Computationele wetenschapsfilosofie

Kuipers, Theodorus

Published in:
Algemeen Nederlands tijdschrift voor wijsbegeerte

Publication date:
1993

REVIEW-ARTIKEL

COMPUTATIONELE WETENSCHAPSFILOSOFIE

Bespreking van:
P. Langley, H.Simon, G.Bradshaw, J.Zytkow, *Scientific discovery. Computational explorations of the creative mind*, MIT-press, Cambridge, 1987.
P. Thagard, *Computational philosophy of science*, MIT-press, Cambridge, 1988.
P. Thagard, *Conceptual revolutions*, Princeton UP, Princeton 1992.
J. Shrager & P. Langley (eds.), *Computational models of scientific discovery and theory formation*, Kaufmann, San Mateo, 1990.

1. Inleiding

Computationele wetenschapsfilosofie is een co-productie van wetenschapsfilosofie en cognitiewetenschap. Cognitiewetenschap is zelf een co-productie van cognitieve psychologie, artificiële intelligentie, taalwetenschap, neurowetenschap en filosofie (Stillings et al., 1987), die allerlei interessante problemen oproept waar wetenschapsfilosofen iets over menen te kunnen zeggen. Bechtel (1988) geeft een toegankelijke inleiding in de wetenschapsfilosofie die speciaal bedoeld is voor cognitiewetenschappers. Hij besteedt onder andere veel aandacht aan reductionistische en niet-reductionistische samenwerkingsmogelijkheden tussen de betrokken disciplines. Kortom, Bechtel demonstreert in zijn boek op cognitiewetenschap toegepaste wetenschapsfilosofie.
In de computationele wetenschapsfilosofie zijn de rollen omgedraaid. Het betreft op wetenschapsfilosofie toegepaste cognitiewetenwetenschap. Geprobeerd wordt om klassieke wetenschapsfilosofische problemen op te lossen met middelen die in het bijzonder ontwikkeld zijn in de artificiële intelligentie en de cognitieve psychologie. Het soort resultaten dat daarbij voor ogen staat zijn computerprogramma's die bepaalde dingen kunnen, zoals wetten (her-)ontdekken, verklarende hypothesen opstellen, begrippen vormen, theorieën ontwikkelen en reviseren, theorieën vergelijken en selecteren, experimenten voorstellen, etc.
Computationele wetenschapsfilosofie kan worden gekenmerkt door drie uitgangspunten.
Het wetenschapsfilosofisch meest revolutionaire uitgangspunt is dat niet alleen het evaluatieproces, de *Context of Justification* (CoJ), maar ook het ontdekkingsproces, de *Context of Discovery* (CoD), onderzoekbaar en in elk geval ten dele programmeerbaar wordt geacht. Sommige onderzoekers houden daarbij de klassieke scheiding van deze contexten zelfs niet voor zinvol omdat in de CoD geanticipeerd wordt op de CoJ en omdat de CoJ allerlei sporen van de CoD draagt. Voorts wordt algemeen aangenomen dat de traditionele CoJ-doelen van de logisch-empiristen afgezwakt moeten worden in Popperiaanse richting: zekerheid over kennisaan-

spraken is nooit te verkrijgen, alle beoordelingen zijn voorlopig, en de hypothese die gevormd wordt hoeft niet de enige te zijn die verenigbaar is met de experimentele gegevens.
Het tweede uitgangspunt van de computationele wetenschapsfilosofie is dat wetenschappelijk onderzoek wordt opgevat als een vorm van probleemoplossen. In de samenwerking tussen cognitieve psychologie en artificiële intelligentie is voor probleemoplossen een algemeen paradigma ontwikkeld door Allen Newell en Herbert Simon dat bekend staat als *heuristisch zoeken*. Schaakcomputerprogramma's zijn typisch op dit paradigma gebaseerd en de volgende abstracte karakterisering laat zich dan ook gemakkelijk invullen in eenvoudige schaaktermen. Eerst moet een probleemruimte van mogelijke toestanden worden gedefinieerd. Vervolgens wordt een probleem gekarakteriseerd als het verschil tussen een begintoestand en een doeltoestand. Toestandsovergangen worden gemaakt met behulp van ''als/*doe* dan''-(of productie- of computatie- of conditie-actie)regels, waarin heuristische overwegingen zijn ingebakken, reden waarom ze heuristische operaties worden genoemd. Een computerprogramma voor een bepaald type probleem bestaat derhalve uit een passende specificatie van de probleemruimte, in termen waarvan de begintoestand (gegevens) en de doelstand kunnen worden ingevoerd, en een geheel van heuristische operaties.
Het derde uitgangspunt van de computationele wetenschapsfilosofie is dat men, afgezien van de eerder genoemde direct beoogde resultaten, één of meer van de volgende doelstellingen voor ogen heeft met zulke computerprogramma's: * filosofische adequaatheid: de beste theorie opstellen en selecteren, of zelfs de theorie die het dichtste bij de waarheid is, * historische adequaatheid: wetenschapshistorische hoogtepunten reproduceren, * psychologische of zelfs neuropsychologische adequaatheid: de processen die omgaan in de hoofden van wetenschappers in detail simuleren, * sociologische adequaatheid: wetenschapsbeoefening als groepsproces simuleren, * practische relevantie: het ontdekken en evalueren in lopend onderzoek ondersteunen. Het is duidelijk dat deze vijf doelstellingen onafhankelijk van elkaar kunnen zijn en dat men dus in beginsel maar één van deze doelstellingen expliciet hoeft na te streven.

In dit artikel zal ik twee hoofdlijnen uit de computationele wetenschapsfilosofie bespreken. In de eerste hoofdlijn (Par. 2), die is ingezet door Simon, gaat het om programma's die klassiek computationeel kunnen worden genoemd omdat zogenoemde activatie-spreiding geen rol speelt. Het betreft in de eerste plaats de BACON-programma's. Deze zijn gericht op het herontdekken van kwantitatieve wetten in met name de natuurkunde, bijvoorbeeld de wetten van Kepler en de ideale gaswet. Voorts gaat het om de programma's GLAUBER, STAHL en DALTON, welke zijn gericht op het herontdekken van kwalitatieve wetten uit de klassieke scheikunde. Beide groepen programma's worden uitvoerig behandeld in Langley et al. (1987). Steeds geldt dat de evaluatie van de gevormde hypothese, op grond van de beschikbare data, is ingebakken in de 'berekening' van de hypothese, de hypothesevorming, op grond van die data.
De tweede hoofdlijn (Par. 3) betreft het werk van Paul Thagard (1988, 1992). Bij hem spelen complexe begrippen een centrale rol en hij gebruikt daarbij het idee

van activatie-spreiding dat stamt uit de cognitieve psychologie. Voorts houdt hij bij het programmeren van ontdekken en evalueren de klassieke scheiding tussen beide contexten in acht. In het programma PI (Processes of Induction) van 1988 staan allerlei soorten inductie en abductie centraal. In zijn boek van 1992 richt Thagard zich met het programma ECHO vooral op theorieselectie, waarbij wel de observationele begrippen van de in het geding zijnde theorieën overeen moeten komen, maar niet de theoretische begrippen. Thagard heeft ECHO met succes getoetst aan de hand van een aantal historische voorbeelden, te weten Lavoisier, Darwin en Wegener. Bij de presentatie hiervan zal ik de vraag bespreken of ECHO niet gecompliceerder is dan nodig is, althans voorzover het deze voorbeelden betreft.
In Par. 4 zal ik nog zeer korte impressies geven van de zes hoofstukken uit Shrager en Langley (eds., 1990) waarin het reviseren van theorieën centraal staat.
Tot slot zal ik in Par. 5 een beknopte aanduiding geven van de analyses uit de (neo-)klassieke wetenschapsfilosofie die zich bij uitstek lenen voor computationele implementatie.
In dit artikel zal de specificatie van programma's niet aan de orde komen. Wat *programmeertalen* betreft volsta ik met te vermelden dat de te gebruiken taal in eerste instantie erg afhankelijk is van het specifieke doel, maar dat in tweede instantie het belang van talen die verschillende doelen kunnen dienen steeds groter wordt omdat de programma's op elkaar moeten kunnen aansluiten. Varianten van de computertaal LISP blijken in dit verband zeer geschikt. Wat *technieken van kennisrepresentatie* betreft is er veel ontwikkeld voor expert-systemen. Davis (1990) geeft een goede en brede introductie, o.a. van Forbus' representatie van kwalitatieve processen. Het is zeker niet beperkt tot alledaagse kennis, zoals de titel van het boek van Davis suggereert.
Het is onmogelijk in de volgende schetsen recht te doen aan alle 'mitsen en maren' die de auteurs aan hun claims (zouden moeten) verbinden. De uiteenzetting heeft daardoor een optimistischere toon dan strikt genomen gerechtvaardigd zou zijn. Niettemin mag zeker wel geconcludeerd worden uit deze toonzetting dat ik persoonlijk vind dat het om 'belangwekkende ontwikkelingen' gaat.

2. Herontdekken van wetten

Scientific discovery (Langley et al., 1987) is ongetwijfeld het meest bekende boek. Zoals gezegd, het presenteert de oudste programma's, de BACON-programma's, welke gericht zijn op kwantitatieve natuurkundige voorbeelden. Voorts worden enkele programma's beschreven waarin kwalitatieve scheikundige voorbeelden centraal staan.

2.1. Op zoek naar kwantitatieve wetten met de BACON-programma's

Het specifieke doel van de BACON-programma's is het ontdekken van kwantitatieve wetten die de beschikbare data bij benadering samenvatten. Uitgangspunt is een verzameling getalswaarden van een aantal variabelen. Het doel wordt benaderd door middel van expliciete definitie van nieuwe termen op basis van heuristische operaties en het toetsen of een constante functie is bereikt dan wel een lineair verband.

De heuristische basis-operaties zijn de volgende: als twee termen omgekeerd gerelateerd zijn, d.w.z. als de ene stijgt daalt de andere, dan wordt het product gevormd, en als twee termen direct gerelateerd zijn, d.w.z. tegelijkertijd dalen of stijgen, dan wordt het quotiënt gevormd. Bij de toepassing van deze operaties krijgen later gedefinieerde termen voorrang.
In de volgende tabellen worden de relevante stappen weergegeven voor het opstellen van de derde wet van Kepler op basis van gefingeerde gegevens die de wet gehoorzamen (Tab. 1) en op basis van de originele gegevens op grond waarvan Newton controleerde of de manen van Jupiter aan de wet voldeden (Tab. 2). In beide gevallen wordt eerst geconstateerd dat de omloopperiode (P) van een planeet resp. maan toeneemt als de afstand (A) van het betreffende hemellichaam tot de zon resp. Jupiter toeneemt. Op grond van de tweede heuristische operatie wordt derhalve het quotiënt A/P als nieuwe term gedefinieerd. Dit quotiënt blijkt af te nemen als de afstand toeneemt, dus wordt volgens de eerste heuristische operatie het product Ax(A/P) gedefinieerd. Dat product blijkt toe te nemen als het eerder gedefinieerde quotiënt afneemt, dus wordt weer het product van die twee termen gedefinieerd: (A/P) x Ax(A/P) = A^3/P^2. De derde wet van Kepler zegt dat precies deze term een constante is, en dat constateert het programma dan ook direct in het eerste geval en stelt dat als wet voor. In het tweede geval concludeert het programma, indien de benaderingsmarge ruim is afgesteld, dat die laatste term bij (grove) benadering constant is en stelt als wet voor dat die term gelijk is aan de gemiddelde waarde + of – de benaderingsmarge.

Planeet	Afstand A	Periode P	Term-1=A/P	Term-2=AxT1	Term-3=T1xT2
A	1.0	1.0	1.0	1.0	1.0
B	4.0	8.0	0.5	2.0	1.0
C	9.0	27.0	0.333	3.0	1.0

Tabel 1

Maan	Afstand A	Periode P	Term-1=A/P	Term-2=AxT1	Term-3=T1xT2
A	5.67	1.769	3.203	18.153	58.15
B	8.67	3.571	2.427	21.035	51.06
C	14.00	7.155	1.957	27.395	53.61
D	24.67	16.689	1.478	36.459	53.89

Tabel 2

Zoals uit het tweede voorbeeld duidelijk is gebleken kan een programma dat met 'echte data' werkt niet zonder afrondingskriteria, omdat echte kwantitatieve gegevens om uiteenlopende redenen nou eenmaal nooit puntgaaf aan wetten beantwoorden. De afrondingsproblematiek krijgt in het onderhavige boek overigens minder aandacht dan nodig is. In Shrager & Langley (1990), het vierde te bespreken boek, is er wel een heel hoofdstuk aan gewijd.
Aan het geschetste basisprogramma van BACON zijn steeds meer heuristische operaties en andere verfijningen toegevoegd. Deze worden kort aangestipt in de volgende opsomming van herontdekte wetten:
– met het geschetste basisprogramma: de wetten van Kepler, Boyle, Ohm en Galilei,

- na verruiming tot meer dan twee variabelen en toevoeging van een methode om de data gelaagd af te zoeken: de ideale gaswet, en de wetten van Coulomb, Kepler (verfijnd) en Ohm (verfijnd),
- na invoering van de mogelijkheid tot het voorstellen van intrinsieke eigenschappen van objecten
  - door middel van een 'probeertruc': de wetten van Ohm (nog verfijnder), met weerstand als intrinsiek begrip, Archimedes (volume), Snellius (brekingsindices), Black (soortelijke warmte), impulsbehoud (trage massa) en de gravitatiewet (zware massa),
  - of door middel van het zoeken van een gemeenschappelijke deler: de resultaten van Cannizaro (atoomgewichten) en Millikan (lading electron).
- uitgaande van een theoretisch gezichtspunt, zoals het zoeken naar symmetrieën en naar behoudswetten: verfijnde versies van de wetten van Snellius en Black, en de wetten van impuls- en energiebehoud.

De presentatie in *Scientific Discovery* is meestal vrij helder, maar bij enkele voorbeelden blijft (mij) de werking van de laatstgenoemde verfijningen onduidelijk, te weten het derde stadium voor de wet van Snellius (p. 177) en de wijze waarop J wordt ingevoerd bij de wet van Black.

2.2. Op zoek naar kwalitatieve wetten en modellen

Het is vanzelfsprekend dat de kwantitatieve BACON-programma's in de eerste plaats natuurkundige voorbeelden hebben. Het is evenmin verbazingwekkend dat programma's gericht op kwalitatieve wetten en modellen eerder scheikundige voorbeelden hebben. De drie bekendste 'kwalitatieve' programma's zijn dan ook genoemd naar hoofdfiguren uit de geschiedenis van de scheikunde.

GLAUBER gaat uit van een aantal reactievergelijkingen en eigenschappen (smaak etc.) van de deelnemende stoffen, en vormt klassen van stoffen en reactiewetten in termen van die klassen van stoffen. Zo werden, uitgaande van de smaak van acht stoffen en vier reactievergelijkingen daartussen, de klassen 'zuren', 'basen' en 'zouten' gevormd en werd de wet ''zuren en basen vormen zouten'' opgesteld. De heuristische basisoperaties in het programma kunnen als volgt worden gekarakteriseerd: vorm steeds de grootst mogelijke klasse, en kwantificeer *zo mogelijk* universeel en anders existentieel.

STAHL gaat ook uit van een aantal reactievergelijkingen, maar is gericht op de ontleding van verbindingen in hun componenten. De heuristische operaties zijn even simpel als voor de handliggend: schrap stoffen die links en rechts optreden in een vergelijking; als aan een kant maar één stof staat, concludeer dan dat aan de andere kant zijn componenten staan; en vervang een stof door zijn componenten in de resterende vergelijkingen.

Ook DALTON gaat uit van reactievergelijkingen, in het bijzonder in de vorm van betrokken gasvolumina, en is in eerste instantie gericht op de bepaling van het aantal moleculen van elke stof in de reactievergelijkingen, de zogenoemde moleculaire reactievergelijkingen, en in tweede instantie op de bepaling van het aantal atomen in de moleculen van de deelnemende stoffen, de zogenoemde molecuulformules. De heuristische operaties volgen vrij nauwkeurig de principes die Avogadro al introduceerde: de aantallen moleculen moeten in overeenstemming zijn met zijn

beroemde hypothese, die zegt dat gelijke volumina van verschillende gassen bij gelijke temperatuur en druk evenveel moleculen moeten bevatten; voorts moet het aantal atomen links en rechts gelijk zijn; voor het overige moeten molecuulformules zo eenvoudig mogelijk zijn. Het is duidelijk dat DALTON zeer theoriegestuurd te werk gaat en voorts dat nieuwe reactievergelijkingen kunnen dwingen tot herziening van de molecuulformules.
Geheel in de lijn van een klassiek voorbeeld van Avogadro is het eenvoudig om na te gaan dat uit de vergelijking ''2 liter waterstof + 1 liter zuurstof --> 2 liter waterdamp'' en de heuristische operaties de moleculaire vergelijking ''2 waterstofmoleculen + 1 zuurstofmolecuul --> 2 watermoleculen'' volgt en daaruit drie molecuulformules, waarvan er twee afwijken van onze huidige inzichten: ''1 waterstofmolecuul = 1 (nu 2!) waterstofatoom'', ''1 zuurstofmolecuul = 2 zuurstofatomen'' en ''1 watermolecuul = 1 (nu 2!) waterstofatoom + 1 zuurstofatoom''. Iets moeilijker is het om vervolgens na te gaan dat, als de vergelijking van ammoniakproductie (''3 liter waterstof + 1 liter stikstof --> 2 liter ammoniak'') wordt toegevoegd, de voorgestelde molecuulformules voor waterstof en water niet meer aan alle voorwaarden voldoen en dat dan de ons bekende molecuulformules daarvoor in de plaats worden voorgesteld.

De meest voor de hand liggende kritiek op de tot nu toe besproken programma's is ongetwijfeld dat de probleemstelling steeds uitgaat van een gegeven conceptualisering, terwijl het vinden van een geschikte conceptualisering meestal een belangrijk onderdeel uitmaakt van het ontdekkingsproces. De auteurs maken echter overtuigend aannemelijk dat de conceptualisering in werkelijkheid vaak geruime tijd voor de ontdekking van de betreffende wet bekend was, zodat het betreffende type (her-)ontdekking wel degelijk van groot wetenschapshistorisch belang is.
Tot slot zij vermeld dat in het onderhavige boek regelmatig aardige uitstapjes worden gemaakt naar toevallige ontdekkingen en de programmeerbaarheid daarvan. Zeker met de technieken die verderop aan de orde komen, lijken de opmerkingen daarover in tweede instantie veel realistischer dan op het eerste gezicht: met name (deel-)programma's voor het opsporen van anomalieën en het opstellen van analoge verklaringen kunnen steeds de vinger aan de pols houden voor ongezochte vondsten.

3. Hypothesevorming volgens PI en selectie volgens ECHO

Paul Thagard geeft in zijn boek van 1988 een uitvoerige presentatie van het programma PI (Processes of Induction), dat zich zowel richt op hypothesevorming als op evaluatie en selectie. In zijn boek van 1992 staat de verandering van conceptuele systemen centraal. Voor de ontwikkeling van een nieuw conceptueel systeem en een daarbinnen geformuleerde theorie blijft PI van kracht, maar voor de selectie van theorieën, al dan niet binnen hetzelfde conceptuele kader, introduceert hij het nieuwe programma ECHO. In deze bespreking zullen we de kennismaking met PI beperken tot hypothesevorming.

3.1. Vormen van inductie

De aanpak van Thagard sluit veel directer aan bij wat gangbaar is in de cognitieve

psychologie dan de in de vorige paragraaf behandelde programma's, welke primair vanuit de artificiële intelligentie ontwikkeld zijn. Uitgangspunt vormen standaardrepresentaties voor boodschappen (messages), begrippen, regels (voor wetten en hypothesen zonder theoretische termen), theorieën (met theoretische termen) en, tot slot, verklaringsproblemen. Een *boodschap* komt neer op een atomaire bewering aangevuld met een waarheidswaarde of betrouwbaarheidsgraad, zijn status en een naam. Een *begrip* (b.v. geluid) is een geordend geheel van boven- en ondergeschikte begrippen, instanties van het begrip en regels die het begrip activeren. In het volgende schema is het geluidsvoorbeeld uitgewerkt:

Naam:	geluid
Type:	begrip
Activatie:	0.75
Bovengeschikt:	fysisch verschijnsel, gewaarwording
Ondergeschikt:	stem, muziek, fluittoon, diergeluiden
Instanties:	het klokgelui van vanmiddag twaalf uur

Geactiveerd door de regels:

1. als x gehoord wordt dan is x een geluid
2. als x een geluid is dan plant het zich voort in lucht
3. als x een geluid is en wordt tegengehouden dan weerkaatst het
4. als x een geluid is en y een persoon en als x dichtbij y is dan hoort y x
5. als x een geluid is dan verspreidt het zich bolvormig

Schema 1

In bovenstaand schema worden regels genoemd die elders in het kennisbestand in extenso zijn gerepresenteerd. Zo'n *regel* postuleert in de vorm van een ''conditie-actie''-regel een empirische verband tussen twee of meer begrippen (geluid plant zich voort/ weerkaatst) en wordt voorzien van een boekhouding van tegenvoorbeelden en verklaringssuccessen, en een hierop gebaseerde betrouwbaarheidsgraad. Een *theorie* is een geheel van observatiebegrippen (geluid, golf) en theoretische begrippen (geluidsgolf), bijbehorende regels, tegenvoorbeelden en verklaringssuccessen. Tot slot worden twee typen *verklarings-problemen* in termen van begin- en doeltoestand onderscheiden, namelijk verklaringsproblemen van individuele en algemene aard: de begintoestand betreft een individueel of algemeen object (een bepaald geluid resp. een willekeurig geluid) en de doeltoestand specificeert de gedragsaspecten (voortplanten en weerkaatsen) waarvoor een verklaring wordt gezocht.

Zoals alle problemen in veel cognitief psychologische modellen worden verklaringsproblemen door PI aangepakt door activatie van boodschappen en begrippen in het probleem, gevolgd door activatie van de daaraan gekoppelde begrippen en regels. Geactiveerde regels kunnen vuren (rule-firing) als een drempelwaarde overschreden wordt. Het probleem is opgelost als er een verbinding (matching) tot stand komt tussen begin- en doeltoestand. Het bijzondere van PI is dat, als de verbinding niet tot stand komt, verschillende soorten *inductie* gericht op hypothesevoming geprobeerd worden, gevolgd door evaluatie, te weten inductieve generalisatie, diverse vormen van abductie en, tot slot, begrips- en bijbehorende theorievorming.

Inductie is het zetten van niet-deductieve stappen. Bij *inductieve generalisatie*

wordt, indien Ga te verklaren is bij begintoestand Fa en er geen verbinding tot stand komt, de regel ''als Fx dan Gx'' gevormd en geëvalueerd aan de hand van andere informatie die verbonden is met de begrippen F en G. Bij voldoende gevarieerde steun leidt dit tot opname van de regel in het kennisbestand, met bijbehorende betrouwbaarheidsgraad (voor verdere details, zie Holland et al., 1986). In het algemeen zal de nieuwe regel daarna meteen vuren en komt de verbinding tussen Fa en Ga tot stand.

Abductie is een paraplu-benaming voor vier andere vormen van hypothese-vorming, die in PI alle gevolgd worden door eenzelfde evaluatie-methode, welke in het boek van 1922 door ECHO vervangen wordt. Basisidee achter abductie (Peirce) is het voorstellen van een hypothese als deze het te verklaren probleem zou kunnen verklaren. De vier specifieke abductie-methoden kunnen als volgt kort getypeerd worden:

Bij *individuele abductie* wordt, indien boodschap Ga te verklaren is en de regel ''als Fx dan Gx'' in het kennisbestand zit (en dus geactiveerd wordt), de hypothese Fa voorgesteld, vanwege het feit dat Ga dan met de genoemde regel verklaard kan worden. Als bijvoorbeeld te verklaren is dat een bepaald geluid zich voortplant en als de wet ''golven planten zich voort'' in het kennisbestand zit, dan wordt de hypothese voorgesteld dat het betreffende geluid een golf is.

Bij *regel-abductie* wordt ter verklaring van de wet ''als Fx dan Gx'', indien de wet ''als Hx dan Gx'' in het kennisbestand zit, de hypothese ''als Fx dan Hx'' gevormd. Om bijvoorbeeld ''geluid plant zich voort'' te verklaren als ''golven planten zich voort'' in het kennisbestand zit, wordt de hypothese ''geluiden zijn golven'' gevormd.

Regel-abductie is min of meer een algemene variant van individuele abductie. Beide lijken voorbeelden van de drogreden die bekend staat als het 'bevestigen van de consequent', ware het niet dat het slechts om hypothese-voorstellen gaat die nog geëvalueerd moeten worden. De derde vorm van abductie, existentiële abductie, behoort tot dezelfde categorie ongeldige redeneringen, de vierde, analoge abductie, is van heel andere aard.

Bij *existentiële abductie* wordt ter verklaring van boodschap Ga, indien de regel ''als S(x,y) dan Gx'' in het kennisbestand zit, de hypothese ''er is een y zodanig dat S(a,y)'' gevormd. Bekende voorbeelden zijn het postuleren van de planeet Neptunus en het postuleren van flogiston, en later zuurstof. Deze vorm van abductie komt ook in de metafysica voor, denk bijvoorbeeld aan het postuleren van een mensonafhankelijke en dus ongeconceptualiseerde werkelijkheid.

De minst gefundeerde hypothesen worden gevormd met *analoge abductie* (AA): stel een individueel of algemeen feit F* is te verklaren en stel dat de activatiepatronen van F en F* op elkaar lijken en F wordt volgens het kennisbestand verklaard door G, construeer dan een maximale afbeelding van de activatiepatronen rond F en G in *-termen en postuleer het bijbehorend analogon G* van G als hypothese.

Analoge abductie geeft aanleiding tot enkele korte uitweidingen. Ten eerste, indien G* de evaluatie met succes doorloopt, wordt vervolgens een abstract schema gevormd, met de paren <F,G> en <F*,G*> als instanties, waardoor een gegeneraliseerde vorm van analoge inductie (GAA) mogelijk wordt in de toekomst. Ten tweede, er is ook een specifieke causale vorm van AA (CAA): als F en F* op

elkaar lijken, F en G causaal gerelateerd zijn, en G en G* lijken ook op elkaar dan wordt de hypothese dat F* en G* causaal gerelateerd zijn gevormd. Uiteraard is hier ook weer een gegeneraliseerde vorm van mogelijk.
De derde en laatste vorm van inductie betreft de *vorming van een theoretische begrip en bijbehorende theorie*. Stel dat bij een algemeen verklaringsprobleem een initieel geactiveerd begrip leidt tot activatie van een tweede begrip voor dezelfde instantie(s), zodanig dat de aan de twee begrippen gekoppelde regels ten dele strijdig zijn. In dat geval wordt een combinatie-begrip gevormd met de volgende regels: alle regels van het initieel begrip en voorts de daarmee niet-strijdige regels van het secundair geactiveerde begrip. De bijbehorende theorie zegt dat instanties van het initieel begrip ook instanties van het combinatie-begrip zijn.
Bij het voorbeeld van regel-abductie werd het begrip geluid in verband gebracht met het begrip (water)golf, hetgeen leidt tot strijdige verwachtingen, namelijk bolvormige resp. vlakvormige voortplanting. In dit geval wordt het begrip geluidsgolf gevormd, met daaraan gekoppeld bolvormige voortplanting, en de golftheorie van geluid ''geluiden zijn geluidsgolven'' met o.a. als verklaringssuccessen dat geluiden zich voortplanten en weerkaatsen.

Resten in deze subparagraaf nog enige opmerkingen over de rijkdom aan ideeën in het onderhavige boek over wetenschapsfilosofische onderwerpen vanuit computationeel perspectief.
Het evaluatie-fragment van PI betreft een implementatie van zogenoemde Inference to the best explanation (IBE). IBE roept zelf de vraag op naar zijn rechtvaardiging. Thagard besteedt daaraan maar liefst twee hoofdstukken. Eerst ontleent hij drie kriteria voor evaluatie van argumentatiesystemen, bestaande uit principes, praktijken, achtergrondtheorieën en doelen, aan andere argumentatiesystemen, o.a. de methode van het 'reflective equilibrium' in de sociale ethiek van Rawls. De drie kriteria zijn 'robuustheid', 'accomodatie' en 'doelmatigheid' en leiden voor IBE-argumentatie tot de volgende conclusies. IBE is robuust omdat het welbeschouwd zeer veel voorkomt (o.a. blijkens de eerder genoemde voorbeelden). Ook bij filosofische theorieën wordt IBE veel gebruikt. Afwijkingen van IBE zijn verklaarbaar (accomodeerbaar) met behulp van psychologische motivatietheorieën. En IBE is doelmatig gemeten aan drie doelen: verklaren, voorspellen en waarheidsbenadering. Het eerste is evident, dat het tweede doel gehaald wordt volgt omdat voorspellen een vorm van verklaren is. Voor het derde doel, waarheidsbenadering, bewandelt Thagard weer een omweg door eerst tot realisme te besluiten, althans voor de natuurwetenschappen. Dat doet hij met een (volgens hem niet-circulaire) toepassing van IBE op de relevante filosofische theorieën ter verklaring van technologische toepassingen, kennisaccumulatie en consensus over kennisaanspraken. Volgens mij is deze laatste omweg, waarvan het zeer de vraag is of deze daadwerkelijk niet-circulair is, niet nodig, omdat IBE, mits gecorrigeerd tot 'inference to the best theory as the closest to the truth', aantoonbaar functioneel is voor waarheidsbenadering (zie verderop, zie ook Kuipers, 1991).
Evolutionaire epistemologie is volgens Thagard gebaseerd op een slechte analogie: op de drie hoofdproblemen (verklaringen voor variatie, selectie en transmissie) geeft deze epistemologie verkeerde antwoorden. Thagard maakt aannemelijk dat

dat bijzonder duidelijk wordt bij het ontwikkelen van computerprogramma's als PI. Die stelling spoort overigens ook goed met Goldman (1986), volgens wie de wetenschapsfilosofie behoort tot de secundaire epistemologie. Terwijl de primaire epistemologie zich bezighoudt met aangeboren kenproceseen, gaat de secundaire epistemologie over aangeleerde kenprocessen, i.e. methoden van kennisverwerving. Voor het eerste type processen is een naturalistisch-evolutionaire benadering heel plausibel, maar niet voor het tweede type, omdat daar allerlei normatieve overwegingen een belangrijke rol spelen.

Thagard presenteert profielen voor wetenschap en pseudo-wetenschap van Wittgensteiniaanse aard, dus met tussenvormen. Belangrijke accentverschillen tussen wetenschap en pseudowetenschap betreffen correlatie- versus gelijkenisdenken, wel of niet gericht zijn op empirische evaluatie, wel of niet vergelijken met andere theorieën, wel of niet streven naar eenvoudige theorieën, wel of niet streven naar verbetering van theorieën. Dit soort kenmerken worden expliciet bij het programmeren van taken van wetenschappelijk onderzoek.

De computationele benadering in de cognitiewetenschap kan volgens Thagard leiden tot 'echte' wetenschap, met 'realistische successen'. Daarvoor is het wel nodig dat de beperking van PI tot inductie (en evaluatie) van hypothesen in het kader van probleemoplossen wordt opgeheven en experimenteren in de beschouwing betrokken wordt. De meeste verbindingspijlen tussen de drie hoofdactiviteiten, probleemoplossen, inductie, experimenteren, moeten echter nog worden ingevuld.

Thagard benadrukt tot slot het belang van parallelle computationele systemen, omdat het leidt tot betere programma's. Bovendien stelt Thagard voor groepsrationalietit te bestuderen door een wetenschappelijke gemeenschap op te vatten als een geheel van parallelle computationele systemen. Deze analogie roept bijvoorbeeld de vraag op naar de mogelijkheid en wenselijkheid van verdergaande taakverdeling dan de bekende verdeling tussen theoretici en experimentatoren. Daarbij denkt hij vooral aan die tussen constructieve dogmatici en sceptische kritici, om de termen te gebruiken waarin ik zelf dit idee eerder heb geopperd (Kuipers, 1992).

3.2 Selectie van theorieën

Thagard behandelt in zijn boek van 1992 in de eerste plaats uitvoerig zijn visie op conceptuele systemen en hun verandering. Volgens hem worden conceptuele systemen, en theorieën daarbinnen, stapsgewijs (piece-meal) gevormd, maar vindt hun vervanging holistisch plaats. Dit geldt zowel voor de pionierende onderzoeker(s), als voor degenen die later volgen.

Bij de vorming van conceptuele systemen kunnen programma's als PI en BACON een belangrijke rol vervullen. Thagard voegt in dit opzicht geen nieuwe deelprogramma's toe, maar beperkt zich tot een zeer toegankelijke en plausibele behandeling van de structuur van conceptuele systemen en hun verandering in het verlengde van zijn boek van 1988. Ik beperk me hier tot het basisidee van een conceptueel systeem en van conceptuele verandering. Een conceptueel systeem bestaat uit een geheel van begrippen, georganiseerd in soort- en deel-hiërarchieën, en verder onderling verbonden door regels die wetmatige verbanden representeren, en bovendien verbonden met concrete individuen: de laatste kunnen instanties zijn van een

bepaald soort-begrip of er kan een eigenschapsbegrip op van toepassing zijn. Uit deze opbouw van een conceptueel systeem is direct af te leiden welke meer en minder ingrijpende veranderingen er mogelijk zijn van een conceptueel systeem. Van alleen maar verandering van opvattingen over wetmatige verbanden in termen van ongewijzigde concepten (belief revision), tot reorganisatie van specifieke soort- en deel-hiërarchieën of zelfs verandering van algemene organisatieprincipes.

Zoals gezegd, de (her-)vorming van zulke systemen gebeurt stukje bij beetje, terwijl de uiteindelijke vervanging het resultaat is van een vergelijkende evaluatie van beide systemen als gehelen. Thagard pretendeert met het programma ECHO een simulatie van vervanging van theorieën inclusief bijbehorende conceptuele systemen te bieden die historisch adequaat is. De theorie achter ECHO heet TEC: Theory of explanatory coherence. Het gaat in feite om een verfijnde versie van de theorie achter het hypothese-selectie fragment in PI. Dat betreft zogenoemde 'Inference to the best explanation' (IBE). In PI wordt de beste verklaring gedefinieerd als de hypothese met de hoogste waarde, waarbij de waarde wordt bepaald door het product van het verklaringssucces (consilience) en de eenvoud van de hypothese.
TEC bestaat uit zeven principes die relaties van verklaringscoherentie ('explanatory coherence') vaststellen en de beoordeling mogelijk maken van de aanvaardbaarheid van de hypothesen van een verklaringssysteem. In de zeven principes spelen verklaringsrelaties en contradicties tussen een of meer hypothesen en de data een belangrijke rol. Daarnaast zijn verklaringsrelaties tussen hypothesen onderling van invloed. In alle typen verklaringsrelaties zijn eenvoudsoverwegingen ingebouwd: hoe minder verklarende premissen hoe sterker de verklaringscoherentie tussen de premissen onderling en tussen de premissen en de conclusie. Tot slot, kunnen ook analogie-overwegingen aanleiding geven tot vergroting van de verklaringscoherentie. Volgens TEC moet de oude theorie vervangen worden door de nieuwe theorie als de nieuwe een (veel) grotere verklaringscoherentie heeft dan de oude.
Terwijl TEC een verfijnde versie is van IBE, is ECHO, het implementatieprogramma van TEC, allesbehalve een verfijnde versie van het klassiek computationele fragment van PI dat IBE implementeert. Zoals veel andere fragmenten van PI quasi-connectionistisch zijn is ook ECHO quasi-connectionistisch. Het gebruikt activatie-spreiding en parallelle vervulling van beperkingen (parallel constraint satisfaction). Het is niet echt connectionistisch omdat de representatie van proposities niet gedistribueerd is: proposities treden zelf op als de knooppunten van het netwerk. Hoe dit ook zij, selectie is met behulp van ECHO, zoals verwacht kan worden, een ondoorzichtig proces van herhaalde herwaardering van (systemen van) proposities. Dat neemt niet weg dat dit proces gemakkelijk kan convergeren naar een ondubbelzinnige conclusie ten gunste van de ene of de andere theorie.
Helaas is het onmogelijk in dit bestek dieper in te gaan op de werking van ECHO. Het is echter wel mogelijk om aan te geven waarom TEC en ECHO vooralsnog ingewikkelder lijken dan nodig is en wellicht fundamenteel op het verkeerde spoor zitten. Elders (Kuipers, 1993b) heb ik dit uitvoerig uitgewerkt.
Mijn claim is dat TEC en ECHO qua opbouw verkeerd in elkaar zitten. Uitgangspunt is een mengeling van overwegingen: verklaringsrelaties, eenvoud, an-

alogie etc. In mijn optiek dient theorie-selectie nadrukkelijk gelaagd te zijn opgebouwd, zodanig dat eerst verklaringssuccessen en -problemen van twee theorieën worden vergeleken. Als dat niet tot de duidelijke conclusie leidt dat de ene theorie (vooralsnog) superieur is in verklaringssucces aan de andere dan kunnen vervolgens pragmatische overwegingen als eenvoud en analogie in stelling worden gebracht. Mijn argumenten voor de gelaagde aanpak betreffen historische en filosofische adequaatheid.

Thagard's hoofddoel is een programma dat in elk geval historisch adequaat is met betrekking tot de conclusies die wetenschappers in het verleden verbonden aan door henzelf gedefinieerde probleemsituaties. In zijn boek presenteert hij in groot detail vijf simulaties van historische theorie-selectie. Het betreft de verwerping van de flogistontheorie door Lavoisier ten gunste van zijn zuurstoftheorie, de verwerping van de creationistische theorie door Darwin ten gunste van zijn theorie der natuurlijke selectie en drie selecties in de geologie, te beginnen bij de verwerping van de contractionistische theorie over de aarde door Wegener ten gunste van zijn theorie over de drift der continenten.

Welnu, uitgaande van de door Thagard gereconstrueerde probleemsituaties van de betrokken onderzoekers in termen van hun elementaire beoordelingen van de verklaringssuccessen en -problemen van de betreffende theorieën, blijkt een eenvoudige 3x3-matrix op inzichtelijke wijze te leiden tot precies dezelfde theorie-selecties als men ervan uitgaat dat de meest succesvolle theorie wordt gekozen. De matrix ontstaat door uit te gaan van de driedeling dat een theorie een bepaald fenomeen kan verklaren, (contrafactisch) het tegendeel kan verklaren of juist geen van beide kan verklaren. Historisch adequate simulatie van de selectie-resultaten vergt dus helemaal geen beroep op de pragmatische overwegingen van eenvoud en analogie. Dat wil uiteraard niet zeggen dat de wetenschappers zelf geen rol daaraan toekenden. Integendeel, bij een historisch adequate simulatie van het selectie-proces, i.c. de selectie-argumentatie, mogen deze factoren niet ontbreken. De kwestie is slechts dat ze feitelijk geen verschil blijken te maken voor het resultaat. Met andere woorden, als we ons beperken tot het doel van historisch adequaatheid met betrekking tot het product dan zijn beide methoden vooralsnog even adequaat. In zo'n situatie dient de voorkeur uit te gaan naar de eenvoudigste methode. Merk op dat dit in feite een meta-toepassing is van mijn gelaagde aanpak.

De gelaagde aanpak heeft echter ook een motivatie vanuit het perspectief van ultieme filosofische adequaatheid: waarheidsbenadering. Boven aangeduide matrix-evaluatie betreft in feite een gecorrigeerde versie van IBE (Inference to the best explanation) die wordt gemotiveerd vanuit dit perspectief (zie o.a. Kuipers, 1991). De eerste correctie betreft het vervangen van de conclusie dat de beste theorie vermoedelijk waar is, door de conclusie dat deze vermoedelijk het dichtste bij de waarheid is, dat wil zeggen, de sterkste ware theorie die formuleerbaar is met de beschikbare begrippen uit beide conceptuele kaders. Deze eerste correctie schept meteen ruimte voor de tweede correctie: de beste theorie mag best reeds gefalsifieerd zijn. De logica van waarheidsbenadering zit nu zo in elkaar dat weliswaar niet is gegarandeerd dat de beste theorie het dichtste bij de waarheid is, maar wel dat er goede redenen zijn voor deze hypothese: deze verklaart namelijk het feit dat deze theorie succesvoller is. De gelaagde benadering is in dat opzicht

dus filosofisch adequaat. Hier tegenover staat dat de theorie die het dichtste bij de waarheid is helemaal niet het eenvoudigst hoeft te zijn noch analoog hoeft te zijn aan succesvolle theorieën op andere terreinen. Er is immers geen reden om te veronderstellen dat de werkelijkheid eenvoudig en, op verschillende gebieden, analoog in elkaar zit. Dat daar onze voorkeur naar uitgaat als we tussen overigens even succesvolle theorieën kunnen kiezen is een heel andere kwestie.
Resteert de vraag of men alleen het resultaat moet proberen te reproduceren of ook het proces. Het is duidelijk dat Thagard daarbij de voorkeur geeft aan het laatste en dat niet alleen de geschetste gelaagde aanpak maar bijvoorbeeld ook de BACON-programma's juist op het eerste mikken, door heuristische operaties te gebruiken die het meest efficiënt zijn, of ze nu wel of niet feitelijk gebruikt zijn door de wetenschappers die oorspronkelijk het werk deden. Uit deze voorbeelden blijkt dat het alleen nastreven van simulatie van het product ruimte laat voor normatieve overwegingen, zoals effectiviteit, het benaderen van de waarheid, en efficiëncy, via een zo kort mogelijke route. Juist om die reden is bij dergelijke benaderingen het tweeledige doel om uiteindelijk te komen tot computerondersteund ontdekken en evalueren dat bovendien filosofisch adequaat is ook het meest plausibel.

4 Vormen van theorierevisie

Shrager en Langley (1990) hebben een bijzonder rijke collectie samengesteld. De eerste vijf hoofdstukken betreffen variaties op en uitwerkingen van hetgeen hierboven reeds aan de orde kwam. Het betreft o.a. een hoofdstuk waarin Thagard ECHO aan de hand van de geologievoorbeelden uit zijn tweede boek presenteert en, zoals reeds opgemerkt werd, een hoofdstuk over de banaderingsproblematiek. In de laatste vijf hoofdstukken staan cognitief psychologische vraagstukken centraal, die niet direct passen bij de hoofdthema's van deze bespreking. De centrale hoofdstukken 6 tot en met 11 betreffen belangrijke varaianten en vernieuwingen ten opzichte van het voorgaande, waarbij het accent komt te liggen op theorierevisie. Uitgangspunt is steeds een kennisbestand van data en theorieën, waarbij de representatie meestal geschiedt met de al eerder genoemde methode van Forbus.
Voor degenen die op de hoogte zijn van het zogenoemde 'Belief revision'-programma van P. Gärdenfors c.s. zij vooraf opgemerkt dat daar weinig gelijkenis mee bestaat. De hoofdoorzaak is ongetwijfeld dat in dat programma geen poging wordt gedaan op wetenschappelijk kennis betrekking te hebben.
De volgende typeringen geven slechts een summiere indicatie over waar het heel globaal gesproken om gaat, hopelijk net voldoende voor iemand die al belangstelling heeft voor iets in deze sfeer.
Falkenhainer heeft een programma (PHINEAS) ontwikkeld dat een vergaande uniformering realiseert van een viertal min of meer standaard verklaringsscenario's, met bijbehorende theorievorming en -revisie, in termen van de mate van gelijkenis tussen explanandum en explanans. De selectie van theorieën vindt plaats aan de hand van de mate van ingrijpendheid van de extra vooronderstellingen: geen toevoegingen, toevoeging van nieuwe eigenschappen van een bekend type aan reeds bekende objecten, toevoeging van nieuwe entiteiten van een bekende soort, toevoeging van nieuwe eigenschappen van een nieuw type aan reeds erkende entiteiten, en tot slot toevoeging van nieuwe entiteiten van een nieuw type.

De 'Abduction Engine' (AbE) van *O'Rorke, Morris en Schulenburg* pretendeert revolutionaire revisie van een theorie mogelijk te maken, uitgaande van een observationele anomalie tussen een theorie in het kennisbestand en nieuwe data. Na het constateren van de anomalie wordt de theorie eerst afgezwakt tot een basistheorie die verenigbaar is met de nieuwe data. Vervolgens wordt deze basistheorie met abductie à la Thagard versterkt tot een theorie die de anomaleuze data kan verklaren. Deze theorie is ontwikkeld aan de hand van de overgang van de flogiston-theorie naar de zuurstoftheorie over verbranding en kan deze episode daadwerkelijk reproduceren.

In het COAST-programma van *Rajamoney* wordt theorierevisie stapsgewijs aangepakt: 1) constateren van een anomalie tussen nieuwe data en een theorie in het kennisbestand, 2) opstellen van revisievoorstellen die de anomaleuze data transformeren in verklaringssuccessen, 3) suggereren van uit te voeren experimenten, 4) eerste selectie op basis van de te verklaren anomalie, de uitkomst van de experimenten en de mate waarin 'exemplarische successen' behouden blijven, 5) de resterende selectie vindt plaats op basis van eenvoud en voorspellend vermogen. Dit programma en de volgende twee zijn ontwikkeld aan de hand van biochemische voorbeelden.

Het programma KEKADA van *Kulkarni en Simon* signaleert verrassende verschijnselen door reguliere experimentele resultaten te vergelijken met verwachtingen op basis van het kennisbestand, en gaat door op zulke verschijnselen, d.w.z. het gaat de betreffende theorie reviseren en experimenten die daarop gericht zijn voorstellen. Hiervoor zijn vijf strategieën ingebouwd, elk bestaande uit een hypothese-generator, een experiment-voorsteller en een evaluator. Het kernidee van de eerste strategie is bijvoorbeeld te proberen het verrassende verschijnsel te versterken door de systeemvariabelen afzonderlijk te manipuleren

Karp (HYPGENE) vat hypothese-vorming en -revisie heel plausibel op als een ontwerpprobleem, met constraints (een profiel van gewenste eigenschappen) en operatoren die het voorlopige ontwerp (het profiel van feitelijke eigenschappen) steeds beter aan de constraints laten voldoen.

Darden, tot slot, vat het oplossen van een anomalie van een theorie op als een taak tot diagnostisch redeneren, dat speciaal ontwikkeld is voor expertsystemen, naar het model van het opsporen van een fout in een technisch systeem. Zij laat met Mendeliaanse voorbeelden zien dat decompositie van alle vooronderstellingen van de theorie de aangrijpingspunten kan leveren voor de oplossing van een anomalie, en dat de oplossing, zoals in de beroemde analyses van Lakatos, al dan niet fundamentele theorierevisie kan inhouden. (Zie ook Darden, 1991).

5. Enkele verbindingen met neo-klassieke wetenschapsfilosofie

Voorzover moderne wetenschapsfilosofie zich richt op cognitief-heuristische structuren in kennis en methode (Kuipers, 1993a) zijn er veel aanknopingspunten voor de aansluiting met onderzoek ten behoeve van computationele wetenschapsfilosofie. Enerzijds houden de aansluitingen direct verband met allerlei aspecten van het moderne waarheidsbenadering-onderzoek. Anderzijds gaat het om cognitieve patronen die in beginsel als heuristieken in computerprogramma's kunnen worden ingebouwd.

In de computationele literatuur over theorie-evaluatie en -revisie blijkt grote onduidelijkheid over wat nu precies de eenheden voor successen en problemen zijn en hoe ze samenhangen. Uit het waarheidsbenaderingonderzoek volgen directe aanwijzingen voor wat precies als verklarende successen en wat als descriptieve problemen van theorieën moeten worden opgevat en hoe deze gerepresenteerd kunnen worden. De plausibele doelstelling voor theorie-revisie, toename van verklarend succes en afname van descriptieve problemen, wordt daarmee ook gepreciseerd. In bovenstaande bespreking zijn hier en daar formuleringen opgenomen die al enigszins zijn aangescherpt vanuit dit perspectief.
Er werd reeds gezinspeeld op de door waarheidsbenaderingonderzoek gesuggereerde correctie van 'Inference to the best explanation' en de mogelijkheid deze gecorrigeerde versie te rechtvaardigen als functioneel voor waarheidsbenadering.
Tot slot is het inderdaad verhelderend om theorie-vorming en -revisie met Karp te zien als ontwerponderzoek, met als uiteindelijke doel het opstellen van de sterkste ware theorie die hoort bij de context. Hierbij zijn de verschillen met gewoon ontwerponderzoek minstens even interessant als de overeenkomsten (cf. Kuipers, Vos & Sie, 1992).

De cognitieve patronen die in beginsel als heuristieken in computerprogramma's kunnen worden ingebouwd betreffen in de eerste plaats gestandaardiseerde stappendecomposities van verschillende soorten verklaringen. Deze kunnen enerzijds gemakkelijk geherinterpreteerd worden als gegeneraliseerde abductieschema's in de zin van Thagard. Voorts kunnen ze gebruikt worden voor de diagnostische fouten-analyse bij anomalieën in de zin van Darden.
Andere globale heuristieken worden geleverd door analyses van de structuur en ontwikkeling van onderzoeksprogramma's en hun mogelijkheden tot interactie, het theorie-relatieve onderscheid tussen observationele en theoretische termen, en de interne structuur van theorieën. Bovendien heb ik sterk de indruk dat het laatste soort analyse, met name de structuralistische analyse van theorieën, vrijwel direct gebruikt kan worden voor de computationele representatie van theorieën.
In het ANTW-themanummer over artifiële intelligentie van 1990 (Aflevering 1) komen in de bijdragen van Van Benthem, Schopman, Tan en Visser niet alleen nog andere aanknopingspunten aan de orde tussen wetenschapsfilosofie en computationele benaderingen, maar ook met betrekking tot andere delen van de filosofie, met name logica en kentheorie.

Uit de hiergegeven schets van aanknopingspunten moge al duidelijk zijn dat er weinig aanleiding is tot concurrentie tussen computationele benaderingen en de wetenschapsfilosofie zodra men bereid is het dogma van de (filosofisch-methodologische) *on*onderzoekbaarheid van de *Context of Discovery* op te geven.
Wel is er interessante concurrentie mogelijk tussen computationele wetenschapsfilosofie en de moderne wetenschapssociologie, getuige bijvoorbeeld de titel: ''Scientific discovery by computer as empirical refutation of the strong programme'' (Slezak, 1989). Immers, waar blijft de volgens veel wetenschapssociologen dominante rol van sociaal-culturele factoren bij computationeel ontdekken en evalueren? Naast dergelijke competitie-mogelijkheden zijn er ook hier van beide kan-

ten niet-extreme standpunten denkbaar die kunnen leiden tot vruchtbare samenwerking. Thagard doet daartoe suggesties verspreid in beide boeken. Één daarvan, in zijn eerste boek, noemden we al, het simuleren en optimaliseren van groepsrationaliteit met parallelle programma's met taak- en rolverdelingen. Bijvoorbeeld een ongeduldig Popperiaans programma en een meer behoudend Kuhniaans programma. Ook hier is de vraag natuurlijk of primair sociologische adequaatheid wordt nagestreefd dan wel uiteindelijk zo effectief en efficiënt mogelijke computationele ondersteuning van nieuw onderzoek.

Verwijzingen:

W. Bechtel, *Philosophy of science. An overview for cognitive science*, Erlbaum, Hillsdale, 1988.

L. Darden, *Theory change in science: strategies from Mendelian genetics*, Oxford UP, Oxford, 1991.

E. Davis, *Representations of common sense knowledge*, Kaufmann, San Mateo, 1990.

A. Goldman, *Epistemology and cognition*, Harvard UP, Cambridge Ma, 1986.

J. H. Holland et al., *Induction. Processes of inference, learning and discovery*, MIT-press, Cambridge Ma, 1986.

T. Kuipers, ''Realisme en convergentie'' in: *Realisme en waarheid* (red. J. van Brakel en D. Raven), Van Gorcum, Assen, 1991, 61-83.

T. Kuipers, ''Methodologische grondslagen voor kritisch dogmatisme'' in *Het vooroordeel in de wetenschap* (red. J. W. Nienhuis), Skepsis, Utrecht, 1992, 43-51.

T. Kuipers, *Structures in Science. Heuristic patterns based on cognitive structures*, Filosofisch Instituut RUG, 1993a.

T. Kuipers, ''On the architecture of computational theory selection'', in: *Philosophy and the cognitive sciences* (eds. R. Casati en G. White), 16th Wittgenstein symposium, Kirchberg /Wenen, 1993b.

T. Kuipers, R. Vos & H. Sie, ''Design research programs and the logic of their development'', *Erkenntnis*, 37, 37-63, 1992.

P. Slezak: ''Scientific discovery by computer as empirical refutation of the strong programme'', gevolgd door discussie, *Social Studies of Science*, 19.4, 1989, 563-695, voortzetting discussie door o.a. Simon, 21.1, 1991, 143-156.

N. Stillings et al., *Cognitive science. An introduction*, MIT-press, Cambridge, 1987.

Groningen Theo A.F. Kuipers