De opkomst van het orakelend brein

Sander Van de Cruys
Laboratory of Experimental Psychology, KU Leuven

Dit artikel verscheen eerder in het Karakter tijdschrift van wetenschap.

Wist je dat wijngisten in hun vat de volgende stap van het wijnmaakproces al kunnen voorspellen? Elke stap in het proces van simpele druif naar gebotteld edel vocht komt met een aantal specifieke stresserende omstandigheden voor een gistcel. Onderzoek wees uit dat, als je zo'n cel blootstelt aan een milde stress van het type dat gewoonlijk in de eerdere fasen van het wijnproces voorkomt (bv. warmte of ethanolshocks), dit reeds een anticiperende reactie teweegbrengt waardoor de cel beter bestand wordt tegen alle stressoren die pas later in het proces zullen optreden (bv. oxidatieve stress). Geen klein bier voor zo'n simpel, eencellig organisme. Terwijl wij ons wijnmaakproces optimaliseren, passen die levende wezentjes zich ook aan aan onze tradities. Een mooi voorbeeld van culturele en biologische co-evolutie.

Natuurlijk is het niet de individuele gistcel die hier door *trial and error* de juiste voorspelling geleerd heeft. In een hele populatie van gistcellen is het gewoon zo dat die gistcellen met een moleculaire, regulerende keten die zo'n vooruitziende aanpassing realiseren beter overleven en zich dus meer voortplanten. In plaats van puur op de onmiddellijke bedreiging voor hun overleving te reageren, ontwikkelden deze micro-organismen het vermogen om te reageren op de *betekenis* van een bepaalde verstoring in de context van hun 'natuurlijke' habitat (het wijnmaakproces). Zo heeft een gistcel heeft een erg rudimentair maar efficiënt *model* van haar werkelijkheid, door evolutie afgestemd op de dynamische structuur van haar niche. Gistcellen zijn daarin zeker niet alleen; de biologische wereld zit vol van zulke anticiperende processen. Zelfs een boom laat zijn bladeren vallen als de dagen korten omdat dit een goede voorspeller is van aanstaand koud weer. Zonder toverij en met de woorden van cyberneticus Heinz von Foerster kunnen we stellen dat in de biologie de oorzaak vaak in de toekomst ligt.

Anticiperende, moleculaire ketens in micro-organismen stellen dus iets aanwezig dat er (nog) niet is. Dankzij die netwerken wordt een zintuiglijke prikkel voor het organisme een rudimentair symbool van latere gebeurtenissen. Men zou kunnen zeggen dat de essentie van cognitie hier reeds in de kiem aanwezig is. 'Denken' is immers het kunnen voorbijgaan aan of loskomen van de zintuiglijke input van het hier en nu. Natuurlijk is de stap van de erg specifieke en starre vorm van voorspellen door een gistcel naar de menselijke souplesse in het mentale 'tijdreizen' en het jongleren met symbolen enorm groot. Toch legt de analogie de kracht bloot van een theorie die in toenemende mate wordt verdedigd en toegepast in filosofische, psychologische en (neuro)biologische kringen. Deze theorie van de 'predictieve verwerking' heeft een verklarende, eenmakende ambitie die we zelden zien in de biologie of psychologie, met uitzondering misschien van de evolutietheorie. Zeker in de psychologie is men terecht argwanend voor theorieën die pretenderen heel het mentale domein te verklaren. Het eisenlijstje is dan ook lang: zowel de waarneming als de actie moeten erin passen, zowel emotie als ratio, zowel het leren als het dromen. Liever spit iedere onderzoeker haar of zijn eigen paar vierkante meters van het veld om op zoek naar vruchtbare inzichten dan haar of zijn vingers te branden aan voorbarige, hoogmoedige voorstellen die zo omvattend zijn dat ze specialistische kennis vereisen in alle genoemde domeinen.

Maar hoe riskant die laatste weg ook mag zijn, het is belangrijk dat wetenschappers die zich willen focussen op een diepe integratie tussen disciplines in het huidige wetenschappelijke klimaat niet verdrongen worden door even belangrijke, maar eerder specialistische en feitvergarende vorsers. Die laatsten kunnen immers vaker rekenen op snellere, verzekerde resultaten, op het ritme van de projectaanvragen. Het is in dat verband tekenend dat de grote synthese onder de theorie van de predictieve verwerking voor een aanzienlijk deel door filosofen, met name Andy Clark en Jakob Hohwy, wordt gemaakt, in plaats van door psychologen of biologen. Clark schreef in de laatste jaren een rist boeken en wetenschappelijke papers die de aanpak van predictieve verwerking bredere ingang hebben doen vinden, zelfs bij eerder specialistische biologen en psychologen.

Dat de gepredikte interdisciplinariteit in onze onderzoeksvoorstellen niet louter voor de bühne mag blijven, maakt onze onderzoeksmaterie zelf voortdurend duidelijk. Zo blijkt meer en meer dat de hersenen onze nette opdeling in subdomeinen van de psychologie niet respecteren. Ratio en emotie zitten niet in aparte hersendelen, alle populaire theorieën over reptielbreinen ten spijt. Hersendelen die instaan voor de verwerking van visuele prikkels blijken een bijverdienste te hebben in het verwerken van de emotionele betekenis van

opgevangen prikkels. Ons geheugen en onze dromen zijn niet in één regio te vatten, maar betrekken het hele brein. De mooie indeling in functies en regio's is dan wel overzichtelijk, ze haalt haar inspiratie te veel uit de computerwereld, waar precair biologisch bestaan en overleving geen punt zijn. In plaats daarvan redeneert de theorie van de predictieve verwerking vanuit eerste principes. Zij vraagt zich allereerst af wat er nodig is voor een organisme om te (blijven) bestaan als eenheid. Ze doet beroep op de cybernetica, de evolutietheorie en artificiële intelligentie om hierop een antwoord te formuleren met verstrekkende gevolgen. Veel van de puzzelstukken zijn dan ook niet nieuw, maar ze lijken de laatste jaren wel vaker in elkaar te passen.

In de rest van dit essay bekijken we de theorie in vogelvlucht. Het spreekt voor zich dat dit de omvattende theorie oneer aandoet en enkel als *amuse-gueule* kan fungeren voor wie zich er dieper in wil onderdompelen. Wat dat betreft is Andy Clarks werk een uitstekend startpunt.

Aan de oorsprong ligt de strijd om het bestaan, wat in concrete termen verstaan moet worden als het opzwemmen tegen de 'kracht' van de tweede wet van de thermodynamica. Die wet beschrijft niet meer dan dat er veel meer manieren zijn om ongeordend te zijn dan om geordend te zijn, dus dat alles statistisch gezien neigt naar wanorde, naar meer (thermodynamische) entropie of minder organisatie. Om te bestaan moet een 'georganiseerd' organisme dus actief zijn eigen organisatie creëren en in stand houden. Het moet ervoor zorgen dat de 'spreiding' van zijn toestanden (informatie-entropie) in toom gehouden wordt, gezien slechts een beperkt aantal toestanden consistent zijn met een verdere overleving. Dit kennen we natuurlijk als homeostase. Welke set van toestanden dat is, werd door evolutie 'ontdekt', simpelweg omdat een organisme zonder die restricties minder goed overleefde. De set van verwachte toestanden karakteriseert in essentie het organisme. Goede voorbeelden van ingebakken toestanden voor stabiele organisatie zijn *interoceptieve* verwachtingen zoals het bloedsuikergehalte of de lichaamstemperatuur bij zoogdieren. Merk op dat een organisme die verwachtingen meestal niet op eigen houtje kan vervullen, het moet daarvoor via de omgeving gaan. Het moet selectief open zijn — via *exteroceptie* — voor de omgeving en de regelmatigheden die erin gelden in zoverre die een impact hebben op het interne milieu. Een goede regulator van een systeem moet een goed model zijn van dat systeem, stelde cyberneticus Ashby. Wat hij daarmee bedoelt is dat een organisme een model moet bevatten van hoe de verlangde interoceptieve toestanden gegenereerd kunnen worden, hetzij via processen in de omgeving, hetzij via eigen acties, en meestal in een combinatie van beide. Zo'n *generatief model* moet de latente oorzaken van de beoogde interne staten vatten, zodat die gerealiseerd kunnen worden via de acties waarover

het organisme beschikt. Het model kan, zoals we eerder bij de gistcel zagen, rudimentair zijn en vastgelegd door evolutie. Cruciaal is dat het model voorspellingen toelaat die betere compensatie van externe verstoringen van interne toestanden mogelijk maken: predictie is beter dan reactie.

Het aanleren van nieuwe, flexibele voorspellingen wordt pas belangrijk voor een organisme wanneer het tot stand brengen van de beoogde toestanden afhankelijk is van regelmatigheden en voorwaarden die veranderlijk zijn tijdens het leven van dat organisme. Pas dan wordt leren de moeite. Dit is bij uitstek het geval bij mensen, die een uitgebreid repertoire aan mogelijke acties hebben en vaak steunen op sociale organisatie voor het voorzien in voedsel en bescherming. Ons generatief model is grotendeels aangeleerd, zodat voedselvergaring mogelijk is in volstrekt andere sociale omstandigheden (jagers-verzamelaars vs loonverdieners-supermarktgangers).

Elk organisme moet dus voorkomen dat het zich bevindt in zintuiglijke toestanden die verrassend zijn voor het soort systeem dat het organisme is (bv. een vis op het droge). Verrassing geeft hier aan hoe onwaarschijnlijk een zintuiglijke toestand is ten aanzien van de verwachte set van *attractor* toestanden van het organisme. We moeten het dus zien als 'voorspellingsfout', of het verschil tussen een verwachte of voorspelde toestand en de feitelijke zintuiglijke data. Het centrale dictum voor elk organisme is dus het verminderen van voorspellingsfouten oftewel het maximaliseren van het empirisch bewijs voor het model dat een biologisch organisme is (technisch gezien gaat het om het minimaliseren van *variational free energy,* wat onder bepaalde voorwaarden equivalent is aan de voorspellingsfout). Dóór zijn bestaan levert het organisme dus letterlijk en figuurlijk bewijsmateriaal voor zijn eigen bestaan. Deze opzettelijke tautologie is de kern van de theorie.

Voorspellingsfouten verminderen kan je doen op twee manieren. Je kan je voorspelling aanpassen aan de wereld of je kan de wereld aanpassen aan je voorspellingen. In het tweede geval ga je ingrijpen op de wereld via je acties zodat die wereld zich meer conformeert aan je model ervan. Het eerste geval, je voorspelling updaten, kennen we dan weer beter als 'leren', maar het omvat ook al onze waarnemingen. Wat we waarnemen is immers het resultaat van een voortdurend confronteren van voorspellingen met zintuiglijke inputs. Dat onze waarneming door en door gekleurd wordt door voorspellingen wordt voor ons vaak enkel duidelijk als het misloopt, bijvoorbeeld als iemand de kamerplant verplaatste of zijn haarsnit veranderde. Deze opvatting betekent een hele omslag in ons denken over de waarneming. Waar klassieke modellen stellen dat we passief zintuiglijke data opvangen die

door het perceptueel systeem van de grond af opgebouwd worden tot de objecten die we kennen, gaat de predictieve aanpak ervan uit dat we de buitenwereld voortdurend proactief construeren. De zintuiglijke data, in de vorm van voorspellingsfouten, dienen daarbij 'enkel' ter correctie. Ze zorgen ervoor dat onze constructies geüpdatet worden door actuele inputs. Vanuit dit oogpunt is perceptie een vorm van gecontroleerde hallucinatie. Die opvatting verklaart niet alleen heel wat hardnekkige perceptuele illusies, maar ook waarom hallucinaties frequent voorkomen als we mensen langdurig afsluiten van zintuiglijke prikkels of bij mensen met schizofrenie. Het klopt ook met de bevinding dat hersenactiviteit tijdens het fantaseren of dromen lang niet zo verschillend is van die tijdens daadwerkelijke waarneming.

De theorie van de predictieve verwerking stelt een concreet schema van informatieverwerking in de hersenen voor: een hiërarchische verwerking die op elk niveau berust op de vergelijking van twee versies. Een regio die hoger ligt in de hiërarchie stuurt haar verwachting omtrent het patroon van neurale activiteit in een lagere regio naar die regio en kan aldaar, als het patroon klopt, de consistente neurale activiteit onderdrukken of 'wegverklaren'. De activiteit die overblijft is de nieuwe info, de predictiefout, die op haar beurt hogerop gestuurd kan worden om de toekomstige voorspellingen bij te sturen, zodat het model beter wordt. Dat tweerichtingsverkeer van neurale boodschappen ontspint zich op alle niveaus van de hiërarchie in de hersenschors, waarbij hoger gelegen regio's voorspellingen of patronen genereren die steeds abstracter zijn, en steeds meer tijd en ruimte omvatten. Neem het voorbeeld van het lezen van een boek, waarbij je verwachtingen over opeenvolgingen van gebeurtenissen in het verhaal op het hoogste niveau leven, daaronder verwachtingen met betrekking tot betekenissen van zinnen, en weer daaronder voorspelde woorden binnen een gestarte zin. Helemaal onderaan voorspellen we specifieke perceptuele kenmerken binnen letters. Daar onderaan worden voorspellingen gegenereerd over contrasten en oriëntaties die erg snel veranderen en uiteindelijk op het niveau van de retina 'beantwoord' kunnen worden. Uiteraard vindt die cascade grotendeels onbewust plaats: de voorspellingen blijven impliciet, maar zorgen er wel voor dat lezen enorm efficiënt kan gebeuren, eenmaal de heersende patronen goed geleerd zijn. Daardoor moeten we niet eens alle letters lezen maar kunnen onze ogen doorheen woorden en zinnen springen. Soms lezen we wel eens iets wat er, bij nader inzien, niet stond, maar uit eigen hoofd — voorspellingen dus — ontsproten was. Dat dit niet vaker gebeurt, geeft aan hoe goed onze aangeleerde voorspellingen de redundanties in een tekst uitbuiten.

Merk op dat goed voorspellen ook betekent dat je weet wanneer je *niet* kan voorspellen, wanneer er meer data nodig zijn. Het impliceert het scherpstellen van het ritme van de

dataverzameling op basis van een context (hier: het lezen van een tekst). Als ik de taal van de tekst in kwestie niet erg machtig ben (andere context), dan zal ik mijn ritme van het monsteren van zinnen via oogbewegingen moeten aanpassen. Zo ook voor zinnen die niet samenhangen in een coherent verhaal. Het gaat hier dus over het vormen van een (meta)-voorspelling over de kwaliteit en veranderlijkheid (de onzekerheid) van de zintuiglijke informatie én van mijn eigen voorspellingen. Het laat ons toe om, wanneer nodig, een relatief groter gewicht te geven aan voorspellingsfouten (de data) in plaats van aan onze voorspelling. Dit wordt gevat met het concept van onzekerheidsverwachtingen (technisch 'precisie' genoemd). Stel: je neemt elke dag op een vast tijdstip een bus. Na enkele dagen heb je een goede voorspelling van wanneer je bus aankomt (het gemiddelde van je ervaringen), maar evenzeer een verwachting over de variabiliteit of onzekerheid van die voorspelling. Als je bus de volgende dag niet aankomt op het voorspelde tijdstip en dus een voorspellingsfout genereert, kan je die fout een gewicht toekennen op basis van de ingeschatte betrouwbaarheid. In het licht van die onzekerheid kan een lichte afwijking nog binnen het verwachte bereik liggen. Dit zal bepalen of we nog wat blijven wachten op de bus (geen gevolg geven aan de voorspellingsfout), in plaats van ons model onmiddellijk te gaan aanpassen (is het busschema veranderd?) of een actie te gaan stellen (dan zoek ik maar een ander vervoersmiddel) om de voorspellingsfout te reduceren.

Wat geldt voor het busvoorbeeld en het leesvoorbeeld geldt voor heel onze waarneming. Dit schema voor de werking van de hersenen, *predictive coding* genoemd, wordt ondertussen volop onderworpen aan empirische toetsen. Een aantal experimentele testen zijn reeds gebeurd en zijn positief uitgedraaid. Zo blijken onvoorspelde zintuiglijke prikkels (een voorspellingsfout, bijvoorbeeld in de vorm van een hogere toon na een reeks lagere tonen) inderdaad een sterker hersensignaal uit te lokken, zelfs als het om een onvoorspeld *gebrek* aan prikkels gaat (een weggelaten toon in een patroon van tonen): geen zintuiglijke input, maar wél een voorspellingsfout en dus een hersensignaal. Ook het bestaan van een hiërarchie van voorspellingen in de hersenschors is bevestigd. Andere experimenten wezen uit dat voorspelde prikkels (bijvoorbeeld een prikkel op het juiste moment in de baan van een herhaalde, dus voorspelbare beweging van een object) onderdrukt worden, zodat ze slechter opgemerkt worden door proefpersonen. Nog meer empirisch bewijs komt uit het domein van de controle van actie, waar bevindingen al langer suggereren dat acties niet zozeer als commando's maar eerder als verwachte zintuiglijke ervaringen worden gecodeerd. Wanneer we acties opvatten als voorspellingen van hun zintuiglijke, proprioceptieve gevolgen — over de positie van spieren en het lichaam —passen ze mooi binnen dit kader. Een actie wordt gerepresenteerd als de bereikbare, verwachte toestand die ze vervult. Ook hier speelt natuurlijk een hiërarchie waarbij abstracte langeretermijndoelen en verlangens vertaald

kunnen worden in actiepatronen en verwachte tussentijdse zintuiglijke uitkomsten. Maar op het laagste niveau kunnen we die actie steeds 'uitpakken' tot proprioceptieve verwachtingen die vervuld kunnen worden door spierbewegingen: hier verwekt de voorspellingsfout de reflexboog.

Dit begrip van actie levert ons twee voordelen op. Ten eerste kunnen voorspellingsfouten de universele taal voor de hersenen vormen: ze reguleren zowel perceptie als actie. Ten tweede kunnen we de mentale modellen die we opgebouwd hebben om onze eigen acties te voorspellen, hergebruiken om de handelingen van anderen te voorspellen. Het bestaan van spiegelneuronen, die zowel vuren wanneer ik zelf een actie doe als wanneer ik diezelfde actie door iemand anders zie doen, past als gegoten binnen dit kader. Intenties en verlangens zijn 'slechts' diepere latente oorzaken die we gebruiken om (de zintuiglijke inputs veroorzaakt door) acties van anderen te verklaren, net zoals ze dat voor zintuiglijke inputs van eigen handelingen doen. Onderzoek wijst inderdaad uit dat we acties meteen als intenties of verlangens zien in plaats van als de ruwe zintuiglijke data die we eigenlijk binnenkrijgen, omdat dit ons gewoon toelaat om onszelf en anderen efficiënter te voorspellen. Schopenhauer wist het al wanneer hij schreef: 'door het observeren van mijn eigen daden doorheen mijn hele leven, ontdek ik wie ik ben'. Inderdaad, zelfs ons 'zelf' is deel van ons model, het verschijnt als beste verklaring (latente oorzaak) voor gecorreleerde patronen in zelf-gegenereerde inputs vanuit de verschillende zintuigen.

Ondanks positieve voortekenen blijft er nog een enorme hoop empirisch werk voor de theorie van de predictieve verwerking. Internationale onderzoeksgroepen met erg verschillende invalshoeken werken hieraan, van de robotica en de artificiële intelligentie over de neurowetenschappen en de psychologie tot de psychiatrie en de filosofie.

Want de theorie leidt ook filosofisch tot een aantal prikkelende vaststellingen. Dat we bijvoorbeeld 'de wereld daarbuiten' nooit rechtstreeks kunnen kennen, maar altijd slechts vanuit onze modellen en onze constructies ervan. We ontmoeten de werkelijkheid enkel in ons falen, met name onze voorspellingsfouten. Een bemoedigende boodschap voor wetenschappers, me dunkt. Let wel: dat onze realiteit geconstrueerd is, zet niet noodzakelijk de deur open voor een ongebreideld relativisme: een observatie wordt nog steeds bepaald door ruwe data uit de wereld, maar wat we kunnen observeren wordt bepaald door hoe we de wereld 'porren', aan welke vragen (voorspellingen) we ze kunnen onderwerpen, wat op zijn beurt weer afhangt van onze modellen. Merk de parallel met het wetenschapsproces zelf, waar nieuwe theorieën pas nieuwe data ontsluiten. Een tikfoutje zal irrelevante ruis zijn als we de inhoud van een tekst willen kennen, maar hetzelfde foutje wordt een informatief

signaal als we de vraag stellen naar de zorgvuldigheid van de schrijver. De betekenis —of iets signaal of ruis is— is geen functie van de data maar van onze vragen, dus van de constructie die we erop leggen. Om nieuwe betekenis te verkrijgen leggen we telkens twee versies op elkaar: het patroon zoals we het verwachten volgens ons model en het patroon zoals de sensorische data het ons aanleveren. Het resulterend verschil of de voorspellingsfout is dan, met de woorden van antropoloog Gregory Bateson, een verschil dat een verschil maakt.

Het 'verschil maken' gaat hier in rechtstreekse zin over de nood aan het updaten van onze voorspellingen of overtuigingen. Maar het feit dat heel ons predictieve apparaat geënt is op de verwachte, 'levensvatbare' toestanden die het organisme wil waarmaken, zorgt ervoor dat voorspellingsfouten ook het verschil uitmaken in deze core business. Volgens de theorie van de predictieve verwerking is dit de ultieme en enige functie van de hersenen (en bij uitbreiding het hele organisme): het vertalen van inputs uit de wereld naar een formaat dat iets zegt over het verdere bestaan van het systeem zelf dat die vertaling maakt. Het gaat hier om het vertalen in actiemogelijkheden die voorspellingsfouten verminderen en daarmee overlevingskansen vergroten. Een vertaler die zichzelf het bestaan in vertaalt. Het orakel dat zichzelf de wereld in orakelt. Ik voorspel (of denk) dus ik ben, als ik voorspel wat ik ben. Deze circulaire causaliteit is het fundament van biologische systemen en verklaart hun kenmerkende doelgerichtheid, het 'streven' dat hen onderscheidt van de niet-levende wereld.

Maar botst de leuze van de vermindering van voorspellingsfouten niet met onze verkenningszucht, onze zin voor creativiteit, onze drang naar volstrekte verrassing in de kunst en de moppen? Zie het tweede deel van dit essay.

Meer lezen?

Clark, A. (2015). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford/New York: Oxford University Press.

Colombo, M., Irvine, E., & Stapleton, M. (2019). *Andy Clark and His Critics*. Oxford/New York: Oxford University Press.