

Experimentele filosofie, kunstmatige intelligentie en cognitieve neurowetenschap

ARTIKEL

PIM HASELAGER

1. Inleiding

De filosoof Knobe (2003a, 2003b, 2006) heeft experimenteel onderzoek verricht (zie elders in dit nummer voor meer details) dat enige opwinding heeft veroorzaakt. Die opwinding is wellicht het best verwoord in The New York Times (van 9 December, 2007) in een artikel getiteld 'The new philosophy': *"perhaps the most striking thing is this: The study was conducted by a philosopher, as a philosopher, in order to produce a piece of ... philosophy"*.

Een belangrijke krant schrijft erover, en de puntjes lijken aan te geven dat er verder niets hoeft te worden uitgelegd: Filosofen die, als filosoof, experimenteren met filosofische doeleinden. Dat is nieuws. Knobe en diverse anderen manifesteren zich als onderdeel van een nieuwe ontwikkeling in de filosofie, die zich 'X-phi' of 'Experimentele filosofie' (XF) noemt. De beweging verspreidt haar boodschap via publicaties en internetpagina's, compleet met professioneel licht provocerende filmpjes op YouTube met animaties van een brandende leunstoel (zie <http://pantheon.yale.edu/~jk762/ExperimentalPhilosophy.html>, http://experimentalphilosophy.typepad.com/experimental_philosophy/http://www.youtube.com/user/xphilosophy)

In hun 'An experimental philosophy manifesto' vatten Knobe en Nichols (2008b: 3) de essentie van XF als volgt samen: *"unlike the philosophers of centuries past, we think that a critical method for figuring out how human beings think is to go out and actually run systematic empirical studies."*

Hieronder wil ik proberen aan te geven waarom ik de experimentele filosofie weliswaar als goed nieuws beschouw, maar niet half zo goed als zou kunnen, en ook niet als echt nieuw. Qua filosofische experimenten beperkt de XF-beweging zich vooralsnog onnodig door haar focus op het eliciteren van intuïties, terwijl er reeds lang diverse goede voorbeelden van XF bestaan op het gebied van de kunstmatige intelligentie en de cognitieve neurowetenschap.

2. Drie vormen van filosofie

In de discussie over experimentele filosofie lijkt het me goed drie verschillende vormen van filosofische analyse te onderscheiden. Er zijn filosofen die niets van wetenschap willen of hoeven te weten. Andere filosofen gebruiken bij hun analyses de resultaten van wetenschappelijk onderzoek. Zij lezen en

schrijven erover. Tenslotte zijn er degenen die, ten behoeve van hun filosofisch werk, zelf wetenschappelijk onderzoek verrichten.

De eerste vorm, misschien het best te betitelen als 'pure' filosofie, heeft uiteraard het volste bestaansrecht. Zeker in verband met onderwerpen waarvoor geen relevante empirische resultaten beschikbaar zijn, zoals op het terrein van de metafysica, ethiek, of geschiedenis van de filosofie. Er zijn echter veel filosofische thema's waarvoor empirisch onderzoek relevant en beschikbaar is, bijvoorbeeld op het terrein van de filosofie van de geest (het lichaam-geest probleem, de mogelijk causale rol van bewuste wil, zie bijvoorbeeld (Wegner 2002)), epistemologie (welke rol spelen psychologische mechanismen in de fundering van epistemologische inzichten (Goldman 1986, 1993); wat is een verklaring, wanneer wordt die als normatief bevredigend geëvalueerd (Trout 2002, Bishop en Trout 2005)) of de ethiek (berusten morele oordelen op redeneren, emotie of op een combinatie van beide (Knobe 2003a, Casebeer 2003, Haidt, 2001)). En als er relevant empirisch onderzoek bestaat of te doen valt, waarom zou dat dan geen onderwerp of onderdeel van wijsgerige reflectie mogen zijn? Er is voor zover ik kan inzien geen reden waarom de filosofie zich intrinsiek afzijdig zou *willen* of zelfs *moeten* houden van empirisch onderzoek.

De tweede vorm, filosofische analyse die kennis neemt van wetenschappelijk werk, wordt door Prinz (2008) empirische filosofie (EF) genoemd. Deze vorm is, in ieder geval in de context van de filosofie van de geest, tamelijk gangbaar, denk bijvoorbeeld aan de invloed van wetenschappelijke observaties van *split-brain* patiënten (Sperry et al. 1963) op filosofische analyses over bewustzijn en identiteit. In het werk van filosofen als Fodor, Dennett, Churchland en Clark zijn besprekingen en analyses van cognitief-wetenschappelijk onderzoek zeer gebruikelijk.

Toch wordt empirische filosofie wel eens smalend weggezet als 'wetenschapsjournalisme' en worden empirische filosofen soms vergeleken met priesters die zich uiten over het huwelijk: ze weten niet echt waar het over gaat. Hoewel hier meer dan een kern van waarheid in zit, lijkt smalen me in zoverre onterecht dat goede empirische filosofie niet stil houdt bij louter beschrijving van de resultaten van het wetenschappelijk onderzoek van derden. Het kan uiterst productief zijn een aantal empirische artikelen in een breder kader te plaatsen en daar een theoretische analyse over op te bouwen door het analyseren van het gebruik (i.c. de verklarende of te verklaren rol) en betekenis van sleuteltermen, door het stellen van een diagnose over de huidige wetenschappelijke stand van zaken, hoe die gerelateerd is tot de vooronderstellingen van het besproken onderzoek, en/of over de mogelijke implicaties daarvan. Zie bijvoorbeeld een recent artikel van Jacob (2009) (zie ook Uithol et al. 2008) waarin hij zijn filosofische visie geeft op de empirische ontwikkelingen rondom 'spiegelneuronen'. Zo'n theoretische analyse kan toekomstig onderzoek sterk beïnvloeden (denk aan filosofische discussies over *modularity theory* (Fodor 1983), *frame problem* (Pylyshyn 1987), en *embodied embedded cognition* (Clark 1997) in de cognitiewetenschap) en zelfs leiden tot nieuwe

experimenten (Clark en Toribio 1994, van Rooij et al. 2002, Haselager et al. 2003). Kortom, empirische filosofie is meer dan een versimpelde weergave van wetenschappelijk onderzoek zoals bij wetenschapsjournalisme het geval is, of kan dat in elk geval zijn.

Prinz (2008) reserveert de term 'experimentele filosofie' (XF) voor het werk van degenen die zich daadwerkelijk zelf bezig houden (alleen, dan wel in samenwerking) met empirisch onderzoek. Hoewel experimenteren niet de eerste associatie is die men bij filosofie zal hebben lijkt de stelling dat XF *nieuw* is ietwat overdreven. In het verleden hebben filosofen nooit erg hun best gedaan om de filosofie met waterdichte schotten af te scheiden van empirisch onderzoek. Knobe & Nichols geven zelf aan dat er door de geschiedenis heen filosofen zijn geweest die gereflecteerd hebben op en soms ook deel genomen hebben aan empirisch onderzoek. Het combineren van filosofisch en empirisch onderzoek vormt een krachtig element in het werk van bijvoorbeeld Aristoteles en William James. Een belangrijk verschil tussen EF en XF is dat degenen die tot de laatste categorie behoren soms mogelijkheden zien voor filosofisch relevant empirisch onderzoek die door wetenschappers niet worden opgemerkt of als urgent worden beschouwd.

Opvallend is de nadruk die op dit moment in de XF wordt gelegd op de, ook door Knobe gehanteerde, praktijk van het afnemen van *vragenlijsten* om alledaagse intuïties (*folk intuitions*) beter in kaart te brengen. Dergelijke enquêtes kunnen gebruikt worden om filosofische intuïties (normaliter beschouwd als het product van het wijsgerig achterover hangen in de fameuze leunstoel) bij te stellen. De mogelijkheid van bijstelling is uiteraard niet gebaseerd op het principe dat 'de meest gedeelde intuïtie wint'. Een filosoof die goed heeft nagedacht over een bepaalde, bijvoorbeeld morele, kwestie is uiteraard niet gedwongen haar analyse en conclusies op te geven als na experimenteel filosofisch onderzoek blijkt dat de meerderheid van de bevolking andere intuïties of oordelen heeft. Onderzoek naar alledaagse morele opvattingen levert inzicht op in de psychologische mechanismen die intuïties mee helpen veroorzaken en dat inzicht kan van belang zijn bij de (filosofische) beoordeling van hun waarde (Knobe en Nichols 2008b: 6-8). Onderzoek naar intuïties lijkt me een nuttige, hoewel misschien vooral door filosofen soms overschatte bezigheid, maar er is al veel over gedebatteerd (zie bijvoorbeeld Nichols 2004, Kauppinen 2007, Nadelhoffer en Nahmias 2007, Schroeter 2008, Knobe en Nichols, 2008a) en ik zal er hier verder niet op ingaan. Belangrijker lijkt mij de vraag waarom filosofen zich alleen met intuïties bezig zouden mogen houden, en niet met meer structurele, werkelijkheidsgerelateerde, vragen. Waarom zouden de traditionele vragen van de filosofie, over de aard van de geest, over causaliteit, verklaring, moraliteit, e.d. louter via leunstoel of vragenlijsten verkregen intuïties beantwoord mogen worden en niet op basis van andere vormen van evidentie? Ik wil in het onderstaande me dan ook richten op andere vormen van XF, geïnspireerd door de Kunstmatige Intelligentie (KI) en de Cognitieve Neurowetenschap (CN).

3. XF en KI: een 60-jarig bestaan?

Zes jaar voordat een bijeenkomst te Dartmouth in 1956 het officieuze begin van de Kunstmatige Intelligentie zou vormen (Gardner 1984), hanteerde Turing in zijn artikel 'Computing machinery and intelligence' (1950) reeds een methode die we nu als een combinatie van KI en XF zouden kunnen kenschetsen. In plaats van uit te zoeken hoe mensen de term 'intelligentie' zouden definiëren stelde hij zijn roemruchte Turing Test (TT) voor. Turing's argument is in deze context de moeite waard om volledig te citeren:

I propose to consider the question "Can machines think?" This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. (Turing, 1950: 433)

In plaats van een conceptueel onderzoek (middels enquête) over de betekenis van de sleutelwoorden 'machine' en 'denken', stelde Turing een empirische test voor. Zijn idee was om onder goed gecontroleerde omstandigheden vast te stellen of mensen op basis van nauwkeurig gereguleerde interacties (vraag en antwoord sessies) een computationeel systeem kan worden onderscheiden van een normaal intelligent mens. Hoewel ik de TT als een vroeg voorbeeld van door KI geïnspireerde XF wil beschouwen, zijn er argumenten om dat niet te doen. Mede omdat ze relevant zijn voor de discussie over XF in het algemeen, wil ik er een paar bespreken.

Een tegenargument zou kunnen zijn dat de TT een beroep doet op menselijke intuïties en dus feitelijk een gedachte-experiment is. Uiteraard zullen mensen hun intuïtie gebruiken tijdens deze test (bijvoorbeeld bij het bedenken van vragen, en het beoordelen van de antwoorden). Maar het essentiële verschil met de enquêtenadering bestaat eruit dat menselijke beoordelaars moeten interageren met een daadwerkelijk bestaand computationeel systeem, in plaats van nadenken over hoe ze termen als 'intelligentie' zouden willen omschrijven, of hoe een mogelijke interactie met een hypothetisch systeem zou kunnen verlopen. Bovendien moet de uitkomst gemeten worden en kan het resultaat specifieke hypothesen (bijvoorbeeld over hoe menselijke talige interactie het best gemodelleerd zou kunnen worden) ondersteunen of ontkrachten. Dit verschil wordt goed geïllustreerd door de Turing test te contrasteren met Searle's (1980) beruchte 'Chinese kamer' gedachte-experiment, waarbij het voor iedereen duidelijk is dat dit een praktisch onmogelijk uitvoerbare test behelste. Turing deed geen gedachte-experiment, maar stelde een concrete, empirische test voor, compleet met voorspellingen van een mogelijk resultaat rond het jaar 2000.

Een ander tegenargument is de stellingname dat computationele modellering geen vorm van empirisch onderzoek is. Zo zegt Dennett (1978: 117):

AI programs are not empirical experiments but thought-experiments prosthetically regulated by computers. Some AI people have recently become fond of describing their discipline as “experimental epistemology”. This unfortunate term should make a philosopher’s blood boil, but if AI called itself thought-experimental epistemology (or even better: Gedanken-experimental epistemology) philosophers ought to be reassured.

Later heeft Dennett nog expliciet toegelicht waarom hij computer modellen als gedachte-experimenten zou willen beschouwen:

Most AI projects are explorations of ways things might be done, and as such are more like thought experiments than empirical experiments. They differ from philosophical thought experiments not primarily in their content, but in their methodology: they replace some – not all – of the “intuitive”, “plausible” hand-waving background assumptions of philosophical thought experiments by constraints dictated by the demand that the model be made to run on the computer.” (Dennett 1998: 271, zie ook 2003: 128).

Dennett’s standpunt lijkt me niet sterk. Het is juist de stringentere methodologie die een groot verschil maakt tussen gedachte- en echte experimenten. Iedereen die wel eens modellerwerk verricht heeft weet dat de computationele werkelijkheid weerbarstig is en dat falsificaties van voorspellingen over het gedrag van het model regelmatig voor komen. Daarbij komt dat een op psychologische (of biologische, etc.) plausibiliteit gerichte AI juist niet probeert te modelleren ‘hoe het zou kunnen’, maar ‘hoe het is’ (bij mens of dier). De systematische vergelijking tussen de resultaten van computer simulaties en gedragsobservaties (al dan niet tijdens experimenten) garandeert dat er van de vrijblijvendheid van gedachte-experimenten weinig over blijft. Hieruit volgt uiteraard niet dat computer modellering in alle opzichten overeen komt met traditionele experimentering. Zo wijst bijvoorbeeld Guala (2005) op een interessant verschil. In de systemen waarmee geëxperimenteerd wordt spelen dezelfde causale oorzaken een rol als in de systemen die men tracht te begrijpen. Bij computer simulaties is daar geen sprake van en bestaat er slechts een formele correspondentie tussen het model en het gemodelleerde. Parker (2009: 486-487) benadrukt dat in essentie een model een representatieve relatie heeft tot het gemodelleerde, terwijl het bij experimenten gaat om een interventie in het bestudeerde systeem om de reacties van het systeem daarop in kaart te brengen. Net als Morgan (2003) volgt hier volgens Parker echter niet uit dat experimenten en modellen tot elkaar wederzijds uitsluitende categorieën behoren. Parker geeft aan dat het ook bij computer simulaties mogelijk is te interveniëren, bijvoorbeeld door een bepaalde aanvangstoestand te creëren om te bezien hoe het systeem zich van daaruit ontwikkelt. Een andere, in de praktijk veel voorkomende, methode is het computationele systeem te perturberen middels specifieke input (die vaak ‘natuurlijk’ kan zijn in de zin van rechtstreeks uit de werkelijkheid afkomstig, bijvoorbeeld door het systeem te voorzien van sensoren, ofwel daar sterk mee correlerend) om de reacties van het systeem te peilen. Deze vorm van simulatie is experimenteel in die zin dat er sprake is van materiële interventies in een model dat in relevante opzichten correspondeert met een materieel systeem. Ook voor Morgan is de mogelijk-

heid van een dergelijke (vooral natuurlijke) interventie in een computationeel model cruciaal, en zij spreekt dan van 'hybride experimenten' die tussen de twee uitersten van pure experimentatie enerzijds en zuiver mathematische modellering anderzijds gelokaliseerd kunnen worden.

Al met al is er dus behoorlijk wat speelruimte tussen experimenten en gedachte-experimenten, en mijns inziens voldoende om computermodellering toe te kunnen voegen aan het arsenaal van de experimentele filosoof. Verder is het opmerkelijk dat Dennett geen enkele reden geeft waarom filosofenbloed zou moeten gaan koken bij een label als 'experimentele epistemologie' (of, in de huidige context relevanter, 'experimentele filosofie van de geest'). Ik denk dat de XF een gezonde afkoeling van filosofenbloed inhoudt: filosofische experimentatie *an sich* is niets om je over op te winden.

4. Wijsgerig computationeel modelleren

Sinds Turing's artikel hebben diverse concrete computationele modellen een rol gespeeld in de discussie over filosofische problemen. Een van de duidelijkste voorbeelden is wellicht de rol van connectionistische modellen (kunstmatige neurale netwerken) in de discussie over eliminatief materialisme (EM) (zie Feyerabend 1963, Churchland 1989). Kort gezegd komt EM erop neer dat het lichaam-geest probleem opgelost kan worden door te onderkennen dat onze traditionele interpretatie van onze beweegredenen, i.c. meningen en verlangens (propositionele attitudes) die ons gedrag veroorzaken, fundamenteel onjuist is. Deze alledaagse taxonomie van mentale toestanden dient vervangen te worden door een betere, op de computationele neurowetenschap (een combinatie van connectionistische modellen in de KI en inzichten uit de neurowetenschap) gebaseerde taxonomie. Een dergelijk wetenschappelijk gefundeerd zelfbeeld zal, zo claimt het EM, niet alleen een beter beeld geven van de oorzaken van ons alledaags gedrag, maar ook de menselijke zelfkennis verdiepen. Voor Churchland (1989) vormde het connectionisme, dat halverwege de jaren '80 opkwam (Rumelhart en McClelland 1986) een belangrijke empirische ondersteuning voor zijn EM.

Tot zover echter is er nog 'slechts' sprake van EF, niet van XF. Dit veranderde echter na de frontale aanval van Fodor en Pylyshyn (1988) op het connectionisme. Zij betoogden dat het connectionistische modellen zowel representatief als computationeel ongeschikt waren om als levensvatbaar alternatief voor de klassieke symbool-regel modellen (berustend op wat door Fodor (1975, 1987) omschreven is als een 'denktaal') te fungeren. Connectionistische modellen bieden geen verklaringsmogelijkheid voor cognitieve fenomenen zoals systematiciteit. Systematiciteit betreft het kenmerk dat het vermogen om bepaalde gedachten te kunnen hebben intrinsiek verbonden is met het vermogen om bepaalde andere gedachten te hebben. Wie de zin 'Jan houdt van Marie' kan begrijpen, moet ook in staat zijn de zin 'Marie houdt van Jan' te snappen. Connectionistische modellen, die niet een implementatie vormen van de klassieke symbool-regel modellen, ontberen de syntactisch gerelateerde constituenten alsmede de syntactische verwerkingsprocessen die voor de modellering van systematiciteit noodzakelijk zijn.

Als dit waar zou zijn, zou het connectionisme geen sterke steun vormen voor EM. Daardoor zou het EM als positie in het lichaam-geest debat aan kracht verliezen (Haselager en van Rappard 1998, Haselager 1999). Diverse pogingen werden ondernomen om aan te tonen dat connectionistische modellen wel degelijk in staat waren systematiciteit te vertonen. Onder hen, en dat is in de huidige context het belangrijkste, bevonden zich diverse filosofen, o.a. David Chalmers (1990, 1993) en Tim van Gelder (Niklasson en van Gelder 1994). Die discussie wordt tot op heden gevoerd op basis van zowel filosofische als empirische argumenten, vaak in combinatie (zie bijvoorbeeld Frank et al. 2009), en de modelontwikkeling komt voor veel onderzoekers voort uit de filosofische relevantie van de discussie. Het is belangrijk om op te merken dat het hier vaak niet alleen gaat om de in principe kwestie of neurale netwerken al dan niet tot systematiciteit in staat zijn, maar ook om de empirische vraag hoe natuurlijke systemen systematiciteit kunnen vertonen. Zo hebben Frank et al. (2008) zich uitdrukkelijk tot doel gesteld de input van hun neurale netwerk te baseren op de structuur van de externe wereld (waarin zich gebeurtenissen afspelen waarover uitspraken gedaan worden die het netwerk moet kunnen verwerken). Het model beoogt meer te doen dan alleen te demonstreren dat 'het kan', door tevens te laten zien hoe systematiciteit op een *psychologisch plausibele* manier tot stand kan komen, namelijk door internalisering van een extern aanwezige structuur.

Kortom, filosofen hebben zich bezig gehouden met computationele modellering om een empirisch geïnformeerde discussie te voeren over de aard van mentale representaties en computationele processen, een discussie die potentieel consequenties heeft voor theorieën over het lichaam-geest probleem. Een computationele vorm van XF bestaat, zo niet al sinds de jaren '50, dan toch zeker sinds de jaren '90.

5. Neurowetenschap, ethiek en de naturalistische drogredenering

Een reactie op XF zou eruit kunnen bestaan te wijzen op een mogelijk essentieel verschil tussen filosofie en wetenschap: Filosofie houdt zich bezig met zaken die zich principieel onttrekken aan empirische evidentie. Dat dit zo kan zijn staat buiten kijf. Dat het zo moet zijn lijkt mij zeer aanvechtbaar. Deze kwestie speelt op dit moment in de ethiek in verband met de cognitieve neurowetenschap en de morele psychologie. In toenemende mate zijn er pogingen om op neurowetenschappelijk onderzoek gebaseerde uitspraken te doen over ethiek (zie bijvoorbeeld Zahn et al. 2009, Takahashi et al. 2008). Zo probeert Casebeer (2003) te beargumenteren dat recent neurowetenschappelijk onderzoek een ondersteuning vormt voor de Aristoteliaanse deugden theorie. Zijn bewijsmateriaal is mijns inziens niet zo overtuigend (voornamelijk gebaseerd op het feit dat veel verschillende hersengebieden bij moreel oordelen betrokken zijn), maar dat is in de huidige context niet zo belangrijk. Essentiëler is de tegenwerping van Greene (2003: 847): "*how does it follow from the proposition that Aristotelian moral thought is more natural than Kant's or Mill's that Aristotle's is better?*" Dit raakt uiteraard aan de discussie over de naturalistische drogredenen, een term gebruikt door Moore (1903) om aan te geven dat ethische

eigenschappen niet in termen van natuurlijke eigenschappen gedefinieerd kunnen worden. Greenes tegenwerping baseert zich op Humes stelling (1739, III, I, i, p.469) dat hoe iets zou moeten zijn niet deductief kan worden afgeleid uit hoe zaken feitelijk zijn. In een redenering met een ethische conclusie lijkt, naast een feitelijke premisse, een ethische premisse noodzakelijk (Wilson et al. 2003). Zo kan bijvoorbeeld uit de premisse dat het martelen van mensen leed veroorzaakt niet geconcludeerd worden dat het martelen van mensen verkeerd is (ethische conclusie), tenzij er een ethische premisse wordt toegevoegd (het is verkeerd om lijden te veroorzaken). Net zo kan er uit feitelijke observaties over hersenprocessen niet afgeleid worden wat ethisch (in)correct gedrag zou zijn. Met name vanuit de evolutionaire psychologie is er echter op gewezen dat “*the facts of the world do have ethical implications, which the naturalistic fallacy was never intended to deny.*” (Wilson et al 2003: 670). Empirische gegevens kunnen meehelpen te verklaren waarom mensen bepaalde ethische premissen als uitgangspunt accepteren, bijvoorbeeld door de belangrijke rol van intuïties in morele oordelen te verduidelijken (Haidt 2001, Casebeer 2003). Een op neurowetenschappelijk en evolutionair-psychologisch onderzoek gebaseerd begrip van de factoren die van invloed zijn op het ontstaan en verdere verwerking van dergelijke intuïties kan van grote invloed zijn op de filosofische discussie over moraliteit. Dergelijke (descriptieve) kennis kan gebruikt worden in de discussie over normatieve ethiek, ook al kan ze niet deductief dwingend tot normatieve uitspraken leiden (Mesoudi en Danielson 2008: 229; zie ook Wilson et al. 2003, Greene 2003, Singer 2005). Dergelijk onderzoek kan dus de moeite waard zijn voor filosofen om zelf (al dan niet in samenwerking) ter hand te nemen.

6. Conclusie

De XF is een goede ontwikkeling en een nuttige correctie van een twintigste-eeuwse taalwending in de filosofie die een onnodige en contraproductieve kloof met de wetenschap heeft laten ontstaan. Dat wil niet zeggen dat pure en empirische vormen van filosofie niet eveneens belangrijk en op zichzelf gerechtvaardigd zijn. Ik heb betoogd dat de manier waarop de XF zich momenteel manifesteert wat te eenzijdig de nadruk legt op het via vragenlijsten e.d. systematisch verzamelen van intuïties. Hoewel zinvol, is dit zeker niet de enige vorm van XF. Ook is de XF niet nieuw, maar heeft het met name in de KI reeds stevige wortels. De KI en de cognitieve neurowetenschap bieden werkterreinen waarop empirisch geïnteresseerde filosofen zelf het onderzoek kunnen doen waar hun wijsgerige analyses om vragen.

Literatuur

- Bishop, M.A. en Trout, J.D. (2005) *Epistemology and the psychology of human judgment*. Oxford: Oxford University Press.
- Casebeer, W.D. (2003) Moral cognition and its neural constituents, *Nature Reviews Neuroscience* 4, pp. 841–847.
- Chalmers, D.J. (1990) Syntactic transformations on distributed representation, *Connection Science* 2, pp. 53-62.

- Chalmers, D.J. (1993) Connectionism and compositionality: why Fodor and Pylyshyn were wrong, *Philosophical Psychology*, 6(3), 305-319.
- Churchland, P.M. (1989) Eliminative materialism and the propositional attitudes, in P.M. Churchland (red.) *A neurocomputational perspective*. Cambridge, MA: MIT Press.
- Clark, A. (1997) *Being there: Putting mind, body and world together again*. Cambridge, MA: MIT Press.
- Clark, A. en Toribio, J. (1994) Doing without representing?, *Synthese* 101, pp. 401-431.
- Dennett, D. (1978) *Brainstorms*. Cambridge, MA: MIT Press.
- Dennett, D. (1998) *Brainchildren: Essays on designing minds*. Cambridge, MA: MIT Press.
- Dennett, D. (2003) *Freedom evolves*. New York: Viking.
- Feyerabend, P. (1963) Mental Events and the Brain, *Journal of Philosophy* 40, pp. 295-6.
- Fodor, J.A. (1983) *The modularity of mind*. Cambridge, MA: MIT Press.
- Fodor, J.A. (1987) *Psychosemantics*. Cambridge, MA: MIT Press.
- Fodor, J.A., & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: A critical analysis, *Cognition* 28, pp. 3-71.
- Frank, S., van Rooij, I. en Haselager, W.F.G. (2009) Connectionist semantic systematicity, *Cognition* 110, pp. 358-379.
- Gardner, H. (1985) *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.
- Goldman, A.I. (1986) *Epistemology and Cognition*. Harvard University Press.
- Goldman, A.I. (1993) *Philosophical Applications of Cognitive Science*. Westview Press.
- Greene, J. (2003) From neural 'is' to moral 'ought': what are the moral implications of neuroscientific moral psychology?, *Nature Reviews Neuroscience* 4, pp. 846-850.
- Guala, F. (2005) *The methodology of experimental economics*. Cambridge: Cambridge University Press.
- Haidt, J. (2001) The emotional dog and its rational tail: a social intuitionist approach to moral judgment, *Psychological Review* 108, pp. 814-834.
- Haselager, W.F.G. (1997) *Cognitive science and folk psychology: The right frame of mind*. London: Sage.
- Haselager, W.F.G. (1999) On the potential of non-classical constituency, *Acta Analytica* 22, pp. 23-42.
- Haselager, W.F.G. en Van Rappard, J.F.H. (1998) Connectionism, systematicity, and the frame problem, *Minds and Machines* 8, pp. 161-179.
- Haselager, W.F.G., Bongers, R.M. en van Rooij, I. (2003) Cognitive science, representations and dynamical systems theory, in W. Tschacher en J-P. Dauwalder (red.) *The dynamical systems approach to cognition*. Singapore: World Scientific, pp. 229-242.
- Hume, D. (1739/1978) *A Treatise of Human Nature* (2nd Selby-Bigge/Nidditch edition). Oxford: Oxford University Press.
- <http://pantheon.yale.edu/~jk762/ExperimentalPhilosophy.html>, accessed at 24-7-2009.
- http://experimentalphilosophy.typepad.com/experimental_philosophy/, accessed at 15-7-2009.
- <http://www.youtube.com/user/xphilosophy>, accessed at 15-7-2009.
- Jacob, P. (2009) A philosopher's reflections on the discovery of mirror neurons, *Topics in Cognitive Science* 1, pp. 570-595.
- Kauppinen, A. (2007) The rise and fall of experimental philosophy, *Philosophical Explorations* 10, pp. 95 - 118.
- Knobe, J. (2003a) Intentional action and side effects in ordinary language, *Analysis* 63, pp. 190-193.
- Knobe, J. (2003b) Intentional action in folk psychology: An experimental investigation. *Philosophical Psychology* 16, pp. 309-324.
- Knobe, J. (2006) The concept of intentional action: A case study in the uses of folk psychology, *Philosophical Studies* 130, pp. 203-231.
- Knobe, J. en Nichols, S. (red.) (2008a) *Experimental Philosophy*. Oxford: Oxford University Press.
- Knobe, J. en Nichols, S. (2008b) An experimental philosophy manifesto, in J. Knobe en S. Nichols (red.) *Experimental philosophy*. Oxford: Oxford University Press, pp. 3-14.
- Mesoudi, A. en Danielson, P. (2008) Ethics, evolution and culture, *Theory in Biosciences*, 127, pp. 229-240.
- Morgan, M. (2003) Experiments without material intervention: Model experiments, virtual experiments and virtually experiments, in H. Radder (red.) *The philosophy of scientific experimentation*. Pittsburgh: University of Pittsburgh Press, pp. 216-235.

- Morgan, M. (2005) Experiments versus models: New phenomena, inference, and surprise, *Journal of Economic Methodology* 12, pp. 317-329.
- Moore, G.E. (1903) *Principia Ethica*. Cambridge: Cambridge University Press.
- Nadelhoffer, T. en Nahmias, E. (2007) The past and future of experimental philosophy, *Philosophical Explorations* 10, pp. 123-149.
- Nichols, S. (2004) Folk concepts and intuitions: from philosophy to cognitive science, *Trends in Cognitive Sciences* 8, pp. 514-518.
- Niklasson, L.F. en van Gelder, T. (1994) On being systematically connectionist, *Mind & Language* 9, pp. 288-302.
- Parker, W.S. (2009) Does matter really matter? Computer simulations, experiments, and materiality, *Synthese* 169, pp. 483-496.
- Prinz, J.J. (2008) Empirical philosophy and experimental philosophy, in J. Knobe en S. Nichols (red.) (2008a) *Experimental Philosophy*. Oxford: Oxford University Press.
- Pylyshyn, Z.W. (1987) *The robot's dilemma*. Norwood: Ablex Publishing Cooperation.
- Rumelhart, D.E. en McClelland, J.L. (red.) (1986) *Parallel distributed processing: Explorations in the microstructure of cognition, vol.1: Foundations*. Cambridge, MA: MIT Press.
- Schroeter, F. (2008) Experimental philosophers, conceptual analysts, and the rest of us, *Philosophical Explorations* 11, pp. 143-149.
- Singer, P. (2005) Ethics and intuitions, *The Journal of Ethics* 9, pp.331-352.
- Sperry, R.W., Gazzaniga, M.S. en Bogen, J.E. (1969) Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection, in: P. J. Vinken & G.W. Bruyn (red.) *Handbook of clinical neurology*. Amsterdam: North-Holland Publishing Co. pp. 273-290.
- The New York Times (2007) The new new philosophy. <http://www.nytimes.com/2007/12/09/magazine/09wwln-idealab-t.html>. Accessed at 31-3-2009.
- Takahashi, H., Kato, M., Matsuura, M., Koeda, M., Yahata, N., Suhara, T. en Okubo, Y. (2008) Neural Correlates of Human Virtue Judgment, *Cerebral Cortex* 18, pp. 1886-1891.
- Turing, A. (1950) Computing machinery and intelligence, *Mind* 59, pp. 433-460.
- Trout, J.D. (2002) Scientific explanation and the sense of understanding, *Philosophy of Science*, 69, 212-233
- Uithol, S., Haselager, W.F.G. en Bekkering, H. (2008) When do we stop calling them mirror neurons?, *Proceedings of the 30th annual meeting of the Cognitive Science Society*. Washington, USA.
- van Rooij, I., Bongers, R.M. en Haselager, W.F.G. (2002) A non-representational approach to imagined action, *Cognitive Science* 26, pp. 345-375.
- Wegner, D. (2002) *The illusion of conscious will*. Cambridge, MA: MIT Press.
- Wilson, D.S., Dietrich, E. en Clark, A.B. (2003) On the inappropriate use of the naturalistic fallacy in evolutionary psychology, *Biology and Philosophy* 18, pp. 669-682.
- Zahn, R., Moll, J., Paiva, M., Garrido, G., Krueger, F., Huey, E.D. en Grafman, J. (2009) The neural basis of human social values: Evidence from functional MRI, *Cerebral Cortex* 19, pp. 276-83.