

De Simulatiehypothese

drs. Michel van Dartel, IKAT, Universiteit Maastricht
prof. dr. Eric Postma, IKAT, Universiteit Maastricht

Een belangrijk onderdeel van het bewustzijn is het voorstellingsvermogen. De simulatiehypothese stelt dat bewust denken het resultaat is van gesimuleerde interactie met de omgeving. Met behulp van de gesitueerde robotica wordt deze hypothese onderzocht. Van Dartel, Postma en Van den Herik deden onderzoek naar het effect van simulatiemechanismen in robots en hun anticiperende vermogen.

Even denken. Je maakt een bakje van bladerdeeg, dan gaan er wat plakjes aubergine in, laagje mozzarella er op, laagje aubergine, laagje mozzarella, knoflook natuurlijk, maar die moest eigenlijk onderop, en... wat

deed je daar ook alweer voor kruiden op? Wild gebarend sta je voor het Silvo vak van de supermarkt in gedachten je quiche te bereiden. De schappen zie je niet eens meer, in gedachten sta je gewoon thuis aan het aanrecht en pak je de verschillende ingrediënten uit de kasten en lades. Op een plankje in het keukenkastje recht voor je, staan de kruiden. Je mentale oog ziet de Italiaanse kruiden vlak naast elkaar staan. Oregano! Natuurlijk, die is laatst opgegaan bij het bereiden van een pastasalade. ‘Hallo?!’ klinkt het vanuit de verte. Even kijk je nog om je heen in je keuken, maar plots ben je weer terug in de supermarkt en staat je gast-aan-tafel naast je. ‘Hebben we alles?’ vraagt ze.

Het voorstellingsvermogen is een belangrijk deel van het bewustzijn. Een mentale voorstelling maakt het mogelijk om in gedachten handelingen te verrichten in omgevingen en met objecten die niet fysiek aanwezig hoeven te zijn. Hierdoor ben je in staat om te anticiperen op wat er in een bepaalde situatie gaat gebeuren. Door je bijvoorbeeld de procedure van het bakken van een quiche voor te stellen, kun je anticiperen op de benodigde ingrediënten. Op een zelfde wijze anticipeert een volleerde schaker op reeksen van mogelijke zetten van de tegenstander. Het vermogen om te anticiperen maakt het mogelijk gedrag te plannen. Het doorgronden van de neurale basis van het plannen van gedrag vormt momenteel een grote uitdaging in de cognitiewetenschappen en, met name, in het robotonderzoek. De uitdaging wordt aangegaan door Hesslow, die recentelijk een neurowetenschappelijke hypothese over het bewustzijn opstelde, de simulatiehypothese.

De simulatiehypothese

De simulatiehypothese is gebaseerd op het basisprincipe van de empirische filosofie, namelijk dat de verklaring voor bewustzijn gezocht moet worden in de waarneming. Volgens Hesslow is ons bewuste denken het resultaat van gesimuleerde interactie met de omgeving. Onder gesimuleerde interactie verstaat hij: het in het brein ‘nadoen’ van perceptuele en motorische processen. Dit houdt in dat, wanneer je denkt aan bijvoorbeeld een aubergine, er zich in je brein dezelfde perceptuele en motorische processen afspelen die zich normaal voordoen wanneer je daadwerkelijk een aubergine bekijkt.

Drie aannames liggen ten grondslag aan de simulatiehypothese:

1. Simulatie van actie. De aanname dat activatie van de pre-motor gebieden in de hersenen niet noodzakelijk leidt tot een actie, maar wel tot de verbeelding daarvan.
2. Simulatie van perceptie. De aanname dat activatie van sensorische gebieden in de hersenen niet noodzakelijk wordt veroorzaakt door externe stimuli, maar wel leidt tot de verbeelding daarvan.
3. Anticipatie. De aanname dat simulatie van een actie activatie van sensorische gebieden kan veroorzaken, hetgeen aanleiding geeft tot de verbeelding van de perceptuele consequenties van de actie.

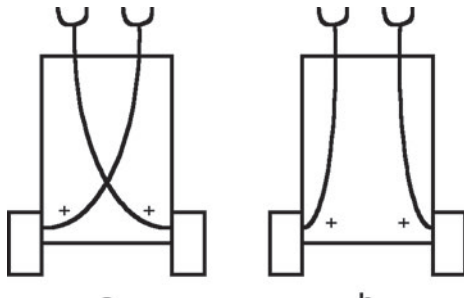
De simulatiehypothese wordt ondersteund door neurowetenschappelijk onderzoek waaruit blijkt dat dezelfde breingebieden actief zijn bij gesimuleerde en daadwerkelijke percepties en acties. Ook de gesitueerde robotica biedt ondersteuning voor Hesslow’s hypothese. Bovendien maakt de gesitueerde robotica een nadere bestudering van het simulatiemechanisme mogelijk.

Gesitueerde robots

Gesitueerde robots zijn robots die interacteren met hun omgeving door middel van sensoren en actuatoren. Het belang van directe interactie met de omgeving voor artificiële intelligentie werd benadrukt door de bekende robotonderzoeker Rodney Brooks. Lang voor Brooks was al gebleken dat de klassieke robotica ontoereikend was voor de ontwikkeling van intelligente robots. In de klassieke robotica bestaat een sterke scheiding tussen sensoren (bijvoorbeeld camera’s), centrale verwerking (een computer), en actuatoren (een manipulator). De beperkingen van de klassieke robotica laten zich goed illustreren aan de hand van de experimenten die in de jaren tachtig met de robot Shakey zijn uitgevoerd. Shakey werd ontwikkeld door Nilson en maakte gebruik van een op scripts gebaseerd planningmechanisme. Shakey kon succesvol door een kantoorruimte manoeuvreren via een vooraf bepaalde route. Hoewel, in eerste instantie, het gedrag van Shakey intelligent leek, bleek bij nadere beschouwing de intelligentie uiterst beperkt. Zo kon bijvoorbeeld een kleine verandering in de opstelling van de kantoorruimte Shakey’s planning zodanig verstoren dat het doel niet meer bereikt werd. De gevoeligheid voor kleine veranderingen maakte de toepassing van Shakey in een natuurlijke omgeving onmogelijk.

Brooks, en na hem vele anderen, demonstreerde dat door

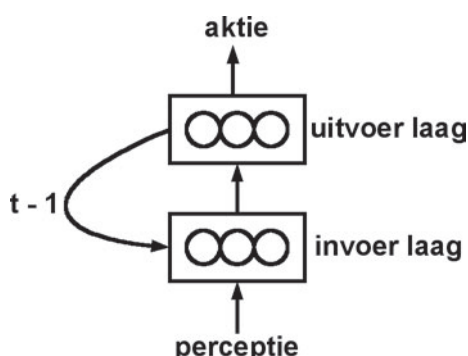
een directe koppeling van sensoren en actuatoren, dus zonder tussenkomst van een expliciet model van de wereld zoals vervat in Shakey's scripts, een robot veel beter in staat is om te gaan met de dynamiek van de werkelijke wereld. In essentie implementeerde Brooks daarmee de gedachte-experimenten van Valentino Braitenberg. Deze Duitse onderzoeker betoogde middels gedachte-experimenten dat simpel cognitief gedrag, zoals het ontwijken (angst) of benaderen (agressie) van objecten, kan voortkomen uit een directe koppeling tussen de sensoren en actuatoren van een robot. Twee voorbeelden van Braitenberg's imaginaire robots, 'vehicles' genaamd, zijn hieronder afgebeeld. Vehicle 'a' zal, wanneer zijn sensoren (de U-vormige uitsteeksels aan de voorzijde) worden geactiveerd, aangetrokken worden door de bron van de sensorische activiteit, terwijl vehicle 'b' bij activiteit van zijn sensoren juist van de bron weg zal draaien.



In moderne robots wordt de koppeling tussen sensoren en actuatoren gerealiseerd door een kunstmatig neurale netwerk. De optimale koppeling wordt verkregen door het automatisch aanpassen van de parameters (de gewichten) van het neurale netwerk. Hiertoe wordt doorgaans gebruik gemaakt van evolutionaire algoritmen of van 'reinforcement learning' algoritmen.

Interne simulatie in robots

De realisatie van een eenvoudig intern simulatiemechanisme in een neurale netwerk verloopt als volgt: een standaard neurale netwerk wordt voorzien van een extra uitvoer en bijbehorende terugkoppelingsslus naar de invoer van het neurale netwerk. Deze terugkoppeling is hieronder geïllustreerd.



Deze simulatie-architectuur verschilt van standaard neurale netwerk architecturen met een terugkoppelingsslus, zoals simpele recurrente netwerken, omdat de terugkoppeling plaats vindt op het niveau van uitvoer en invoer, in plaats van op het niveau van tussengelegen lagen. Hierdoor voldoet de simulatie-architectuur aan de drie aannames van de simulatiehypothese: (1) acties kunnen worden gesimuleerd zonder te worden uitgevoerd, (2) teruggekoppelde uitvoer kan dienen als gesimuleerde perceptie, en (3) de robot kan anticiperen op wat de invoer zal zijn wanneer een bepaalde actie zou worden uitgevoerd. Kortom, door de uitvoer van het netwerk terug te koppelen als invoer, kan een robot, in principe, perceptie en gedrag nabootsen en daardoor 'intern simuleren'.

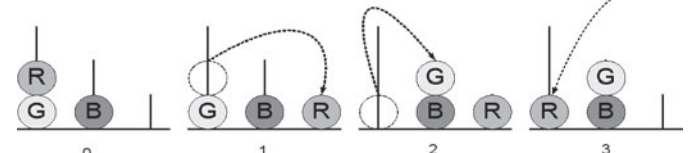
Recentelijk trainden Ziemke, Jirnhed en Hesslow een robot om, aangestuurd door een neurale netwerk voorzien van een simulatie-mechanisme, een doolhof te bewandelen zonder tegen de randen te botsen. Nadat de robot de doolhof had leren te bewandelen op basis van de normale en gesimuleerde invoer, werd de robot getest zonder de normale invoer. Tot op zekere hoogte was de robot in staat de doolhof te bewandelen op basis van enkel de gesimuleerde invoer. Volgens Ziemke et al. betekent dit dat de robot op de toekomstige invoer anticipeerde en dat de robot dus had geleerd intern te simuleren.

Van Dartel, Postma en Van den Herik toonden aan dat een intern-simulerende robot beter presteert op een categorisatietask dan een niet intern-simulerende robot. Uit een analyse van de van uitvoer naar invoer teruggekoppelde activiteit bleek dat er geen sprake was van op invoer gelijkende activiteiten.

De Tower of London taak

In een vervolgonderzoek maakten van Dartel, Postma en Van den Herik gebruik van een robotmodel van een taak waarbij een beroep wordt gedaan op het anticiperend vermogen van de robot, de Tower of London (ToL) taak. De Tower of London taak wordt in de neuropsychologische praktijk gebruikt om het planningsvermogen van patiënten met frontaal hersenletsel te testen.

In de ToL taak wordt aan een proefpersoon gevraagd een beginconfiguratie van drie ballen op drie stokken te veranderen in een doelconfiguratie in het minst mogelijk aantal zetten, waarbij enkel gebruik gemaakt mag worden van 'vrije' plaatsen op de stokken. In de figuur hieronder wordt een voorbeeld gegeven van een optimale oplossing voor een ToL probleem waarbij de beginconfiguratie gelabeld is met een '0' en de doelconfiguratie met een '3'. De letters op de ballen in de figuur representeren de kleur van de desbetreffende bal, waarbij 'G' voor geel staat, 'B' voor blauw en 'R' voor rood.



Plannen speelt bij het oplossen van een ToL probleem een belangrijke rol. Om tot de optimale oplossing te komen, moeten zetten, voordat ze daadwerkelijk worden uitgevoerd, namelijk eerst mentaal uitgetoetst (gesimuleerd) en vergeleken worden. Volgens Hesslow wordt dit mentaal plannen van zetten mogelijk door het vermogen om intern zetten te kunnen simuleren. Hierdoor is de Tower of London taak een typisch voorbeeld van een taak waarbij interne simulatie vereist is.

Shallice⁹ heeft een standaard testset ontwikkeld voor de klinische praktijk. Deze testset is gebruikt door Van Darter et al. om robots met en zonder een simulatiemechanisme te trainen op de ToL taak. De eerste resultaten uit dit onderzoek laten zien dat beide typen robots evenveel problemen van de set leerden oplossen. Wel bleek dat de robots met een intern simulatiemechanisme minder zetten nodig hadden om de ToL problemen op te lossen dan de robots zonder dit mechanisme. Kennelijk is het planningsvermogen van robots daadwerkelijk gebaat bij een intern simulatiemechanisme. Een analyse van intern simulerende robots liet zien dat de robots zetten intern uitprobeerden, alvorens ze al dan niet uit te voeren. Het simulatiegedrag van deze robots ondersteunt Hesslow's simulatiehypothese.

Interne simulatie en bewustzijn

Hesslow claimt dat de simulatiehypothese een verklaring biedt voor het gevoel van een 'interne wereld', dat een belangrijk aspect van het bewustzijn vormt.

Een bredere discussie met betrekking tot interne simulatie en bewustzijn komt van eigen bodem. Phaf en Wolters ontwikkelden namelijk, nog vóórdat Hesslow zijn hypothese publiceerde, een soortgelijke hypothese. Phaf en Wolters menen echter dat interne simulatie alleen niet genoeg is voor het verklaren van bewustzijn. Menselijk bewustzijn, zo stellen zij, is het resultaat van het kunnen manipuleren van een omgeving met alle fylogenetische en ontogenetische vermogens die mensen hebben. Daarnaast benadrukken ze het belang van (verbale) interactie met andere individuen.

Met de beperkte fylogenetische en ontogenetische vermogens van de hedendaagse robots ligt bewustzijn vooralsnog buiten het bereik van de kunstmatige intelligentie. Desalniettemin zou een simulatiemechanisme wel eens de belangrijkste bouwsteen kunnen vormen in de ontwikkeling van bewuste robots.

Nee, we moeten nog oregano en... o ja, zout en peper natuurlijk, maar dat heb ik thuis nog wel. Wacht eens even, zout en peper doe ik nooit direct in de quiche, maar roer ik altijd eerst door een eitje dat vervolgens door de quiche gaat. Ook nog eieren dus, daarna moeten we alles wel hebben. \emptyset

1. Hesslow, G. (2002) Conscious thought as simulation of behaviour and perception. *Trends in Cognitive Sciences*, 6, 242-247.
2. Jeannerod, M. (1994) The representing brain. Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17, 187-245.
3. Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001) Neural foundations of imagery. *Nature Reviews Neuroscience*, 2, 635-642.
4. Brooks, R. A. (1991). Intelligence without Representation. *Artificial Intelligence*, 47, 139-159.
5. Nilsson, N. J. (1984). Shakey the robot. Technical Report 323, AI Center. SRI International, Menlo Park, CA.
6. Braitenberg, V. (1986). *Vehicles: Experiments in Synthetic Psychology*. MIT Press, Cambridge, MA.
7. Elman, J. L. (1991) Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, 7, 195-224.
8. Ziemke, T., Jirnhed, D.-A. & Hesslow, G. (submitted). Toward Internal Simulation of Perception in Mobile Robots.
9. van Darter, M. F., Postma, E. O., & van den Herik, H. J. (2004). Categorisation through internal simulation of perception and behaviour. *Proceedings of the 16th Belgium-Netherlands Conference on Artificial Intelligence* (eds. L. Schomaker, N. Taatgen, and R. Verbrugge), 187-194. Groningen, The Netherlands.
10. Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society London B*, 298, 199-209.
11. van Darter, M. F., Postma E. O., & van den Herik H. J. (in preparation) Symbol manipulation in situated robots.
12. Phaf, R. H. & Wolters, G. (1997). A constructivist and connectionist view on conscious and unconscious processes. *Philosophical Psychology*, 10, 287-307.