



Radboud Universiteit

Werkstuk ter afronding van de Bacheloropleiding Wijsbegeerte aan de Faculteit der Filosofie,
Theologie en Religiewetenschappen van de Radboud Universiteit Nijmegen.

Een ontologische basis voor 1-op-1 relaties tussen cognitieve concepten en hun neurale correlaten

Naam: Tanja Mourachova

Studentnummer: s1027979

Eerste lezer: dr. Frank van Caspel

Datum van inleveren: 10-04-22

Woordaantal: 6185

*Hierbij verklaar en verzeker ik, Tanja Mourachova,
dat voorliggende eindwerkstuk getiteld 'Een ontologische basis voor 1-op-1
relaties tussen cognitieve concepten en hun neurale correlaten', zelfstandig
door mij is opgesteld, dat geen andere bronnen en hulpmiddelen dan die door
mij zijn vermeld zijn gebruikt en dat de passages in het werk waarvan de
woordelijke inhoud of betekenis uit andere werken – ook elektronische media –
is genomen door bronvermelding als ontlening kenbaar gemaakt worden.*

Inhoudsopgave

Inleiding	4
1 Het theoretisch kader van de cognitieve wetenschappen	6
1.1 Modulariteit.....	6
1.2 Correlaties tussen algemene cognitieve functies en regio's van het brein.....	7
2 De problematische interpretaties van neuroimaging onderzoeken volgens Francken en Slors	9
2.1 Cognitive functies of cognitieve concepten?.....	9
2.2 De onderliggende oorzaak van de problematische interpretaties volgens Francken en Slors..	10
3 Poldrack's methode	12
3.1 De twee kanten van het cognitief systeem in kaart gebracht.....	12
3.2 Een neurale netwerk dat het missende ingrediënt voor 1-op-1 relaties verschaft.....	13
4 Een ontologische basis voor Poldrack's methode	16
4.1 Concepten weerspiegelen de structuur van de realiteit.....	16
4.2 Structuren in de realiteit die een rol spelen in concepten kunnen worden geïsoleerd.....	17
Conclusie	19
Bibliografie	21

Inleiding

Een actueel debat op het terrein van cognitiefilosofie en neurowetenschappen gaat over de verhouding tussen cognitieve concepten zoals aandacht, geheugen en taal en hun neurale basis. Een van de grootste velden waar cognitieve concepten en hersengebieden aan elkaar worden gekoppeld is die van de neuroimaging. Zowel in de cognitiewetenschappen als vanuit de cognitiefilosofie wordt er kritiek geleverd op de interpretatie van dit soort neuroimaging onderzoeken. Het probleem is dat de resultaten inconsistent zijn en de interpretaties elkaar soms zelfs tegenspreken.

Cognitiefilosofen Jolien Francken en Marc Slors claimen dat deze problemen het gevolg zijn van dat neurowetenschappers onbewust een achterhaalde filosofische theorie aanhangen, namelijk metaal realisme. Ik zal echter beargumenteren dat wetenschappers die neuroimaging onderzoeken interpreteren niet persé deze theorie aanhangen. In de methode van wetenschapper Russell Poldrack, waar Francken en Slors in hun paper deels naar verwijzen, kan er zelfs een andere theoretische basis worden gezien. Mijn doelstelling voor dit paper is om Poldrack's methode verder onder de aandacht te brengen bij filosofen. Mijn bijdrage zal zijn om de ontologische aannames die zijn methode veronderstelt expliciet te maken en zo een ontologische basis te geven aan de 1-op-1 relaties die zijn methode produceert tussen cognitieve concepten en regio's van het brein.

In de eerste sectie zal ik stilstaan bij de manier waarop kennis in de cognitieve wetenschappen wordt gestructureerd en wat er verwacht wordt als men spreekt van kennis of begrip van het cognitief systeem als geheel. Dit zal de lezer helpen een beeld te vormen van het algemeen streven van wetenschappers die zich bezighouden met cognitie, namelijk om deze te begrijpen zoals een technicus een printer begrijpt en deze kan repareren.

De tweede sectie zal gaan over de problemen die zijn aangekaart door Francken en Slors. Het eerste probleem met de interpretatie van neuroimaging onderzoeken is dat er geen éénduidige relaties worden gelegd tussen specifieke cognitieve concepten en delen van het brein. De correlaten van bijvoorbeeld ADHD in de ene studie, zijn in een andere studie de correlaten van een ander cognitief fenomeen. De onderliggende oorzaak voor dit probleem is, volgens Francken en Slors, dat neurowetenschappers onbewust denken dat cognitieve concepten verwijzen naar iets dat te vinden is in het brein en daarom 1-op-1 relaties proberen te leggen, terwijl cognitieve concepten gedragspatronen beschrijven binnen een sociale dimensie die niet gereduceerd kunnen worden tot alleen patronen in de hersenen.

De derde sectie staat in het teken van Poldrack's onderzoeksmethode. Aangezien het een aanzienlijk groot project was, denk ik dat filosofen de waarde van zijn onderzoek door alle technische informatie over het hoofd hebben gezien. Ik zal kort samenvatten wat hij en zijn collega's precies hebben gedaan. Daaruit zal blijken dat het toch mogelijk is om 1-op-1 relaties te leggen tussen algemene cognitieve concepten en bijvoorbeeld specifieke regio's van het brein.

In de laatste sectie zal ik uitleggen wat voor ontologische aannames Poldrack's methode onderliggen. In plaats van de aanname dat cognitieve concepten in het brein te vinden zijn als breinprocessen, denk ik dat Poldrack structuren van het brein probeert te isoleren die een rol spelen in cognitieve concepten die deze structuren weerspiegelen.

1 Het theoretisch kader van de cognitieve wetenschappen

1.1 Modulariteit

Een belangrijke aanname in de cognitieve wetenschappen is dat het cognitief systeem modulair is. Om dit te begrijpen zouden we de vergelijking kunnen maken met een machine, neem bijvoorbeeld een printer. Een printer heeft als hoofdfunctie een fysieke kopie te maken van een digitaal bestand. De verschillende onderdelen van een printer hebben afzonderlijke sub-functies. De rolletjes bewegen het papier, de cartridges verzorgen de toevoer van inkt, etc. Al deze verschillende onderdelen of modules en hun functies samen vormen het totale printstelsel. Het idee is dat het cognitief systeem ook bestaat uit verschillende onderdelen met afzonderlijke sub-functies.

One key assumption is modularity, meaning the cognitive system consists of numerous modules or processors operating fairly independently or separately of each other (...) each specialised for a given type of processing. (Eysenck en Keane 2020, 8)

We hebben dus aan de ene kant de functies en aan de andere kant de onderdelen die deze functies realiseren, waarbij je in gedachte moet houden dat de realisatie van een bepaalde functie op verschillende manieren kan gebeuren. Sommige printers werken bijvoorbeeld met een toner in plaats van inktcartridges. Je kunt van het totale printstelsel dus een deel schematisch weergeven als een som van verschillende functies waarbij nog niet bepaald is hoe elk onderdeel wordt gerealiseerd. Dit is nuttig als je bijvoorbeeld een nieuwe printer wilt ontwerpen met nieuwe functionaliteiten. Het materiele realisatie proces komt later pas in beeld als de opdracht naar de fabriek gaat.

Een tweede belangrijke aanname in de cognitieve wetenschappen is dat we in onze hersenen modules kunnen terugvinden die de functies die we schematisch hebben vastgelegd realiseren. Dit wordt anatomische modulariteit genoemd. Een bewijs voor deze aanname zijn de resultaten uit studies met proefpersonen waarbij blijkt dat een bepaalde cognitieve functie niet meer werkt als een bepaald deel van de hersenen is beschadigd (Eysenck en Keane 2020, 8). Kennis van de materiele realisatie van functies is niet alleen belangrijk voor het kunnen creëren van een fysieke machine, maar ook om het te kunnen repareren.

Als je een kapotte printer wil repareren moet je heel precies weten welk materieel onderdeel van die specifieke printer welke functie van je schema vervult. Waar het op neerkomt is dat kennis of begrip van het totale cognitieve systeem neerkomt op het kennen van de inhoud van zowel het schematische deel als bekend zijn het materiele deel, waarbij je kunt aanwijzen welke modules in de afzonderlijke materiele instanties corresponderen met welke functies uit het schematische deel.

In de wetenschap hebben verschillende vakgebieden zich gespecialiseerd op het in kaart brengen van deze delen van het cognitief systeem. De cognitieve psychologie houdt zich bezig met het schematische deel. Aan de hand van gedragsexperimenten worden verschillende theorieën gevormd over welke algemene functies kunnen worden gemeten. Neem bijvoorbeeld de algemene functie *geheugen*. De dominante theorie op dit moment is dat deze functie bestaat uit de afzonderlijke sub-functies *opslaan*, *vasthouden* en *ophalen* (Eysenck en Keane 2020, 239). Biologie en anatomie focussen zich op het lichamelijke deel en de neurowetenschappen zoeken naar de connectie ertussen.

Waar ik vooral aandacht op wil vestigen is dat wetenschappers erg gewend zijn aan dit kennis- of begripmodel van het cognitief systeem. Ik merk het zelfs bij mijzelf op, dat ik na een cursus cognitieve psychologie al een schematische weergave in mijn hoofd heb van de verschillende functies die zijn behandeld in de cursus. Het is een bril geworden waarmee ik naar de wereld kijk in alledaagse situaties. Ik kan mij voorstellen dat een neurowetenschapper die onderzoek doet naar de functie *geheugen* misschien bij het ontwerpen van de experimenten denkt hoe haar zoontje over het schoolplein naar haar toe rende, omdat in dat gedrag ze de functies *herkenning*, *geheugen* en *ophalen* ziet.

1.2 Correlaties tussen algemene cognitieve functies en regio's van het brein

In de vorige paragraaf hebben we gezien dat het algemeen wetenschappelijk streven is om alle delen van het cognitief systeem in kaart te brengen en ook precieze kennis te verkrijgen over welke materiele onderdelen een rol spelen in welke cognitieve functies. Hoewel neuroimaging studies precies dit laatste onderzoeken, is de informatie die uit deze studies vloeit nog niet genoeg.

In zijn paper uit 2010 vertelt Poldrack dat, na de bestaande neuroimaging literatuur uitgebreid onderzocht te hebben, hij tot de conclusie is gekomen dat het op dit moment onmogelijk is om aan de hand van meta studies algemene 1-op-1 correlaties te leggen tussen cognitieve functies en brein structuren.

Although popular accounts often imply unique structure–function mappings (e.g., the amygdala is the “fear area,” the anterior cingulate is the “conflict area”), closer examination of nearly every such claim uncovers counterexamples that are difficult to reconcile with a selective structure–function mapping. (Poldrack 2010, 754)

Interpretaties van individuele studies maken soms sterke claims over dat bepaalde functionaliteiten samenhangen met activatie van bepaalde gebieden in het brein. Als je deze studies echter in het licht ziet van de talloze andere studies die worden gedaan, zie je dat dezelfde gebieden ook worden geassocieerd met andere functies.

De reden dat het nog niet lukt om zulke algemene relaties te vinden kan volgens Poldrack zowel aan het model van het schematische deel van het cognitief systeem als aan het model van het materiele deel liggen. Misschien zijn sommige functies verkeerd gecategoriseerd en moeten algemene functies dus in andere sub-functies worden opgedeeld. Het kan echter ook liggen aan de categorisering van de onderdelen in het model van het materiele deel van het cognitief systeem. Het brein kan je namelijk ook op verschillende manieren in stukjes hakken: je kunt de anatomische structuur als leidraad nemen, maar je zou ook netwerken van neurale paden kunnen nemen waardoor je op een andere verzameling onderdelen uitkomt (Poldrack 2010, 755).

De kunst is om aan beide kanten de juiste categorisatie te maken, zodat de twee kanten goed op elkaar aansluiten. Hierbij bedoel ik ‘juist’ niet in objectieve zin, maar in relatie tot elkaar. Neem bijvoorbeeld een pottenbakker die een pot maakt om bijvoorbeeld rijst in op te slaan. Zolang de klei nat is kan de pottenbakker zowel de opening van de pot als de deksel bijvormen op zijn draaischijf, zodat beide delen de juiste afmeting krijgen en goed op elkaar passen. Tegen het eind van dit paper zal ik hebben laten zien dat Poldrack’s methode als een mal werkt, een hulpmiddel dat helpt beide kanten van het wetenschappelijk kleimodel van het cognitief systeem bij te stellen, zodat ze goed op elkaar aansluiten.

Voordat ik dit laat zien wil ik eerst het standpunt van Francken en Slors uitleggen. De ontologische basis voor het bovenstaande idee dat ik ga uitwerken, dient namelijk als tegenvoorbeeld voor de claim die ze maken in hun paper over de achterhaalde ontologische aanname die neurowetenschappers onbewust zouden maken.

2 De problematische interpretaties van neuroimaging onderzoeken volgens Francken en Slors

2.1 Cognitieve functies of cognitieve concepten?

Francken en Slors wijzen op een soortgelijk probleem als Poldrack in hun paper uit 2014: “the translations from CCC/SCC to brain activity and from brain data to SCC or CCC are highly indirect and prone to various interpretations” (blz. 250). Hier wordt weer het probleem benoemd dat er op dit moment geen algemene relaties kunnen worden gelegd tussen cognitieve functies en brein data. Het is interessant dat Francken en Slors echter niet spreken over cognitieve functies maar cognitieve concepten. CCC staat voor ‘commonsense cognitive concept’ en SCC staat voor ‘scientific cognitive concept’ (blz. 248 en 249).

Het verschil tussen de cognitieve concepten van Francken en Slors en de cognitieve functies van Poldrack is erg subtiel, maar niet geheel onbelangrijk. Voor Francken en Slors zijn cognitieve concepten talige beschrijvingen van *gedrag patronen* (Francken en Slors 2014, 249). Neem het voorbeeld van de kinderen op het schoolplein die hun ouders tegenmoet rennen. Als Francken tegen Slors zou zeggen: dat kind herkent haar vader, verwijst ze met ‘herkent’ naar een gedrag patroon dat vaak waargenomen wordt in alledaagse situaties en van toepassing is op deze situatie.

Poldrack heeft het ook over cognitieve concepten in zijn paper: “mental concept, which refers to any concept that describes a mental function, structure, or process” (Poldrack 2010, 757). Concepten zijn voor Poldrack dus ook talige uitingen die onder andere kunnen verwijzen naar cognitieve functies (Poldrack lijkt ‘mentaal’ en ‘cognitief’ als synoniemen te gebruiken). Dus als Poldrack zou zeggen ‘dat kind herkent zijn moeder’ dan verwijst hij met ‘herkent’ naar een functie in het schematische deel van het cognitief systeem waar hij zo bekend mee is dankzij zijn wetenschappelijke achtergrond.

Cognitieve concepten zijn dus termen of woorden waarmee verwezen wordt naar iets abstracts, een gedrag patroon of een cognitieve functie, afhankelijk van welke persoon deze woorden uit en wat diens theoretische achtergrond is.

2.2 De onderliggende oorzaak van de problematische interpretaties volgens Francken en Slors

Het belangrijkste punt dat Francken en Slors maken in hun paper is dat het taalgebruik in wetenschappelijke papers ten opzichte van cognitieve concepten slordig is. Vooral wanneer het gaat over de verhouding tussen deze concepten, functies en breinstructuren, wordt daar niet consistent over gepraat en worden de termen door elkaar gebruikt. Dit is voor een filosoof natuurlijk ondragelijk om aan te zien. En de reden voor dit vage gebruik van termen ligt in “the lack of a systematic view on the relation between CCCs and brain data” (Francken en Slors. 2014, 250). Zoals ik dit citaat begrijp, ontbreekt er niet alleen een systematische manier waarop omgegaan wordt met termen, maar er ontbreekt ook een ontologische basis binnen het theoretisch kader van de wetenschappen waarbij de ontologische status van concepten, functies en breinstructuren en de verhoudingen daartussen duidelijk wordt.

Francken en Slors menen dat neurowetenschappers impliciet een misleidende ontologische aanname maken, namelijk dat cognitieve concepten eenzelfde soort categorieën vormen als brein gerelateerde categorieën:

Apparently the pie of the mind can be cut into pieces in various ways. (...) CCCs as human kinds cannot be naïvely equated with brain states, brain areas or brain processes, which are natural kinds. (Francken en Slors 2014, 253)

Anders gezegd zijn cognitieve concepten contingent: we kunnen mentale fenomenen op verschillende manieren indelen en benoemen. Deze eigenschap heb ik ook al benoemd in de vorige sectie en zien we terug in de cognitieve psychologie waar meerdere verschillende theorieën bestaan over de indeling van dezelfde functies. Het zijn ‘human kinds’: categorieën die we zelf bedacht hebben als mensen. Brein regio’s of processen staan daarentegen vast. De natuur heeft deze ingedeeld en zijn daarom niet contingent. Hoe je het ook wendt of keert, er lopen nou eenmaal diepe groeven door het brein, waardoor het brein ruimtelijk gezien verdeeld wordt in verschillende lobben bijvoorbeeld.

Francken en Slors gaan zelfs zo ver dat ze claimen dat neurowetenschappers per ongeluk een achterhaalde theorie aanhangen dat *mentaal realisme* wordt genoemd.

Realism suggests that ideally there should be a one-on-one mapping of CCCs or SCCs on brain processes, however complex these may turn out to be. When CCCs or SCCs are brain processes, then it seems that each instantiation of a CCC/SCC

can be identified with a (set of) brain process(es), at least in theory. (Francken en Slors 2014, 254)

Dit is een ingewikkeld stukje omdat hier veel wordt gezegd. Het lijkt er in elk geval op dat Francken en Slors hier zeggen dat mensen die mentaal realistisch zijn (onbewust) denken dat cognitieve concepten hetzelfde zijn als breinprocessen: "When CCCs or SCCs are brain processes". Nu is het natuurlijk de vraag wat breinprocessen precies zijn. In verband met de beperkte ruimte die ik heb in dit paper zal ik breinprocessen hier interpreteren als termen die verwijzen naar bijvoorbeeld activatiepatronen van het brein of iets anders dat dit proces representeert. Dus dan interpreteer ik hun stelling als: neurowetenschappers nemen onbewust aan dat termen-voor-patronen-in-het-brein en cognitieve concepten dezelfde dingen zijn, maar met verschillende namen worden aangeduid (wiskundig zou je dat weergeven als: $x = y$). En zoals we al hebben gezien is dit inderdaad onjuist, omdat cognitieve concepten verwijzen naar gedragspatronen of algemene functies en niet naar patronen die alleen voorkomen in het brein.

Ook begrijp ik uit bovenstaand citaat dat Francken en Slors tot deze conclusie komen omdat neurowetenschappers op zoek zijn naar 1-op-1 vertalingen tussen CCCs/SCCs en breinprocessen: "one-on-one mapping of CCCs or SCCs on brain processes". Dit is echter een conclusie die niet zomaar getrokken kan worden. Een 1-op-1 vertaling kan inderdaad zoiets zijn als $x = y$, in dat geval is de vertaling een identiteitsrelatie. Maar een 1-op-1 'mapping' kan ook een wiskundige functie zijn die twee punten uit verschillende verzamelingen of domeinen met elkaar verbindt: $f(x) = y$. In dit laatste geval zijn x en y niet hetzelfde maar kunnen ze wel aan elkaar worden gerelateerd, zo'n relatie wordt in de wiskunde een *bijjectie* genoemd (Ivanova 2011). Aangezien neurowetenschappers erg wiskundig zijn onderlegd lijkt het mij plausibel dat ze ook op zoek zouden kunnen zijn naar bijjecties in plaats van identiteitsrelaties.

3 Poldrack's methode

3.1 De twee kanten van het cognitief systeem in kaart gebracht

Allereerst hebben Russell en zijn team een begin gemaakt aan een cognitieve ontologie, dat ze de Cognitieve Atlas noemen. Het idee is dat het een kennisbank wordt van cognitieve concepten, waarin gedetailleerd wordt beschreven naar welke globale cognitieve functies deze concepten verwijzen en hoe ze zich verhouden tot de verschillende taken die gebruikt worden in wetenschappelijke experimenten. Dit is dus een visualisering van het schematische deel van het cognitieve systeem.

Omdat elke onderzoeker werkt vanuit een eigen theoretisch kader met eigen definities en ideeën over welke globale concepten bij welke experimentele taken een rol spelen, zijn er veel verschillende impliciete indelingen van dit schematische deel. Dit wordt ook wel het *tandenborstel probleem* genoemd: “no self-respecting cognitive psychologist wants to use anyone else’s theory” (Eysenck en Keane 2020, 7). De bedoeling is dat zoveel mogelijk cognitiewetenschappers aan de Cognitieve Atlas bijdragen om deze inconsistenties in kaart te brengen, naast het uitvinden van waar wél consensus over is (Poldrack 2010, 757). Francken en Slors verwijzen ook naar de Cognitieve Atlas en benadrukken hoe belangrijk het is dat dit idee gerealiseerd wordt (Francken en Slors 2014, 252).

Poldrack's methode die ik in dit paper wil belichten maakt echter gebruik van de Cognitieve Atlas als maar één deel van de puzzel. Het andere deel van de puzzel zou in de ideale wereld een model zijn van het materiele deel van het cognitief systeem (het deel dat overeenkomt met de materiele onderdelen die de functies uit het schematische deel realiseren, zoals besproken in sectie één). Voor zover ik weet bestaat daar nog geen uitgebreide visualisatie van zoals het Cognitieve Atlas, waarbij alle soorten indelingen bij elkaar komen. Wel heb je afzonderlijke ‘kaarten’ van het brein. Poldrack maakt dus gebruik van één van deze kaarten als tweede puzzelstuk in zijn methode. De kaart die hij heeft gekozen heet BrainMap.

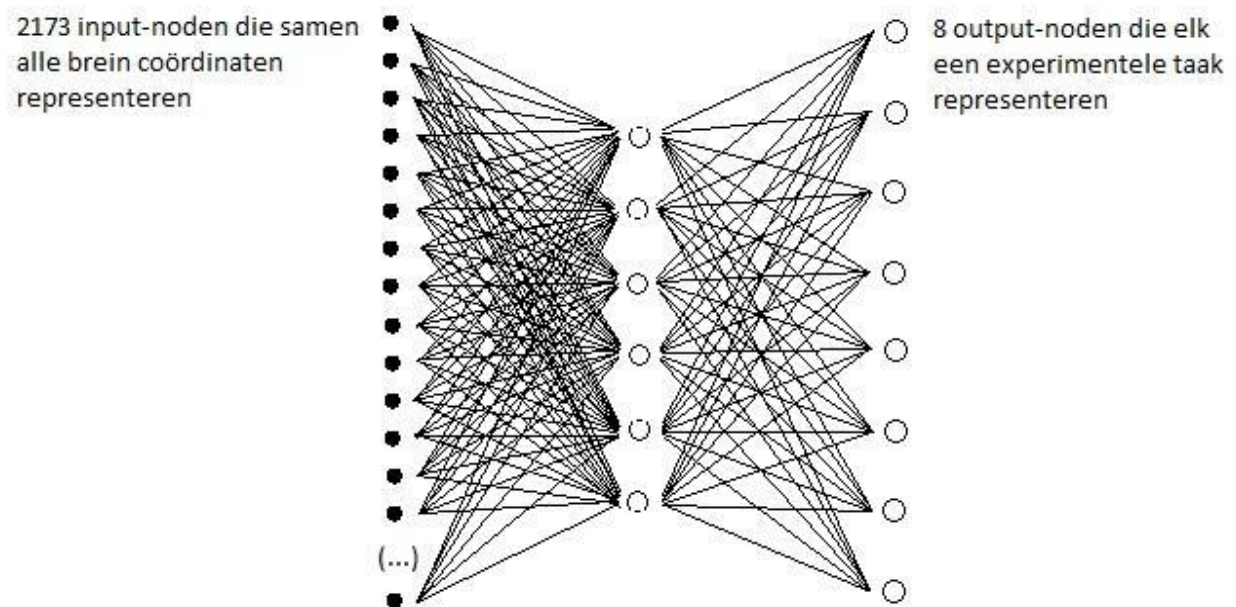
BrainMap is een database die is ontwikkeld aan de Research Imaging Instituut van de Universiteit Texas Health Science Center in San Antonio. Het is een verzameling van data uit zo veel mogelijk individuele neuroimaging studies. Als je zou willen onderzoeken welke delen van het brein *gemiddeld genomen over alle studies* actief zijn bij het uitvoeren van een bepaalde taak, bijvoorbeeld ‘lezen’, dan zou je een meta studie kunnen uitvoeren met de informatie uit deze database. Het bevat echter wel maar één soort indeling van het brein, namelijk de ruimtelijke informatie van waar activiteit

plaats vond tijdens het uitvoeren van een experimentele taak, weergegeven als 3-dimensionale coördinaten van het brein [x,y,z] (Laird, Lancaster en Fox 2005, 66).

Zoals ik eerder heb genoemd, heeft Poldrack geprobeerd zo'n meta studie te doen. Het bleek echter onmogelijk om de gemene delers te vinden in breinactivatie over alle studies. Als er overlappingen zijn, zijn ze voor het menselijke oog onmogelijk waar te nemen in deze enorme berg aan informatie die deze database bevat. Poldrack en zijn collega's kwamen dus met het briljante idee om de meta studie uit te besteden aan een neuraal netwerk dat een veel grotere rekencapaciteit heeft dan een mens.

2.2 Een neuraal netwerk dat het missende ingrediënt voor 1-op-1 relaties verschaft

Poldrack en zijn collega's Halchenko en Hanson, hebben aan de hand van de informatie uit BrainMap een neuraal netwerk getraind dat op kleine schaal kon voorspellen wat voor een taak een proefpersoon aan het doen was (bijvoorbeeld hardop lezen) op basis van een gegeven activatiepatroon (een set van brein coördinaten). In figuur twee zie je een visuele representatie dat ik heb gemaakt van het algoritme op basis van de paper waarin ze uitleggen hoe ze te werk zijn gegaan (Poldrack, Halchenko en Hanson 2009, 1368).



Figuur 2. Dit is een visuele representatie van het neuraal netwerk dat Poldrack, Halchenko en Hanson hebben getraind aan de hand van data uit BrainMap.

Je kan de noden van het neurale netwerk (te zien in figuur 2) vergelijken met lampjes die gaan branden bij activatie. Als er een combinatie van lampjes aan de linkerkant aan wordt gezet, moet het neurale netwerk berekenen welk lampje aan de rechterkant aan moet springen. De input van het algoritme wordt dus gerepresenteerd door de combinatie brandende lampjes van de linker rij, wat overeenkomt met een activatie patroon uit de neuroimaging studies. De output wordt gerepresenteerd door een brandend lampje uit de rechter rij, wat overeenkomt met een experimentele taak die tijdens de meting werd uitgevoerd door een proefpersoon. Er zijn maar 8 output noden, omdat alleen een klein deel van de database is gebruikt als voorbeeld.

Tijdens de trainingsperiode worden de waarden van de verbindingen tussen alle noden willekeurig aangepast. Afhankelijk van of de uitkomst klopt of niet worden deze waarden bevestigd of ontmoedigd, net zo lang tot het neurale netwerk alleen nog maar de juiste uitkomsten produceert. Na de training was het neurale netwerk in staat nieuwe activatiepatronen die het nog niet eerder had 'gezien' te koppelen aan de juiste taken die de proefpersonen uitvoerden tijdens de experimenten waaruit de activatiepatronen waren genomen. Niemand weet wat de 6 noden in het midden precies representeren, alleen dat het netwerk het beste succespercentage behaalde met dit aantal: 80% ten opzichte van 13% als je het aan de toeval overlaat (Poldrack, Halchenko en Hanson 2009, 1364).

Na al dit voorbereidend werk, hebben Poldrack en zijn team uiteindelijk met informatie uit dit neurale netwerk en het Cognitieve Atlas een voorspelling op kleine schaal kunnen geven van welk cognitief concept het sterkst geassocieerd wordt met welk regio van het brein. De berekening hiervan hebben ze gedaan door de verhoudingen tussen 3 ingrediënten met elkaar te vermenigvuldigen, gebruikmakend van lineaire algebra. Het eerste ingrediënt wordt gevormd door de 6 verborgen noden in het neurale netwerk (dat informatie bevat over de verhouding tussen brein coördinaten en 8 experimentele taken). Het tweede en derde ingrediënt worden gevormd door 22 cognitieve concepten en diezelfde 8 experimentele taken die een rol spelen bij deze cognitieve concepten, genomen uit een voorloper op de Cognitieve Atlas. De exacte berekening kun je terugvinden in sectie 2.9 van de bijlage bij het paper van Poldrack, Halchenko en Hanson uit 2009.

Na alle waarden vermenigvuldigd te hebben krijg je een voorspelling in hoe groot de kans is dat een bepaald regio van het brein (een optelsom van specifieke brein coördinaten) een rol speelt bij een bepaald cognitief concept. Als er maar één concept hoog scoort ten opzichte van andere concepten die laag of zelfs negatief scoren, dan is er een 1-op-1 relatie gelegd. Zo'n relatie noemt Poldrack een selectieve associatie. Als we naar de resultaten kijken, zien we inderdaad bijvoorbeeld dat de 'Left Superior Temporal Gyrus posterior' erg hoog scoort op alleen 'audition' (Poldrack 2010, 757). Het is

in dit geval niet meer ambigue welke cognitief concept gecorreleerd wordt aan dat specifieke regio van het brein.

Toch waren er ook gevallen waarin een brein regio niet geassocieerd kon worden met één duidelijk cognitief concept, omdat de waarden van de voorspellingen te dicht bij elkaar lagen. In dit geval ligt het probleem volgens Poldrack in dat de cognitieve ontologie die gebruikt is bij de berekening nog niet helemaal juist is afgesteld. De ontologie wordt namelijk gebruikt als één van de ingrediënten in de laatste berekening. Door dit deel van de input, de ontologie, aan te passen en te kijken wat het doet met het resultaat, kun je na veel 'trial en error' de ontologie steeds verder verbeteren zodat het steeds duidelijkere resultaten geeft. Op deze manier zal de ontologie steeds juister worden, in de zin dat het passend is bij de indeling van het brein dat op dit moment gebruikt wordt door neuroimaging technieken (Poldrack 2010, 754 en 760). Dit proces van bijwerken van de ontologie is vergelijkbaar met de pottenbakker uit sectie één, die de natte klei waaruit de deksel van de pot bestaat, bijvormt, zodat deze goed past op de pot.

4 Een ontologische basis voor Poldrack's methode

Met een ontologische basis voor Poldrack's methode bedoel ik een set van aannames, die mogelijk worden verondersteld door Poldrack, over de ontologische status of bestaanswijze van cognitieve concepten, cognitieve functies en geassocieerde regio's van het brein. De reden dat ik de uitdrukking 'ontologische basis' gebruik is omdat ik een onderscheid wil maken tussen het bestaande theoretische kader van de cognitieve wetenschappen en de manier waarop dit model wordt geïnterpreteerd door individuele wetenschappers. De ontologische basis voor Poldrack's methode kan worden samengevat als de aanname dat er structuren zijn in de werkelijkheid die worden weerspiegeld door concepten, en dat deze structuren te isoleren zijn.

4.1 Concepten weerspiegelen de structuur van de realiteit

Toen ik de methode van Poldrack onderzocht moest ik denken aan de filosofische gedachtegoed van John McDowell. In zijn boek *Mind and World* benadrukt hij de rol van de realiteit in de vorming van concepten. Concepten komen niet zomaar uit de lucht vallen. Ze worden gevormd op basis van ervaring. Ervaring is volgens McDowell een "constraint on thinking from a reality external to it" (McDowell 1996, 29). Oftewel het lijkt misschien dat we met ons denken een oneindige vrijheid hebben om allerlei luchtkastelen te bouwen, maar eigenlijk zijn de bouwstenen die we daarvoor gebruiken om te beginnen altijd abstracties van de werkelijkheid.

Hoe wild onze fantasieën ook zijn, ons verbeeldingsvermogen is niet oneindig in de zin dat we afhankelijk zijn van de realiteit voor de ingrediënten ervoor. Je zou dus kunnen zeggen, op basis van deze gedachtegang, dat concepten altijd iets reflecteren uit de realiteit. Er is dus iets in de realiteit dat wordt gereflecteerd door deze concepten. Dat iets is niet direct waarneembaar als een compleet 'ding' en is waarschijnlijk verspreid over tijd en ruimte. McDowell noemt het de structuur of 'layout' van de realiteit.

conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of reality. Experience enables the layout of reality itself to exert a rational influence on what a subject thinks. (McDowell 1996, 26)

Met andere woorden staat hier, volgens mijn interpretatie van McDowell, dat ervaring het mechanisme is dat het voor mensen mogelijk maakt de structuur van de realiteit te vatten in concepten.

Tot zover is deze analyse van de bestaansvorm van concepten en structuren misschien nog wat onduidelijk. Daarom zal ik proberen aan de hand van een voorbeeld duidelijk te maken wat ik bedoel. Dit voorbeeld zal ook als argument dienen voor het idee dat zulke structuren van de realiteit die een rol spelen in concepten, kunnen worden geïsoleerd.

4.2 Structuren in de realiteit die een rol spelen in concepten kunnen worden geïsoleerd

Ik zou graag je aandacht willen vestigen op een kunstwerk van Tim Noble en Sue Webster gemaakt in 1998 te zien in figuur 3. Het is een hoop afval dat ze hebben verzameld in een periode van zes maanden terwijl ze bezig waren met het maken van andere kunstwerken. Het verschil met een gewone hoop afval is dat ze in dit geval de individuele stukken afval zo hebben geplaatst dat er in de totale hoop een plaatje zit verstoppt. Dat plaatje wordt pas zichtbaar als er vanuit een bepaalde hoek licht schijnt en de afvalhoop zo een schaduw werpt op de muur erachter. In het schaduwbeeld zie je dan ineens twee mensen die ontspannen tegen elkaar aanleunen. De een drinkt een wijntje en de ander is een sigaret aan het roken.



Figuur 3. Het kunstwerk van Tim Noble en Sue Webster

Laten we nu dit kunstwerk als metafoor zien voor het idee dat de structuur van de realiteit te vatten is in concepten. In dat geval representeert de afvalhoop de totale materiele realiteit. Deze totale realiteit heeft een bepaalde vorm. Deze vorm omvat veel verschillende structuren. Eén van deze structuren, in dit geval de omtrek van de afvalhoop gezien vanaf één specifiek zijaanzicht, wordt zichtbaar gemaakt door de lichtbron en de schaduw. Dankzij de schaduw zien we nu alleen één specifieke structuur van de afvalhoop op de muur. Het is een structuur dat eerst onzichtbaar ingebed zat in de afvalhoop samen met allerlei andere structuren. Nu is het geïsoleerd en daarom zichtbaar geworden.

Het licht staat in dit metafoor voor ons denken en de schaduw voor ervaring. De afvalhoop, de realiteit, oefent een restrictie uit op het licht. Daarom bevat de ervaring, de schaduw, een deel van de structuur van de realiteit, de afvalhoop. Aangezien deze specifieke structuur al vaker voorbij is gekomen in onze ervaring hebben we er allerlei associaties bij en interpreteren we dit patroon met concepten als 'mensen', 'dingen' en 'ontspanning'.

Ik hoop dat het mij gelukt is te illustreren op welke manier structuren in de realiteit een rol spelen in concepten op basis van bovenstaande ontologische aannames van McDowell. De laatste stap voor mij is om te laten zien hoe Poldrack's methode betekenis krijgt met deze ontologische basis.

Poldrack en zijn collega's zouden kunnen worden vergeleken met kunstenaars. Ook zij lijken een project op zich te hebben genomen om specifieke structuren in de realiteit, die een rol spelen in concepten, zichtbaar te maken. In plaats van een hoop afval hebben zij gekozen voor een hoop brein coördinaten. In plaats van een lichtbron maken ze gebruik van een neuraal netwerk. In dit geval oefent de hoop brein coördinaten een restrictie uit op de ingevoerde experimentele taken. Zo ontstaan er representaties van de specifieke structuren in de brein coördinaten, in de vorm van waarden opgeslagen in het neuraal netwerk.

Helaas zijn deze waarden in het neuraal netwerk minder goed te interpreteren met het blote oog dan een schaduwvorm op een muur. Poldrack en zijn collega's hebben echter toch een interpretatie kunnen maken door een schematische representatie van algemene concepten, een cognitieve ontologie, te gebruiken in een berekening samen met de waarden uit het neuraal netwerk. En zo is het hen gelukt om een relatie te leggen tussen structuren in de realiteit en concepten.

En zoals Tim en Sue waarschijnlijk de onderdelen van de afvalhoop met vele 'trial en error' pogingen hebben geplaatst, net zo lang tot de schaduw op de muur klopte met het concept 'ontspanning' dat veel mensen delen, zo is ook het kunstwerk van Poldrack en zijn collega's aanpasbaar en kunnen er dankzij 'trial en error' steeds betere relaties worden gelegd.

Conclusie

Met dit paper heb ik geprobeerd een bijdrage te leveren aan het debat over cognitieve concepten en hun verhouding tot delen van het brein waar ze mee worden gecorreleerd.

In de eerste sectie heb ik een schets gegeven van de wetenschappelijke context waarin dit debat zich afspeelt. Daarin heb ik uitgelegd dat modulariteit een belangrijke aanname is in het wetenschappelijk model van het cognitief systeem. Het gevolg is dat inzicht in het cognitief systeem bestaat uit twee delen. Het eerste deel is een model van afzonderlijke cognitieve functies dat de meeste cognitieve wetenschappers in hun hoofd hebben. Het tweede deel bestaat uit het inzicht in hoe deze functies materieel worden gerealiseerd en door welke onderdelen, in materiele instanties. Het globale streven is om met zijn allen het cognitief systeem te begrijpen zoals een technicus die een printer begrijpt en zou kunnen repareren, omdat die zowel bekend is met de materiele onderdelen van de printer als met welke functies ze vervullen.

Hierna heb ik stilgestaan bij het probleem dat het niet lukt om algemene relaties te leggen tussen cognitieve functies en delen van het brein op basis van een meta studies van neuroimaging studies. Volgens Poldrack zou de reden hiervoor kunnen zijn dat de manier waarop het schematische deel van het cognitieve systeem, de functies, op een manier zijn ingedeeld dat niet overeenkomt met de indeling van het breinmodel dat op dit moment gebruikt wordt. Om ze op elkaar te laten passen zouden beide of één van deze indelingen aangepast moeten worden, zoals een pottenbakker dat doet met het bijvormen van natte klei zodat verschillende delen van een pot op elkaar aansluiten.

In de tweede sectie heb ik besproken wat het standpunt van Francken en Slors is in dit debat. Waar het op neerkomt is dat zij het een onmogelijk doel vinden van neurowetenschappers om 1-op-1 vertalingen te vinden tussen algemene cognitieve concepten en breinprocessen. Hierbij interpreteren ze '1-op-1 vertaling' als een identiteitsrelatie, wat zou betekenen dat neurowetenschappers bewust of onbewust zouden geloven dat cognitieve concepten en breinprocessen hetzelfde ding zijn, maar dat er met verschillende termen naar wordt verwezen. Dit kan echter niet, want cognitieve concepten verwijzen volgens Francken en Slors naar gedragspatronen die veel meer sociale informatie bevatten dan alleen dat wat zich afspeelt in het brein van één persoon. Aangezien neurowetenschappers erg wiskundig zijn onderlegd, denk ik dat het hier eerder gaat om een bijctie, een functie dat twee punten uit verschillende domeinen met elkaar verbindt.

In de derde sectie heb ik het technische deel van Poldrack's methode uitgelegd, waarmee hij in staat was 1-op-1 verbanden te leggen tussen cognitieve concepten en regio's van het brein. Daarvoor heeft hij en zijn collega's op basis van data uit BrainMap een neuraal netwerk getraind, waarmee ze als het ware de metastudie hebben uitbesteed aan een algoritme. Na afloop van de training bevatte het neuraal netwerk specifieke waarden. Deze waarden reflecteren de overlappende structuur van activatiepatronen van breinen die geassocieerd zijn met experimentele taken uitgevoerd tijdens neuroimaging studies.

Daarnaast is hij met zijn team begonnen aan een cognitieve ontologie waarin gespecificeerd wordt welke experimentele taken een rol spelen in welke algemene cognitieve concepten. Met de informatie uit het neuraal netwerk en de relaties uit de cognitieve ontologie hebben ze een vervolgens een berekening gemaakt voor een aantal regio's van het brein in hoeverre deze kan worden geassocieerd met alle cognitieve concepten uit de ontologie. Als er één concept duidelijk boven de andere uitsprong met een hoge waarde, was de 1-op-1 relatie gelegd.

In de laatste sectie heb ik een ontologische basis gelegd waarmee Poldrack's methode betekenis krijgt. Daarvoor heb ik eerst de filosofische ideeën van McDowell uitgelegd aan de hand van het kunstwerk van Tim Noble en Sue Webster. Het idee is dat er structuren zijn in de realiteit die worden weerspiegeld door concepten en dat deze structuren te isoleren zijn, net zoals een specifieke omtrek van de afvalhoop van Tim en Sue zichtbaar is op de muur dankzij een lichtbron geplaatst op een strategische plek.

Poldrack's methode lijkt precies dit te doen met brein coördinaten en cognitieve concepten. Het neuraal netwerk dat ze hebben getraind is vergelijkbaar met een lichtbron en de informatie over de overlappende structuur van breinactivatie patronen is vergelijkbaar met de omtrek van de afvalhoop. Samen werpen ze een 'schaduw', oftewel een representatie van een specifieke structuur van de realiteit dat kan worden geïnterpreteerd. Omdat het kunstwerk van Poldrack aanpasbaar is, stelt het wetenschappers instaat, met 'trial en error', steeds juistere structuren in het brein te isoleren die een rol spelen in cognitieve concepten. Of andersom, stelt het wetenschappers instaat steeds betere indelingen te vinden van cognitieve concepten die passen bij de huidige indelingen van het brein.

Bibliografie

Eysenck, Michael W. en Mark T. Keane. 2020. *Cognitive Psychology : A Student's Handbook*. Achtste editie. New York: Psychology Press.

Francken, Jolien C. en Marc Slors. 2014. "From Commonsense to Science, and Back: The Use of Cognitive Concepts in Neuroscience." *Consciousness and Cognition* 29: 248–58. DOI: 10.1016/j.concog.2014.08.019.

Ivanova, O. A. 2011. "Bijection." *Encyclopedia of Mathematics*. Geraadpleegd op 05-04-2022. URL: <http://encyclopediaofmath.org/wiki/Bijection>.

Laird, Angela, Jack Lancaster, en Peter Fox. 2005. "Brainmap." *Neuroinformatics* 3 (1): 65–77.

McDowell, John. 1996. *Mind and World*. First Harvard University Press paperback edition. Cambridge, MA: Harvard University Press.

Noble, Tim en Sue Webster. 1998. "DIRTY WHITE TRASH (WITH GULLS)." Geraadpleegd op 28-03-2022. URL: http://www.timnobleandsuewebster.com/dirty_white_trash_1998.html.

Poldrack, Russell A. 2010. "Mapping Mental Function to Brain Structure: How Can Cognitive Neuroimaging Succeed?" *Perspectives on Psychological Science* 5 (6): 753–61. DOI: [10.1177/1745691610388777](https://doi.org/10.1177/1745691610388777).

Poldrack, Russell A., Yaroslav O. Halchenko en Stephen J. Hanson. 2009. "Decoding the Large-Scale Structure of Brain Function by Classifying Mental States Across Individuals." *Psychological Science* 20 (11): 1364–72. DOI: [10.1111/j.1467-9280.2009.02460.x](https://doi.org/10.1111/j.1467-9280.2009.02460.x).