

## INLEIDING

Toen ik destijds promoveerde op een experimenteel-psychologisch proefschrift over binoculair zien gebeurde dat in de splinternieuwe Leidse Faculteit van Sociale Wetenschappen. Ik weet nog dat me dat mateloos ergerde. Mijn onderzoek had immers niets uit te staan met vakken als sociologie of culturele anthropologie; wat had ik dan in 's hemelsnaam in hun faculteit de doen? Ik voelde me geen sociaal-wetenschappelijk, maar een gedrags-wetenschappelijk onderzoeker – en dat was, in mijn perceptie, iets heel anders. Bovendien was een experimenteel psycholoog nog eens een speciaal soort gedragsonderzoeker. Hij kon, dank zij zijn messcherpe experimentele methodiek, neerkijken op zijn collega's die het met test- en veldonderzoek meenden te moeten doen. Om maar niet te spreken van andere beoefenaars van de sociale wetenschappen die van experimenteren zelfs nog nooit gehoord hadden.

Ben ik dan de aangewezen persoon om aan dit symposium een bijdrage te leveren? Is het niet net zoiets als van president Reagan verwachten dat hij met een ontwapeningsvoorstel zal komen? Dat is inderdaad een onwaarschijnlijk scenario, maar toch blijkt het – onder de druk van omstandigheden – zover te kunnen komen. En zo is het ook mij gegaan. Ik denk dat we met elkaar moeten praten, en ik zal u straks vertellen onder welke presserende omstandigheden ik dat ben gaan denken. Dat zal ik niet doen om u met mijn biografie te belasten, maar omdat die omstandigheden een algemeen karakter hebben waar we wellicht allemaal wat aan kunnen hebben. Daarmee ben ik dan nog op redelijk veilig terrein. Maar het wordt anders wanneer ik vervolgens ga zwichten voor de specifieke druk van Jan Pen, namelijk om mijn nek uit te steken. Ik moet mijn buren, of tenminste één buur, vertellen wat ik van hem vind, wat hij beter zou kunnen doen, en wat ik daarbij voor hulp te bieden heb. Anderzijds mag ik hem ook vertellen waarom ik hem soms zo aardig vind, en wat ik van hem zou willen leren. Dat is onveilig terrein. Ik moge op begrip rekenen.

Maar laten we beginnen bij de status quo. Vooral sinds de tweede wereldoorlog zijn de verschillende gedrags- en maatschappijwetenschappen – ik zal voor het gemak van sociale wetenschappen spreken – aan expansieve groei en verzelfstandiging onderhevig. Elk van die wetenschappen kan bogen op een eigen theoretisch kader van concepten en verklaringsprincipes, op eigen methoden van onderzoek, en op eigen data- en toepassingsdomeinen. Binnen die wetenschappen is er natuurlijk allerlei onenigheid, maar meestal is er ook een orthodoxie, een soort theoretisch-methodologische standaard die men niet ongestraft kan negeren.

Deze verzelfstandiging van de sociale wetenschappen heeft als natuurlijk gevolg dat de buren zich minder met elkaar onderhouden dan in de goede oude tijd het geval was. Er is minder gelegenheid voor, er is minder behoefte aan, en er ontstaan grotere weerstanden wanneer het toch nog eens gebeurt. Dat laatste is deels te wijten aan de verschillende orthodoxieën die zich ont-

wikkeld hebben. Wat norm is in de ene sociale wetenschap is verfoeilijk in de andere.

Liever koesteren we de illusie dat de sociale werkelijkheid geleed is conform de scheidslijnen tussen onze disciplines. Haar wordt dan geen speciaal geweld aangedaan wanneer ieder zich aan zijn eigen abstracties houdt. Van al die sociale wetenschappen zijn er slechts twee waarmee ik enigszins vertrouwd ben, de psychologie en de taalkunde. Van die twee weet ik dat hun scheidslijn buitengewoon arbitrair is, en niets te maken heeft met de geleiding van de werkelijkheid. Het zal tussen andere sociale wetenschappen wel niet anders gesteld zijn. We hebben het veel minder over verschillen in object dan over verschillen in wetenschapscultuur.

Wat dat betreft hebben we ons ver verwijderd van de 'founding fathers'. Voor Wilhelm Wundt, de vader van zowel de experimentele psychologie als van de taalpsychologie, was het even vanzelfsprekend dat hij zich met binoculair zien als met taal en cultuur van natuurvolkeren bezig hield. In beide gevallen bestudeerde hij immers psychische processen, alleen de methoden verschilden noodgedwongen. Waar het om relatief eenvoudige psychische processen zoals visuele en auditieve waarneming gaat, gebruike men het experiment, de ideale natuurwetenschappelijke methode. Die laat directe conclusies toe over oorzakelijke verbanden. Waar het om complexe psychische processen gaat zoals taal en denken, moet men uitwijken naar de observatieve methode. Men moet kijken naar de *uitkomsten*, de produkten van psychische processen. Die produkten zijn buitengewoon variabel. Ze stabiliseren zich pas als massaverschijnsel, als aggregaat. Door dat eindprodukt, taal, zeden, mythen van een samenleving te analyseren, kan men toch theoretiseren over de complexe psychische processen die eraan ten grondslag liggen.

Bij ons is er een hoge schutting opgetrokken tussen deze twee benaderingswijzen. Zelfs als we het allemaal over psychische processen hebben, dan houden sommigen van ons zich van huis uit bezig met de 'directe' experimentele analyse van die processen. Dat zijn de experimentele psychologen onder ons. Anderen volgen de indirecte weg. Ze observeren de uitkomsten van het gedrag, en proberen daar theoretische verklaringen voor te geven. Dat zijn de meeste andere sociaal-wetenschappelijke onderzoekers onder ons. Dat is natuurlijk een beetje te scherp gesteld zo, maar er zit een grond van waarheid in. Het is de grote verworvenheid van wat tegenwoordig 'cognitive science' of cognitiewetenschap heet, dat deze schutting wordt afgebroken. En daarbij heeft de taalkunde een wezenlijke rol gespeeld. Ik wil die situatie nu kort uiteenzetten, met de bedoeling daar lering uit te trekken voor de verhouding tussen onze wetenschappen. Vervolgens zal ik mijn nek uitsteken door avances te maken naar de economie, een vak waar ik niets vanaf weet, maar wel van alles over denk.

#### VERKLARINGSNIVEAUS

Wat tegenwoordig cognitiewetenschap heet, is een interdisciplinair allegaartje van taalkunde, logica, psychologie, anthropologie, neurowetenschappen,

computerwetenschappen, en wie er verder nog maar mee wil doen. Wat ze bindt is de vraag naar kennisrepresentatie. Hoe wordt door een natuurlijk of kunstmatig intelligent systeem kennis verworven, getransformeerd, en gebruikt? Het gaat daarbij steeds om complex gedrag, zoals schaken, een taal verwerven, visuele scenario's interpreteren, muziek van Mozart en Haydn onderscheiden, iemand de weg wijzen, iemand ergens van overtuigen. Noem maar op. Het gaat tevens om een formele representatie van dat gedrag. Het kennisverwerkende systeem wordt behandeld als een formeel systeem, dat volgens expliciete regels een bepaalde invoer transformeert in een bepaalde uitvoer. Niet zelden wordt zo'n formeel systeem geïmplementeerd op een computer; het gaat tenslotte om systemen van vaak zeer grote omvang en complexiteit.

Die cognitiewetenschap is niet uit de lucht komen vallen. Maar van haar interessante voorgeschiedenis haal ik slechts één aspect naar voren dat voor ons onderwerp van belang is, en waar ik zelf bij betrokken ben geweest. Het betreft de relaties tussen taalkunde en psychologie. Aan het einde van de jaren vijftig verkondigde de revolutionaire taalkundige Noam Chomsky de volgende stelling: *Linguistics is a chapter of psychology*. Diezelfde Chomsky had toen een aantal formele systemen bedacht om de welgevormde zinnen van een taal te genereren, generatieve grammatica's. Zo'n generatieve grammatica representeerde, volgens hem, het taalvermogen (de 'competence') van de betreffende taalgebruiker. Het was aan de psychologen om uit te zoeken hoe dat taalvermogen wordt aangewend in de processen van spreken en verstaan, en hoe het door het kind wordt verworven. Zij moesten, kortom, zorgen voor een 'performance' theorie, een theorie van het taalgebruik. Over deze kwestie zijn inmiddels boekenkasten vol geschreven, en ik zal u daar niet verder mee vervelen. Waar het om gaat was dat psychologen plotseling werden geconfronteerd met een gedetailleerde formele karakterisering van een psychisch vermogen, een beschrijving waaraan ze zelf part noch deel hadden gehad. En de confrontatie was des te pijnlijker omdat Chomsky ook nog eens het formele bewijs leverde dat de taalmechanismen die door de dominante behavioristische psychologen van die tijd werden voorgesteld, niet het generatieve vermogen hadden om een natuurlijke taal voort te brengen. Kortom, het zelfgenoegzame psychologische imperium werd ruw bezet door taalkundigen.

Psychologen zijn sindsdien bescheidener geworden. Ze hebben ontdekt dat er 'chapters of psychology' zijn die door niet-psychologen zijn geschreven, dat ze die hoofdstukken vooral heel goed moeten lezen, en er rekening mee moeten houden wanneer ze hun proces-modellen bouwen. Het is dit soort interactie dat thans de basis vormt voor de cognitiewetenschap. De traditionele disciplines verschillen vaak niet zozeer in hun object, als wel in welke aspecten van het object ze bij voorkeur karakteriseren, het beschrijvingsniveau. Ze kunnen echter niet meer zonder elkaar.

David Marr heeft in 1982 voorgesteld een onderscheid te maken tussen drie beschrijvings- of verklaringsniveaus in de cognitiewetenschap. Idealiter

moet elk complex gedrag op alle drie niveaus worden behandeld. Omdat ik denk dat deze indeling ook voor de verhouding van onze wetenschappen van belang is, wil ik haar kort bespreken.

Het eerste beschrijvingsniveau noemt Marr het niveau van de *computational theory*. Dit is een formele karakterisering van het doel en de interne logica van het systeem, een formeel model van de invoer/uitvoer-relaties van het systeem. Een generatieve grammatica is zo'n computationele theorie. Doel is de verzameling van zinnen die samen een taal vormen te karakteriseren, en de interne logica is axiomatisch-deductief. Er zijn ook andere computationele theorieën. Zo zijn er visuele grammatica's die de verzameling van 2-dimensionele patronen karakteriseren welke 3-dimensioneel interpreteerbaar zijn. De cognitiewetenschap maakt verder veel gebruik van allerlei logische systemen die axiomatisch zijn opgebouwd, zoals de epistemische, modale, en actiologica's. Andere computationele theorieën zijn stelsels van vergelijkingen, Fourier transformaties, differentiaalvergelijkingen, of wat maar nodig is om de uitvoer van het onderzochte systeem te karakteriseren. Sommige van die systemen zijn 'chapters of psychology'; zij modelleren wat intuïtief wordt ervaren als welgevormd of rationeel menselijk gedrag. Het zijn computationele theorieën die aanspraak maken op wat we 'psychologische realiteit' zullen noemen. Er zijn wetenschappen die zich traditioneel met 'computational theory' bezig houden. De taalkunde, de logica en – zoals ik straks zal beweren – de economie.

Het tweede beschrijvingsniveau is dat van het *algorithme*. Daar gaat het om de implementering van de computationele theorie. Er zijn tientallen manieren om de zinnen van een taal te genereren, om een stelsel van differentiaalvergelijkingen op te lossen, enzovoorts. De vraag is echter hoe dat wordt gedaan door het systeem dat mijn object van onderzoek is. Wanneer het systeem een mens is die spreekt, keuzen maakt, handelingen verricht, willen we de theorie zo implementeren dat die 'psychologisch realistische' proceseigenschappen krijgt, zoals bijvoorbeeld de volgende: Het proces moet met een zeer beperkt werkgeheugen uit kunnen komen; wat een mens relatief vlug of langzaam doet moet het algoritme in dezelfde mate vlug of langzaam doen; waar de mens fouten maakt moet ook het algoritme fouten maken.

Over het punt van psychologische realiteit bestaat veel verwarring. Wanneer de computationele theorie een 'chapter of psychology' is, dan maakt zij *ipso facto* aanspraak op psychologische realiteit, zoals we zagen. Zij karakteriseert de taal die iemand beheerst, de keuzehandelingen die hij verricht, of wat dies meer zij. Maar ze abstraheert van algorithmische beperkingen. Zo heeft Chomsky het over de 'ideal speaker/listener' die geen last heeft van geheugen- en aandachtsbeperkingen, en die zijn taal volmaakt beheerst. Een computationele theorie is logisch onafhankelijk van haar implementatie. Het gaat alleen om de correcte *uitkomsten*. Een algorithmische theorie daarentegen kan pas bogen op psychologische realiteit wanneer het mentale *proces* dat die uitkomsten voortbrengt correct is beschreven.

Het derde verklaringsniveau betreft de implementatie van het algoritme

in hardware. Hoe wordt het *fysisch gerealiseerd*? Wanneer het systeem een mens of dier is zitten we met deze vraag meteen in de neurowetenschappen. Ook de neuropsychologie is in dit verklaringsniveau geïnteresseerd. Waarom wordt iemand afatisch bij beschadiging van de linker hersenhelft? Daar bevindt zich blijkbaar het weefsel waarin de taal is geïmplementeerd. Ik denk dat dit derde verklaringsniveau voor onze huidige discussie van minder belang is. Ik kan me beperken tot de opmerking dat onder de sociale wetenschappen alleen de psychologie zich met dit verklaringsniveau bezighoudt. Daar zijn haar naaste burens de biologie en de neurologie.

David Marr heeft aanbevolen om in de cognitiewetenschap van boven naar beneden te werken. Maak eerst een computationele theorie, kijk dan naar de algoritmische implementatie ervan, en tenslotte naar de hardware waarop de zaak loopt. Die mening had Chomsky ook. Psychologen kunnen met hun taalproces-theorieën beter even wachten totdat linguïsten hun hoofdstuk geschreven hebben, dan weten ze tenminste wat er geïmplementeerd moet worden. Niet iedereen deelt die mening. Taalkundigen doen er wel erg lang over voordat ze hun hoofdstuk klaar hebben. Bovendien zou het wel eens zo kunnen zijn dat psychologen bij het proces-onderzoek merken dat er iets gekks aan de hand is met de theorie, dat mensen systematisch iets anders doen dan wat de taalkundigen beweren. Tenslotte gebruiken taalkundigen alleen maar hun eigen intuïties als evidentie. Misschien kunnen de hoge heren ook iets van het voetvolk leren. Je weet maar nooit.

Zojuist heb ik beweerd dat de opkomst van de cognitiewetenschap de hoge schutting heeft geslecht tussen psychologen die zich met directe experimentele analyse van psychische processen bezig hielden, en de andere sociaal-wetenschappelijke onderzoekers die theorieën ontwerpen over de uitkomsten van dat gedrag. De theorievorming in de moderne cognitieve psychologie verloopt nogal anders dan vroeger in de experimentele psychologie. Het is een kunst geworden om een werkend model van complex gedrag te maken dat tegelijk aan twee eisen voldoet. Ten eerste moet het model alle en alleen de gedragingen voortbrengen die de computationele theorie toestaat, en ten tweede moet het model als proces-model psychologisch reëel zijn. De computationele theorie wordt vaak hand in hand met het algoritme gebouwd, of zij wordt ten dele aan een andere discipline ontleend. Daar schamen we ons niet meer voor. Cognitieve psychologen besteden veel tijd aan het bouwen van zulke complexe modellen, zonder elke assumptie eerst experimenteel te willen toetsen (zie Kempen, 1987). We beweren niet, zoals Wundt, dat de experimentele toetsing van zulke complexe gedragsmodellen onmogelijk is, maar wel dat niet elk onderdeel ervan toetsbaar is of getoetst hoeft te worden. We zitten wat minder krampachtig op de empirie en kunnen weer profiteren van wat anderen te vertellen hebben over *uitkomsten* van menselijk gedrag, de taal die gesproken wordt, de keuzen die gemaakt worden, enzovoorts. En daarover hebben juist andere sociale wetenschappen soms interessante dingen te zeggen. Kijken we eens naar de economie.

## DE MICRO-ECONOMIE. EEN COMPUTATIONELE GEDRAGSTHEORIE

De micro-economie is een theorie over keuzegedrag van mensen. Het is dus een 'chapter of psychology'. Maar het is een computationele theorie. Het is een axiomatisch-deductief formeel systeem dat een verzameling van welgevormde keuzen genereert. Een welgevormde keuze is er een waarbij, gegeven een stabiele verzameling van voorkeuren, en optimale informatie over kosten en nut van de keuzealternatieven, de eigen satisfactie gemaximaliseerd wordt. Zo'n keuze heet een rationele. De *homo economicus* is een rationele beslisser. Ik zie nu even af van de vele alternatieve definities van rationeel beslissen; in feite vertelt ons slechts het formele systeem als geheel wat een rationele keuze is.

Dit is bij uitstek het type van een computationele theorie. Er is een ideale beslisser, net zoals er bij Chomsky een ideale taalgebruiker is. Het is iemand die volledig geïnformeerd is over de eigen behoeften, en optimale kennis heeft van de relevante keuzealternatieven en hun utiliteiten, net zoals Chomsky's taalgebruiker zijn taal volledig beheerst. Het is bovendien iemand die niet onderhevig is aan beperkingen van aandacht en geheugen; alle relevante informatie heeft hij steeds paraat. Ook dat wordt in de generatieve taalkunde aangenomen. Waar het echter vooral om gaat is dat het formele systeem de uitkomsten van het gedrag voorspelt, maar niet de wijze waarop die uitkomsten worden geproduceerd door het subject. De psychologische realiteit van het formele systeem is computationeel, niet algoritmisch. Het beschrijft psychologische *uitkomsten*, namelijk rationele keuzen, maar geen psychologische *keuze-processen*. Zo is het ook precies in de generatieve taalkunde. Dat is een theorie over welgevormde zinnen, maar niet over zinsproductie.

Ik heb gezegd dat cognitieve psychologen er tegenwoordig niet meer van schrikken wanneer ze weer een 'chapter of psychology' tegenkomen dat ze niet zelf geschreven hebben. Zoals bijna steeds gaat het ook hier weer om een computationele theorie, en we weten nu hoe we reageren moeten. We stellen twee vragen. De eerste is: Zijn de computationele assumpties psychologisch in orde? Of anders gezegd: Zijn menselijke beslissingen inderdaad rationeel in de omschreven zin? De tweede vraag is: Hoe kan de theorie in een algoritme worden geïmplementeerd dat kan bogen op psychologische realiteit? Anders gezegd: Welk mentaal proces brengt de keuzehandeling voort? Net zoals we bij de taalkunde zagen, zijn die twee vragen niet onafhankelijk van elkaar. De perceptie van de keuze-alternatieven, beperkingen van aandacht en geheugen, alsmede emoties, kunnen de uitkomsten van het keuzegedrag systematisch beïnvloeden. Wanneer dat echt systematisch is – dus niet alleen maar een toename van ruis in het keuzegedrag veroorzaakt –, dan zou men die systematiek graag afgebeeld zien in de formele 'computationele' theorie. De twee vragen kunnen natuurlijk niet zomaar worden beantwoord. Er kunnen in eerste instantie hoogstens een paar suggesties worden gedaan. Ook kan er wellicht gesproken worden over methoden om die suggesties te toetsen. Laat ik beginnen met de eerste vraag.

## DE COMPUTATIONELE ASSUMPTIES

Toen ik voor het eerst nadacht over die rationele beslisser die zijn opbrengst maximaliseert, had ik een *déjà vu* ervaring. Waar had ik zoiets toch meer gehoord? Natuurlijk, in de evolutietheorie. Organismen evolueren zo dat de kans op survival wordt gemaximaliseerd. Alternatieve organismen worden geheel volgens toeval gegenereerd, maar alleen de sterksten trekken het. De rationele keuze van de natuur. Later heb ik begrepen dat dit verband niet toevallig is. Darwin had Malthus bestudeerd, en ook de moderne markteconomie hanteert expliciet het 'survival of the fittest' principe. Maar daar wil ik het niet over hebben. Ik wil de gelijkenis alleen maar gebruiken om één ergernis weg te nemen, alvorens tot zaken over te gaan. Biologen hebben wel eens de neiging om hun evolutietheorie tot een ontoetsbare tautologie te reduceren. Welke soort overleeft? De soort die het best is aangepast. Welke soort is het best aangepast? De soort die overleeft. Economen zijn onderhevig aan hetzelfde gevaar. Wat kiest de mens? Datgene wat zijn satisfactie maximaliseert. Hoe weten we wat maximale satisfactie geeft? Kijk naar wat de mens kiest. 'Revealed preference' noemen ze dat. Ik denk soms wel eens dat veel economen zich hun leven lang ophouden binnen de veilige omheining van deze vicieuze cirkel. Ik had althans geen moeite een affreus voorbeeld van deze cirkelredenering te vinden in een bekend inleidend boek over de economie.<sup>1</sup> Maar we zijn hier niet om op elkaar te vitten. Er moet burenhulp komen.

Het is genoegzaam bekend dat er psychologische problemen zijn met de uitgangspunten van de klassieke micro-economie. Het beeld, aldus van Veldhoven (1987), 'klopt van geen kanten'. Veldhoven, van Raaij, van de Poel en andere auteurs in het zeer leerzame boek van Keizer en Soeters (1987) geven daar allerlei argumenten voor. Die zal ik niet herhalen. Het mooiste vind ik nog steeds het ondermijnende werk dat Kahneman en Tversky (1979, 1981) hebben verricht. Zij konden zeer systematische en voorspelbare afwijkingen van rationeel beslisgedrag aantonen. Een van die afwijkingen is, bijvoorbeeld, risicovermijding wanneer er wat te verdienen valt en ri-

<sup>1</sup> Tullock en McKenzie (1985) beschrijven hoe rationeel gedrag, maximalisatie van profijt, bij mens en dier terug te vinden is. In dat verband noemen zij ook een experiment van Plott, een simulatie van een marktsituatie waarin een factor van vervuiling een rol speelt. Door iets te produceren ontstaat er ook een zekere mate van vervuiling waar alle anderen last van hebben. En men weet dat dat gebeurt. De uitkomst is dat ieder naar hartelust vervuult om zijn eigen gewin te maximaliseren. Zie je wel, zeggen Tullock en McKenzie, rationeel gedrag. Het experiment zat echter zo in elkaar dat ieder nog veel meer had kunnen verdienen wanneer gezamenlijk zou zijn besloten de vervuiling terug te dringen. Dát rationele gedrag werd echter niet vertoond. Wat is dan de grond om te zeggen 'zie je wel, rationeel gedrag', zoals de schrijvers doen? Rationeel gedrag is blijkbaar datgene wat mensen doen. Het begrip is tautologisch; er wordt niets mee verklaard. (Het tekstboek – of beter het essay, want dat is het eigenlijk – probeert overigens wel tot substantiële gedragsverklaringen te komen, o.a. van sexueel en crimineel gedrag. Op die verklaringen is ook van alles af te dingen, maar dat is hier niet aan de orde).

sicozoeken wanneer er wat te verliezen is. Anders gezegd, je bent liever helemaal zeker van een kleine winst dan tamelijk zeker van een grotere winst, ook als de utiliteit in het eerste geval geringer is. En zo ben je liever onzeker over een groot verlies dan zeker over een klein verlies, al is de verwachte schade in het eerste geval groter. Ik kom op dat werk straks nog terug.

Gelukkig zijn het niet alleen psychologen die ontdekken dat de assumpties niet kloppen, maar ook economen. Pen bespreekt in 'Over de schutting' (1986) een hele stoet van *homines economici* die op allerlei manieren, al of niet rationeel, met hun schaarse goederen omspringen. Ik denk dat veel van die *homines* psychologisch al net zo invalide zijn als hun voorgelacht, maar één spreekt me aan als psycholoog. Het is de *homo bacharachensis*, die handelt op grond van wat hij *gelooft*. 'Belief systems' vormen een zeer actueel onderwerp in de cognitieve psychologie, en ik denk dat we juist daar moeten kijken naar mogelijkheden tot samenwerking.

Het psychologische uitgangspunt is dat mensen handelen op grond van wat ze weten, geloven, hopen, van plan zijn, enzovoorts. Men noemt die toestanden van de geest *propositionele houdingen*. Het zijn de houdingen die het subject aanneemt ten aanzien van proposities. Jan gelooft dat hij met Marietje gaat trouwen, Jan wenst met Marietje te trouwen, Jan is van plan met Marietje te trouwen, enzovoorts. Elk type propositionele houding heeft zijn eigen interne logica. Zo kan iemand heel goed twee dingen *wensen* die elkaar uitsluiten. Jan kan bijvoorbeeld zowel wensen met Marietje te trouwen als wensen met Sofietje te trouwen. Dat zijn nu net precies de keuzealternatieven. Er is hier geen sprake van incoherentie. Maar je kunt niet *van plan zijn* twee handelingen te verrichten die elkaar uitsluiten. De logica van handelingsintenties vereist consistentie en coherentie (zie Harman, 1986). Dit laatste geldt ook voor de logica van 'beliefs', van meningen. Of precieser gezegd: mensen vermijden meningen te hebben die ze herkennen als zijnde inconsistent met elkaar. En mensen zijn bereid dingen te geloven die ze herkennen als zijnde geïmpliceerd door meningen die ze al hebben (dit zijn Harman's consistentie- en implicatie-principes). Dat 'herkennen als' is hier heel kritisch; ik kom daar nog op terug.

Handelingsintenties hangen ten nauwste samen met propositionele houdingen van wensen en menen. Wanneer we menselijk keuzegedrag willen voorspellen moet er een theorie komen over hoe handelingsintenties voortkomen uit wensen en meningen. Die psychologische theorie bestaat nog niet, maar wel is er het nodige bekend over meningen en een beetje over wensen, en ik denk dat micro-economen daar op den duur niet omheen kunnen.

### *Meningen*

De rationele beslisser is steeds optimaal geïnformeerd. Wat doet hij met nieuwe informatie? Nieuwe informatie kan zijn verwachtingen ten aanzien van uitkomsten wijzigen, en dus zijn beslissingen. Het klassieke model om dit bijstellen van verwachtingen te behandelen is dat van Bayes. Dat model gaat

echter geheel voorbij aan de interne logica van meningen, aan het mechanisme van *rechtvaardiging* van meningen. Een nieuw feit is al of niet een *reden* om een mening te handhaven of te veranderen.<sup>2</sup>

De cognitiewetenschap heeft zich in recente jaren nogal beziggehouden met de menselijke mening-huishouding, en daar ook formele modellen voor ontwikkeld (de theorie van *reasoned assumptions*, niet-monotone logica). Als je een keuze moet maken uit handelingsalternatieven ben je meestal niet volledig geïnformeerd, maar er moeten wel conclusies worden getrokken. Dat kan alleen door logische kortsluitingen te maken, vuistregels te volgen, en wat dies meer zij. Mensen maken 'default assumptions': Men concludeert uit 'Pietje is een vogeltje' dat 'Pietje kan vliegen', zolang er geen evidentie is dat Pietje een struisvogel of pinguïn is, een vleugel gebroken heeft, of om andere redenen niet kan vliegen. Die werkassumpties zijn meningen over hoe de wereld in elkaar zit; ze hebben geen logische status, maar een psychologische. De belangrijke vraag is hoe we aan die meningen komen, en hoe ze gewijzigd kunnen worden, en daar is een hele literatuur over (zie vooral Nisbett en Ross 1980, Doyle en London 1980, Doyle 1985, Rapoport 1986, en Harman 1986).

Mensen hebben redenen om meningen te hebben. De meeste meningen die we hebben steunen niet op eigen observatie, maar op meningen van anderen. Die rechtvaardiging kan volkomen consistent zijn. Bijvoorbeeld: De meester zegt dat de wereld rond is; mijn werkassumptie is dat wat de meester zegt waar is, *ergo* ik ben van mening dat de wereld rond is. Vrijwel ons hele onderwijssysteem is op dit rechtvaardigingsprincipe gebaseerd.

We kunnen ook meningen afleiden uit andere meningen die we al hebben. En dat scheidt een enorm probleem in onze psychologische mening-huishouding. Er is namelijk geen einde aan het aantal afleidingen dat er te maken is. Het principe hier schijnt, ruw geformuleerd, te zijn dat we ons alleen maar meningen vormen over dingen die wij belangrijk vinden, d.w.z. dingen die betrekking hebben op onze behoeften. We zijn dus zuinig met het produceren van meningen – een goed economisch principe.

We zijn ook zuinig met het *opgeven* van meningen. Over het opgeven of veranderen van meningen gaat eigenlijk het grootste deel van het werk waar ik het nu over heb. Zoals gezegd hebben meningen hun rechtvaardiging. Het ligt nu voor de hand om te beweren dat een mening wordt opgegeven als de rechtvaardigingsgrond eraan ontvalt. Dat is echter in strijd met een ander principe van onze mening-huishouding, eveneens een economisch principe. Het is ruw geformuleerd dit: 'Onthoud slechts de conclusies, niet de gronden'. Weet u nog van wie u geleerd hebt dat de wereld rond is? Toch

<sup>2</sup> Er zijn echter ontwikkelingen in de richting van een 'truth-functional' Bayesiaanse theorie van mening-revisie, zie Dempster (1967) en Shafer (1976). Het is echter niet duidelijk hoe zulke theorieën een algorithmische implementatie kunnen krijgen die kan bogen op 'psychologische realiteit', aangezien elke nieuwe evidentie potentieel relevant is voor elke bestaande mening.

was het vertrouwen in die persoon de reden om dat te geloven. Maar als wij de gronden voor onze meningen niet onthouden, dan kunnen we, zonder inconsistent te worden, een mening handhaven waaraan de rechtvaardiging ontvallen is. Misschien was die oude onderwijzer wel een leugenaar, het zal ons niet deren.

Wanneer wordt een mening dan wel opgegeven? Alleen wanneer er positieve redenen zijn om dat te doen, wanneer we over evidentie komen te beschikken die strijdig is met die mening, en die we als zodanig herkennen. Dit laatste is een belangrijke toevoeging. Vaak herkennen we evidentie of een andere opinie niet als relevant voor een mening die we al hebben. We kunnen met tegenstrijdige meningen leven zolang ze in verschillende hokjes zitten, en nooit *tegelijk* in ons aandachtsveld komen. Op die hokjes kom ik straks nog even terug. De formele modellen over revisie van meningen zijn mede gebaseerd op niet-triviale experimentele resultaten, waarvan ik u de details zal besparen. Maar een belangrijke vondst is deze: Wanneer je iemand van alles wijs maakt, en hem achteraf zegt dat je hem onzin verteld hebt, dan blijft er toch een verschuiving van mening over in de richting van wat je beweerd hebt (Ross, Lepper and Hubbard, 1975). De verklaring daarvoor is dat mensen, wanneer ze in goed vertrouwen een mening overnemen, zelf nog extra rechtvaardigingen ervoor proberen te vinden. Valt de hoofdgrond voor de mening weg, namelijk het vertrouwen in de informant, dan blijft nog de extra zelf-gegenereerde rechtvaardiging. Het is dan consistent om de nieuwe mening te handhaven. Het door Hoenkamp (1987) gemaakte formele model voorspelt nu dat de mening in dezelfde richting moet verschuiven wanneer je iemand *van te voren* mededeelt dat hij misleidende informatie zal krijgen. En ja hoor, dat gebeurt ook (Wegner, Coulton en Wenzlaff, 1985).

Een ander punt van groot belang voor keuzegedrag is dat onze meningen niet een lange ongestructureerde lijst vormen. Meningen zijn onderverdeeld naar onderwerp, belangengebied, waardegebied – ‘hokjes’ noemde ik dat zojuist. Wanneer we een keuze overwegen heeft die keuze betrekking op zo'n gebied of onderwerp, en we recruteren dan de meningen die wij over dat onderwerp hebben om tot een besluit te komen. Dat is een kwestie van selectieve aandacht. Het fraaiste voorbeeld dat ik hiervan ben tegengekomen is een experiment van Tversky en Kahneman (1981), dat ik nu even aan de huidige ambiance aanpas. U wilt naar het optreden van Horowitz in het concertgebouw. We nemen aan dat kaartjes van te voren te krijgen zijn, maar ook aan de zaal. Eerste scenario: U hebt van te voren voor 200 gulden een kaartje gekocht, en gaat te bestemder tijde naar het concertgebouw. Daar aangekomen haalt u uw portefeuille uit uw zak, en ontdekt dat u uw kaartje kwijt bent. Koopt u nu voor nog eens 200 gulden een nieuw kaartje? De meeste proefpersonen zeiden ‘nee’. Tweede scenario: U wilt naar hetzelfde concert, maar u hebt nog geen kaartje gekocht. Dat doet u wel aan de zaal. Daar aangekomen haalt u uw portefeuille uit uw zak en ziet tot uw schrik dat u 200 gulden bent verloren. Gaat u nu toch een kaartje kopen? De

meeste proefpersonen zeiden hierop 'ja'. Voor de rationele beslisser zijn de twee scenario's identiek, maar blijkbaar niet voor mensen. In het eerste geval breng je de kosten van tweemaal 200 gulden onder bij één onderwerp: het is wel erg veel om wellieft 400 gulden voor zo'n plaats uit te geven. In het tweede geval verdeel je de kosten over twee onderwerpen. De plaats kost je niet meer dan 200 gulden, de rest gaat op rekening van een andere post in je psychologische boekhouding.

Zover over meningen. Ik denk dat economen met psychologen moeten praten over *belief systems*, de psychologische mening-huishouding. Wat mensen kiezen hangt daarmee ten nauwste samen. Keren we nu nog even naar wensen en emoties.

### *Wensen en Emoties*

Ook wensen zijn propositionele houdingen, maar over hun logica (de zg. boulogomaische logica) is weinig bekend. In elk geval ziet die er anders uit dan die van meningen. Wensen behoeven geen rechtvaardiging of consistentie. Het is dan ook moeilijk ze op te geven. Waar meningen konden buigen voor positieve tegenevidentie, bestaat er zoiets niet voor wensen. Cognitieve psychologen weten niet zo erg veel van wensen af. Het beste wat ik erover ken is te vinden in Frijda's (1986) nieuwe studie over de emoties. Volgens Frijda's theorie spelen emoties een belangrijke rol in de menselijke behoeftehuishouding. Iets wekt een emotie op wanneer het relevantie bezit voor de satisfactie van een behoefte. Er ontstaat een actiebereidheid. Dat kan bijvoorbeeld de bereidheid zijn om zich nieuwe meningen te vormen, zoals we al eerder zagen. Het kan ook, speciaal in het geval van wensen, tot gevolg hebben dat er prioriteit wordt gegeven aan handelingen die die satisfactie kunnen bewerkstelligen. Emoties sturen de aandacht en de prioritering van handelingsintenties. Of die intenties ook worden uitgevoerd is een andere zaak. Dat is stellig een kwestie van afweging. Hoe sterk is de impuls, de behoefte, welke satisfactie kan ik verwachten, en tegen welke kosten?

Economen kunnen, denk ik, veel leren uit Frijda's boek. De menselijke emoties, de behoeften waarop ze betrekking hebben, en de stimuli waardoor ze worden opgewekt zitten heel wat ingewikkelder in elkaar dan die van de doorsnee rationele beslisser. Met name is er geen psychologische grond voor de gangbare economische veronderstelling dat de eigen behoeften bekend zijn aan de beslisser. Maar, zoals Pen mij duidelijk maakte, en inmiddels in een recensie van Frijda's boek in *Vrij Nederland* nader omschreef, misschien kunnen psychologen hier ook veel van de economen leren. Hebben niet juist economen formele theorieën ontworpen over de volgorde waarin behoeften voor satisfactie in aanmerking komen? Is het niet net precies de balans van marginaal nut en marginale kosten die van keuze tot keuze die prioritering bepaalt? Ik denk dat Pen gelijk heeft. Ik denk ook dat dit idee in de handen van Frijda tot interessante psychologische experimenten aanleiding kan geven.

## DE ALGORITHMEN

Het is me nooit opgevallen dat economen – met uitzondering van Herbert Simon – geïnteresseerd zijn in het psychologische verloop van het keuzeproces zelf. Het is net als bij taalkundigen. Die maken ook prachtige computationele theorieën, maar zijn nauwelijks geïnteresseerd of verschrikkelijk naïef als het om de implementatie ervan gaat, althans een psychologisch realistische implementatie. Economen vragen zich niet af hoe mensen keuzealternatieven waarnemen, hoe ze verschillende alternatieven tegelijkertijd in het aandachtsveld houden om ze tegen elkaar af te wegen, hoe ze waarden toekennen aan die alternatieven, en kansen van relevante gebeurtenissen schatten (de *uitkomst* van die schattingen interesseert ze soms wel, maar niet de manier waarop die schatting tot stand komt). Ik kan hiertegen werkelijk geen bezwaar maken. Economen hebben er niet voor geleerd, terwijl zulks voor psychologen gesneden koek is. Laat ieder wat dit betreft bij zijn stiel blijven, dan schieten we het hardste op. Alleen moeten we elkaar heel goed informeren. Zoals ik al zei heeft een algorithmische theorie vaak consequenties voor de computationele theorie. Het zou jammer zijn als een systematische proces-variabele verkeerd werd geïnterpreteerd in de computationele theorie. Laat ik daarvan een voorbeeld geven.

Mensen zijn niet, zoals de ideale economische beslisser, optimaal geïnformeerd. Geen wonder dat ze soms schijnbaar irrationele keuzen maken. Maar geen nood, aldus sommige economen, er zijn kosten verbonden aan zich informeren. Laten we dat gewoon meerekenen, dan is het hele systeem toch weer rationeel. Tversky en Kahneman hebben er echter op gewezen dat dat soms de verkeerde toewijzing is. Je kunt een intelligente beslisser volledig informeren over de fout die hij gaat maken. Dat neemt de neiging om die fout te maken absoluut niet weg. Het is in die gevallen net als bij perceptuele illusies. Het kan je verteld worden, je kunt het na-meten, maar de illusie blijft. Het is geen gebrek aan informatie waar het irrationele gedrag uit voorkomt, maar de manier waarop die informatie bijna reflexmatig wordt gecodeerd door mensen. ‘Framing’ noemen de auteurs dat. Die post moet niet onder informatiekosten worden opgenomen in de computationele theorie.

Er zijn ook allerlei andere proces-variabelen die beslissingen kunnen beïnvloeden. Dezelfde auteurs hebben er in een eerder stuk (1973) op gewezen dat mensen bij het doen van schattingen erg afhankelijk zijn van het gemak waarmee ze zich iets kunnen voorstellen. Dit is de *beschikbaarheids*-variabele. Werklozen schatten het aantal werklozen hoger dan mensen die een baan hebben. Dat komt door het proces van schatten. Je loopt even na van wie je zoal weet dat hij werkeloos is. Als je er vlug een boel kan vinden wordt je schatting hoog. Werklozen weten vlugger voorbeelden te vinden dan werkenden (dat is geen wonder), daarom geven ze hogere schattingen. In ons land heeft Wagenaar (1977, 1983, en vele andere publicaties) schitterend onderzoek gedaan naar schattingsfouten die mensen maken. Die fouten kunnen soms dramatische gevolgen hebben voor hun (keuze)gedrag. Ik denk

dat economen goed op de hoogte moeten worden gehouden van zulke onderzoeksresultaten.

Omgekeerd moeten psychologen economische theorieën over het keuzegedrag vrij goed kennen om tot zinvol onderzoek over keuzegedrag te komen. Het implementeren van een computationele theorie in een psychologisch algoritme vereist zeer gedetailleerde kennis van die theorie. Althans zo ervaar ik het bij de psychologische implementatie van taalkundige theorieën. Maar daaruit weet ik ook dat het mogelijk is, als het boek van Keizer en Soeters me daar al niet van had overtuigd.

Voor ik nu tot een conclusie kom wil ik nog kort één punt van mogelijke burenhulp aan de orde stellen, nl. het psychologische experiment.

#### EXPERIMENTELE ECONOMIE

Sinds Wundt is de orthodoxe psychologie een experimentele psychologie. Ik begon met te schetsen hoe de verhouding tussen experiment en theorievorming in de psychologie is gaan schuiven onder invloed van de cognitiewetenschap. We maken tegenwoordig graag complexe gedragstheorieën zonder daarvan steeds elke assumptie te toetsen. Maar anders dan bij Wundt blijft het experiment onze belangrijkste empirische methode, ook als het om complex gedrag gaat. Bovendien zou het een misvatting zijn te denken dat psychologen alleen experimenten doen over proces-modellen. Met hetzelfde gemak toetsen ze computationele theorieën langs experimentele weg.

De experimentele methode bestaat, zover ik begrijp, ook in de economie, maar er wordt niet mee weggelopen. Het is een marginaal verschijnsel. Cramer (1986) vindt het nodig daar eens wat meer aan te doen, en dat heeft natuurlijk mijn instemming. Toch geloof ik dat hij onderhevig is aan een misvatting. Waarom is het zo moeilijk experimenten te doen in de economie? Omdat 'real-life' experimenten zo verschrikkelijk duur zijn zegt hij. ZWO betaalt dat niet graag. 'Real-life' experimenten zijn inderdaad duur, maar zo moet je het ook helemaal niet doen. Je moet de keuzesituatie al of niet in het laboratorium *simuleren*, en psychologen weten hoe je dat doet. Als een micro-economische theorie, d.w.z. een theorie over het individuele keuzegedrag, al niet blijkt op te gaan in het klein, wat moet je dan nog met een grootschalig experiment? En als die theorie wel op gaat in het klein, wat wil je dan nog meer weten? Het blijkt keer op keer dat het gedrag niet wezenlijk verandert als het om meer geld gaat (Smith, 1962; Tversky en Kahneman, 1981). Maar wat nog erger is: 'real-life' is als het weer, een hopeloze interactie van onvoorspelbare variabelen. Dat is niet de plaats voor experimenten, doch slechts voor statistisch en beschrijvend onderzoek. Daar had Wundt gelijk aan. En Wundt had er ook gelijk aan dat dat allerm minst de mogelijkheid tot het geven en toetsen van verklaringen uitsluit.

#### CONCLUSIE

Pen heeft in zijn inleiding een ontwerp gemaakt voor een 'Social Science Vil-

lage'. Dat dorp ziet er uit als een taart met voor ieder van onze wetenschappen een segment. De psychologie echter bevindt zich met haar woning precies in het midden, aldus grenzend aan alle andere sociale wetenschappen. Dit is geheel conform het beeld dat Wundt een eeuw geleden ontwierp over de plaats van de psychologie tussen de wetenschappen. Ik heb geprobeerd die centrale plaats van de psychologie nader te omschrijven. De ronde woning van de psychologen heeft grote open vensters. Binnen ziet men hen werken aan algorithmische gedragstheorieën, theorieën over mentale processen. Maar ze kunnen dat niet zonder door de vensters, en zelfs buiten onder de platanen, te praten met al die burens die computationele theorieën van het menselijk gedrag ontwerpen, theorieën over *uitkomsten* van het gedrag. Er zijn daar geen terreinafscheidingen; de agora loopt geleidelijk over in de tuinen van de burens.

Maar de woning van de psychologen heeft ook een souterrain. Daar wordt gewerkt aan hardware-implementatie. We staan er oog in oog met de menselijke hersenen, die al dat complexe gedrag voortbrengen. En daar hebben de andere sociale wetenschappen niets te maken. Ik verblijf soms graag in dat souterrain.

### *Dankbetuiging*

Zonder de hulp van Jan Pen en van Eduard Hoenkamp was dit stuk niet geschreven. Mijn hartelijke dank aan beiden.

### LITERATUUR

- Bacharach, M.O.L., The problem of agents' beliefs in economic theory. In: M. Baranzini en R. Scazzieri, *Foundations of economics*. Blackwell, New York, 1986.
- Cramer, J.S., Het economisch onderzoek in de toekomst. In: H.M. Jolles, A.P. Plompen en H. Weijma, *Vijftien wetenschappen vijftien jaar verder*. Staatsuitgeverij Den Haag, 1986.
- Dempster, A., Upper and lower probabilities induced by a multivalued mapping. *Am. Math. Statistics*, 38, 325-339, 1967.
- Doyle, J., *Reasoned assumptions and Pareto Optimally*, 1985.
- Doyle, J. en P. London, A selected descriptor-indexed bibliography to the literature on belief revision. *Sigart Newsletter*, 71, 7-23, 1980.
- Frijda, N., *The emotions*. Cambridge University Press, Cambridge, 1986.
- Harman, G., *Change in view*. MIT Press, Cambridge, Mass., 1986.
- Hoenkamp, E., An analysis of psychological experiments on non-monotonic reasoning. *Proceedings of the tenth international conference on artificial intelligence*. 115-117, Morgan Kaufman, Los Altos, Ca., 1987.
- Kahneman, D. en A. Tversky, Prospect theory: An analysis of decision under risk. *Econometrica*, 47, nr. 2., 1979.
- Kahneman, D. en A. Tversky, The psychology of preferences. *Scientific American*, 136-141, 1981.

- Keizer, P.K. en J. Soeters, *Economie, sociologie en psychologie: visies op integratie*. Van Gorcum, Assen, 1987.
- Kempen, G., *Natuurlijke taal en kunstmatige intelligentie*. Wolters-Noordhoff, Groningen, 1987.
- Marr, D., *Vision*. Freeman, San Francisco, 1982.
- Nisbett, R., en L. Ross, *Human inference: Strategies and shortcomings of social judgment*. Prentice-Hall, Englewood Cliffs, 1980.
- Pen, J., *Over de schutting*. Hollands maandblad. 17-21, 1986.
- Pen, J., *Wat het smijten met borden te vertellen heeft*. Vrij Nederland, augustus, 3, 1987.
- Rapoport, W.J., *Logical foundations for belief representation*. *Cognitive science*, 10, 371-422, 1986.
- Ross, L., M. Lepper en M. Hubbard, *Perseverance in self-perception and social perception. Biased attributional processes in the debriefing paradigm*. *Journal of Personality and Social Psychology*, 32, 880-892, 1975.
- Shafer, S., *A mathematical theory of evidence*. University of Princeton Press, Princeton, NJ., 1976.
- Smith, V., *An experimental study of competitive market behavior*. *Journal of political economy*, 70, 118, 1962.
- Tullock, G. en R.B. McKenzie, *The new world of economics*. Irwin, Homewood, 1985.
- Tversky, A. en D. Kahneman, *Availability. A heuristic for judging frequency and probability*. *Cognitive psychology*, 5, 207-232, 1973.
- Tversky, A. en D. Kahneman, *The framing of decisions and the rationality of choice*. Manuscript 1981.
- Wagenaar, W.A., *De beste stuurlied dempen de put*. Ambo, Baarn, 1977.
- Wagenaar, W.A., *Menselijk falen*. *Nederlands Tijdschrift voor de Psychologie*, 38, 209-222, 1983.
- Wegner, D., Coulton, G. en Wenzlaff, R., *The transparency of denial. Briefing in the debriefing paradigm*. *Journal of Personality and Social Psychology*, 49, 338-346, 1985.