Kunstmatige neurale netwerken in de psychiatrie

Theoretische concepten

door J.M. van Beveren en E.J. Colon

Samenvatting

Een kunstmatig neuraal netwerk is een computersimulatie van enkele samenwerkende neuronen. Men kan een dergelijke simulatie beschouwen als een eenvoudig model van de werking van het zenuwstelsel. Met behulp van zulke modellen kunnen mentale processen bestudeerd worden. De laatste jaren verschijnen in toenemende mate publicaties waarin met behulp van kunstmatige neurale netwerken psychopathologie bestudeerd wordt. In dit artikel worden de theoretische concepten van kunstmatige neurale netwerken besproken aan de hand van twee veel gebruikte netwerken. Getoond wordt hoe het dynamische gedrag van kunstmatige neurale netwerken kan worden beschreven met het concept 'state space' en hoe dit concept kan worden gebruikt om de relatie tussen neurobiologie en mentale processen te verhelderen. De waarde van kunstmatige neurale netwerken als model wordt besproken.

Inleiding

Sedert het begin van deze eeuw is bekend dat de (menselijke) cortex bestaat uit een netwerk van onderling verbonden neuronen (Arbib 1995). Hoe in dit netwerk cognitieve processen kunnen ontstaan is daarna lange tijd onderwerp van speculatie gebleven. In de jaren vijftig maakte de ontwikkeling van de computer het mogelijk het gedrag van neuronen elektronisch te simuleren. Op grond van zulke simulaties werden hypotheses opgesteld over de wijze waarop neuronale activiteit aanleiding kon geven tot het ontstaan van cognitieve processen (Rosenblatt 1958). Onderzoek naar cognitieve processen door middel van computergesimuleerde neurale netwerken leidde gedurende de jaren zestig en zeventig echter een marginaal bestaan op grond van veronderstelde fundamentele beperkingen in de mogelijkheden van kunstmatige netwerken (Gardner 1985). Nieuwe ontwikkelingen in de theorie van de kunstmatige neurale netwerken (Hopfield 1982; Rumelhart en McClelland 1986; McClelland en Rumelhart 1986) hebben sedert het begin van de jaren tachtig echter gezorgd voor hernieuwde belangstelling.

Gegeven een model dat op adequate wijze normaal cognitief functioneren verklaart, is het interessant een dergelijk model te laederen om te komen tot wat men zou kunnen noemen 'kunstmatige psychopatho-

logie'. Crick en Mitchison (1983) suggereerden 'schizofrene' pathologie te bestuderen met behulp van kunstmatige neurale netwerken. In 1987 verscheen de eerste publicatie waarin hiertoe daadwerkelijk een poging werd gedaan (Hoffman 1987). In de jaren daarna zijn in toenemende mate publicaties op dit gebied verschenen.

Het doel van dit artikel is de lezer enig inzicht te geven in de principes van kunstmatige neurale netwerken. In een volgend artikel zal een overzicht worden gegeven van onderzoek waarin met behulp van dergelijke netwerken psychopathologie wordt bestudeerd (Van Beveren en Colon 1997).

Kunstmatige neurale netwerken

Het zenuwstelsel bestaat uit een netwerk van een groot aantal onderling verbonden neuronen (Kalat 1995). Bij een kunstmatig neuraal netwerk gaat het om een computermodel dat geïnspireerd is door de kenmerken van het biologische neurale netwerk. De stroming die met behulp van dergelijke modellen het ontstaan van mentale processen bestudeert, heet connectionisme. Het connectionisme gaat ervan uit dat mentale processen het resultaat zijn van informatieverwerking door talrijke eenvoudige elementen ('neuronen') die door middel van gewogen verbindingen met elkaar zijn verbonden ('synapsen') (Smolensky 1988; Caspar e.a. 1992; Phaf 1994).

Aangezien kunstmatige neurale netwerken gebaseerd zijn op biologische neurale netwerken zal de beschrijving vanuit het biologisch neuron beginnen. Het typische neuron bestaat uit een cellichaam waaruit een axon ontspringt, dat actiepotentialen van het cellichaam af voert. De meeste neuronen bezitten een groot aantal dendrieten, die signalen van andere neuronen ontvangen en naar het cellichaam vervoeren. Deze signalen bereiken de dendrieten via synapsen, die de prikkels chemisch modificeren. Bij een voldoende aantal afferente prikkels per tijdseenheid verandert de basisfrequentie van de actiepotentialen van het neuron (Kalat 1995).

Dit biologisch mechanisme is door McCulloch en Pitts (1943) geformaliseerd. Zij construeerden een kunstmatig neuron dat we, om onderscheid te maken tussen biologische en kunstmatige neuronen, verder zullen aanduiden met de term unit, overeenkomstig conventies (Hertz e.a. 1991). De werking van deze unit is als volgt. Veronderstel een unit met input vanuit twee afferente units u(1) en u(2). (Meerdere afferente units zijn ook mogelijk.) De input vanuit de afferente units kan getalsmatig voorgesteld worden als '0' (geen actiepotentiaal) of '1' (actie-
potentiaal). De verbindingen ('kunstmatige synapsen') tussen de afferente units en ontvangende unit kunnen worden voorgesteld door parameters w(1) en w(2) ('w' van weight, 'gewicht') met een grootte

tussen -1 (maximale inhibitie) en 1 (maximale excitatie), die de modificerende invloed verbeelden van de biologische synapsen. De totale input van de twee afferente units kan nu eenvoudig berekend worden door elke input te vermenigvuldigen met het bijbehorende gewicht en de aldus verkregen twee gewogen inputs bij elkaar op te tellen. Indien deze totale input groter is dan een grenswaarde, genereert de ontvangende unit een actiepotentiaal, dat wil zeggen: zendt zelf een output uit met grootte 1. McCulloch en Pitts vonden dat met behulp van combinaties van dergelijke units logische berekeningen konden worden uitgevoerd.

Dit concept is onder anderen door Rumelhart en McClelland (1986) verder uitgebreid. Zij ontwikkelden een unit waarin een tijdsaspect een rol speelt middels het begrip 'activiteit'. De activiteit is afhankelijk van van de huidige input en de activiteit op een vorig tijdstip. Via de activiteit wordt dan uiteindelijk de output van de unit gevonden, bijvoorbeeld doordat een grenswaarde wordt overschreden. Andere relaties tussen activiteit en output zijn echter ook mogelijk. De gevonden output kan dan weer als input dienen voor andere units.

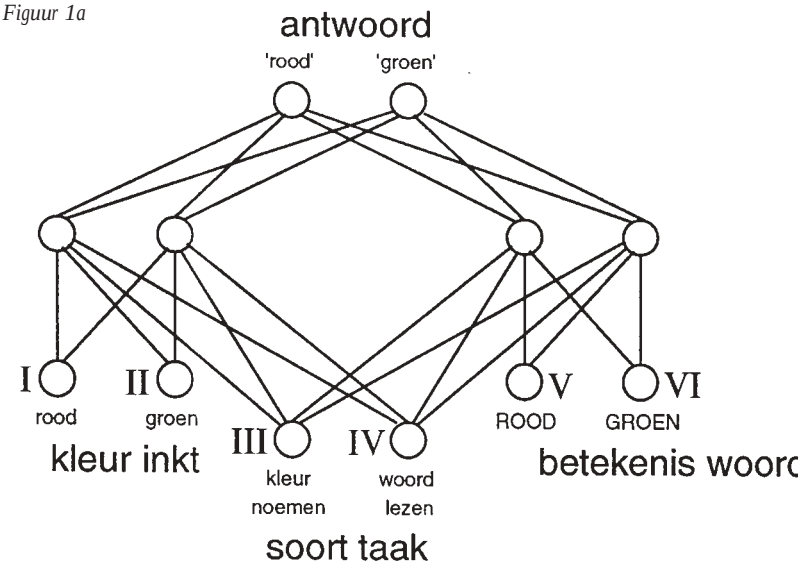
Het functioneren van de aldus geconstrueerde unit en de daaruit op te bouwen kunstmatige neurale netwerken kan met behulp van een computer gesimuleerd worden. Het is voor het verdere betoog echter van het grootste belang dat de lezer zich realiseert dat het connectionisme de computer zélf niet als model voor de menselijke cognitie ziet. Men kan de verhouding tussen biologisch neuraal netwerk, kunstmatig neuraal netwerk en computer vergelijken met die tussen het weer, een meteorologisch programma en een computer: het programma simuleert de weersontwikkeling, de computer is de intermediair waar het programma op draait, maar heeft zelf geen enkele relatie met het weer. Men vindt binnen het connectionisme dus geen ideeën als zou men het menselijk geheugen kunnen voorstellen als een soort hard disk waar informatie ligt opgeslagen die door een centrale verwerker met behulp van een programma wordt opgehaald en bewerkt.

Neurale netwerken kunnen ruwweg in twee categorieën verdeeld worden, namelijk feed-forward- en attractor-netwerken (Amit 1989). De principes van beide netwerken zullen uitgelegd worden.

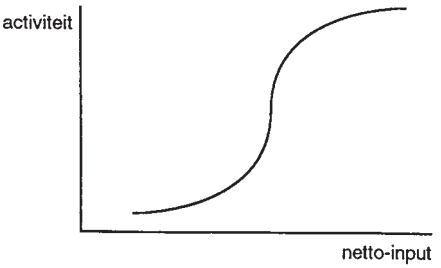
Twee netwerken

Feed-forward-netwerk – De structuur beschreven in figuur 1a is een voorbeeld van een feed-forward-netwerk. Cohen e.a (1990) gebruikten dit netwerk in een onderzoek naar de Stroop-test. De Stroop-test is een neuropsychologische test waarbij twee conflictuerende stimuli aan een proefpersoon worden aangeboden, terwijl aan de proefpersoon gevraagd wordt een van de twee stimuli te veronachtzamen. Het woord ROOD wordt bijvoorbeeld aangeboden, gedrukt in groene inkt. De proefpersoon wordt gevraagd te reageren op kleur of betekenis, terwijl

Figuur 1a



Figuur 1b



Figuur 1: Simulatie van de Stroop-test (naar Cohen, Dunbar en McClelland 1990; gemodificeerd).
1a. Feed-forward-netwerk.
1b. Relatie tussen input en activiteit.

de reactietijd wordt gemeten. Het basisidee bij de te bespreken simulatie is dat het netwerk zo wordt afgesteld dat normaal gedrag wordt gesimuleerd. Onder normaal gedrag wordt verstaan het reageren op de betekenis van een woord in een neutrale kleur inkt, of reageren op kleur zonder dat hieraan een betekenis is gekoppeld. Daarna wordt niets meer aan het netwerk veranderd en wordt gekeken of ook de testresultaten (i.c. conflictuerende stimuli) adequaat gesimuleerd kunnen worden. We zien dat dit netwerk uit drie lagen bestaat, van onderen naar boven: een input-laag, een tussenlaag (aangeduid als 'hidden layer'), en een output-laag, die hier uit twee units bestaat en die de twee mogelijke antwoorden op de test verbeelden. Verder zien we dat de units in de input-laag een verschillende betekenis hebben (bij een

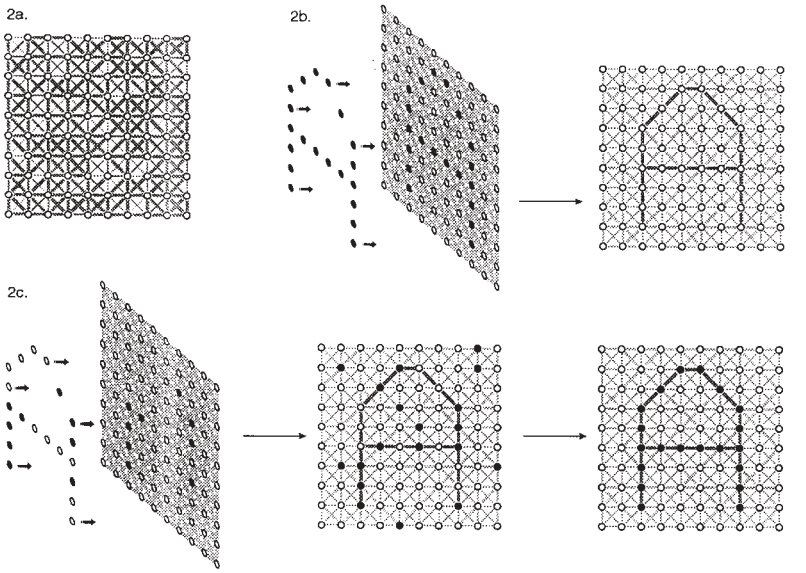
gelijk principe van werking). De units I en II verbeelden de inktkleur van de stimulus, de units V en VI de betekenis van de stimulus, en de units III en IV geven de gevraagde taak aan.

We bieden nu een normale stimulus aan, bijvoorbeeld: 'het woord ROOD in neutrale inkt geschreven met als opdracht te reageren op de woordbetekenis', door de units IV en V te activeren. De units V en VI worden niet geactiveerd, het gaat immers om een woord in neutrale kleur. De bedoeling is dat het activeren van deze units ertoe leidt dat na verloop van tijd de output-unit ROOD actief wordt. Dit is in feite een kwestie van de gewichten tussen de units de juiste grootte laten krijgen. Het 'afstellen' van de gewichten gebeurt bij een netwerk als dit door het herhaald aanbieden van normale stimuli vanuit een beginsituatie waarin de gewichten willekeurige grootte hebben (tussen -1 en 1). Dit zal in eerste instantie tot willekeurig goede en foute antwoorden leiden. Het simulerend computerprogramma bekijkt bij een fout antwoord welke gewichten aan de fout hebben bijgedragen een corrigeert deze gewichten van output- naar inputlaag ('backwards'). Na een (groot) aantal trainingsstimuli zal het netwerk een normale verrichting geleerd hebben. Deze manier van leren noemt men gesuperviseerd leren (namelijk gesuperviseerd door het simulerend computerprogramma), in tegenstelling tot het leren bij het nog te bespreken attractor-netwerk, dat zonder externe supervisor kan geschieden, en de methode van foutcorrectie 'back-propagating of errors' (Rumelhart e.a. 1986). Via hetzelfde principe wordt het netwerk ook geleerd alleen een kleur te benoemen. Na de trainingsfase beschikt men dan over een netwerk dat de betekenis van twee woorden kan herkennen of de kleur inkt kan benoemen, afhankelijk van de gestelde taak. De onderzoekers veranderen nu niets meer aan het netwerk en bieden conflictuerende stimuli aan. Het blijkt dan dat dit netwerk bijzonder goed in staat is de menselijke verrichtingen na te bootsen, zowel in kwantitatieve als in kwalitatieve zin (Cohen e.a. 1990).

Tot slot wordt gewezen op de manier waarop in dit type netwerk de activering van een unit plaatsvindt. Bij het backpropagation-netwerk is er sprake van een S-vormige relatie tussen de netto-input en de activatie van een unit (figuur 1b). Een unit bezit dus altijd enige mate van activatie, en bij toenemende netto-input neemt de activiteit toe.

Attractor-netwerk – Een ander type netwerk is het zogenaamde attractornetwerk (Hopfield 1982; Amit 1989).

In figuur 2a wordt een netwerk bestaande uit honderd units weergegeven. Elke unit is verbonden met elke andere unit door middel van willekeurig sterke verbindingen. (Voor de overzichtelijkheid zijn alleen verbindingen tussen nabij-gelegen units weergegeven.) Er bestaat geen verschil tussen input- en output-laag. Dit betekent ook dat er geen ex-



Figuur 2: Leren van patronen aan een attractor-netwerk. Voor uitgebreide uitleg zie tekst.
2a. Een netwerk van honderd units zonder geleerd patroon.
2b. Het leren van patroon 'A': aanbieden van 'A' (links); verbindingen tussen units behorend bij 'A' zijn versterkt (rechts).
2c. Terugvinden van 'A' door aanbieden van een deel van 'A' (de cue).

terne instantie is die binnen het netwerk units met verschillende functies onderscheidt. Elke unit is actief of inactief, afhankelijk van de waarde van de gesommeerde gewogen input van de 99 andere units en de activatiefunctie. Deze waarde wordt voor iedere unit regelmatig herberekend.

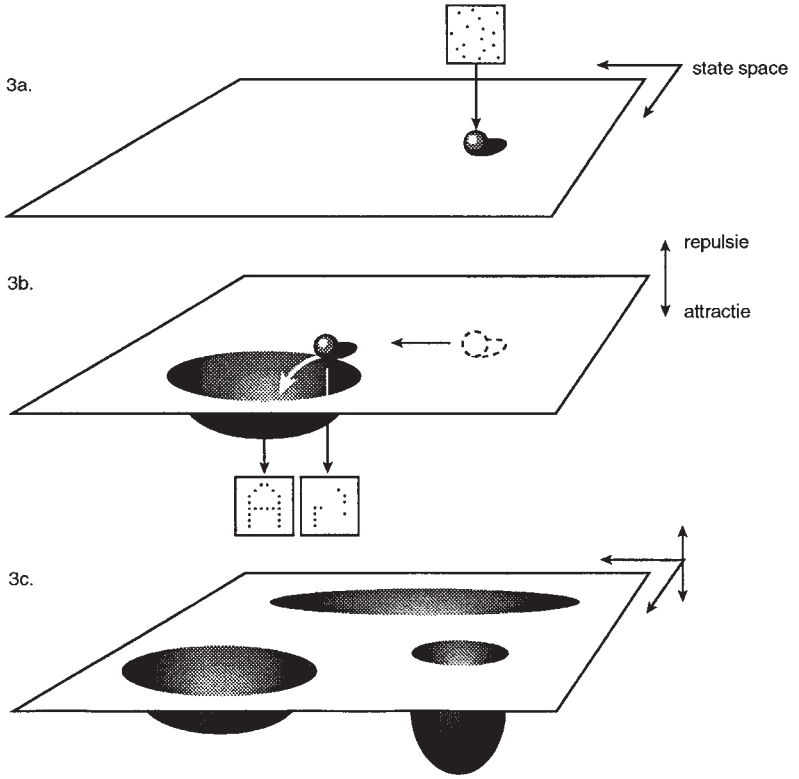
We kunnen nu dit netwerk een zeker patroon aanbieden, bijvoorbeeld een patroon dat de letter A symboliseert (figuur 2b). Het netwerk is zo geconstrueerd dat verbindingen tussen units met gelijke activiteit sterker worden, tussen units met ongelijke activiteit zwakker. Dit mechanisme is afgeleid van Hebb (Hebb 1949), die postuleerde dat verbindingen tussen neuronen die gelijktijdig actief zijn, versterkt worden, en dat dit mechanisme de basis vormt van leerprocessen. Nadat het patroon A een zekere tijd is aangeboden zal door dit mechanisme de sterkte van de verbindingen veranderd zijn en wel zo dat het patroon A verankerd is in de sterkte van de verbindingen. Men zegt nu dat patroon A een 'attractor' van het netwerk is geworden: het netwerk heeft symbool A geleerd. Wat dit betekent, wordt in figuur 2c duidelijk gemaakt. Links wordt een enigszins op A gelijkend patroon aangeboden. Zoals gesteld wordt de activiteit van de units regelmatig herberekend

(figuur 2c, midden). Dit betekent dat er een zekere dynamiek in de activiteit van het netwerk ontstaat die ertoe zal leiden dat het netwerk 'getrokken' wordt naar toestand A, de attractor van het netwerk (figuur 2c, rechts). Dit betekent op zijn beurt dat het netwerk een associatief geheugen bezit: gedeeltelijke informatie (de 'cue') levert de volledige geleerde informatie op. Een netwerk kan meerdere attractors bezitten. We kunnen zeggen dat elke attractor een geheugen is voor een bepaald patroon. Overigens hoeft er geen één-op-éénrelatie te bestaan tussen patroon en attractor. Op elkaar gelijkende patronen kunnen tot dezelfde attractor leiden.

Attractor-netwerken bezitten een aantal interessante eigenschappen (Hopfield 1982). Ze zijn een manier om informatie op te slaan: het netwerk bezit een 'geheugen'; dit geheugen is aanspreekbaar door gedeeltes van de oorspronkelijke informatie aan te bieden; op elkaar gelijkende, maar van het oorspronkelijke geleerde afwijkende input wordt herkend als behorend tot een gemeenschappelijk archetype (als dit netwerk eenmaal de letter A geleerd heeft, herkent het netwerk zonder verdere specificatie ook *A*, A, *A* enzovoort; een praktische toepassing van een dergelijk netwerk is dan ook het geautomatiseerd lezen van postcodes); een attractor-netwerk is bestendig tegen schade: relatief grote aantallen units kunnen disfunctioneel worden zonder dat de functie van het netwerk als geheel snel achteruitgaat.

Aangezien het begrip attractor een grote rol speelt zowel bij netwerken in het algemeen alsook bij de vorming van concepten rond het ontstaan van psychiatrische symptomen (Van Beveren en Colon 1997), willen we hier verder op ingaan. Het begrip attractor komt voort uit de fysica van dynamische systemen. Fysici stellen dat elke toestand waarin een attractor-netwerk kan verkeren een zekere interne energie bezit. Een attractor is een toestand van minimale interne energie. Het aanbieden van een patroon betekent dan in feite het toevoegen van energie aan het netwerk, waarna het netwerk zijn toestand van minimale energie opzoekt (Amit 1989).

We kunnen dit grafisch toelichten (figuur 3a; zie ook Globus en Arpaia 1994). We kunnen alle mogelijke toestanden (toestanden zijn dus: verschillende patronen) waarin een attractor-netwerk zich kan bevinden *gesimplificeerd* voorstellen als een vlak. (In werkelijkheid gaat het hier om een multidimensionele ruimte die zich niet grafisch laat voorstellen). Dit vlak noemt men de 'state space': de verzameling toestanden waarin het netwerk zich kan bevinden. Een bepaalde toestand is dan een punt op dit vlak. Een attractor (een geleerd patroon) kan men zich voorstellen als een kuil in dit vlak (Figuur 3b). De omvang van de attractor verbeeldt de dominantie van de attractor over zijn omgeving; in de Angelsaksische literatuur wordt deze dominantie aangeduid met de term 'basin of attraction'. Verandering van de ene toestand naar de an-



Figuur 3
3a. De verschillende toestanden van een netwerk gesymboliseerd als punten in een vlak: de 'state space'. Aangegeven in een willekeurige toestand.
3b. De mate van attractie gesymboliseerd als een kuil in de 'state space'. Veranderingen in de toestand van een netwerk kunnen worden beschouwd als een traject door de 'state space'.
3c. Attractors van verschillende grootte en diepte.
Voor verdere uitleg zie tekst.

dere kunnen we voorstellen als het rollen van een balletje over het vlak. In de tijd beweegt het netwerk ('rolt het balletje') over het vlak: het netwerk neemt verschillende toestanden aan. Het aanbieden van een deel van een patroon aan het netwerk kunnen we voorstellen als het verschuiven van het balletje in de richting van een kuil. Het balletje zal dan in een kuil rollen, met andere woorden, het netwerk zal een patroon aannemen dat een van zijn attractors is (figuur 3b). Een toestand waarin meerdere attractors zijn aangegeven, is afgebeeld in figuur 3c. Leren kan men opvatten als het verwerven van nieuwe attractors (Hoffman en McGlashan 1993). Er zij opgemerkt dat het herkennen van patronen zich op conceptueel vlak niet hoeft te beperken tot een-

voudige patronen. Ook meer abstracte en complexe 'patronen', bijvoorbeeld de herinnering aan een vakantie of een traumatische gebeurtenis, kunnen als attractor worden geconceptualiseerd. Voor een goed begrip dient men zich het geheel zo dynamisch mogelijk voor te stellen. Dus, de toestand van een netwerk beweegt zich voortdurend door de 'state space' onder invloed van 'cues', de heuvels en dalen in de 'state' space (de 'state space topologie'), en vormt daarbij ook steeds nieuwe sporen door toedoen van het leermechanisme van Hebb (1949).

Beschouwing

Men kan een neuraal netwerk beschouwen als een op de architectuur van het zenuwstelsel gebaseerd model met behulp waarvan mentale processen bestudeerd kunnen worden. Bij een simulatiemodel zoals een kunstmatig neuraal netwerk kan men onderscheid maken tussen product- en processimulatie. Van een productsimulatie is sprake als vooral de resultaten van belang zijn; van een processimulatie is sprake wanneer ook de manier waarop de resultaten behaald worden van belang zijn (Draaisma 1995). Een (product)model benadert de te bestuderen werkelijkheid goed als het aan de volgende voorwaarden voldoet: a) het model levert passende resultaten bij bekende experimentele data; b) het model functioneert vanuit een zekere algemene geldigheid, dat wil zeggen het model is niet (te) expliciet op een dataset toegesneden; c) het model voorziet in een nieuwe interpretatie van gegevens; d) het model doet voorspellingen over nog onbekende verschijnselen; e) deze nog onbekende verschijnselen zijn experimenteel toetsbaar. (Sharkey en Sharkey 1995; Pich en Del Guidice 1992; Globus en Arpaia 1994). Voor een processimulatie dient aan deze eisen nog toegevoegd te worden dat het originele proces op een geloofwaardige manier nagebootst wordt (Draaisma 1995). Van een kunstmatig neuraal netwerk kan men bijvoorbeeld eisen dat er sprake is van een biologisch plausibel werkingsmechanisme.

Wanneer we de eisen voor een productmodel toepassen op de twee gepresenteerde netwerken zien we dat het feed-forward-netwerk voldoet aan de eisen a, c, d, e, en dat kwantitatieve resultaten mogelijk zijn. Volgens Crick (1989) is het feed-forward-netwerk echter in vrijwel geen enkel opzicht biologisch realistisch. Hij geeft hiervoor een aantal argumenten, onder andere dat het backpropagation- leermechanisme een retrograde informatiestroom in neuronen impliceert en dat het feed-forward-netwerk de regel overtreedt dat neuronen ofwel excitatoir zijn ofwel inhibitoir, maar niet beide tegelijk.

Voor de attractor-netwerken geldt dat ze aan alle gestelde eisen voor een productmodel voldoen, maar voornamelijk in kwalitatieve zin. Plaut (1995) stelt: 'Although the specific relation between these net-

works and neurobiology is far from clear, the belief that representation and computation in these networks resembles neural computation at some level remains one of their strongest attractions.'

Hoewel de reciproke connecties het attractor-netwerk niet erg biologisch plausibel maken, zijn er aanwijzingen dat de in het model gevonden processen ook in vivo aanwezig zijn. Friston e.a. (1991) gebruikten PET om patronen van hersenactiviteit te bestuderen, die zij eerst met behulp van aan elkaar gekoppelde attractor-netwerken hadden gesimuleerd. Zij vonden inderdaad de gesimuleerde verschijnselen. Freeman (1991) vond aanwijzingen voor het bestaan van attractors bij dieren. Hertz (1995) presenteerde een overzicht van experimenten in deze richting.

Draaisma (1995) stelt dat connectionistische modellen tot op heden weinig meer doen dan resultaten reproduceren die reeds bekend waren, en dat derhalve de grootste waarde tot nog toe gelegen is in het vormen van concepten over de relatie tussen psychologie en neurobiologie.

Concluderend stellen we dat kunstmatige neurale netwerken en het achterliggende concept van het connectionisme – hoewel eenvoudig en onvolkomen – een interessante ontwikkeling zijn in de poging biologische processen te relateren aan cognitieve processen. In een volgend artikel zal een overzicht worden gegeven van studies waarin kunstmatige neurale netwerken functioneel en/of structureel worden gelaedeerd om zo psychopathologische processen te bestuderen (Van Beveren en Colon 1997).

De auteurs danken de volgende personen: J. Lübeck, bibliothecaris, voor zijn ondersteuning bij het zoeken van de literatuur; C. Heim voor de illustraties; L. de Haan voor zijn nuttige opmerkingen over het manuscript.

Literatuur

Amit, D.J. (1989), *Modeling brain function: the world of attractor neural networks*. Cambridge University Press, Cambridge USA.

Arbib, M.A. (1995), *The handbook of brain theory and neural networks*. MIT Press, Cambridge USA.

Beveren, J.M. van, en E.J. Colon (1997), Kunstmatige neurale netwerken in de psychiatrie. Een literatuurstudie naar neuropsychiatrische toepassingen. Aangeboden voor publicatie.

Caspar, F., T. Rohenfluh en Z. Segal (1992), The appeal of connectionism for clinical psychology. *Clinical Psychological Review*, 12, 719-762.

Cohen, J.D., K. Dunbar en J.L. McClelland (1990), On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332-361.

Crick, F., en G. Mitchison (1983), The function of dream sleep. *Nature*, 304, 111-114.

Crick, F. (1989), The recent excitement about neural networks. *Nature*, 337, 129-132.

Draaisma, D. (1995), *De metaforenmachine: een geschiedenis van het geheugen*. Historische Uitgeverij, Groningen.

Freeman, W.J. (1991), The physiology of perception. *Scientific American*, 264, 2, 78-85.

Friston, K.J., C.D. Frith, P.F. Liddle e.a. (1991), Investigating a network model of word generation with positron emission tomography. *Proceedings of the Royal Society of London B*, 244, 101-106.

Gardner, H. (1985), *The mind's new science. A history of the cognitive revolution*. Basic Books Inc., New York.

Globus, G.G., en J.P. Arpaia (1994), Psychiatry and the new dynamics. *Biological Psychiatry*, 35, 352-364.

Hebb, D.O. (1949), *The organisation of behavior*. Wiley, New York.

Hertz, J., A. Krogh en R.G. Palmer (1991), *Introduction to the theory of neural computation*. Addison-Wesley Publishing Company, USA.

Hertz, J. (1995), Computing with attractors. In: M.A. Arbib (red.), *The handbook of brain theory and neural networks*. MIT Press, Cambridge USA, p. 230-234.

Hoffman, R.E. (1987), Computer simulations of neural information processing and the schizofrenia-mania dichotomy. *Archives of General Psychiatry*, 44, 178-188.

Hoffman, R.E., en T.H. McGlashan (1993), Parallel distributed processing and the emergence of schizophrenic symptoms. *Schizophrenia Bulletin*, 19, 1, 119-140.

Hopfield, J.J. (1982), Neural networks and physical systems with emergent collective computational properties. *Proceedings of the National Academy of Science of the USA*, 79, 2554-2558.

Kalat, J.W. (1995), *Biological Psychology*, fifth edition. Brooks/Cole Publishing Company, Pacific Grove USA.

McClelland, J.L., en D.E. Rumelhart (1986), *Parallel Distributed Processing: Explorations in the microstructure of cognition. Vol. 2: Psychological and biological models*. MIT Press, Cambridge USA.

McCulloch, W.S., en W.H. Pitts (1943), A logical calculus of the ideas immanent in neurons activity. *Bulletin of Mathematics and Biophysics*, 5, 115-133.

Phaf, R.H. (1994), *Learning in natural and connectionist systems*. Kluwer Academic Publishers, Dordrecht.

Pich, E.M., en P. del Giudice (1992), Can neural networks be used as models for neuropsychological dysfunctions? *International Journal of Neural Systems*, 2, 3 (suppl.), 163-168.

Plaut, D.C. (1995), Lesioned attractor networks as models of neuropsychological deficits. In: M.A. Arbib (red.), *The handbook of brain theory and neural networks*. MIT Press, Cambridge USA.

Rosenblatt, F. (1958), The perceptron: A probabilistic model for information storage and organisation in the brain. *Psychological Review*, 65, 386-408.

Rumelhart, D.E., G.E. Hinton en R.J. Williams (1986), Learning representations by back-propagating errors. *Nature*, 323, 533-536.

Rumelhart, D.E., en J.L. McClelland (1986), *Parallel Distributed Processing: Explorations in the microstructure of cognition. Vol. 1: Foundations*. MIT Press, Cambridge USA.

Sharkey, A.J.C., en N.E. Sharkey (1995), Cognitive modeling: Psychology and connectionism. In: M.A. Arbib (red.), *The handbook of brain theory and neural networks*. MIT Press, Cambridge USA, p. 200-203.

Smolensky, P. (1988), On the proper treatment of connectionism. *Behavioral and Brain*

Sciences, 11, 1-74.

Summary: Artificial neural networks in psychiatry. Theoretical concepts

An artificial neural network is a computersimulation of a small number of co-operating neurons. A simulation like that can be regarded as a simple model of the functioning of the nervous system. Mental processes can be studied with such models. Recent years have seen an increasing number of publications in which psychopathology has been studied with artificial neural networks. In this article the theoretical concepts of artificial neural networks will be discussed, guided by two types of commonly used networks. It will be shown that the dynamic behaviour of artificial neural networks can be described with the concept 'state space' and that this concept can be used to illuminate the relation between neurobiology and mental processes. The value of artificial neural networks as models is discussed.

De auteurs zijn respectievelijk arts-assistent in opleiding tot psychiater, en zenuwarts, A-opleider, hoogleraar gerontopsychiatrie (RUGent-B). Beiden zijn verbonden aan het Sint Elisabeth Gasthuis te Deventer (Salland) en Almelo (Twente). Correspondentieadres: drs. J.M. van Beveren, Psychiatrisch Ziekenhuis Brinkgreven, Rielerweg