toegelicht heb met een verwijzing naar gezelschapsspelen, gaat uit van een intensioneel begrip van regels, en introduceert een geforceerde eindigheid in de mogelijk relevante, betekenisvolle situaties waarin de regel geldt door alle oneindigheid weg te werken in de stilzwijgend vooronderstelde achtergrondkennis. De tweede manier, die ik toegelicht heb met een verwijzing naar het toneelspel, gaat uit van een extensioneel begrip van regels, en introduceert een geforceerde eindigheid in de handelingen die onder de regel vallen, door alle oneindigheid weg te werken achter kunstmatige spatio-temporele grenzen.

Het is precies op dit punt dat verdedigers van de aanname dat handelingen sociale fenomenen zijn de nadruk zullen leggen in hun poging het probleem van de status van normativiteit op te lossen. Zij zullen stellen dat het normatieve lijkt te bestaan bij de gratie van regels. Regels, echter, zo zullen zij gesteund door

² Zie pp. 101-107.

Kripke's interpretatie van Wittgenstein betogen, kunnen zelf niet bestaan, tenzij – en dit is in hun redenering een cruciale tenzij – tenzij mensen er gezamenlijk in slagen conventies in stand te houden: (stilzwijgende) afspraken die normativiteit verlenen aan regels, niet doordat er werkelijk normatieve bepalingen zijn, maar doordat in strikt te identificeren uitgangssituaties sommige opties de expliciete instemming krijgen van allen, of, althans, van de meesten, of, althans, van de relevante autoriteiten. Zo leert een kind schaken, niet doordat er schaakregels zijn die hij voor het geestesoog heeft en tracht te volgen, maar doordat hij leert welke gedragingen in specifieke situaties instemming krijgen van een op het terrein deskundige autoriteit. Deze autoriteit is geen autoriteit doordat hij diepgaand kennis heeft van de regels, maar doordat hij door voldoende relevante anderen gerespecteerd wordt als autoriteit. Deze anderen zijn natuurlijk zelf op geen andere wijze relevante anderen dan doordat ook zij door voldoende relevante anderen gerespecteerd worden als autoriteit. Er is hier geen sprake van een kwalijke circulariteit, maar slechts van het feit dat regels steunen op conventies, d.i. op afspraken tussen gelijkgestemden, en van het feit dat deze conventies zelf op niets anders gebaseerd zijn. Het is precies op dit punt dat we kunnen inzien dat, zoals Finn Collin (1997) betoogt, de Sociale Contracttheorie, die teruggaat op Hobbes, historisch gezien ten grondslag ligt aan de aanname die ik in dit hoofdstuk bespreek.

De stelling van de verdedigers van de aanname dat handelingen sociale fenomenen zijn luidt: sociale conventies liggen onder regels, en niets ligt onder deze conventies. De normatieve structurering van handelingssituaties heeft geen ander bestaan dan een sociaal bestaan. Een handeling is daarom een sociaal fenomeen, omdat het een uiting is van een rol, die gekarakteriseerd wordt door de regels die horen bij een positie. Deze positie is een locatie in een sociaal netwerk, een kwestie van relaties, en het zijn de (stilzwijgende) afspraken die aan dit netwerk ten grondslag liggen (de gedeelde achtergrondkennis en de vaststelling van de cruciale spatio-temporele grenzen) die de in de desbetreffende praktijk geldende regels hun inhoud geven.

Voor ik dit met een voorbeeld zal trachten toe te lichten wil ik er alvast op wijzen dat we verderop³ nog zullen zien dat er een zwakke schakel in bovenstaande redenering zit, die gelokaliseerd kan worden in het onbekommerd functioneel gelijkschakelen van gedeelde achtergrondkennis met het expliciet gezamenlijk vaststellen van te respecteren spatio-temporele grenzen. Voorlopig echter aan dit bezwaar voorbijgaand kan een voorbeeld de pretentie van deze positie verduidelijken. Het heffen van een arm telt tijdens een vergadering op een bepaald moment als een handeling, namelijk als bijvoorbeeld het instemmen met een bepaalde beleidsmaatregel. Later op de gang telt precies dezelfde lichaamsbeweging niet als een handeling, of eventueel niet als het instemmen met een bepaald beleid, maar bijvoorbeeld als een groet, of een poging iemands aandacht te trekken. Het zijn enerzijds de spatio-temporele grenzen (de vergaderzaal, en het tijdstip waarop een stemming aan de orde is) die van cruciaal belang zijn voor de betekenis, en daarmee het bestaan, van deze handeling als handeling. Deze grenzen zijn slechts kunstmatig: gebaseerd op afspraken die zelf zonder fundament zijn. Anderzijds is er de in dit beperkte domein geldende achtergrondkennis die alle betrokkenen laat begrijpen dat de omhoog zwiepende arm een handeling is, het instemmen met het voorgestelde beleid, en niet één of andere ongecontroleerde reflex, zonder betekenis, zoals het zwiepen van de takken van de bomen.

Over rolconflicten en rolafstand

Het tot nu toe ontwikkelde model van handelen als sociaal fenomeen is nog verre van volledig. Wat vooral nog niet voldoende aan bod is gekomen, is een adequate rekenschap van de rol van de actor in het geheel. Het is immers niet voldoende dat de actor gedrag vertoont dat vervolgens iedereen kan begrijpen als in overeenstemming met de regels die horen bij de sociale positie van de actor. We hebben niet te maken met gebeurtenissen waaraan we, als een soort optioneel extra, een sociale betekenis kunnen

³ Onder het kopje "Over regelvolgen en regelvolgers", pp. 103-107.

hechten, maar met handelingen die er slechts zijn dankzij het feit dat ze met het oog op die sociale betekenis verricht worden, of althans getuigen van een inzicht in, het herkennen van en het instemmen met deze sociale betekenis. De actor doet, met andere woorden, wat hij doet vanwege het feit dat hij de rol heeft die hij heeft. En het hebben van een rol is méér dan het bezetten van een positie; het impliceert ook dat de actor de regels kent die horen bij de positie (niet per sé in een zeer expliciete zin van het woord), en dat hij doet wat hij doet omdat (het hoge woord moet er toch uit), omdat hij die regels tracht te volgen, omdat hij de normatieve structuur respecteert die in deze regels hun articulatie vinden.

Daarmee stuiten we op een tweetal problemen die niet zo gemakkelijk op te lossen zullen blijken te zijn. Het eerste is het probleem van het rolconflict. Stel ik ben een politieagent, en 's nachts op wacht betrap ik een inbreker op heterdaad, maar, ai, het is mijn zoon! Wat zal ik doen? Laat ik hem lopen? Geef ik hem een afranseling waarvoor iedere agent langdurig geschorst zou worden? Behandel ik hem als willekeurig iedere andere inbreker die ik arresteer? Er is sprake van een rolconflict. Twee verschillende regels doen tegelijkertijd een beroep op mij, omdat ik in één enkele situatie twee verschillende sociale posities bekleed. Hoewel ik het probleem praktisch zal oplossen door, al delibererend, een inschatting te maken van de kracht en de reikwijdte van de verschillende regels, en zo te komen tot een intentie en bijbehorend gedrag, zal ik het probleem daarmee theoretisch toch nog niet opgelost hebben. En dat voelen we doorgaans ook in situaties waarin er sprake is van een rolconflict. Gegeven een dergelijk rolconflict, en gegeven de aanname dat handelingen sociale fenomenen zijn, zullen we ons op ontologisch niveau realiseren dat het niet anders kan dan dat er in deze situatie noodzakelijkerwijs sprake is van hybride entiteiten. Wat ik ook doe: mijn handeling zal niet eenvoudig begrepen kunnen worden als een uiting van de rol die ik als vader heb, en ook niet als een uiting van de rol die ik als politieagent heb. Het is bij uitstek in dit soort situaties dat we ons ineens pijnlijk bewust worden van onze 'human condition': dat al ons gedrag natuurlijk kunstmatig is, en kunstmatig natuurlijk.

Of, om het zo te zeggen: de regels die we respecteren lijken we juist altijd ook te creëren door ze te respecteren, en in dit soort situaties zijn er geen regels die ik zou kunnen respecteren, zelfs al zou ik het willen.

Zonder onze sociale identiteit lukt het ons niet te zijn wie we zijn, maar soms voelen we ineens de afstand die er is tussen deze sociale identiteit en onze persoonlijke identiteit, waarbij het besef dat we niet samenvallen met onze sociale identiteit ons natuurlijk nog geen inzicht verschaft in onze persoonlijke identiteit. Deze persoonlijke identiteit is slechts een abstracte ruimte die zich toont in het rolconflict, in de incongruentie die er bestaat tussen de regels die twee verschillende rollen ons opleggen.

We zijn hiermee echter wel gestuit op een tweede probleem in de vorm van het thema van de rolafstand. We hebben dit thema eerder al even ontmoet⁴ in de vorm van het niet te elimineren onderscheid tussen enerzijds het personage en anderzijds de acteur. Een persoon is binnen de contouren van het dramaturgisch model een samenstel van beiden: personage en acteur. De persoon valt, met andere woorden, niet samen met het personage dat hij in sociale termen is. De rol die hij heeft is een rol die hij speelt maar niet noodzakelijkerwijs moet vertolken, of moet *zijn*. De manier waarop ik gestalte geef aan mijn rol als vader, of filosoof, is een andere dan de manier waarop mijn buurman dat doet. De rol dicteert de handeling blijkbaar niet, althans niet volstrekt, maar creëert een 'ruimte' waarin iedere acteur zijn of haar eigen versie van de handeling kan realiseren. Dit hoeft overigens niet te betekenen dat de persoon iets anders is achter of voorbij het personage dat hij realiseert. Het gegeven van de rolafstand impliceert niet dat mensen eerst en vooral begrepen moeten worden als acteurs die werkelijk zichzelf zijn als ze geen toneel spelen. Zouden we dat wel willen denken dan zouden we in een sterke zin het in dit hoofdstuk te bespreken model vaarwel moeten zeggen, en moeten concluderen dat handelingen geen sociale fenomenen zijn, maar verricht worden door volstrekt zelfstandige actoren.

⁴ Zie p. 90.

Immers, als mensen zichzelf zijn volstrekt buiten de rollen om die ze spelen, dan zouden handelingen eerst en vooral bestaan uit het aannemen van een rol. Die oorspronkelijke daad (namelijk het *aannemen* van een rol) zou dan de fundamenteel te verklaren handeling zijn, en niet de verrichtingen van het personage die we kunnen begrijpen in termen van de rol die iemand heeft.

Maar hoe kunnen we dan wel een adequate rekenschap geven van wat het is om een personage-en-acteur te zijn? Of, om de aandacht te blijven vestigen op ons onderwerp, hoe kunnen we handelingen begrijpen als fenomenen die gekarakteriseerd moeten worden in termen van sociale regels maar niettemin tevens ruimte laten voor een eigenaardig, persoonlijk accent?

Een mogelijk antwoord voor de verdedigers van de aanname dat handelingen sociale fenomenen zijn maakt gebruik van onze bevindingen betreffende het probleem van het rolconflict. Dit probleem was het gevolg van tegenstrijdige eisen die verschillende, onverenigbare rollen in één en dezelfde situatie aan ons kunnen opleggen. Maar stel nu eens dat mensen allemaal hun eigen manier vinden om verschillende rollen, die meer of minder verenigbaar met elkaar zijn, te integreren. Dan is rolafstand eenvoudig te verklaren in termen van de ruimte die een bepaalde rol laat om tegelijkertijd andere rollen te vertolken. Een actor is een personage-en-acteur omdat hij op zijn eigen wijze een knooppunt vormt in een netwerk van al de verschillende rollen die hij heeft. Als ik met tegenzin instem met een bepaalde beleidsmaatregel, en daartoe met zichtbaar misbaar mijn arm hef tijdens de vergadering, dan telt mijn handeling bijvoorbeeld als een bijdrage aan twee verschillende toneelstukken. Enerzijds is er de vergadering waarin ik een duidelijke handeling heb verricht: ik stem voor de maatregel. In deze zin verschil ik van de tegenstemmers, maar kom ik overeen met een collega van mij die wellicht met volle overtuiging en blijmoedig zijn arm heft omdat hij het een uitstekende beleidsmaatregel vindt. De verschillen tussen mij en deze enthousiaste collega zijn als het ware onzichtbaar als we alleen maar rekening zouden houden met de regels die definiëren wat telt als voorstemmen, en wat als tegenstemmen. Maar deze ver-

schillen kunnen we eenvoudig begrijpen als we ze opvatten als onder andere regels vallend: bijvoorbeeld onder de regels die we volgen als we aan groepsvorming doen, of de regels die we volgen als we willen manipuleren, of de regels die we volgen als we onder onze verantwoordelijkheid uit proberen te komen. Van al deze regels kan gedacht worden dat conventies er aan ten grondslag liggen, dat het de collectieve instemming is die deze regels hun normativiteit geeft. Zo kunnen we handelingen eenvoudig begrijpen als nog steeds volstrekt gedefinieerd in termen van de rollen waar ze een uitdrukking van zijn, zodat het eigenaardige persoonlijke accent dat handelingen onmiskenbaar hebben niets te maken heeft met niet-sociale kenmerken, maar slechts een uitdrukking is van het feit dat de handeling tegelijkertijd onder meerdere, congruente regels valt, omdat ze een uitdrukking is van de verschillende rollen die iemand heeft.

Dit antwoord negeert echter één cruciaal aspect van het thema van de rolafstand. Rolafstand bestaat niet alleen omdat mensen niet samenvallen met de rol die ze hebben, in die zin dat de rol hun handelen zou onderbepalen. Rolafstand bestaat ook daarin dat mensen niet als vanzelf rollen hebben, maar zich evaluatief tot die rollen verhouden. We hebben dit punt al eerder ontmoet bij mijn bespreking van een bepaalde oplossing van het 'onvolledigheids'-probleem, die daarop neerkwam dat een actor zelfstandig kan worden genoemd als hij in staat is een oordeel te vellen over de waarde van zijn eigen verlangens.⁵ Het punt is niet dat actoren volstrekt autonoom hun rollen aannemen of kiezen; maar het punt is wel dat actoren in het spelen van hun rollen zich actief, intentioneel, en (stilzwijgend) instemmend zullen moeten verhouden tot de regels die ze in hun rollenspel trachten te volgen, opdat ze hun rollen kunnen spelen.

In de volgende paragraaf zal ik enkele consequenties hiervan uitwerken en betogen dat het bestaan van regels en het bestaan van regelvolgers twee kanten zijn van dezelfde medaille. Daarmee

⁵ Zie pp. 21-25 onder het kopje "Waarde-oordelen".

zal ik een genuanceerder beeld kunnen geven van wat het nu precies wil zeggen dat handelingen sociale fenomenen zijn, en een ontkoppeling kunnen realiseren tussen het normatieve enerzijds en het sociale anderzijds. Hoewel regels wellicht *qualitate qua* sociale entiteiten zijn, is de normatieve structuur van scenario's die dankzij die regels inzichtelijk gemaakt kunnen worden dat niet. Dit betekent, zoals ik zal betogen, dat de these dat handelingen sociale fenomenen zijn niet impliceert dat handelingen slechts een conventioneel bestaan hebben, een bestaan dat herleid kan worden tot (stilzwijgende) afspraken tussen gelijkgestemden.

2. *Individualisme, holisme, en het normatieve*

Inleiding: Succes en sociaal succes

De strekking van het betoog dat ik in de vorige paragraaf heb gepresenteerd, lijkt te zijn geweest dat het normatieve aspect van handelingen, het aspect dat zichtbaar wordt in het gegeven dat het tot de essentie van handelingen behoort dat ze succesvol kunnen zijn, maar ook kunnen mislukken, gefundeerd is in het sociale karakter van handelingen. De redenering van de verdedigers van de aanname dat handelingen sociale fenomenen zijn heb ik als volgt gepresenteerd: (1) de normativiteit van handelingen bestaat in het feit dat handelingen uitingen zijn van rollen; (2) deze rollen worden gekarakteriseerd door de regels die mensen in hun handelen (impliciet) volgen; en (3) deze regels zijn het resultaat van conventies, van de (stilzwijgende) instemming van relevante anderen met het 'doorgaan' op een bepaalde manier, met het kiezen voor een bepaalde optie in een bepaalde situatie (namelijk, de juiste optie, d.w.z. de optie die relevante anderen ook zouden kiezen).

Dit betoog identificeert het succes van een handeling met het *sociale* succes van een handeling, omdat de regel die een actor volgt in het uitvoeren van een handeling slechts zou bestaan bij de gratie van een conventie, hoe impliciet dan ook. Mij lijkt dit uiteindelijk onverdedigbaar. Immers, om een voorbeeld te gebruiken,

als ik mijn handen was, dan is het succes van deze handeling te meten in termen van het verschil in de hoeveelheid vuil die voor en na het wassen aan mijn handen zit. Dit is uiteindelijk meer dan alleen maar een sociaal feit, zelfs als we een heel sterke interpretatie zouden willen verdedigen van de aanname dat handelingen sociale fenomenen zijn. Zo'n sterke interpretatie zou gebaseerd kunnen zijn op de volgende redenering. (1) De term 'wassen' ontleent zijn betekenis volledig aan de conventionaliteit van de Nederlandse taal. (2) Wat daarom met recht 'wassen' genoemd kan worden, zijn slechts die activiteiten die in overeenstemming zijn met de regels van het Nederlandse taalspel – meer algemeen, met de regels van de Nederlandse levensvorm. (3) Deze regels te herleiden tot de betekenisvolle relaties tussen actoren en hun natuurlijke omgeving, bijvoorbeeld door te spreken over de hoeveelheid vuil op iemands handen en over het vuiloplossend vermogen van water en zeep, werkt niet, omdat daarmee de kwestie slechts verlegd wordt naar andere termen en de daar onder liggende taalspelen en levensvormen. Voor 'vuil' geldt immers wat voor 'wassen' geldt, namelijk dat het zijn betekenis volledig ontleent aan de conventionaliteit van de Nederlandse taal.

Een dergelijke redenering werkt echter niet, omdat ze het al dan niet succesvol zijn van het *benoemen* van een gebeurtenis als een bepaalde handeling niet meer kan onderscheiden van het al dan niet succesvol zijn van de handeling zelf. Als ik mijn handen was, en dit mislukt omdat ze nog steeds vies zijn, dan is het niet zo dat het enige dat mislukt is, het als 'wassen' benoemen is van iets anders dat ik wel heb gedaan, en met succes, bijvoorbeeld mijn handen natgemaakt. Een acteur die een rol slecht speelt, doet iets anders dan een acteur die de verkeerde rol (goed) speelt. Iemand die reglementair correct schaakt, maar erg onverantwoord omgaat met zijn koningin, schaakt gewoon slecht, en is niet erg goed in een ander spel ("Hoe verpruts ik mijn koningin zo snel mogelijk"). Het is één ding om van mening te verschillen over het feit of een bepaalde gebeurtenis 'handen wassen' genoemd kan worden, en een ander feit of een bepaalde instantiatie van 'handen wassen' een goede was of niet. Dit wil eigenlijk niet meer zeggen

dan dat er over het succes van het handen wassen van mening verschild kan worden, wat eigenlijk alleen maar kan als er een feit is waarover van mening wordt verschild. En zo'n feit is er: na het wassen van mijn handen zal de toestand waarin mijn handen verkeren een bepaalde 'vuilwaarde' hebben, en zelfs als ook deze 'vuilwaarde' een intrinsiek relationele eigenschap van mijn handen is (een eigenschap die mijn handen alleen maar hebben omdat ze zich verhouden tot bepaalde maatstaven die in onze cultuur gelden), zal de 'oppervlaktekwaliteit' van mijn handen een bepaalde interne structuur hebben dankzij welke mijn handen in staat zijn de relevante relationele eigenschap te hebben. Al deze opmerkingen komen slechts hierop neer: *dat* er van mening kan worden verschild over het al dan niet succesvol zijn van het wassen van de handen, impliceert dat er meerdere betrokkenen zijn die in principe ter zake kundig zijn (omdat ze kennis kunnen nemen van de relevante interne structuur van de 'oppervlaktekwaliteit' van mijn handen, en van de relevante in onze cultuur voor de 'vuilwaarde' van handen geldende maatstaven), *onafhankelijk* van de instemming van anderen. Of, om het zo te zeggen, en daarover zullen de volgende bladzijden gaan, de instemming van anderen kan niet logisch voorafgaan aan de identificatie van de regel die men tracht te volgen.

Over regelvolgen en regelvolgers

In de vorige paragraaf heb ik het probleem van het regelvolgen geïntroduceerd, en heb ik een argument gegeven op grond waarvan we zouden moeten concluderen dat (1) regels niet kunnen bestaan, en (2) regels niet te volgen zijn.⁶ We zijn, dankzij de interventie van het sociale, ontkomen aan de eerste conclusie. Althans zo leek het: omdat mensen afspraken kunnen maken bestaan er binnen bepaalde spatio-temporele grenzen waarin een bepaald aantal uitgangssituaties als betekenisvol erkend wordt, regels die in alle (een eindig aantal) gevallen aangeven wat de correcte manier is om 'door te gaan'. Deze sociaal gerealiseerde 'regels' kunnen we vervolgens gemakkelijk volgen, mits – en dit is

⁶ Zie hierboven, pp. 91-94.

een cruciale mits – mits we de geïnstitutionaliseerde spatio-temporele grenzen en de geïnstitutionaliseerde uitgangssituaties als zodanig herkennen. Maar daarin zit hem natuurlijk het probleem: hoe die te herkennen gegeven het oneindig aantal mogelijke uitgangssituaties en de oneindige uitgebreidheid en deelbaarheid van het spatio-temporele domein. Het lijkt erop alsof we regels moeten kunnen volgen om afspraken te kunnen maken en in stand te kunnen houden.

Dit is het werkelijke probleem van het regelvolgen. En een interessante oplossing er voor is bedacht door Philip Pettit. Zijn oplossing, waarvan ik hier alleen maar de strekking in een zeer vrije interpretatie weergeef, vertrekt vanuit de aanname dat wij, mensen, eigenaardige subjecten zijn, in die zin dat iedere situatie waarin we verkeren weliswaar in principe een instantiatie is van een oneindig aantal regels, maar voor ons een exemplificatie is van slechts een beperkt aantal regels. De oneindigheid wordt met andere woorden niet buiten de deur gehouden doordat we, geconfronteerd met oneindig veel mogelijkheden, met elkaar afspraken maken. De oneindigheid wordt als het ware door onze constitutie al buiten de deur gehouden, doordat we om te beginnen helemaal niet met oneindigheid geconfronteerd worden, maar met een eindigheid die past bij onze aard.

Deze oplossing is in feite op Kant geïnspireerd, waarbij overigens Humeaanse aspecten een hoofdrol spelen. De idee is dat mensen nu eenmaal de neiging hebben op een bepaalde, min of meer stereotype manier te reageren op een reeks voorbeelden, en wel op een dusdanige manier dat we iedere reeks voorbeelden al snel behandelen alsof ze slechts één bepaalde regel exemplificeren. We zijn, om het zo te zeggen, eerst en vooral regelvolgers. Geconfronteerd met een tomaat, een brandweerauto, een druppel bloed, en een oude Engelse telefooncel zullen we al gauw de gewoonte ontwikkelen deze volgens kleur te classificeren, en zullen we al gauw de neiging krijgen andere rode objecten te herkennen als behorend tot dezelfde klasse. Natuurlijk zijn er in principe oneindig veel regels denkbaar, naast de regel ‘rood’, met behulp waarvan deze reeks op dezelfde wijze geclassificeerd zou kunnen worden,

maar de moeilijkheid die ik nu heb om ook maar één van dit oneindig aantal regels ter plekke te verzinnen, is al een duidelijk teken dat er op wijst dat sommige gewoonten zo ontzettend veel meer voor de hand liggen dan andere.

Niets garandeert mij natuurlijk dat ik me niet vergis in het ontwikkelen van mijn gewoonten, of in het spontaan krijgen van een bepaalde neiging in een bepaald geval. In het begin van mijn carrière als regelvolger hier op aarde accepteer ik de correcties die anderen (mijn ouders, de juf, de docent) maken, en zoek ook expliciet hun instemming. Ik kan, bijvoorbeeld, al zelf mijn handen wassen (!) wat ik natuurlijk pas geloof als ik de instemming krijg van mijn ouders. Maar na verloop van tijd krijg ik als het ware zelf de smaak te pakken, en besef dat mijn inzicht in de betreffende regels, d.i. de waarde van mijn neigingen, niet afhankelijk is van de instemming van anderen, maar zelf een voorwaarde is voor het feit dat instemming instemming is. Ik ga me na verloop van tijd realiseren dat mijn neigingen in principe gezaghebbend en relevant zijn. Dit is een proces dat langzaam verloopt. Wie herinnert zich niet de aarzeling bij het ontdekken van een drukfout in een buitenlands boek. Je zult wel gewoon niet begrijpen wat er staat, want wie ben jij om te denken dat er een fout staat in een Duits leerboek. Echt een boek (!) waar vele volwassenen aan hebben bijgedragen, dat ver weg van hier uitgegeven is, uitgezocht en besteld door mensen die er verstand van hebben, en dan denk jij ... Dus je zoekt er de woordenboeken en je schriften nog eens op na. Als die geen uitsluitel bieden, zul je op je eigen oordeel af moeten gaan, en dan volgt ongetwijfeld die eerste keer dat je er van overtuigd bent, zonder het na te hoeven slaan: daar staat een drukfout! (Na de val van het boek en de leraar Duits volgt, natuurlijk, al spoedig de ondergang van je juf en je ouders – in die achterwaartse volgorde.)

Met de ontwikkeling van je neigingen als, in principe, gezaghebbend en relevant ontwikkelt zich ook een tweede vaardigheid die kenmerkend is voor ons als regelvolgers: we leren met elkaar onderhandelen over onze verschillende inzichten in de regels, waarbij we gericht zullen zijn op de convergentie van onze neigingen.

Deze vaardigheid zal zich moeten ontwikkelen, omdat we zullen gaan ontdekken dat al onze neigingen (dus niet alleen maar mijn neigingen, maar ook die van al mijn medemensen) gebaseerd zijn op de vooronderstelling dat ons perspectief (wederom, niet alleen maar mijn perspectief, maar ook dat van anderen) normaal is, wat niet alleen maar een prescriptieve, maar ook een descriptieve betekenis heeft. In prescriptieve zin nemen we aan dat ons perspectief 'normaal' is in de zin van normstellend. In descriptieve zin nemen we aan dat ons perspectief 'normaal' is in de zin van gewoontegetrouw. Het ontdekken van divergentie (Is turkoois nu blauw of groen? Lacht ze me uit of toe? Zijn mijn handen schoon of niet?) impliceert daarom het ontdekken van een normatief significante ruimte tussen wat ik gewoontegetrouw juist vind en wat normstellend juist is. Zodra we tot convergentie komen, blijken de neigingen die ik gewoontegetrouw heb, ook weer te getuigen van de regels en hun normstellende karakter. Zodra we het eens zijn, zijn er regels die we volgen, omdat, en voor zover, we neigingen hebbende regelvolgers zijn. Het is in deze zin dat convergentie ten grondslag ligt aan de regels die van onze handelingen sociale fenomenen, en daarmee handelingen, maken.

Laat het echter duidelijk zijn dat het niet de conventies zijn die ten grondslag liggen aan het bestaan van regels, of aan het kunnen volgen van regels. Wat ten grondslag ligt aan de regels, en aan de conventies, is het feit dat wij wezenlijk regelvolgers zijn. Het is de systematiek in onze neigingen, en de mogelijkheid die we hebben ons evaluatief te verhouden tot deze systematiek (gegeven de ruimte die we ontdekken tussen wat gewoontegetrouw juist en wat normstellend juist is), die ten grondslag ligt aan het normatieve dat zo bepalend is voor de eigen aard van ons handelen.

Ik kan dit ook zo zeggen: het is verdedigbaar te beweren dat regels *qualitate qua* sociaal zijn. Pas als er convergentie is, en wellicht betekent dat: pas als er op basis van onderhandelingen afspraken gemaakt zijn over de juiste manier om in een bepaalde situatie 'door te gaan', is er sprake van regels die een constituerende functie hebben, en pas dankzij deze regels kunnen handelingen verricht worden die uitdrukking geven aan de rol die een

actor heeft. Dit betekent echter niet dat ook de normatieve structuur van handelingssituaties *qualitate qua* sociaal is. Deze normatieve structuur is gelegen in de kwaliteit van een situatie die tot uitdrukking komt in de neiging die iemand heeft die met deze kwaliteit geconfronteerd wordt. Deze kwaliteit kan niet sociaal zijn in de zin dat ze pas tot realiteit komt als gevolg van een conventie of afspraak. Of, wat eigenlijk juister is, niet iedere kwaliteit die een bepaalde situatie haar normatieve structuur geeft, kan op deze wijze haar oorsprong hebben in conventies of afspraken. Dat zou de hele onderneming immers beroven van zijn mogelijkhedenvoorwaarde. Sommige neigingen moeten, als articulaties van de normatieve structuur van een bepaalde situatie, ten grondslag liggen aan de ontdekking van divergentie die eerst pas onderhandelingen, en conventies, nodig, en mogelijk, maken.

Daarmee hebben we de stelling dat handelingen sociale fenomenen zijn een voldoende genuanceerde inhoud gegeven om een laatste kwestie aan de orde te stellen. Dit is de kwestie of we in het verklaren van handelingen volgens dit model slechts een beroep behoren of behoeven te doen op de anonieme sociale structuren die ik in de vorige paragraaf beschreven heb met behulp van de begrippen 'rol', 'script' en 'regel', of altijd ook een beroep moeten of kunnen doen op de individuele psyche van de betrokken actoren.

Verklaringen in de sociale wetenschappen

De kwestie van de aard van de verklaringen die in de sociale wetenschappen op hun plaats zijn, is een kwestie die we in verschillende gedaanten al een aantal malen tegengekomen zijn. Ze heeft te maken met het moeizame onderscheid tussen motivatie en rechtvaardiging in een analyse van de rol die deliberatie speelt in het tot stand komen van handelingen.⁷ En ze speelt ook een hoofdrol in de *Erklären-Verstehen* controverse die ik in de eerste paragraaf van het vorige hoofdstuk al kort besproken heb.⁸ Hoewel de kwestie een methodologische is (namelijk: hoe moeten we

⁷ Zie pp. 50-54 onder het kopje "Motiveren en rechtvaardigen".

⁸ Zie pp. 60-62.

sociale wetenschap bedrijven?), lijkt er in de literatuur voortdurend over gediscussieerd te worden als gaat het om een ontologische kwestie (namelijk: wat zijn handelingen als het sociale fenomenen zijn?). Ik zal in deze paragraaf betogen dat het belangrijk is de kwestie te beschouwen als slechts een methodologische, om juist daardoor een aantal ontologische implicaties van de these dat handelingen sociale fenomenen zijn, recht te kunnen doen.

Te zeggen dat de kwestie een methodologische is, betekent eerst en vooral te wijzen op de epistemische behoeften van de sociale wetenschapper. Immers, verklaringen zijn altijd gesitueerde zaken, gebonden aan specifieke kennisbelangen. Wat voor de één een verklaring is, is dat voor de ander niet. Dit komt doordat een verklaring nooit alleen maar al dan niet accuraat is (op een al dan niet correcte manier gaat over het te verklaren fenomeen), maar ook altijd al dan niet gepast (al dan niet afgestemd is op de specifieke kennisbehoeften van degene voor wie de verklaring een verklaring pretendeert te zijn). Het onderscheid tussen de accuratesse van een verklaring en haar gepastheid is een cruciaal onderscheid als we meer willen begrijpen van het onderscheid tussen de ontologische kwestie (welke noties moeten in ieder geval gebruikt worden in een juiste beschrijving van het te verklaren fenomeen?) en de methodologische kwestie (welke onderdelen uit de causale geschiedenis van het fenomeen mogen niet ontbreken in een geldige verklaring van het te verklaren fenomeen?) die vaak verward worden in de discussies over de aard van sociaal-wetenschappelijke verklaringen.

Met betrekking tot de ontologische kwestie heb ik met het oog op de stelling die in dit hoofdstuk ter discussie staat in de vorige paragraaf een aantal begrippen geïntroduceerd: rol, personage-en-acteur, regels, script. Maar ook heb ik duidelijk gemaakt dat deze begrippen, die de nadruk leggen op de cruciale betekenis van sociale structuren, het functioneren van regelvolgende, delibererende actoren vooronderstellen.⁹ Daarmee blijken we niet te kunnen

⁹ Zie de discussies onder de kopjes "Over regelvolgen en regelvolgers" en "Over rolconflicten en rolafstand", resp. pp. 103-107 en pp. 96-101.

uitsluiten dat ook de begrippen die ik in de tweede paragraaf van Hoofdstuk 1 geïntroduceerd heb, relevant zijn: rationele principes, opvattingen, pro-attitudes, intenties, motivaties, rechtvaardigingen.

Al deze begrippen lijkt een plaats toe te komen in correcte sociaal-wetenschappelijke verklaringen, en als we het daarbij laten gaan we begrijpen waarom het kan lijken alsof deze overwegingen ons dwingen een debat aan te gaan over het relatieve belang van deze verschillende begrippen. Immers, het lijkt er op alsof we niet al deze begrippen tegelijkertijd even belangrijk kunnen vinden. Het lijkt erop alsof we te maken hebben met verschillende verzamelingen, waarbij de ene verzameling gebruikt kan worden om de functie van de andere te verklaren, en *vice versa*. Het lijkt erop, alsof we enerzijds sociale structuren nodig hebben om handelingen te verklaren, maar anderzijds handelende actoren nodig hebben om die sociale structuren te verklaren.

Deze tegenstelling is echter het resultaat van het feit dat men zich niet realiseert dat verklaringen methodologische zaken zijn, en als zodanig hun oorsprong altijd hebben in een gesitueerde epistemische behoefte. Het Hegeliaanse idee van de ultieme verklaring, die niets meer en niets minder is dan een volstreekte recapitulatie van de wordingsgeschiedenis van een fenomeen (het zijn van het fenomeen present gesteld in het denken van het idee), is wellicht debet aan deze vergissing. Want zodra we ons realiseren dat een verklaring aan accuratesse niet meer hoeft te bevatten dan nodig is om gepast te zijn; zodra we ons realiseren dat een verklaring altijd een antwoord is op een specifieke vraag, die als vraag allerlei zaken als onproblematisch gegeven impliceert; zodra we beseffen dat de kwestie een methodologische is, zullen we inzien dat we helemaal niet hoeven kiezen.

Dit inzicht kan duidelijk gemaakt worden door een onderscheid te maken tussen vergelijkende en contrasterende informatie. Soms hebben we het meest aan informatie die vergelijkend van aard is, die datgene in beeld brengt waarin de causale geschiedenis die geleid heeft tot het fenomeen in kwestie overeenkomt met de causale geschiedenis die leidt tot soortgelijke fenomenen. Dus, als we willen weten waarom Klaas tegenwoordig een stropdas draagt,

zou het kunnen dat we de beste verklaring geven (de meest gepaste informatie verschaffen) door het te hebben over de ongeschreven wetten waaraan de kleding moet voldoen van personeel werkzaam in de automatisering. Dat Klaas tegenwoordig als IT-specialist werkt, en nu dus een bepaalde *rol* heeft, verklaart zijn gedrag doordat het een significante vergelijking mogelijk maakt tussen het gedrag van Klaas en dat van zijn talrijke collegae.

Het zou echter ook kunnen dat we willen weten waarom juist *Klaas* een stropdas is gaan dragen – Klaas die altijd zo wars was van kleetconventies. En dan zullen we de beste verklaring geven door het verstrekken van contrasterende informatie: informatie over het verschil tussen de beweegredenen die Klaas nu heeft en de beweegredenen die Klaas al die tijd daarvoor had, of tussen zijn beweegredenen en die van anderen in eenzelfde situatie. Dan geven we een verklaring die geen beroep doet op het (in de causale geschiedenis wel degelijk relevante) gegeven dat Klaas tegenwoordig een bepaalde rol heeft, maar die juist een beroep doet op de (eveneens in diezelfde causale geschiedenis relevante) overwegingen die Klaas ertoe brengen een bepaalde rol te spelen, d.w.z. bepaalde regels te (willen) volgen. En het zal duidelijk zijn dat het in deze verklaring niet zal gaan over de algemene rol van IT-specialist, maar over de meer specifieke rol van een qua kleding onopvallende, zich conformerende IT-specialist.

Het spreekt voor zich dat talrijke andere contrasterende en vergelijkende informatie gegeven kan worden over het gedrag van Klaas. En eenzelfde reeks verhalen laat zich eenvoudig vertellen over de geschiedenis van Karel die naar de keuken ging om een pilsje te halen.¹⁰ Er is, als we het in het algemeen over verklaringen in de sociale wetenschappen hebben, geen enkele noodzaak te kiezen voor een eenzijdig individualisme of een eenzijdig holisme.

Zo zijn een aantal lijnen die ik in de dit en de voorgaande twee hoofdstukken uitgezet heb, in deze laatste subparagraaf bij elkaar gekomen. Ik hoop dat er door duidelijk is geworden dat de

¹⁰ Zie pp. 53-54.

verhouding tussen het sociale en het normatieve om belangrijke nuances vraagt. Het normatieve lijkt me cruciaal voor een plausible filosofie van het handelen, en het lijkt me duidelijk dat het niet samenvalt met het sociale.

Ik verwacht dat belangrijke ontwikkelingen in de analytische filosofie van het handelen in de nabije toekomst gedragen zullen worden door een hernieuwde, en geïntegreerde, belangstelling voor de normatieve dimensie van het handelen. Het is eigenlijk deze dimensie die zichtbaar wordt in het onderscheid tussen motiveren en rechtvaardigen, en die in een discussie over de sociale aspecten van het handelen niet voldoende tot haar recht komt, net zo min als in de discussies over de fysieke en mentale aspecten van het handelen.

Nawoord

In de voorgaande hoofdstukken heb ik drie modellen van het menselijk handelen besproken, die articulaties zijn van drie aannames: dat handelingen (1) gebeurtenissen zijn, (2) verricht worden door actoren, en (3) sociale fenomenen zijn. Ik heb deze aannames in de inleiding inzichten genoemd die zowel aan het begin als aan het eind staan van een intellectuele reflectie op de dagelijkse praktijk, inzichten die aanvankelijk vergezeld gaan van een uitroepteken maar eindigen met een vraagteken. Natuurlijk was dat beeldspraak die we niet zomaar, zonder argumentatieve onderbouwing, moeten accepteren. En natuurlijk klonk dat mooi – het wekte immers de indruk dat ook de filosofie van het handelen een nooit eindigende onderneming is waarvan de vruchten niet aan het eind van de route geïncasseerd kunnen worden, maar onderweg genuttigd moeten worden, voordat ze rotten.

Maar is het nu ook gebleken dat deze drie aannames inderdaad aanleiding hebben gegeven tot onderzoeksprogramma's die resulteren in het ter discussie stellen van de aannames zelf? Te beweren dat dit inderdaad *gebleken* is, zou teveel gezegd zijn. Het zou te veel afstand tot de thematiek en de problematiek vooronderstellen. Het zou te veel arrogante vrijblijvendheid impliceren, alsof men het allemaal al eens gezien heeft en inmiddels wel weet dat er niets nieuws onder de zon is, zelfs niet, of vooral niet, in de analytische filosofie van het handelen.

Het lijkt mij dat duidelijk is geworden dat de drie modellen nog steeds een uitroepteken verdienen. Waarschijnlijk zijn de drie aannames waarop ze gebaseerd zijn ook gewoon alle drie waar. Natuurlijk zijn handelingen gebeurtenissen, worden ze verricht door actoren, en zijn ze sociale fenomenen! Maar juist als we ze op deze vanzelfsprekende wijze accepteren, blijkt dat de discussies over de implicaties en consequenties van deze aannames ons, behalve heel veel verheldering, ook vraagtekens opleveren.

Want wat is verrichten? Wat wil het zeggen dat een actor zelf iets doet? Wat behelst de zelfstandigheid van actoren? Hoe verhouden motivaties en rechtvaardigingen zich tot elkaar? Wat is deliberatie?

En, vervolgens, wat zijn gebeurtenissen? Hoe verhouden gebeurtenissen zich precies tot hun causale verklaringen? Wat zijn de implicaties van naturalisme en determinisme? Wat zijn redenen als het vooral oorzaken zijn?

Tenslotte, wanneer is iets sociaal? Hoe fundamenteel zijn conventies? Wat ligt er achter conventies? Wat maakt convergentie mogelijk? Wat zijn regels? Wanneer telt iets als succes?

Het is de precisering van onze vragen die het succes van een filosofische onderneming uitmaakt. Daarin bestaat dan ook het belang van een aanname: dat ze ons epistemische verlangen stuurt, dat ze ons helpt meer vragen te stellen dan alleen maar die ene grote: wat zijn handelingen?

Daarbij is het belangrijk dat er verschillende modellen zijn, modellen die niet alleen maar verschillende aspecten van het menselijk handelen in kaart trachten te brengen, maar modellen die op een heel fundamenteel niveau incommensurabel zijn, modellen die echte concurrenten van elkaar zijn omdat ze op één of andere manier over hetzelfde gaan maar tevens niet in elkaar vertaalbaar zijn. Dit is belangrijk omdat de enige toegang die we tot handelingen hebben een toegang is via een model. Anders dan in het geval van iemand die van luciferhoutjes een model van de Utrechtse Domtoren wil maken, is er in het geval van de filosofie van het handelen geen sprake van een model dat letterlijk naast het origineel kan worden gelegd. We kunnen het model in die zin nooit vergelijken met het origineel om te beoordelen in hoeverre het geslaagd is. Een model van het handelen lijkt dan wellicht meer op een model dat architecten soms maken van een nog niet bestaand gebouw, een model waarmee ze de opdrachtgevers trachten te verleiden met hun in zee te gaan. Ook in zo'n geval is het succes van het model geen kwestie van de mate waarin het lijkt op het origineel. Maar ook in zo'n geval is het verstandig verschillende architecten verschillende modellen te laten ontwerpen,

omdat het succes van het model een kwestie zal zijn van de mate waarin het model tegemoet komt aan de wensen van de opdrachtgevers. De concurrentie tussen de modellen helpt de opdrachtgever wellicht niet in het vergemakkelijken van zijn keuze, maar helpt hem wel bij het beter articuleren van zijn eigen wensen en behoeften.

Deze vergelijking werkt voor zover ze duidelijk maakt wat de verdienste is van het feit dat er verschillende modellen zijn. Maar ze werkt wellicht niet zo heel goed voor zover we ons realiseren dat er, anders dan in het geval van een nieuw te bouwen huis, in de filosofie van het handelen sprake lijkt te zijn van een object waar we nu al overal om ons heen voortdurend mee geconfronteerd worden. Hoe is het mogelijk dat iets er op zo evidente wijze gewoon is, en we toch overtuigend kunnen beweren dat de modellen die we ervan hebben geen weergaven zijn maar oorspronkelijke structureringen, of wellicht asymptotische benaderingen?

Dit heeft alles te maken met een kwestie waar we al een aantal malen op gestoten zijn¹: de kwestie van de verhouding tussen de ontologische vooronderstellingen en de ontologische implicaties van verklaringen. De kwestie is deze: verklaringen zijn altijd gesitueerd, zijn altijd reacties op een bepaalde epistemische behoefte. Zo'n behoefte ontstaat altijd in een situatie waarin er van alles voorondersteld wordt, waarin er allerlei aannames als vanzelfsprekend geaccepteerd worden, met inbegrip van de wellicht niet altijd zo doorzichtige ontologische implicaties die die aannames met zich meebrengen. Een verklaring is daarom altijd alleen maar betekenisvol tegen de achtergrond van een wereld die niet ter discussie staat. Maar deze achtergrondwereld is natuurlijk in de filosofie nooit verheven boven iedere twijfel, wat sommigen ook mogen denken over de betekenis van het onderscheid tussen het Manifeste en het Wetenschappelijke Wereldbeeld. Vraagtekens bij een gegeven verklaring kunnen daarom altijd ineens weer vraagtekens

¹ Ik doel op een kwestie die in ieder hoofdstuk al aan de orde is gekomen. Zie de discussies op pp. 50-54 onder het kopje "Motiveren en rechtvaardigen", pp. 81-83 onder het kopje "Verklaringen en oorzaken 'zijn' één", en pp. 107-111 onder het kopje "Verklaringen in de sociale wetenschappen".

blijken te zijn bij de aannames, en hun implicaties, die de verklaring eerst als verklaring mogelijk hebben gemaakt. We kunnen dit, met een knipoog naar Plato, ook zo zeggen: in de filosofie zijn er geen originelen, alleen maar fenomenen. In de filosofie is het alsof we ons op een oneindig groot en leeg strand bevinden en spelen dat het een huis is: stel dat hier de deur is, en daar de bank staat, waar is dan de keuken? Of, andersom, in de filosofie is het alsof we ons in een groot huis vinden, met allemaal deuren, en ramen, en muren, en kamertjes, en we niets anders doen dan ons afvragen of we al die onderscheidingen niet kunnen missen: stel dat we deze muur nu eens weghalen, en dan daar een deur maken, zouden we hier dan niet beter wonen?

Vandaar dat we in de analytische filosofie van het handelen ons voordeel moeten, en kunnen, doen met het bestaan van verschillende modellen. Want het is best mogelijk dat we in een poging het probleem van mentale veroorzaking op te lossen ineens stuiten op een hoopvolle conceptie van 'actor-veroorzaking', waarna het evenzeer weer goed denkbaar is dat we in een poging de cruciale rol van deliberatie in kaart te brengen ineens stuiten op de veronachtzaamde relevantie van intermenselijke verhoudingen en de daarbij behorende normativiteit, waarna het evenzeer weer goed denkbaar is dat we in een poging grip te krijgen op de sociale aard van regels ineens weer stuiten op een veelbelovend inzicht in de cruciale rol van de vaste patronen waar onze lichamelijke constitutie aanleiding toe geeft, waarna het evenzeer weer goed denkbaar is dat we in een poging de systematiek van onze gewoonten te doorgronden ineens weer stuiten op onze individuele zelfstandigheid, waarna ... etc.

Literatuursuggesties

In voorgaand overzicht zijn nauwelijks literatuurverwijzingen opgenomen. Dit lijkt mij een slechte zaak voor een tekst met wetenschappelijke pretenties, maar een goede zaak voor een inleidende tekst als deze, zeker waar het de analytische filosofie betreft waarin filosofen en hoofdwerken niet erg prominent functioneren als dominante herkenningpunten. In wat nu volgt zal ik allerlei literatuursuggesties doen waarbij ik nauw zal aansluiten bij de opdeling van de voorgaande hoofdstukken. Daarmee heb ik twee doelen op het oog: (1) wie op een specifiek punt meer wil lezen, krijgt een duidelijke handleiding; (2) wie geïnteresseerd is in de ‘bemsing’ van het ideeënlandschap dat ik hierboven geschetst heb, krijgt op deze manier een aardig beeld van welke filosofen zich nu waar bevinden in de hedendaagse analytische filosofie van het handelen.

Inleiding: Wat gebeurt er als iemand handelt

In de tekst heb ik verwezen naar:

Anscombe, G.E.M., *Intention*, Oxford: Blackwell, 1957.

Ryle, Gilbert, *The Concept of Mind*, London: Hutchinson, 1949.

Velleman, David, “What Happens When Someone Acts”, in *Mind*, Vol. 101, pp. 461-481, 1992.

Davidson, Donald, “Actions, Reasons and Causes”, in *Journal of Philosophy*, Vol. 60, pp. 685-700, 1963.

Klassieke teksten die naast die van Davidson genoemd moeten worden, omdat ze een groot effect hebben gehad op de ontwikkeling van het tweede en derde model van handelingen zijn resp.:

Frankfurt, Harry, “Freedom of the Will and the Concept of a Person”, in *Journal of Philosophy*, Vol. 68, pp. 5-20, 1971.

Strawson, Peter, “Freedom and Resentment”, in *Proceedings of the British Academy*, pp. 1-25, 1962.

De teksten van Strawson en Frankfurt zijn, samen met nog 10 zeer invloedrijke artikelen, herdrukt in de uitstekende bundel:

Watson, Gary (ed), *Free Will*, Oxford: Oxford University Press, 1982.

Een heel erg goede inleiding in de hedendaagse analytische filosofie van het handelen, ook al is hij wat eenzijdig gecentreerd rondom het probleem van de vrije wil en natuurlijk ook zeer duidelijk herkenbaar als een typisch product van zijn auteur, vind ik nog altijd:

Dennett, Daniel C., *Elbow Room. The Varieties of Free Will Worth Wanting*, Oxford: Clarendon Press, 1984.

Een goed en actueel overzicht wordt geboden door:

Mele, Alfred L. (ed) *The Philosophy of Action*, Oxford: Oxford University Press, 1997.

Compatibilisme

In de tekst heb ik verwezen naar:

Nagel, Thomas, "Moral Luck", in *Proceedings of the Aristotelian Society*, Supp. Vol 50 (1976). Herdrukt in Thomas Nagel, *Mortal Questions*, Cambridge: Cambridge University Press, 1979

Chisholm, Roderick, "Human Freedom and the Self", in *The Lindley Lecture*. University of Kansas, 1964.

Beide teksten zijn herdrukt in de uitstekende bundel:

Watson, Gary (ed), *Free Will*, Oxford: Oxford University Press, 1982.

Een tweetal recente publicaties waarin goede overzichten te vinden zijn van de omvangrijke discussie over de vrije wil/determinisme kwestie, zijn:

Fisher, John M. & **Ravizza**, Mark (eds.), *Perspectives on Moral Responsibility*. Ithaca: Cornell University Press, 1993.

Wallace, R. Jay, *Responsibility and the Moral Sentiments*. Cambridge (Mass.): Harvard University Press, 1994.

Het hiërarchische model van zelfstandigheid

In de tekst heb ik verwezen naar:

Frankfurt, Harry, "Freedom of the Will and the Concept of a Person", *Journal of Philosophy*, Vol. 68, pp. 5-20, 1971. Herdrukt in de bundel van Watson.

Danto, Arthur, "Basic Actions", in *American Philosophical Quarterly*, pp. 141-48, 1965.

Christman, John (ed), *The Inner Citadel. Essays on Individual Autonomy*, Oxford: Oxford University Press, 1989.

Dworkin, Gerald, "The Concept of Autonomy", in R. Haller, *Science and Ethics*, Amsterdam: Rodopi Press, 1981. Herdrukt in de bundel van Christman.

Ook van groot belang zijn de artikelen gebundeld in:

Frankfurt, Harry, *The importance of what we care about*, Cambridge: Cambridge University Press, 1988.

Waarde-oordelen

In de tekst heb ik verwezen naar:

Watson, Gary, "Free Agency", in *Journal of Philosophy*, pp. 205-220, 1975. Herdrukt in de bundel van Watson.

Taylor, Charles, *Human Agency and Language. Philosophical Papers I*, Cambridge: Cambridge University Press, 1985.

Friedman, Marilyn, "Autonomy and the Split-Level Self", in *The Southern Journal of Philosophy*, pp. 19-35, 1986.

Over de compartimentalisering van de persoon treft men ook veel waardevoels aan in:

Davidson, Donald, "Paradoxes of Irrationality", in Wollheim, R. et.al. (eds.) *Philosophical Essays on Freud*. Cambridge: Cambridge University Press, 1982.

Pears, David, *Motivated Irrationality*, Oxford: Clarendon Press. 1984.

Een goede Nederlandstalige monografie over het werk van Charles Taylor, één van de grootste filosofen van dit moment, is:

Cuypers, Stefaan & **Lemmens**, Willem (eds.), *Charles Taylor. Een mozaiek van zijn denken*. Kapellen/Kampen: Pelckmans/Kok Agora, 1997.

Coherentie

In de tekst heb ik verwezen naar:

Taylor, Charles, *Sources of the Self*, Cambridge: Cambridge University Press, 1989.

Gadamer, Hans-Georg, *Wahrheit und Methode*, Tübingen: Mohr, 1960.

Rawls, John, *A Theory of Justice*, Oxford: Oxford University Press, 1972.

Een informatieve bundel over de rol van het reflectief equilibrium in het debat over de structuur van normatieve argumentatie is:

Van der Burg, Wibren, & **Van Willigenburg**, Theo, (eds.), *Reflective Equilibrium*, Dordrecht: Kluwer Academic Publishers, 1998.

De besproken positie wordt verdedigd in:

Ekstrom, Laura, "A Coherence Theory of Autonomy", *Philosophy and Phenomenological Research*, 53(3), 599-616, 1993.

Christman, John, "Autonomy: a Defense of the Split-Level Self", in *The Southern Journal of Philosophy*, pp. 281-293, 1987.

Friedman, Marilyn, "Autonomy and the Split-Level Self", in *The Southern Journal of Philosophy*, pp. 19-35, 1986.

Het argument tegen de positie, en de vergelijking met de problemen waar een procedurele ethiek mee worstelt, zijn gebaseerd op Pettits bespreking van het contract-denken in:

Pettit, Philip, *The Common Mind. An Essay on Psychology, Society, and Politics*, Oxford: Oxford University Press, 1993.

De actor-identificerende handeling

Zie voor een uitwerking van de besproken positie:

Frankfurt, Harry, *The importance of what we care about*, Cambridge: Cambridge University Press, 1988.

De bespreking is gebaseerd op:

Bransen, Jan, "Identification and the Idea of an Alternative of Oneself", in *European Journal of Philosophy*, pp. 1-16, 1996.

Hoe belangrijk is zelfstandigheid?

De term 'space of reasons' komt van Wilfrid Sellars en is hernieuwd onder de aandacht gebracht door John McDowell:

Sellars, Wilfrid, *Science, Perception and Reality*, London: Routledge & Kegan Paul, 1963.

McDowell, John, *Mind and World*, Cambridge MA: Harvard University Press, 1994.

Praktische rationaliteit

Een uitstekende inleiding in de geschiedenis en de systematiek van het praktisch redeneren is te vinden in:

Audi, Robert, *Practical Reasoning*, London: Routledge, 1989.

Zie ook van dezelfde auteur:

Audi, Robert, *Action, Intention, and Reason*, Ithaca: Cornell University Press, 1993.

Een toegankelijke, zeer originele en boeiende studie is:

Velleman, David, *Practical Reflection*, Princeton: Princeton University Press, 1989.

Een Nederlandstalige studie waarin Davidson een onuitputtelijke bron van inspiratie blijkt is:

Buekens, Filip, *De mens en zijn intenties*, Kapellen/Kampen: Pelckmans/Kok Agora, 1994.

Rationele principes en opvattingen

In de achtergrond van deze paragraaf is aanwezig:

Davidson, Donald, "Actions, Reasons and Causes", in *Journal of Philosophy*, Vol. 60, pp. 685-700, 1963.

Het onderscheid tussen 'mind-to-world' en 'world-to-mind' komt van:

Anscombe, G.E.M., *Intention*, Oxford: Blackwell, 1957.

Een heel heldere uiteenzetting hiervan is te vinden in hoofdstuk 1 van:

Searle, John, *Intentionality. An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press, 1983.

Een uitstekend overzicht over de discussie over rationaliteit en relativisme is te vinden in:

Hollis, Martin & **Lukes**, Steven (eds), *Rationality and Relativism*, Oxford: Basil Blackwell, 1982.

Pro-attitudes

In de tekst wordt verwezen naar:

Velleman, David, *Practical Reflection*, Princeton: Princeton University Press, 1989.

Over de internalisme-externalisme discussie vindt men veel goeds in:

Dancy, Jonathan, *Moral Reasons*, Oxford: Blackwell Publishers, 1993.

Garrard, Eve, "Motivation, Reasons and Causes", in Bransen & Slors (eds), *The Problematic Reality of Values*, Assen: Van Gorcum, 1996, pp. 32-38.

Over de redelijkheid van de pro-attitudes vindt men veel goeds in een reeks publicaties van Pettit & Smith, bijvoorbeeld:

Pettit, Philip, & **Smith**, Michael, "Backgrounding Desire", in *Philosophical Review*, Vol. 99, pp. 565-592, 1990.

Pettit, Philip, & **Smith**, Michael, "Practical Unreason", in *Mind*, Vol. 102, pp. 53-79, 1993.

Over het al dan niet propositionele karakter van de pro-attitudes, en over wilszwakte en zelfbedrog vindt men veel goeds in:

Macdonald, Cynthia & Graham, *Philosophy of Psychology*, Oxford: Blackwell Publishers, 1995.

Mele, Alfred, *Irrationality*. Oxford: Oxford University Press, 1987.

Pears, David, *Motivated Irrationality*. Oxford: Clarendon Press, 1984.

Intenties

In de tekst heb ik verwezen naar:

Anscombe, G.E.M., *Intention*, Oxford: Blackwell, 1957.

Verder is er eigenlijk geen enkel werk beter over dit onderwerp dan:

Bratman, Michael, *Intentions, Plans, and Practical Reason*, Cambridge (Mass.): Harvard University Press, 1987.

De 'counterfactual test' en de bespreking van de voorziene maar onbedoelde effecten zijn aan dit werk ontleend.

Motiveren en rechtvaardigen

Het betoog over deliberaties als 'basic actions' is, gebruikmakend van een heel eigenzinnige interpretatie, geïnspireerd op:

Pettit, Philip, *The Common Mind. An Essay on Psychology, Society, and Politics*, Oxford: Oxford University Press, 1993.

Toonaangevend over dit onderwerp is:

Smith, Michael, *The Moral Problem*, Oxford: Blackwell Publishers, 1994.

Naturalisme en Determinisme

In de tekst heb ik verwezen naar:

Strawson, Peter, *Skepticism & Naturalism. Some Varieties*, London: Methuen & Co., 1985.

Een vergelijkbaar onderscheid tussen sterk en zwak naturalisme is vandaag de dag in veel publicaties aan te treffen, bijvoorbeeld in het invloedrijke:

McDowell, John, *Mind and World*, Cambridge, Mass.: Harvard University Press, 1994.

Wie belangstelling heeft voor de manier waarop er in de Angelsaksische traditie gedacht wordt over de verhouding tussen wetenschap en filosofie, zal veel interessants aantreffen in het werk van en over **Quine**.

Verhelderend is in dit verband ook:

Putnam, Hilary, *Reason, Truth, and History*, Cambridge: Cambridge University Press, 1984 (het achtste hoofdstuk heet: "The impact of science on modern conceptions of rationality").

Wie dieper op het naturalisme in wil gaan, kan terecht bij de volgende verdedigers van een tamelijk sterk naturalisme:

Dretske, Fred, *Naturalizing the Mind*, Cambridge Mass.: MIT Press, 1995.

Papineau, David, *Philosophical Naturalism*, Oxford: Blackwell, 1993.

bij de volgende bestrijder van dit naturalisme:

Tye, Michael, "Naturalism and the Mental", in *Mind*, Vol. 101, pp. 421-441, 1992.

en bij de volgende overzichtsartikelen:

Armstrong, David, "Naturalism, Materialism and First Philosophy, in *Philosophia* (Israel), pp. 261-276, 1978.

Maffie, James, "Recent Work on Naturalized Epistemology", in *American Philosophical Quarterly*, pp. 281-293, 1990.

Kim, Jaegwon, "What is 'Naturalized Epistemology'?", in *Philosophical Perspectives*, pp. 381-405, 1988.

Wie meer wil weten over het determinisme kan uitstekend terecht bij Ted Honderich, die daarover uitgebreid gepubliceerd heeft, en onlangs een aardig inleiding daarover schreef:

Honderich, Ted, *How Free Are You? The Determinism Problem*, Oxford: Oxford University Press, 1993.

Zeer veel valt er over causaliteit te leren uit het overigens zeer academische (en beslist niet gemakkelijk toegankelijke):

Mackie, John, *The Cement of the Universe*, Oxford: Oxford University Press, 1974.

en meer specifiek gericht op de plaats van causaliteit in het denken over het menselijk handelen vindt men heel veel in:

Bransen, Jan & **Cuyppers**, Stefaan (eds), *Human Action, Deliberation and Causation*, Dordrecht: Kluwer Academic Publishers, 1998.

Over de *Erklären-Verstehen* controverse gelden de volgende twee boeken als uitstekende klassiekers:

Taylor, Charles, *The Explanation of Behaviour*, London: Routledge & Kegan Paul, 1964.

Wright, G.H. von, *Explanation and Understanding*, Ithaca: Cornell University Press, 1971.

Een goed overzicht over deze kwestie treft men ook aan in de uitstekende en zeer toegankelijke inleiding:

Hollis, Martin, *The Philosophy of Social Science. An Introduction*, Cambridge: Cambridge University Press, 1994.

Het probleem van mentale veroorzaking

Een zeer recent overzicht over de literatuur is te vinden in het state-of-the-art artikel: **Jackson**, Frank, "Mental Causation", in *Mind*, Vol. 105, pp. 377-413, 1996.

en een goede verzamelbundel is:

Heil, John en **Mele**, Alfred (eds), *Mental Causation*, Oxford: Oxford University Press, 1993.

Het probleem: twee oorzaken voor één effect

De manier waarop ik het probleem beschreven heb, is gebaseerd op het uitstekende artikel:

Kim, Jaegwon, "Explanatory Exclusion and the Problem of Mental Causation", oorspronkelijk verschenen in 1990, maar herdrukt in de zeer interessante bundel *Philosophy of Psychology*, samengesteld door Cynthia & Graham Macdonald, Oxford: Blackwell Publishers, 1995.

Eén effect? Dan één oorzaak!

De positie die ik in deze paragraaf beschrijf wordt o.a. verdedigd door Donald Davidson, en doorgaans "Anomalous Monism" genoemd. Zie: **Davidson**, Donald, *Essays on Actions and Events*, Oxford: Clarendon Press, 1980.

Twee oorzaken? Dan twee effecten!

Dretske heeft deze invloedrijke positie uiteengezet in zijn:

Dretske, Fred, *Explaining Behavior. Reasons in a World of Causes*, Cambridge Mass.: MIT Press, 1988.

Er is onder hen die zich sterk maken voor het project van het naturaliseren van het mentale veel aandacht voor de rol van het evolutieproces als een ontwerpgeschiedenis zonder ontwerper. Zie hierover bijvoorbeeld ook:

Dennett, Daniel, *Darwin's Dangerous Idea: evolution and the meanings of life*. New York: Simon & Schuster, 1995.

Twee oorzaken? Dat is één te veel!

Bridge-law reductionisme heeft zijn oorsprong in:

Nagel, Ernest, *The Structure of Science*, New York: Harcourt, Brace and World, 1961.

en is op klassieke wijze verdedigd in:

Smart, J.J.C. *Philosophy and Scientific Realism*, London: Routledge & Kegan Paul, 1963.

Armstrong, David, "An argument for the Identity Theory", in *Journal of Philosophy*, 1966.

Goede (kritische) besprekingen van het specifieke mind-brain reductionisme zijn te vinden in:

Macdonald, Cynthia, *Mind-Body Identity Theories*, London: Routledge, 1989.

Sterelny, Kim, *The Representational Theory of Mind. An Introduction*, Oxford: Blackwell Publishers, 1990.

Nagel, Thomas, "What is it Like to be a Bat?", *Philosophical Review*, Vol. 83, 1974. Talloos maal herdrukt, ook in een Nederlandse vertaling in Dennett's en Hofstadter's *De Spiegel van de Geest*. Amsterdam: Contact, 1986.

Eliminatief reductionisme wordt vooral verdedigd door:

Stich, Stephen, *From Folk Psychology to Cognitive Science: the Case against Belief*. Cambridge (Mass.): MIT Press, 1983.

Churchland, Paul, *Matter and Consciousness: A Contemporary Introduction to the Philosophy of Mind*. Cambridge (Mass.): MIT Press, 1984.

Oorzaken en verklaringen zijn twee.

Deze positie wordt min of meer verdedigd door:

Jackson, Frank & **Pettit**, Philip, "Functionalism and Broad Content", in *Mind*, Vol. 97, 1988.

Jackson, Frank & **Pettit**, Philip, "Causation in the Philosophy of Mind", in *Philosophy and Phenomenological Research*, Vol. 50, 1990.

Macdonald, C. & G., "Mental Causes and the Explanation of Action", in *Philosophical Quarterly*, Vol. 36, 1986.

Macdonald, Cynthia, "Weak Externalism and Mind-Body Identity", in *Mind*, Vol. 99, 1990.

Zie hierover ook de discussie in deel 1 van:

Macdonald, Cynthia & **Graham**, *Philosophy of Psychology*, Oxford: Blackwell Publishers, 1995.

Verklaringen en oorzaken 'zijn' één.

Beweringen die tot deze positie zouden kunnen leiden komt men o.a. tegen in:

Crane, Tim, "Mental Causation, I" in *Proceedings of the Aristotelian Society*, 1995.

McDowell, John, *Mind and World*. Lectures IV-VI. Cambridge (Mass.): Harvard University Press, 1994.

Zoals door mij weergegeven is de positie gebaseerd op argumenten ontwikkeld in:

Baker, Lynne Rudder, *Explaining Attitudes. A Practical Approach to the Mind*, Cambridge: Cambridge University Press, 1995.

Het sui generis karakter van het sociale

Zeer goede teksten van Hollis en Gilbert, waarin het sociale karakter van de normativiteit van het menselijk handelen expliciet gethematiseerd wordt, zijn:

Hollis, Martin, *Reason in Action: Essays in the Philosophy of Social Science*, Cambridge: Cambridge University Press, 1996.

Gilbert, Margaret, *On Social Facts*, Princeton: Princeton University Press, 1992.

Ik ben voor wat betreft de wijze waarop ik in dit hoofdstuk aandacht besteed aan het intrinsiek sociale proces dat 'rechtvaardigen' heet en dat constituerend is voor het bestaan van handelingen als entiteiten met een essentieel normatief karakter, geïnspireerd door Alisdair MacIntyre's prachtige cultuurfilosofische studie:

MacIntyre, Alisdair, *After Virtue*, London: Duckworth, 1981.

Het dramaturgisch model

John Searle is zeker niet de enige die het sociale meent te kunnen begrijpen in termen van het individuele handelen van mensen, maar hij heeft dit reductionisme wel zeer helder verwoord in zijn:

Searle, John, *The Construction of Social Reality*, New York: The Free Press, 1995.

Het dramaturgisch model is ontstaan uit het werk van ethnomethodologen. Een hoofdwerk van de grondlegger is:

Goffman, Erving, *The Presentation of Self in Everyday Life*, New York: Doubleday, 1959.

Goffman heeft op zijn beurt weer erg veel te danken aan het onovertroffen en zeer creatieve pionierswerk van George Herbert Mead:

Mead, George Herbert, *Mind, Self, & Society*, Chicago: University of Chicago Press, 1934.

Hollis en Harré hebben veel gedaan om het dramaturgisch model een rol te geven in de filosofie van de sociale wetenschap:

Hollis, Martin, *Models of Man*, Cambridge: Cambridge University Press, 1977.
 Harré, Rom, *Social Being. A Theory for Social Psychology*. Oxford: Blackwell, 1979.

Zie ook de uitstekende inleiding:

Hollis, Martin, *The Philosophy of Social Science. An Introduction*, Cambridge: Cambridge University Press, 1994.

Er is niet zo heel veel creativiteit nodig om in te zien dat het dramaturgisch model op de achtergrond een hoofdrol speelt in het werk van één van de grootste heden-daagse filosofen: Charles Taylor. Zie vooral het eerste deel van het (eigenlijk veel te dikke):

Taylor, Charles, *Sources of the Self. The Making of the Modern Identity*, Cambridge: Cambridge University Press, 1989.

en het boekje waarin een aantal radiolezingen van Taylor gebundeld zijn:

Taylor, Charles, *The Ethics of Authenticity*. Cambridge (Mass): Harvard University Press, 1991. Ook in Nederlandse vertaling verschenen onder de titel: *De malaise van de moderniteit*, Kampen: Kok Agora, 1994.

Over regels en sociale conventies

De actuele filosofische aandacht voor het fenomeen van het volgen van regels was er nooit geweest zonder de verkenningen van Ludwig Wittgenstein:

Wittgenstein, Ludwig, *Philosophical Investigations*, 2nd. ed, Oxford: Basil Blackwell, 1958.

Ik volg Saul Kripke in de presentatie van het probleem, dat ook dankzij Kripke zo zichtbaar is geworden:

Kripke, Saul, *Wittgenstein on Rules and Private Language*, Oxford: Basil Blackwell, 1982.

In mijn presentatie laat ik me overigens ook leiden door Philip Pettits werk over het volgen van regels. Zie met name:

Pettit, Philip, "The Reality of Rule-Following", in *Mind*, Vol. 99, pp. 1-21, 1990.

en het tweede en vierde hoofdstuk van het over de gehele linie uitmuntende:

Pettit, Philip, *The Common Mind. An Essay in Psychology, Society, and Politics*. Oxford: Oxford University Press, 1993.

Kripke heeft zich sterk gemaakt voor een oplossing van het probleem van regelvolgen waarin normativiteit begrepen wordt in termen van sociale conventies. Een standaardwerk over het conventionele karakter van semantische regels is:

Lewis, David, *Convention*, Cambridge (Mass.): Harvard University Press, 1969.

De toepassing van dit conventionalisme op het terrein van morele regels is verkend in het boeiende, en zeer invloedrijke:

Mackie, John, *Ethics. Inventing Right and Wrong*, Harmondsworth: Pelican, 1977.

Conventionalisme heeft ook zijn weg gevonden in de economie, al betreft het daar niet meer de constituerende functie van conventies voor normativiteit überhaupt:

Sugden, Robert, *The Economics of Rights, Co-operation and Welfare*, Oxford: Basil Blackwell, 1986.

Steun voor de conventionalistische these kan ook gezocht worden bij het hoofdwerk van een van de grootste hedendaagse contractdenkers, John Rawls:

Rawls, John, *A Theory of Justice*, Oxford: Oxford University Press, 1971.

En een verdediging van deze positie is voorts te vinden in:

Scanlon, T.M., "Contractualism and Utilitarianism", in Sen & Williams (eds.) *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 1982.

Een uitstekend overzicht over de gehele problematiek van de rol van conventies in ons begrip van normativiteit is te vinden in:

Collin, Finn, *Social Reality*, London: Routledge, 1997.

Over rolconflicten en rolafstand

De thematiek van onze natuurlijke kunstmatigheid c.q. kunstmatige natuurlijkheid is voor het eerst verkend door een Duitse wijsgerig antropoloog die beslist niet in de analytische traditie staat hoewel zijn werk er zeker relevant voor is:

Plessner, Helmut, *Die Stufen des Organischen und der Mensch*, Berlin: De Gruyter, 1928.

Zie ook:

Krüger, Hans-Peter, "The Second Nature of Human Beings: an Invitation for John McDowell to discuss Helmuth Plessner's Philosophical Anthropology", in *Philosophical Explorations*, Vol. 1(2), pp. 107-119, 1998.

De thematiek van de problematische relatie tussen de persoonlijke en de sociale identiteit werd voor het eerst verkend door:

Mead, George Herbert, *Mind, Self, & Society*, Chicago: University of Chicago Press, 1934.

Zie ook hoofdstuk 5 van:

Hollis, Martin, *Models of Man*, Cambridge: Cambridge University Press, 1977.
Succes en sociaal succes

Het voorbeeld van het handen wassen is ontleend aan:

Macdonald, Cynthia, "Externalism and Norms", in *Proceedings of the Royal Institute of Philosophy, Mind Series*, Cambridge: Cambridge University Press, 1998.

De argumentatie die in deze paragraaf wordt gepresenteerd, is geïnspireerd op besprekingen van wat tegenwoordig een *hot topic* is: de *response-dependency account* van relationele eigenschappen. McDowell, die in zijn artikelen overigens een zeer ontoegankelijke stijl hanteert, was hier de pionier:

McDowell, John, "Values and Secondary Qualities", in T. Honderich (ed), *Morality and Objectivity*, London: Routledge & Kegan Paul, 1985.

Johnston, Mark, "Dispositional Theories of Value", in *Proceedings of the Aristotelian Society*, 1989.

Pettit, Philip, "Realism and Response-Dependence", in *Mind*, 1991.

Over regelvolgen en regelvolgers

Pettit heeft zijn oplossing eerst gepresenteerd in:

Pettit, Philip, "The Reality of Rule-Following", in *Mind*, Vol. 99, pp. 1-21, 1990.

Zie voor de verschillende betekenissen van 'normaal' in Pettit's analyse:

Bransen, Jan, "Anthropocentrism in Favourable Circumstances", in *Inquiry*, 1994.

Verklaringen in de sociale wetenschappen

Klassieke werken over de kwestie van de aard van verklaringen in de sociale wetenschappen zijn die van Taylor, die het eigen karakter van dit type verklaringen verdedigt, en Von Wright, die de tegenstelling tussen 'verklaren' en 'begrijpen' sterk relateert:

Taylor, Charles, *The Explanation of Behaviour*, London: Routledge & Kegan Paul, 1964.

Wright, G.H. von, *Explanation and Understanding*, Ithaca: Cornell University Press, 1971.

Een algemene inleiding van analytische snit is:

Trigg, Roger, *Understanding Social Science*, Oxford: Basil Blackwell, 1985.

Het onderscheid tussen de accuratesse en de gepastheid van verklaringen is ontleend aan hoofdstuk 5 van:

Pettit, Philip, *The Common Mind. An Essay in Psychology, Society, and Politics*. Oxford: Oxford University Press, 1993.

De problematische relatie tussen sociale structuren en handelende actoren is het *Leitmotiv* van de uitstekende inleiding:

Hollis, Martin, *The Philosophy of Social Science. An Introduction*, Cambridge: Cambridge University Press, 1994.

Een bijzondere hoofdrol in het debat wordt gespeeld door Jon Elster, die buitengewoon veel publicaties op zijn naam heeft staan en een methodologisch individualisme verdedigt. Zie bijvoorbeeld:

Elster, Jon, *Ulysses and the Sirens. Studies in Rationality and Irrationality*, Cambridge: Cambridge University Press, 1979.

Elster, Jon, *Nuts and Bolts for the Social Sciences*, Cambridge: Cambridge University Press, 1989.

Zie ook de Nederlandstalige monografie over Elsters werk:

Cuypers, Stefaan (ed), *Indirecte Rede. Jon Elster over Rationaliteit en Irrationaliteit*, Leuven: Acco, 1995.

Nawoord

De visie op filosofie die uit dit nawoord spreekt is het onderwerp van:

Bransen, Jan, *Filosofie en ironie. Fantastische opmerkingen over de toekomst van een traditie*, Kampen: Kok Agora, 1992.

Zie voor soortgelijke visies de eerste hoofdstukken van:

Hollis, Martin, *Invitation to Philosophy*, Oxford, Blackwell, 1985.

Nagel, Thomas, *The View from Nowhere*, Oxford: Oxford University Press, 1986.

