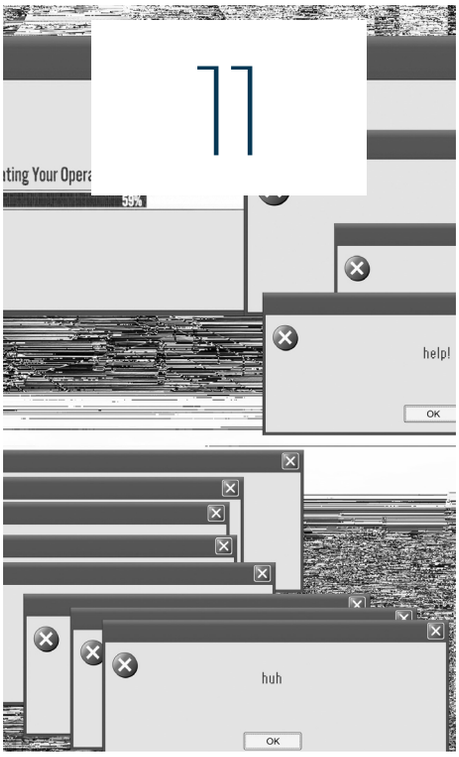
De Filosoof

Periodieke uitgave van de Faculteitsvereniging van Utrechtse Filosofiestudenten voor het Departement Filosofie en Religiewetenschap van de Universiteit Utrecht - Nummer 88

THEMA Zelfreferentie



Zelfreferentie in de AI
Jan Broersen

Renormalisatie
Willem van der Feltz

Connecting with animals
Evelyne Beuriot

Colofon

De Filosoof is een periodieke uitgave van de Faculteitsvereniging van Utrechtse Filosofiestudenten voor het departement Filosofie en Religiewetenschap van de Universiteit Utrecht

HOOFDREDACTIE
Sam Langelaan

REDACTIE
Sam Langelaan
Giulia Grosskop
Jeroen Verkade
Linde van Wingerden
Loes de Groen
Can Polat
Mette van Liempd
Julia Valstar
Victor Smit

VORMGEVING
Giulia Grosskop
Omslag door Giulia Grosskop

OPLAGE
500

MAAND VAN UITGAVE
Maart, 2023

ADRES
Janskerkof 13A,
4512 BL Utrecht

REDACTIE
redactie.filosoof@gmail.com

REDACTIE
http://fufexpluribusunum.nl/links/de-filosoof/

KOPIJ
De redactie behoudt zich het recht artikelen te wijzigen of in te korten

Geachte lezers,

Deze uitgave van *De Filosoof* heeft het fascinerende thema *zelfreferentie*. Zelfreferentie betekent dat een zaak naar zichzelf verwijst. Binnen de filosofie brengt dit vaak raadsels met zich mee. Een historische voorbeeld is de leugenaarsparadox: *"Alle Kretenzers liegen", zei de Kretenzer.* Doordat de spreker hier met zijn zin ook naar zichzelf en de waarheid van zijn uitspraken refereert, wordt het geheel een paradox. Zoals dit klassieke voorbeeld toont, bezit zelfreferentie iets raadselachtigs. Het heeft de merkwaardige eigenschap ogenschijnlijk onmogelijke cirkels te veroorzaken, puzzels van spiegels, waarin een beeld soms helder lijkt, soms juist vaag toeschijnt.

Als redactie hebben ervoor gekozen om dit thema primair te benaderen vanuit de analytische filosofie en vanuit een literair perspectief. Dit leek ons interessant vanwege het volgende. Aan de ene kant is zelfreferentie een concept dat bij uitstek in de analytische filosofie terugkomt. Het leek ons een perfecte kans om ruimte te geven aan de expertise binnen deze stroming die op ons departement aanwezig is. Maar door zelfreferentie ook vanuit een literair perspectief te benaderen hoopten wij ook een mysterieus, humoreus en persoonlijk aspect van zelfreferentie te vatten, dat de filosofie enkel aanraakt. Bovendien leek het ons ook een uitzonderlijke kans voor schrijvers om zich dit merkwaardige thema eigen te maken.

We zijn trots op het resultaat. Toegankelijke theoretische filosofie wordt afgewisseld met originele literaire creaties. We openen met **Albert Visser** die ons meeneemt met een speels taalfilosofisch experiment. Hierop volgt een verhaal over een schrijver die over een schrijver schrijft, geschreven door **Wijcher van Dijk.** Dan wordt het eerste gedicht uit de reeks *Loops* van **Linde van Wingerden** gepresenteerd. In deze reeks wordt op exceptionele wijze het analytische, komische en raadselachtige van zelfreferentie vermengd met het persoonlijke. Drie andere gedichten van deze reeks vindt u verspreid over het blad. **Prof.dr.ig. Jan Broersen** schijnt licht op dit thema vanuit de computerwetenschap en betrekt dit op kunstmatige intelligentie. Zodoende maant hij onze koortsdromen over superintelligentie tot rust. **Willem van der Felz** schrijft met een absurdistische tint over zelfreferentie, wat mooi opgevolgd wordt door de dialoog van **dr. Brandt van der Gaast** en **dr. Jesse Mulder**, die zowel inhoudelijk als humoristisch is. **Evelyne Beuriot** levert een bijdrage over de dierenethiek en **Cas van de Laar** heeft een poëtisch stuk geschreven waarin het "zelf" van zelfreferentie centraal staat. Alles wordt afgesloten door een venijnig stuk van onze columnist **Victor Smit** en een amusante strip van **Hannah Waayers.**

Tenslotte willen wij u er ook op attenderen dat onze vormgever **Giulia Grosskop** hard heeft gewerkt aan een fantastische nieuwe vormgeving om de leeservaring nog meer te verrijken. We hopen dat dit in combinatie met de uitzonderlijke eenheid van literair en analytisch werk u zal fascineren en inspireren.
Namens de redactie,

Sam Langelaan

COCOHUAY

DE NAAM DIE ZICHZELF NOEMT

Albert Vissers werk betreft de studie van theorieën met veel coderingsmogelijkheden, logica's voor bewijsbaarheid en interpreteerbaarheid, constructivisme, semantische paradoxen en dynamische semantiek. Tegenwoordig denkt hij na over zowel de filosofische als de technische kanten van de onvolledigheidsstellingen. Hij wilde altijd een galante amateur zijn in wetenschap en filosofie. Echter, als je er lang in zit, word je vanzelf wel specialist.

Wellicht verschaft reflectie over dit ultiem nutteloze object een goede ingang tot een aantal centrale vragen van de taalfilosofie.

Outline

Het moet zo rond 1978 geweest zijn dat ik Saul Kripke's *Outline of a Theory of Truth* las. Het artikel was een openbaring voor me. Kripke had een, voor mij zeer verrassende, nieuwe manier om Tarski's werk over waarheid te bekritiseren. Ook wist hij op prachtige wijze de theorie van inductieve definities in te zetten om een zeer elegante (maar wel slechts gedeeltelijke) oplossing van het probleem van de Leugenaarsparadox te geven. En dan waren er nog die voetnoten met allerhande filosofische inzichten . . . Kripke's artikel deed mij voor het eerst filosofisch nadenken over zelf-referentie. Kripke schrijft:

> Let 'Jack' be a name of the sentence 'Jack is short' and we have a sentence that says of itself that it is short. I can see nothing wrong with "direct" self-reference of this type. If 'Jack' is not already a name in the language, why can we not introduce it as a name of any entity we please? In particular, why can it not be a name of the (uninterpreted) finite sequence of marks 'Jack is short'. (Would it be permissible to call this sequence of marks "Harry", but not "Jack"? Surely prohibitions on naming are arbitrary here.) There is no vicious circle in our procedure, since we need not interpret the sequence of marks 'Jack is short' before we name it. Yet if we name it "Jack", it at once becomes meaningful and true. (Note that I am speaking of self-referential sentences, not self-referential propositions.) (Kripke 1975, 690-712)

In dit artikel bestudeert **Albert Visser** de enigmatische kwestie of er een naam kan bestaan die zichzelf benoemt.

Kripke's argument lijkt spijkerhard. De syntactische entiteit 'Jack is short', hoe we ook over zijn ontologische status denken, is iets stabiels dat onafhankelijk van onze 'naam-gevende' activiteit bestaat. Wat zou ons kunnen weerhouden die entiteit wat voor naam dan ook te geven? Toch begon ik, na een jaar of wat, problemen te zien met het argument.

Ik twijfel er niet aan dat er zoiets is als een vooraf bestaande Engelse zin 'Jack is short'. Ik geloof niet, zoals de Kripke van het citaat, dat het een *sequence of marks* is, maar een syntactische structuur met woorden. Dat scholastische verschil van mening is voor deze discussie niet relevant. Echter het is niet die vooraf bestaande zin die van zichzelf zegt dat hij kort is. De zin die iets zegt is de betekenisvolle zin. De betekenisvolle zin heeft de vooraf bestaande zin – een syntactisch object uit het Engels – als uiterlijke vorm, maar is niet identiek aan die vooraf bestaande zin.[1]

In dit stuk wil ik de discussie vereenvoudigen tot de vraag of een naam zichzelf benoemen kan. Daarmee hoeven we niet meer na te denken over de vraag wat nu precies een zin is. Natuurlijk lijkt niets nuttelozer dan een naam die zichzelf benoemt. Daar kun je verder toch niets mee zeggen? Zeker, maar wellicht verschaft reflectie over dit ultiem nutteloze object een goede ingang tot een aantal centrale vragen van de taalfilosofie. De naam die zich noemt is een wijsgerige *drosophila*.

2. De Naam en de Naam

Namen van namen zijn relatief zeldzaam. Een voorbeeld van zo'n naam van een naam is de naam van Godsnaam 'YHWH', te weten: het tetragrammaton. Die schaarsheid betekent natuurlijk niet dat er een filosofisch probleem is met namen van namen. Er is gewoon meestal geen behoefte die leidt tot invoering van zulk soort namen. Dat is anders in het geval van de Godsnaam.

Om de vraag of een naam zichzelf kan benoemen in focus te krijgen, moeten we wat meer helderheid hebben over wat een naam nu eigenlijk is. Helaas zijn er meerdere theoriën van namen op de markt. Ik ga in dit stuk gewoon mijn eigen theorie schetsen zonder te proberen die met argumenten te verdedigen. Daarna probeer ik de vraag te beantwoorden of een naam, zoals die er volgens mij uitziet, zichzelf kan benoemen. Mijn idee van de aard van de naam is beïnvloed door David Kaplan's leuke artikel *Words* (Zie [Kap90]) – met de kanttekening dat ik het in veel opzichten niet met David eens ben..

We moeten onderscheid maken tussen een naam als faciliteit die door onze taal geboden wordt en de individuele naam. ' Faciliteitnamen' kun je in naamwoordenboeken opzoeken. Die vertellen dan iets over oorsprong en betekenis. Bijvoorbeeld: er zijn twee homophone en homographische Nederlandse namen *Anne*. De eerste is een meisjesnaam die komt van het Latijnse *Anna*, dat op zijn beurt weer afgeleid is van het Hebreeuwse *Hannah*. De geassocïeerde betekenis is zoiets als *gunst* of *gratie*. De tweede is een Friese naam, verwant aan het Germaanse *Arne*. Het betekent *adelaar*. Natuurlijk hoeft niet iedereen met een faciliteitnaam benoemd te worden. Je kunt je kind ook een fantasienaam geven, zoals Maan, Skylar of Darth. Laten we faciliteitsnamen en fantasienamen in dit stuk '*labels*' noemen. Een label is dus nog geen naam van een specifiek individu.

Dan is er de individuele naam. Zo ken ik Jan Bergstra en Jan Broersen. Ik denk dat Bergstra's naam *Jan* niet dezelfde naam is als Broersen's naam *Jan*. Die namen zijn homoniemen, net als *bank*, waar je geld is, en *bank*, waar je je 's avonds op ontspant.[2] De individuele naam brengt label en benoemde bijeen maar is meer dan alleen maar label + benoemde.

3. Die Naam noemt Zich

Goed. Maar dan nu: kunnen we ons voorstellen dat een individuele naam zichzelf benoemt? Er lijkt geen principiëel probleem te zijn met het statische idee van een al bestaande naam die zich noemt. Maar hoe zou zo'n naam kunnen ontstaan? Om die vraag te beantwoorden stellen we ons een prototypische naamgeving voor. Iemand introduceert een naam voor iets: hierbij noem ik *dit zus*. Het lijkt er op dat we eerst het *dit* moeten hebben om het *zus* eraan te kunnen verbinden om zo de naam te krijgen die het *dit* en het *zus* verbindt.[3] Maar hoe kan de naam zelf het *dit* zijn als de naam pas met de act van naamgeving tot stand komt? De naam in ons scenario is slechts een toekomstige geïntendeerde entiteit. Het *zus*, ofwel het label, daarentegen is Kripke's pre-existente ding. Met *zus* zit het dus wel goed.

Kunnen we een toekomstige entiteit benoemen? Een bekend voorbeeld gaat ruwweg als volgt. "We noemen de eerste persoon die in de 22e eeuw geboren wordt '*Apocatequil.*'" Volgens mij is dat geen legitieme act van naamgeving. Als de zaken zus lopen is Apocatequil deze persoon en als ze anders lopen die ander. We slagen er niet in één bepaalde persoon te benoemen. De expressie '*Apocatequil* ' functioneert daarom als beschrijving en niet als naam.

We benoemen dan datgene wat uit de pre-naam, die als 'naam-embryo' fungeert, gaat worden.

Toch denk ik dat er ook werkende voorbeelden zijn van het benoemen van toekomstige existenten. Stel je bijvoorbeeld een lopende band voor waarop zeer luxe en mooie auto's gemaakt worden, uniek handwerk en zo. De werkenden geven elke auto die ze maken, zodra de assemblage begint, een naam. Op een gegeven ogenblik moet de fabriek sluiten en stopt het assemblageproces terwijl net *Pitahaya* gemaakt wordt. Ik denk dat de werknemers erin slaagden om de toekomstige entiteit Pitahaya te benoemen. Zelfs nu de arme Pitahaya het bestaan onthouden wordt, kunnen ze zeggen: "Wat jammer dat Pitahaya niet afgemaakt kon worden. Ze zou zo mooi zijn geweest." Je kunt je zelfs voorstellen dat ze bij een reünie vele jaren later Pitahaya alsnog uit de oorspronkelijke onderdelen >>

maken. Het cruciale hier is dat er een intentioneel en min of meer deterministisch proces is dat naar Pitahaya leidt.

"Ik noem de naam die ik nu aan het produceren ben *Cocohuay*."[4] Dat zou toch warempel moeten kunnen lukken. Of niet dan? Mijn vaste voornemen de individuele naam te produceren verschaft het benodigde determinisme om die productie ook daadwerkelijk te volbrengen.

Nou ja, helemaal zeker ben ik er nu ook weer niet van. Ik blijf zitten met een knagend gevoel van ongefundeerdheid. Als het scenario leidt tot een naam, dan werkt het en, als niet, niet. Eén manier om de onvrede te benaderen gaat als volgt. Is het benoemde van een naam niet deel van het *materiaal* waar de naam van gemaakt is, net als de onderdelen op de lopende band materiaal zijn voor Pitahaya? Moet dat materiaal er niet al zijn voor we überhaupt iets kunnen maken? Er zijn twee manieren om naar deze tegenwerping te kijken. De eerste is te constateren dat het 'zitten' van de naam in de naam is geregeld via intentionele gerichtheid. Dat vraagt wellicht niet het soort simultane aanwezigheid als de motor in Pitahaya. Ten tweede zouden we het naamgevingsproces misschien ook zo kunnen zien. Onze intentie tot een definiete naam creëert al een flinterdunne pre-naam, die de naam gaat worden zodra het label en het benoemde ermee verbonden zijn. Dat object is zoiets als de gereïficeerde intentie. (Lijkt dit op Husserl's *noema*?) We benoemen dan datgene wat uit de pre-naam, die als 'naam-embryo' fungeert, gaat worden. De pre-naam vervult de rol van het materiaal.[5]

Dat Von-Münchhausen-gevoel van je aan je vlecht optrekken blijft me nog steeds bekruipen.

Dat Von-Münchhausen-gevoel van je aan je vlecht optrekken blijft me nog steeds bekruipen. Om mezelf over de streep te krijgen, heb ik, bij wijze van experiment, net *Cocohuay* daadwerkelijk gecreëerd als zichzelf noemende naam. Welkom, Cocohuay! En mooi dat je je zelf benoemt. Daar slagen er niet veel in. En sorry voor mijn twijfels aan jouw mogelijkheid.

En jullie, lezers, kunnen, via de intentionele link, Cocohuay nu ook gebruiken![6] Probeer het maar. Nota bene: het overnemen van Cocohuay is wel iets anders dan zelf het naamgevingsritueel uitvoeren en een eigen Cocohuay maken. Die Cocohuay zou alleen maar homoniem zijn aan de mijne, niet de identieke naam. Dus wel graag de intentionele link gebruiken. ■

Illustratie door Mette van Liempd

Eindnoten

1 Zoals wel vaker gebeurt, kan een filosofisch incorrect idee tot valide nieuwe technologie leiden. In dit geval zijn dat de zelf-referentiële Gödelnummeringen. Het idee van zulke Gödelnummeringen is afkomstig van Saul Kripke. Zie [Kri75, Voetnoot 6]. Ik heb later het idee uitgewerkt in [Vis89]. Saul heeft zijn voetnoot nader uitgewerkt in [Kri21]. Voor verdere discussie zie ook [GV20]. Balthasar Grabmayr geeft een leuke toepassing van zelfrerentiële Gödelnummeringen in [Gra21]. Ik ben ervan overtuigd dat er nog meer toepassingen mogelijk zijn omdat dezelfreferentiële Gödelnummeringen kunnen komen waar Gödel's Fixed Point Lemma niet kan gaan.

2 In feite geloof ik dat, in principe, dezelfde persoon twee namen zou kunnen

Referenties
[Gra21] B. Grabmayr. On the invariance of Gödel's second theorem with regard to numberings. *The Review of Symbolic Logic*, 14(1):51–84, 2021.
[GV20] B. Grabmayr and A. Visser. Self-reference upfront: a study of self-referential Gödel numberings. *The Review of Symbolic Logic*, pages 1–40, 2020.
[Kap90] D. Kaplan. Words. **Aristotelian Society**, Supp. 64:93–119, 1990.
[Kri75] S.A. Kripke. Outline of a Theory of Truth. **Journal of Philosophy**, 72:690–712, 1975. [Kri21] S.A. Kripke. Gödel's theorem and direct self-reference. **The Review of Symbolic Logic**, pages 1–5, 2021.
[Vis89] A. Visser. Semantics and the liar paradox. In D. Gabbay and F. Guenthner, editors, Handbook of Philosophical Logic, Topics in the Philosophy of Language, volume IV, pages 617–706. Reidel, Dordrecht, 1989.

hebben die allebei verbonden zijn met het zelfde label. Iemand zou twee keer Jan kunnen heten. Jesse Mulder wijst me op de filosoof McTaggart die voluit 'John McTaggart Ellis McTaggart' heette. Zie: https://en.wikipedia.org/wiki/J._M._E._McTaggart.

3 Ik bedoel hier niet mee te zeggen dat de naam niets meer is dan slechts het paar (dit,zus).

4 In het voorgestelde scenario is het naamgevingsproces deiktisch betrokken op de naam die uit dat zelfde proces gaat komen. We zouden ons een alternatief scenario kunnen voorstellen dat we een naam-die-zich-noemt maken, een ab initio zelf-reflexief proces.

5 Er is wellicht nog een derde weg. Die zou echter te ver voeren om hier te behandelen. Het is het idee dat de sociale ontologie aan een soort maximalisatieprincipe gehoorzaamt. Wat kan bestaan, dat bestaat ook. (In de logica is er veel studie van dat soort maximalisatieprincipes, maar niet van de mogelijke toepassing in de sociale ontologie.) Als het scenario leidt tot een naam, dan werkt het, als niet, niet. Precies, daarom werkt het wél op grond van maximalisatie.

6 We moeten gewoonlijk aanhalingstekens of een ander grafisch middel gebruiken om bijvoorbeeld de individuele naam 'Kant' van de filosoof zelf te onderscheiden. Hier hebben we zulke hulpmiddelen niet nodig! Cocohuay = 'Cocohuay'.

ANALOGIE

"U ziet een tor die kijkt naar een spiegel." Niet in de spiegel. Torren hebben volgens de schrijver niet zulke goede ogen maar hij weet het niet zeker. Het maakt de tor gelukkig niet uit wat de schrijver wel of niet denkt te weten. Het hoort/ruikt/proeft/voelt geen tegenovergestelde tor. De schrijver wil graag meer vertellen over deze tor. Deze tor is een keer uit de boom gevallen en heeft sindsdien een zielig pootje, midden-rechts. De tor prefereert dit pootje in rust een klein beetje boven de grond te houden, voor de zekerheid. Voor het echte werk worden echter alle hens aan dek (bast, mossige verticale steenzijde, robuuste stengel) gezet, ook midden-rechts.
Proportioneel is de tor beter in klimmen dan de schrijver. In de absolute zin ook: de schrijver zegt de hele tijd weer eens te willen gaan boulderen maar plant dat vervolgens niet in. Zijn hoofd is slecht geïrrigeerd en ideeën zijn van zijn buien afhankelijk. Wat de schrijver wel doet, zelfs ongepland, is radeloos in de spiegel staren tot hij besluit dat op bed wakker liggen beter is dan niks, hoewel niks ook aanlokkelijk klinkt.
Goed, de kever heeft daar geen last van. Of de tor. Is er een verschil? Een mier dan. Mieren hebben tenminste vrienden. Een mier krijgt geen rugpijn als deze eindelijk toekomt aan dat hoopje hemden opvouwen dat al een maand lang afwisselend op een mierenbed en mierenbureaustoel heeft gelegen.

Oké kijk, het ding is, insecten zijn klein en nietig en mensen kunnen daar een mooie duistere les over leren en dan een beetje diepzinnig fronsen en "hmmm" zeggen en denken "wauw! die schrijver heeft het volgens mij heel zwaar maar op een coole manier. Misschien kan ik hem een keer uitnodigen voor mijn verjaardag?" Als deze mensen dat zouden doen dan zou de schrijver waarschijnlijk komen met een fles net-niet-goedkope port en een zak pittige nootjes. Houdt u van port? Het maakt niet uit, u zult het niet proeven. U zal zijn woorden niet lezen. Hij zal u niet vervelen in het kringgesprek. Hoe dan ook, een rups is niet in staat om over koetjes en kalfjes te praten want die weet niet wat dat zijn. Ze zijn te groot om waar te nemen voor de rups. Misschien is dat ook waarom niemand met de schrijver wil praten, omdat hij zo intimiderend en indrukwekkend is. De schrijver is een koe en de vlieg is zijn

Proportioneel is de tor beter in klimmen dan de schrijver. In absolute zin ook.

Illustratie door Mette van Liempd

therapeut die felicitaties om zijn hoofd zoemt wanneer de koe zijn tanden poetst. Zijn melk is inkt en zijn kalf papier. Het gras is de wereld en daar leven de insecten die hem zachtjes maar toch hoorbaar uitlachen in de rij voor de apotheek. Desondanks is de koe met aambeienzalf en al van de berg gekomen en is, zoals de bijen met hun honing, zijn zwarte melk overvloedig zat. Hij wil de grasproleten ermee bederven. Hij wil ze zijn druifjes toewerpen en wraakzuchtig lachen als de insecten ze opeten.
Maar niemand, behalve kleuters, drinkt inkt. Deze insecten zijn geheelonthouders en durven rotte druifjes niet eens te besnuffelen. Ze zijn allergisch voor bijen. De tor ziet niets in de spiegel. De koe, de schrijver, wordt geslagen om zijn zure zwarte melk weer zoet te schrikken. Hij loeit en niemand kijkt naar hem om en een vogeltje roept "poo-tee-weet?" Best wel zielig, die schrijver. Althans, dat vind ik. ■

Wijcher van Dijk, die voorheen filosofie studeerde aan de UU, is een vrolijke ambtenaar die niet van zeuren of port houdt. Hij refereert zo goed als nooit naar zichzelf in de derde persoon tenzij dat expliciet van hem gevraagd wordt.

LOOPS I

EEN POËTISCH ONDERZOEK NAAR ZELFREFERENTIE

WE SPELEN EEN SPEL

we spelen een spel

kiezertje

kiezertje is:
jij kiest een spel

alleen de spellen met een einde
je mag dus kiezen: vlaggenroven
je kiest niet: tikkertje

dan spelen we dat spel

ja

uh

doe niet zo moeilijk

nu heb ik er geen zin meer in

Wat voor spel?

Kiezertje?

Van alle spellen?

O.

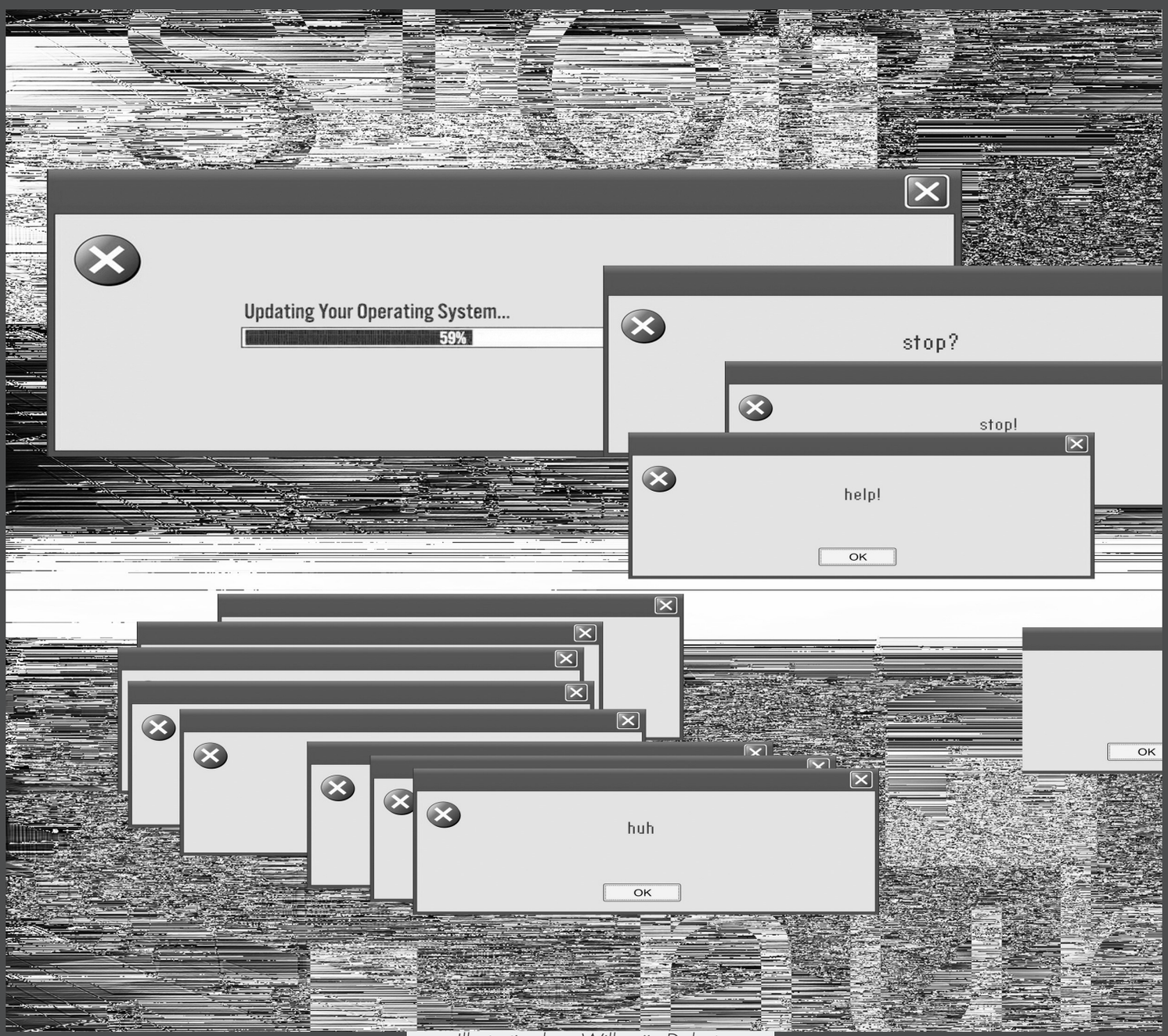
Kiezertje zelf heeft dus ook een einde.

Stel je voor,
we spelen kiezertje
en ik kies een spel.
Stel je voor, ik kies kiezertje,
want dat heeft een einde,
dan moet jij een spel kiezen
en dan kies jij ook kiezertje
en dan kies ik ook kiezertje.

Kies jij dan weer kiezertje,
dat spel met een einde?

Wat?

Redacteur **Linde van Wingerden** is student Filosofie en Wiskunde aan de *Universteit Utrecht*. Tot voor kort was die ook student aan de *HKU* bij de opleiding Writing for Performance.



Illustratie door Willemijn Debets

ZELFREFERENTIE IN DE AI

Als zelfreferentie een probleem vormt binnen de logica en de computerwetenschap, hoe vertaalt dit probleem zich dan naar de kunstmatige intelligentie (AI)? In dit artikel legt **Jan Broersen** uit wat het stop-probleem betekent voor computers en wat wij daar als filosofen van kunnen leren.

[het stop-probleem]

Velen kennen het probleem: je bent het operating system van je computer aan het updaten en krijgt spaarzame en onduidelijk informatie over hoelang het nog gaat duren voor het apparaat klaar is. Na een uur ga je je serieus afvragen of het installatieprogramma ooit nog zal stoppen. Wat te doen? De installatie afbreken komt met het risico van een half-afgemaakte installatie en het niet meer kunnen opstarten van je computer. Misschien raak je zelfs al je data kwijt, omdat je je harddisk opnieuw zult moeten formatteren. Maar afwachten is ook een onaantrekkelijke optie. Hoelang ga je nog wachten? Nog een uur? Een dag? Een week? Wat nu als je een dag hebt gewacht en je besluit morgen alsnog de installatie af te breken? Als je filosoof bent, dan schieten allerlei overwegingen rond diachronic rationality door je hoofd en probeer je een inschatting te maken van de beslissing die je "toekomstige ik" zal nemen, totdat je dan toch maar besluit dat het beter is to bite the bullet.

Wanneer je naast filosoof ook informaticus bent, dan vraag je je na verloop van tijd af waarom de softwaremaker eigenlijk geen hulpprogramma heeft bijgeleverd dat met zekerheid kan melden of je installatieprogramma zal gaan stoppen of niet. Dat moet toch niet heel moeilijk zijn? Een programma dat niet stopt is of vroegtijdig vastgelopen, of draait rond in een cirkel. En of dat gaat gebeuren, kun je wellicht met zekerheid voorspellen. Goed, heel eenvoudig zal het niet zijn. Je zult eerst het gehele installatieprogramma, mogelijk het serienummer van je computer en de relevante data op je harddisk als input moeten geven aan het hulpprogramma, zodat het alle informatie heeft om tot de conclusie te komen dat de combinatie (correct) stoppende is. En de berekening van het hulpprogramma zal een andere strategie moeten volgen dan het simpelweg runnen van het installatieprogramma, anders hebben we een situatie die je als filosoof zult karakteriseren als 'begging the question'. Het hulpprogramma zal dus naar structurele meta-kenmerken van het installatieprogramma moeten kijken, waardoor het zich een oordeel kan vormen over het ontstaan van zaken als cirkelberekeningen en voortijdige vastlooppunten. Ongetwijfeld erg ingewikkeld allemaal, maar niet onmogelijk?

> Als het niet stopt, dan zal het toch moeten stoppen om mede te kunnen delen dat het niet stopt? En als het wel stopt dan kan het toch nooit mededelen dat het niet stopt?

Maar dan realiseer je je iets onaangenaams. Wat nu als het hulpprogramma zelf niet stopt? Dan komt er geen antwoord op je belangrijke vraag over het installatieprogramma. Want hoe kunnen we nu weten of garanderen dat het hulpprogramma zelf stopt? Aha, denk je, wellicht heb je de reden gevonden waarom de softwarefabrikant zo'n hulpprogramma niet heeft meegeleverd. Deze heeft ongetwijfeld uitvoerig geprobeerd dergelijke hulpprogramma's te ontwikkelen, maar bij het testen bleek het probleem simpelweg te worden verschoven van installatieprogramma naar hulpprogramma.

Op dat moment krijg je een ingeving. Het voelt als een lichte vorm van kortsluiting in je hoofd, alsof je gedachten even stoppen. Als we zo'n 'stop-analyse-hulpprogramma' ontwikkelen, kunnen we het dan niet vragen om zichzelf te analyseren? Op dit punt aangekomen, moet je hopen dat je naast filosoof en informaticus ook logicus bent, want voor je ogen doemen enkele eigenaardige logische problemen op. Als het niet stopt, dan zal het toch moeten stoppen om mede te kunnen delen dat het niet stopt? En als het wel stopt, dan kan het toch nooit mededelen dat het niet stopt? Kan zo'n 'stop-analyse-hulpprogramma' dus eigenlijk wel bestaan?

> De functie die bij elke mogelijke combinatie van input en programma beantwoordt of de berekening stopt, kan blijkbaar niet door een computer worden berekend.

Wanneer je, zoals Alan Turing, ook nog eens wiskundige bent, dan kun je de bovengeschetste intuïties en observaties gebruiken als inspiratie voor een exact bewijs voor het niet kunnen bestaan van een programma dat van elk mogelijk programma berekent of het stopt. (Het bekendste formele bewijs is niet analoog aan de bovenstaande schets, maar is te maken als een minimale toevoeging aan het beroemde kruisbewijs van Cantor voor het bestaan van over-aftelbaar oneindige verzamelingen.) En dan hebben we een verrassend resultaat: de functie die bij elke mogelijke combinatie van input en programma beantwoordt of de berekening stopt, kan blijkbaar niet door een computer worden berekend. Daarmee hebben we een concrete beperking van computers aan het licht gebracht. En misschien is datgene wat we intelligentie noemen ook wel zo'n onberekenbare functie.

[consistentie]

Een AI-onderzoeker die gebruikmaakt van geautomatiseerde logische afleidingssystemen om AI te programmeren (bekend onder de naam GOFAI voor 'good old-fashioned AI') heeft nog een ander probleem dan het probleem dat hij nooit zeker kan weten of zijn algoritmen niet tot in de eeuwigheid blijven rekenen. Onderzoekers willen ook dat wanneer hun AI een afleiding maakt, die afleiding dan correct en veilig is. In het bijzonder wil een ontwerper van logische AI de garantie dat bepaalde desastreuze afleidingen nooit kunnen voorkomen. Het afleiden van een inconsistentie in een logisch systeem leidt er bijvoorbeeld toe dat daarna alles afleidbaar is. Het systeem wordt onbruikbaar en de kennis die het op logische wijze representeerde kan onmogelijk een correcte afspiegeling van de werkelijkheid zijn geweest. Maar hoe garanderen we nu dat een sterk en complex logisch afleidingssysteem een bepaalde formule niet kan afleiden? We kunnen natuurlijk op het idee komen om een tweede afleidingssysteem te maken dat afleidingen maakt die kunnen worden geïnterpreteerd als uitspraken over het andere systeem en dan in het bijzonder of in dat andere afleidingssyteem zaken niet afleidbaar zijn. De lezer voelt al aan dat dit dezelfde kant op gaat als in het verhaal hierboven, waarin de vraag werd opgeworpen of een programma van andere programma's kan berekenen of ze stoppen. Ook hier, in het verhaal over consistentie, dreigen we het probleem nu slechts te verschuiven naar een ander systeem, waarvan we dan zouden willen kunnen bewijzen

> En misschien is datgene wat we intelligentie noemen, ook wel zo'n onberekenbare functie.

of het consistent is. En ook hier kunnen we, om de opdoemende oneindige regressie het hoofd te bieden, dan de ingeving krijgen om een systeem uitspraken over zijn eigen afleidingen te laten afleiden en te vragen of het van zichzelf zou kunnen bewijzen dat het consistent is. U voelt het antwoord al. En Kurt Gödel, in zijn tweede onvolledigheidsstelling, bewees het formeel: dit kan niet. Een formeel afleidingssysteem dat sterk genoeg is om uitspraken over zichzelf af te leiden, kan helaas niet van zichzelf bewijzen dat het consistent is.

[zelfmodificerende programma's]

We hebben nu twee voorbeelden gezien die mogelijk een beperking vormen voor GOFAI: intelligentie is wellicht niet berekenbaar en wanneer we ervan uitgaan dat intelligentie zit verscholen in het kunnen maken van logische afleidingen, dan kunnen we niet garanderen dat die afleidingen consistent zijn. In beide gevallen is er een hoofdrol weggelegd voor zelfreferentie. En het is een ondermijnende rol: de zelfreferentie haalt een streep door het idee dat alles berekenbaar is, of te vangen is in logische formules. En dat gaat op dezelfde manier als waarop de bekende leugenaarszin "deze zin is onwaar" je ontredderd achterlaat bij het nadenken over het toekennen van waarheidswaarden aan betekenisvolle zinnen en de imperatieve zin "volg geen geboden op!" je diep in het duister laat tasten over wat je nu moet doen. Het is daarom wellicht verrassend dat er momenteel veel AI-onderzoekers zijn die voor zelfreferentie geen ondermijnende rol maar juist een versterkende rol zien weggelegd in de AI. En het is in mijn ogen niet toevallig dat dit vaak onderzoekers zijn die met op logica gebaseerde AI (symbolische AI, GOFAI) niet veel op hebben. Ik doel natuurlijk op de onderzoekers die vrezen dat AI in staat zal zijn zichzelf te herprogrammeren en zich op die manier zodanig zal kunnen versterken dat er serieus rekening moet worden gehouden met een situatie die uit de hand loopt: een AI die zijn intelligentie heeft versterkt zal daarna namelijk nog beter in staat zijn zichzelf verder te versterken. Een versterkingsproces dat zichzelf steeds verder versterkt zal divergeren, exploderen of imploderen (denk bijvoorbeeld aan zwarte gaten). In wiskundige beschrijvingen van dit soort fenomenen spreken we vaak van het optreden van een 'singulariteit'. De angst bij sommigen is dus dat zichzelf herprogammerende en versterkende AI zal leiden tot een intelligentie-singulariteit.

Ik wil hier toch proberen een flinke scheut koud water over dat idee te gooien. Daarbij denk ik terug aan pakweg 30 jaar geleden, toen ik als student Wiskunde & Informatica in Delft de PDP-11 van de Amerikaanse fabrikant DEC programmeerde in assemblertaal. Assemblertaal is niet heel ver verwijderd van het programmeren van instructietabellen van Turing-machines. En omdat je bij het programmeren in assembler zo dicht op Turings universele model van berekening zit, krijg je een goed gevoel voor wat rekenen is. Je hebt maximale vrijheid om te doen wat je wilt, omdat je direct de fysieke registers van de processor en de geheugenplaatsen van de opslag aanstuurt en manipuleert. Elke hogere programmeertaal zal uiteindelijk terugvertaald moeten worden naar deze elementaire machinetaal.

Nu realiseren velen zich waarschijnlijk niet dat computerprogramma's op heel veel niveaus werken volgens principes die je als zelfmodificerend kunt aanmerken. Een subroutine die zichzelf steeds opnieuw recursief aanroept doet dat telkens met een andere instantiatie van variabelen en 'modificeert' daarmee zijn werking. En dit type zelfreferentie kun je zo complex maken als je wil. Sommige talen laten inderdaad expliciet toe dat programmacode wordt geherinterpreteerd of zelfs herschreven door andere code. Maar daar wordt een programma natuurlijk niet begrijpelijker of leesbaarder op. Uiteindelijk moeten al deze zelfrefererende taalconstructies worden platgeslagen in een vertaling naar het machineniveau. En dat is het belangrijke punt. Alle programmataal-constructies die je bouwt boven op het machineniveau, inclusief de zelfrefererende, recursieve en 'zelfversterkende', zijn uiteindelijk alleen maar beperkingen die je jezelf oplegt. De beperkingen maken het handiger om te kunnen programmeren in termen van recursieve en zelfrefererende concepten die aansluiten bij je probleemdomein, maar het zijn geen versterkingen in de zin dat je er zaken mee kunt programmeren die je op het eenvoudige machineniveau niet zou kunnen doen. Hogere orde zelfrefererende programmaconstructies voegen dus uiteindelijk niets essentieels toe.

Het is daarom wellicht verrassend dat er momenteel veel AI-onderzoekers zijn die voor zelfreferentie geen ondermijnende rol maar juist een versterkende rol zien weggelegd in de AI.

Prof. dr. ir. Jan Broersen is hoogleraar logische methoden in de AI aan het departement Filosofie en Religiewetenschappen aan de *UU*. Hij onderzoekt hoe je met behulp van op logica gebaseerde methoden en modellering, vormen van Kunstmatige Intelligentie kunt ontwikkelen die zich verantwoordelijk gedragen, die controleerbaar zijn, en die in dienst staan van de mens.

[conclusie]

Onderzoekers die menen dat zelfrefererende programmaconstructies wel iets toevoegen, iets wat essentieel is voor intelligentie, maken misschien een begrijpelijke vergissing. Wanneer je al begint met denken dat er zoiets als een 'zelf' in een computer kan opstaan en je denkt dat dat kan worden gerealiseerd zonder iets aan het basisconcept van een computer (Turing-machine) te veranderen, dan kun je daarna denken dat zo'n 'zelf' vervolgens in staat zou moeten kunnen zijn zich te herprogrammeren en versterken. Maar in die redenering ben je dan dus al fout begonnen. Het lijkt mij duidelijk dat we helemaal niet mogen beginnen met de aanname dat er een 'zelf' in een computer kan opstaan. We zouden natuurlijk juist moeten beginnen met het proberen aannemelijk te maken dat dat überhaupt mogelijk is. Tot die tijd zijn alle vormen van zelfreferentie die we waarnemen en construeren slechts de schaduwen van ons eigen vermogen tot zelfreflectie—schaduwen die ons vooral veel vertellen over wat niet mogelijk is en minder over wat wel mogelijk is.

EINDE ■

LOOPS II

ONGEDIAGNOSTICEERDE MOEDERS EN NEURODIVERGENTE KINDEREN

Mijn moeder
tegenover mij
kijkt me aan met die blik

“Begrijp je het echt niet?”
“Nee”

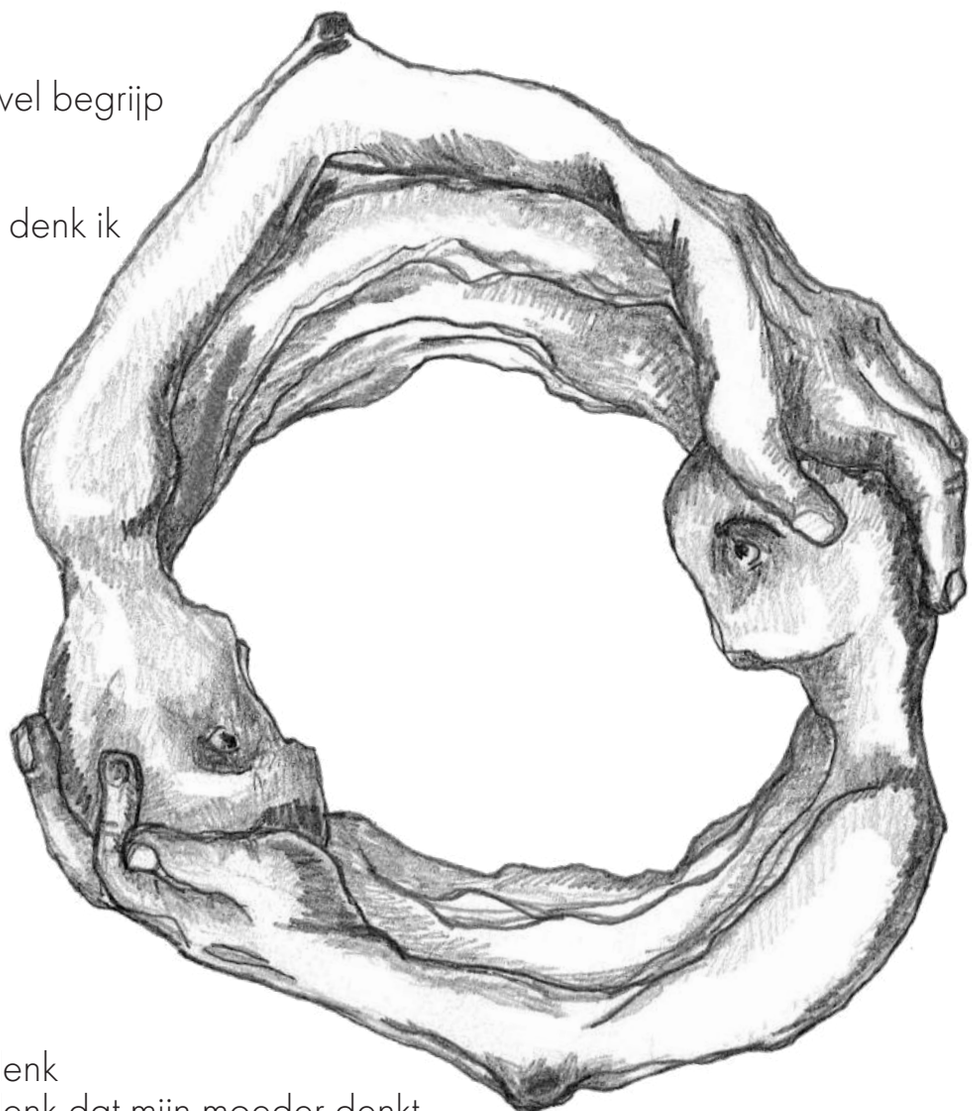
Mijn moeder denkt dat ik het wel begrijp
Ze zwijgt

Er valt niets meer uit te leggen, denk ik

Mijn moeder denkt
dat ze weet waar ik aan denk

Ik denk
dat mijn moeder denkt
dat ik niet aan haar denk

Mijn moeder denkt
dat ik denk
dat mijn moeder
denkt dat ik
denk
dat
mijn moeder denkt dat
ik
denk dat
mijn moeder denkt dat ik
denk
dat mijn moeder denkt dat ik denk
dat mijn moeder denkt dat ik denk dat mijn moeder denkt
dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat
ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder denkt dat ik denk dat mijn moeder



Illustratie door Willemijn Debets

RENORMALISATIE

In dit verhaal van Willem van der Feltz worstelt Kareltje met zijn writer's block. Het lijkt hem niet te gaan lukken om zijn deadline te halen, tot hij een geniaal plan verzint.

Ietwat onderuitgezakt en met zijn kin op zijn hand geleund keek Kareltje uit het raam. Eigenlijk keek hij niet uit het raam, maar naar de druppels die daarop hun lome dronkemansweg aflegden, tot zij een buur tegenkwamen en dan plotseling snel weggleden. Het computerscherm scheen in zijn ooghoek als een groot blauw-wit vlak, alleen onderbroken door het kleine hoopje zinnen dat hij kort tevoren had getypt. Kareltje had gehoopt dat de vele uren peinzen hem hadden gevuld met de juiste woorden en dat die eerste zinnen op papier als het ware een innerlijke dam zouden openen, waardoor al die woorden nu moeiteloos uit hem zouden stromen. In plaats daarvan vond hij de aanblik van zijn essay nu nog armetieriger dan toen het nog leeg was.
Waarom moet dit toch telkens weer zo ingewikkeld zijn, dacht hij. Dit is toch talloze malen eerder gelukt. Hij had wel succes gehad met zijn verhalen, maar die waren altijd terloops tot hem gekomen, zonder dat hij daar zo nadrukkelijk voor was gaan zitten. Nu was hem gevraagd iets te schrijven voor een tijdschrift en was er van spontaniteit geen sprake. Was hij met deze opdracht ingestemd om zichzelf uit te dagen? Voor de eer? Of was hij weer te laf geweest eens "nee" te zeggen?

Was hij met deze opdracht ingestemd om zichzelf uit te dagen? Voor de eer? Of was hij weer te laf geweest eens "nee" te zeggen?

Willem van der Feltz is opgegroeid in Maastricht, deed een bachelor Natuurkunde in Utrecht en een master Natuurkunde in Keulen, waar hij zich specialiseerde in kwantumzwaartekracht. Voor zijn verhalen haalt hij dan ook vaak inspiratie uit natuurkundige thema's. Verder is hij graag bezig met houtbewerking en meubelmaken.

Met dat soort gedachten dwaalde Kareltje steeds verder af, terwijl zijn hoofd steeds wat verder zakte en zijn kin rood werd van het lange drukken in zijn handpalm. In een vlaag van frustratie stond hij op om thee te gaan zetten, om dan ten minste iets te doen. In zijn keukentje vulde hij de waterkoker tot het onderste streepje, 0,5 liter, zette hem in zijn houder en schakelde hem aan. Het vlijtige ruisen van de waterkoker stelde hem wat gerust. Ik heb immers nog anderhalve week, dacht hij, en wanneer de tijd echt gaat dringen, dan kom ik altijd wel op dreef. Deze zogenaamd verspilde uren voegen alleen maar toe aan die opgehoopte mooie woorden, die met een beetje stress moeiteloos uit hem zouden vloeien.
Toen het water na ongeveer twee minuten kookte, nam hij de koker van de houder en goot de nog borrelende inhoud in een kopje waarin hij al een theezakje had gehangen. Terwijl hij het theezakje wat op en neer bewoog voelde hij zich toch weer wat somber worden. Hij had toch speciaal deze middag vrij genomen en kon hij dan zo slecht van zichzelf op aan? Was dat schrijven niet iets wat op een gegeven moment met het volwassen worden vanzelf zou moeten gaan? Maar terwijl hij onnadenkend de nog veel te warme thee naar zijn mond bracht, viel hem plotseling een geweldig plan in. Geschrokken door dit plotselinge inzicht nam hij een te grote slok en brandde zijn mond, maar hij was te opgewonden om erbij stil te staan.
Hij had besloten met Marietje te fuseren. Zij is altijd zo verstandig, heeft lol in haar werk en maakt haar dingen ruim op tijd af, dacht hij. Als wij nou een twee-eenheid zouden vormen, dan zou dat een schepsel creëren dat grootser is dan de som van haar delen. Kareltje rilde en zweette van opwinding over zijn eigen geniale ingeving en zijn thee vergetend zocht hij zijn telefoon om een afspraak met Marietje te maken. >>

Het geschiedde zoals hij zich had voorgenomen. Marietje was enthousiast over het plan en omdat Kareltjes deadline al dichtbij kwam besloten ze de volgende ochtend vroeg direct te fuseren. Zo ontstond de twee-eenheid, die meteen vol goede moed achter het bureau ging zitten. Alles leek, ondanks de nog korte samenwerking, verrassend harmonieus te gaan: de pc werd opgestart, Kareltje voerde de toegangscode in en Marietje downloadde een nieuwe tekstbewerker met minder afleidende knopjes en kleurtjes dan Word. Maar helaas: kort nadat het lege blad uitnodigend tegenover de twee oplichtte ontstond er onenigheid. Kareltje kwam met wat ideeën en schreef zelfs het een en ander op, maar steeds was Marietje het er niet mee eens en maakte dan weer schoon schip. Toegegeven, de ideeën waren wat flauw en de zinnen slordig, maar het was ten minste een opzet die zou kunnen rijpen tot een meesterwerkje, aldus Kareltje. Marietje vond dit een rommelige houding en was van mening dat een goed werk eerst in de grote lijnen moet worden uitgedacht: eerst denken, dan schrijven. Ook de herhaalde voorstellen van Kareltje om dan maar thee te gaan zetten "om even de geest te verzetten" vond ze niet professioneel. Toen de twee-eenheid na een uur nog niet echt verder was gekomen, werd er dan toch maar een knoop doorgehakt: de geest moest echt even worden verzet en thee zetten was ook weer geen ónprofessionele manier om dat te bereiken. De twee-eenheid vulde de waterkoker met 1 liter water, zette hem in zijn houder en schakelde hem aan. Het vlijtige geborrel gaf Kareltje nieuwe moed en om de samenwerking constructief een nieuwe dynamiek in te blazen begon hij enthousiast nogmaals zijn ideeën uiteen te zetten, maar nu in samenhangender volgorde. Marietje, die al lang door had wat die ideeën waren, luisterde met ergernis en ongeduld,

Kareltje rilde en zweette van opwinding over zijn eigen geniale ingeving en zijn thee vergetend zocht hij zijn telefoon om een afspraak met Marietje te maken.

maar hield zich beleefd koest. Dat gaf Kareltje nog meer moed en zelfs terwijl het gevaarlijk borrelende water werd ingeschonken en het zakje heen en weer werd gedoopt bleef hij maar doorratelen. Toen hij tussen twee zinnen door de nog veel te warme thee naar hun mond probeerde te brengen, kon Marietje zich niet meer beheersen. Met een ongecontroleerd ruwe schok rukte ze hun hand weer terug, zodat de bijna kokende thee over de rand van het kopje klotste en direct op de kousenvoeten van de twee-eenheid viel.
Luid vloekend en tierend (dat wel harmonieus en dubbel zo krachtig) hinkte de twee-eenheid door het appartement. Gelukkig was de twee-eenheid volwassen genoeg om, na enigszins bedaard te zijn, er even goed over te praten. Het was duidelijk, er was maar één verstandige oplossing mogelijk...
Wim, een gemeenschappelijke vriend van de twee-eenheid, was altijd rustig en evenwichtig, conflictvermijdend, maar toch doelgericht. Met hém moest de twee-eenheid fuseren. Gelukkig liet Wim zich makkelijk overhalen (ook Wim kon moeilijk "nee" zeggen) en de volgende dag was de drie-eenheid een feit. Wim voegde best wat toe aan het geheel: de conflicten waren een stuk minder fel en alles liep bedaarder en harmonieuzer.

Illustratie door Willem van der Feltz

Het continu dienen als stootkussen, of buffer, eiste echter zijn tol van Wim: hij werd doodvermoeid. Bovendien, als hij zich met éigen standpunten in de discussies wilde mengen, deed hij dit zo omslachtig en overdreven genuanceerd dat Marietje en Kareltje al snel niet meer naar hem luisterden. Zo kwam het dat Wim zicht steeds vaker tijdens discussies afzijdig hield en dan zelfs van slaap begon te knikkebollen. Kortom, Wim was nog niet genoeg.
We slaan nu in het verhaal wat tijd over: met betrekking tot het tijdschrift zelf waarin dit verhaaltje gepubliceerd wordt, is het verstandig er wat vaart in te brengen en het behapbaar te houden. U, beste lezer, begint nu vast de crux door te krijgen en hebt misschien ook andere dingen te doen.

Na tomeloos staren en eindeloos denken, stond de tienduizend-eenheid op om thee te gaan zetten. Het vulde de waterkoker tot het onderste streepje, 5000 liter.

In elk geval was daar, na enkele maanden, de tienduizend-eenheid. Ietwat onderuitgezakt, de kin in de handpalm rustend, staarde het voor zicht uit. Gedachten vlogen als wolken zonder randen door het hoofd, van hot naar her, tegen kant noch wal. Er waren weliswaar prangende zaken die steeds als speldenprikken in hun gedachten staken, maar het kwam er niet toe daar al te concreet over na te denken. Na tomeloos staren en eindeloos denken, stond de tienduizend-eenheid op om thee te gaan zetten, om dan ten minste iets te doen. Het vulde de waterkoker tot het onderste streepje, 5000 liter. Het vlijtig ruisen van de waterkoker gaf een vaag gevoel van tevredenheid. Toen het water na ongeveer 14 dagen kookte, goot de tienduizend-eenheid deze in een mok waarin al een theezakje hing. Het bracht de mok richting de mond en blies geduldig de damp van het nog veel te warme theewater, tevreden starend in de verte. Al die prangende zaken hadden immers geen haast en wanneer de nood écht aan de man kwam, zouden ze tegen die tijd zeker wel op dreef raken. ■

"DENK JIJ ECHT DAT ELK DENKEN DENKEN DENKT?"

Als je denkt, denk je dan altijd dat je denkt? Is dat een gedachte over denken hetzelfde als denken over een kat? In een emailcorrespondentie bevragen **dr. Brandt van der Gaast** en **dr. Jesse Mulder** elkaar over dit zelfreferentionele aspect van het begrip 'denken.

Illustratie door Sam Langelaan

Brandt van der Gaast is docent theoretische filosofie aan de Universiteit Utrecht. Na een studie filosofie aan de VU in Amsterdam, promoveerde hij in de taalfilosofie aan de University of Massachusetts Amherst. Brandt zit graag op de racefiets, waaruit volgt (via Griceaanse conversational implicature) dat hij dan ook een koffiesnob is. Zijn interesses liggen bij vraagstukken rond mentale representatie, zoals de vraag wat er nodig is voordat je een gedachte over een object kunt hebben.

Brandt: Ha, die Jesse! Alles goed? Ik moest zonet denken aan iets wat jij in een recent artikel schreef en waar ik je even over aan de tand wil voelen. Je schreef:

> "A realist considers the human being with her power for thought 'from sideways on', as an object that is 'out there' to be investigated, among various other objects. But, the idealists insist, this is a fantasy: there is no stepping outside thought; thought can be understood only 'from within'." (Mulder 2022, 472)

Iets later in de tekst bespreek je een voorbeeld over katten. Je schrijft daar:

> "[A]lthough I can think of cats without thinking of gold, I can think of neither without thereby already thinking of thought." (Mulder 2022, 482)

Dit is gerelateerd aan zelfreferentie. Jij stelt hier, volgens mij, dat denken altijd naar zichzelf refereert. Zelfreferentie zit ingebakken, zeg maar. Laat ik proberen tegen de bovenstaande claim in te gaan. Ik denk dat iemand over katten kan nadenken, zonder daarbij tegelijkertijd over denken na te denken. Het woord 'kat' drukt het concept KAT uit en ik ben geneigd bepaalde wezens in mijn omgeving als kat te classificeren. Als ik dit doe, dan denk ik kat-gedachtes (die normaal gesproken bewust zijn). Maar deze gedachtes gaan niet over het denken zelf.

Ik kan kat-gedachtes denken, maar deze gedachtes gaan niet over het denken zelf.

Jesse: Ha Brandt! Leuk dat mijn artikel je aan het denken heeft gezet. Jij stelt je 'denken over denken' voor, precies zoals je je 'denken over een kat' voorstelt. En dat is nou juist het probleem: die vooronderstelling gaat niet op. Om begrippen te kunnen hebben, moet je kunnen denken; kunnen denken betekent tot een gedachte kunnen komen; tot een gedachte kunnen komen betekent weten dat je denkt. Begrippen als 'begrip', 'gedachte', 'negatie', etc., zijn begrippen die al in je vermogen om te denken liggen, onafhankelijk van waar je in een gegeven geval over denkt.

Het woord 'gedachte' wordt te pas en te onpas gebruikt; daar schieten we filosofisch weinig mee op.

Brandt: Toch heb ik de indruk dat iemand gedachten over katten kan hebben zonder gedachten over denken te hebben. Waarom denk ik dit? Nou, kijk hoe we het woord 'kat' gebruiken en hoe we het woord 'gedachte' gebruiken. De term 'kat' gebruiken we voor dieren met een bepaalde verschijningsvorm. De term 'gedachte' gebruiken we voor interne staten van anderen, met bepaalde inhouden. Hoe weten we iets over de interne staten van anderen? Op basis van zichtbaar gedrag, en op basis van causale relaties met de omgeving, maar ook op basis van de aanname dat mensen rationele wezens zijn. We zoeken voortdurend naar de meest redelijke interpretatie van elkaar. Dit proces is extreem holistisch en daardoor ook extreem voorlopig (de meest rationele interpretatie die ik gister maakte, kan drastisch verschillen van die van vandaag). >>

Jesse Mulder is docent theoretische filosofie aan de *Universiteit Utrecht*. Eerder studeerde hij daar LAS en filosofie, deed aansluitend de RMA Philosophy en kon vervolgens in 2014 op ditzelfde instituut zijn promotie voltooien met een proefschrift getiteld *Conceptual Realism*, waarvan de centrale stelling luidde: essenties zijn begrippen. Zijn onderzoeksinteresse gaat vooral uit naar de combinatie van analytische filosofie (vooral metafysica) met inzichten uit het Duits Idealisme. Jesse woont met zijn vrouw en twee dochters in Driebergen.

Jesse: Mij interesseert "de wijze waarop we het woord 'gedachte' gebruiken" eigenlijk niet zozeer. Het woord wordt te pas en te onpas gebruikt; daar schieten we filosofisch weinig mee op. Ik baken een heel duidelijk onderwerp af: het komen tot een vaststelling, oordeel, uitspraak, statement, gedachte 'dat P', waarbij 'P' dus waar of onwaar is – dat is wat ik met onder 'denken' wil verstaan. (Hoe eventuele andere zogeheten 'mental states' zich daartoe verhouden is dan een vervolgvraag, die op dit punt afleidt van waar het om gaat.)

Brandt: Ik ben juist wél geïnteresseerd in hoe het woord 'denken' wordt gebruikt. Zoals ik net zei, schrijven mensen elkaar gedachtes toe op basis van gedrag en de impact van de omgeving. Folk psychology is een proto-theorie die hier iets over te zeggen heeft.

Tegelijkertijd moeten we ook oppassen dat we niet teveel op taal gaan leunen bij het nadenken over gedachtes, omdat uitspraken zoals 'S denkt dat P' nooit de hele lading van een gedachte kunnen dekken (vanwege de beperktheid van taal, maar ook contextuele zaken, zoals de kennis van de luisteraar).

Jesse: 'Denken' is geen theoretisch concept, maar 'denken' is het begrip waarin je je zelfbegrip als denker vat, en dat dus in elke gedachte al gegeven is. (Je past het natuurlijk ook toe op anderen etc., dat is dan weer een kwestie op zich.)

'Denken' is geen theoretisch concept, maar 'denken' is het begrip waarin je je zelfbegrip als denker vat.

Brandt: Eerder zei je dat iemand die denkt tegelijkertijd weet dat hij denkt en wat hij denkt. Ik ben het met je eens dat iemand die over X denkt normaal gesproken ook weet dat hij over X denkt. Ik beschouw dit als een soort niet-problematische oneindige regressie. Iemand die P denkt, denkt ook dat hij denkt 'dat P', en denkt ook dat hij denkt dat hij P denkt... Maar deze regressie geeft geen problemen. Het vergt niet een hersenpan van gigantische afmetingen.

Jesse: Dat is precies ook wat ik wil zeggen! Maar let op: dat kan natuurlijk alleen als je het begrip 'denken' hebt. Dus als wat jij hier schrijft klopt, dan kun je alleen P denken als je het begrip 'denken' hebt. Bovendien volgt hier ook uit dat dat begrip – 'denken' – in elke gedachte steekt, en dus precies op de wijze die ik eerder claimde verschilt van het begrip 'kat.' :)

Dat betekent niet dat je meteen ook het vocabulaire hebt om die zaken helder en duidelijk in uit te drukken. Het probleem is precies dat ons vocabulaire, wat we gewend zijn te zeggen, primair gericht is op de diverse onderwerpen waar we ons mee bezighouden – katten, goud – en niet op het denken zelf. Vandaar de neiging om van het denken zelf op dezelfde wijze een onderwerp te maken.

Brandt: Ik snap je kritiek op mijn positie. Aan de ene kant zeg ik dat iemand aan katten kan denken zonder daarbij tegelijkertijd aan denken te denken en aan de andere kant zeg ik dat iemand die X denkt normaalgesproken ook denkt dat hij X denkt... Zijn deze twee niet in strijd met elkaar? Toch wil ik aan beide claims vasthouden.

Wat mijn eigen ervaring betreft, lijkt het zo te zijn dat ik mezelf potentieel bewust kan maken van mijn huidige gedachtes, maar dat dat niet altijd gebeurt. Als ik op de automatische piloot door de stad fiets, ben ik me niet bewust van mijn denkproces. Als ik rondom Janskerkhof een vuilniswagen probeer te ontwijken, dan speelt in mijn hoofd niet de gedachte: ik ben nu aan het inschatten hoe snel ik moet fietsen om niet regelrecht in die laadklep te fietsen. Is dit geen voorbeeld waarin ik denk 'dat P', maar niet denk dat ik denk 'dat P'?

Jesse: Mijn punt is dat dat geen aanvullende gedachte is. Je kunt niet X denken zonder daarmee te weten dat je X denkt. Vergelijk 'assertie', het doen van een uitspraak met een waarheidswaarde. Als je de uitspraak doet dat X het geval is, dan weet je dat je dat aan het zeggen bent. Als je niet weet wat je zegt, dan is er ook geen assertie, hoogstens een inhoudsloos 'napraten' o.i.d. En hetzelfde geldt dus, meen ik, voor denken/oordelen.

Brandt: Betekent dit dat kinderen en dieren geen gedachtes kunnen hebben? Zij lijken het concept 'denken' nog niet paraat te hebben. Waarom zouden zij niet het soort gedachte kunnen hebben dat ik zonet beschreef: gedachtes die je gebruikt bij het navigeren van de omgeving waarin we ons bevinden? Dit doet me ook denken aan het onderscheid tussen 'animal knowledge' en 'reflective knowledge', ooit gemaakt door de epistemoloog Ernest Sosa.

Als ik rondom Janskerkhof een vuilniswagen probeer te ontwijken, dan speelt in mijn hoofd niet de gedachte: ik ben nu aan het inschatten hoe snel ik moet fietsen om niet regelrecht in die laadklep te fietsen.

Jesse: Inderdaad, mij gaat het om 'reflective knowledge' en niet om 'animal 'knowledge' – maar ik denk niet dat kinderen 'slechts' animal knowledge hebben. Ik twijfel er zelfs aan of dat überhaupt wel bestaat!

Brandt: Dan stel ik voor dat we het daar de volgende keer over gaan hebben! ■

Referenties
Mulder, Jesse M. 2022. 'Absolute Idealist Powers'. Australasian Journal of Philosophy 100 (3): 471–84.

CONNECTING W
ANALYZING K

In this essay **Evelyne** argues that the concept of empathy should be incorporated and analyzed when arguing about the value of animals relative to humans. She takes the theory of Korsgaard into account, who argues that there is no objective point from which you could argue that humans are more important than animals.

In this essay I will analyze Korsgaard's answer to the question 'Are humans more important than animals?' She tells us that because all value is tethered, there is no objective point from which you could argue that humans are more important than animals. The lives of humans and animals, creatures, are final goods, which means that they have value for the sake of themselves. It seems to me that it does not follow that humans should care about the intrinsic value of animals. I will try to solve this disconnection. I will argue that empathy between individuals and between species can be the solution and I end this essay by arguing that Korsgaard should have focused on incorporating and analyzing the concept of empathy, because it seems essential to the theory.

Animal Selves and the Good

In "Animal Selves and the Good", Korsgaard tries to answer the question of whether humans are more important than animals, because we do tend to have the intuition that we should save a person over an animal, or even that we can eat animals and use their skin for leather.

She argues that all value is tethered, which means that when something is good, it is always good to someone. If all that is good is good to some creature, then there is no point of view from which we could evaluate one creature to be more important than another in an absolute way (Korsgaard 2018: 1-5). She argues that animals have some sense of good at least because they can sense pain and pleasure and because they have a continuous self. For different animals it can range based on how much of a sense of self they have (Korsgaard 2018: 24-28).

But then, if all value is tethered, does this mean there is no absolute good? Korsgaard tries to avoid this outcome and says that there still are some things that are good to everyone and that "There is no difference between being absolute and being relative to everyone." (Korsgaard 2018: 6) What is good to every creature is, briefly put, maintaining itself. She explains that, while all value is tethered, there are two types of good, the functional and the final good. A knife can be a good knife, only in a functional way, because when it is sharper, it is easier to cut with and functions better (Korsgaard 2018: 8-18). A creature is a good creature when it keeps themselves functioning as the being that it is:

> "There is a kind of self-referential character to an organism's functioning, for its function is more or less to preserve a certain way of functioning, the way that is characteristic of its kind, and nothing more." (Korsgaard 2018: 14)

It is self-referential in the sense that the existence of a creature is not good because it is a means to something else.

This makes the lives of creatures a final good, Korsgaard says. This means that it is good for its own sake. It is self-referential in the sense that the existence of a creature is not good because it is a means to something else. So what would be good for all creatures is to keep living and functioning as the type of creature that they are (Korsgaard 2018: 8-18).

In this short paragraph I will explain what it means to live as the 'type of creature one is.' What is good for humans is what helps them function and what enables humans to create the conditions that keep them functioning. The same goes for any other creatures. This does not merely mean that the creatures function if they survive.

If a creature is a predator, it functions well when it hunts their prey. If a creature does not remember the pain it experiences at all, it would not be worse to inflict the pain twice rather than once.

She discusses the objection one might have that this type of teleology, there being a purpose for the existence of creatures, namely to keep existing, is based upon some religious fantasy. She responds that creatures with this type of purpose do not have to be designed by a sort of god. The theory of evolution does not imply purposelessness, but it shows us that creatures like this have been able to evolve. (Korsgaard 2018: 18-19)

WITH ANIMALS: KORSGAARD

Illustraties door Mette van Liempd

Final goods came into existence once creatures did, as they are the only beings that have value in themselves, for themselves, creating in their turn all functional value as what is functional is in the end functional to a creature(Korsgaard 2018: 8-18) This shows how all value is tethered. It can be tethered to humans, but also to animals and that makes that the point cannot be made that humans are objectively more important than animals.

The connection problem

I agree with Korsgaard's conclusion. If we recognize the fact that we can truly only judge and value from our own perspective, and that animals must have their own perspective which includes their own pains and pleasures that makes possible that there can be things good for them, then we must conclude that all value is tethered. There is no true objective position from which we can judge that humans are objectively more important.

But, I do not think that what follows is that we should treat animals accordingly. I think that what is being said in "Animal Selves and the Good" leaves open the possibility that it would be morally right to treat animals as if humans are more important than them, in the way that humans are more important to themselves as they are ends in themselves. This would make Korsgaard's ideas true in theory but unimportant in practice.

In her paper Korsgaard established that there is something good to any specific type of creature. But it is not established that what is important or valuable to animals, is also important to humans, or even that it is worth respecting. Korsgaard seems to imply that it is, that because all creatures are ends in themselves, they should be respected

as the type of being that they are. She tries to form a type of objective value by saying that what is good to all is the same as being objectively good.

But, from the fact that a creature is a final good in themselves, it does not follow that that creature is valuable to all others. There is a problem of disconnection between what is important to each creature individually. I can see that it is harmful to the pig to eat them, but from what Korsgaard is saying, it does not follow that I as a human being should respond to that harm.

Empathy

In this section I will argue that empathy should play a bigger role in Korsgaard's theory and that it even stands or falls with it because it can solve the disconnection. As it stands Korsgaard never explicitly mentions empathy as the deciding factor in the matter of valuing animal lives like our own. She does not appeal to the concept of empathy as a possible solution for the problem of disconnection.

How would one solve this problem of connection? If we are to take this problem seriously it becomes hard to explain how much we as humans care for each other at all. If all value is subject bound wouldn't we just care for our own interests? It is obvious that we don't and because it is obvious we should think about

But, from the fact that a creature is a final good in themselves, it does not follow that that creature is valuable to all others. There is a problem of disconnection between what is important to any type of creature individually.

what else could be at play when we engage in moral thinking.

When it comes to how we as humans should treat other humans, most answers come naturally and intuitively. I am lucky to have been born as myself, but I could imagine myself being born in a different situation, with perhaps abusive parents, in a totalitarian regime, or in a sexist cult. In virtue of being human I can imagine myself as being any other human, which makes the old idea of, 'don't do unto others what you would not do unto yourself' tangible. Korsgaard emphasizes the importance of the question, whether or not an animal has a continued sense of self. I think that an animal having a continued sense of self enables us to empathize with them.. It is hard to imagine being someone with no continued sense of self. It seems that the animals we value most are easiest to empathize with, be it because of nearness with our pets, or because we recognize intelligence in octopuses.

If a different species were to take over the position of the humans, by which these beings would have the same relation to us as humans now have to animals, I would want to be treated the way Korsgaard describes, in accordance with the type of being that I would be. I would want to receive the type of moral consideration that would fit the experiences that I would have as that being.

Korsgaard does mention empathy in the paper when she questions "By what standard are we evaluating a creature's life or her ends when we say that they are good, in the final sense of good?". She answers that the standard is deployed

from the standpoint of empathy. (Korsgaard 2018: 18) However she does not explicitly use it so solve the issue of disconnection.

She does not tell us why we should take this empathetic standpoint. She should have elaborated more on what it means to have empathy. In the way I described it, empathy connects our self-interest with the interest of the other, although it is not entirely clear what she means.

The only reason that a human should treat other creatures in the way that Korsgaard has in mind, by respecting the way that creatures functioning as that specific creature is maintained, is if it is part of human functioning to care about this. I propose that empathy is part of what makes us human.

If empathy is essential to human functioning, it wouldn't just be correct for humans to be empathetic and act empathetically

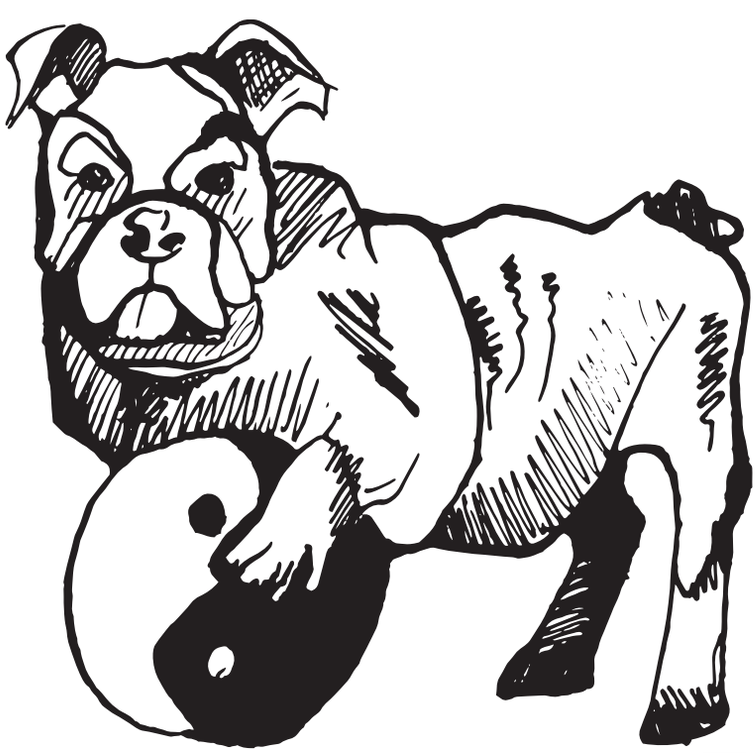
Empathy connects our self-interest with the interest of the other.

because it is our purpose given the type of creature that we are, but the characteristic of having empathy intrinsically means that what I, as a human, value, is connected to what others value. With empathy, protecting others is connected to helping myself. I can imagine myself being born as the other, and when I do that, their pain becomes my pain and their value becomes my value. The point of saying all this, is that for Korsgaard's theory to mean anything, there must be much more focus on analyzing the concept of empathy.

Conclusion

In this essay I have assessed Korsgaard's Animals Selves and the Good, in which she argues that all value is tethered and because of that there is no objective point from which you could argue that humans are more important than animals. The lives of humans and animals, creatures, are final goods, which means that they have value for the sake of themselves. I have argued that it does not follow that humans should care about the intrinsic value of animals. I tried to solve this disconnection. I argued that empathy between individuals and between species can be the solution and I ended this essay by arguing that Korsgaard should have focused on incorporating and analyzing the concept of empathy, because it seems essential to the theory. ■

Evelyne Beuriot is a third year BA philosophy student. She is interested in ethical and political philosophy, particularly the question of whether we can use ethical theory to give a certain meaning to our existence. Korsgaard's explanation of what value is in the first place, could be a step in the right direction towards answering this question.



Bibliography
Korsgaard, Christine M. Animal Selves and the Good. Oxford University Press, 2018.

STUKJES VAN MIJ

Cas van de Laar onderzoekt in deze poëtische tekst thema's als zelfidentificatie, opgroeien en verandering.

Er is een man, hij was vroeger een jongen, toen ik hem nog kende. Ik ken hem nu niet meer.
Als jongen had hij de meest bruine ogen, de meest bruine haren en de meest lieve moeder.
Zijn vader was vreemd, hij lag vaak te slapen. Hij leek op een pitbull en liet zijn dochter vaak schrikken.
Hun huis was erg bruin, zo bruin als zijn ogen. Er stond een trampoline vergeten in het gras.
In zijn buik zaten darmen, grote en kleine. En daar tussenin, een stukje van mij.
Het stukje kan niet erg groot zijn geweest. Of hoekig of warm, of juist heel erg koud.
Elk tienjarig kind had dan moeten huilen, maar huilen dat deed de toen-jongen nooit.
Toen, in het dorp, wist ik al niet hoe het daar kwam. Was ik het vergeten? Had hij het gejat?
Of had ik per ongeluk, misschien onvoorzichtig, dat stukje van mij zelf aan hem gegeven.
Hoe het ook zij, ik mis het nog steeds. Ik miste het toen al en nu des te meer.
Het was een kleuterig stukje, vol trekkers en vogels. Altijd besmeurd en ruikend naar gras.
Het stukje dacht vaak aan jongens uit groep zeven. Ze rookten vast veel, ze droegen Adidas.
Soms kwam het stukje bij mij op bezoek, als de jongen dat wilde en het mocht van zijn moeder.
Dan bouwden we blaaspijpen en met felgele tape bonden we er dan soms wel vier aan elkaar.
Het schoot er niet beter van maar toch moest het zo. Dat wisten de jongen, het stukje, en ik.
De jongen heb ik ergens vroeger voor de laatste keer gezien. Zijn moeder is dood nu, het hele dorp huilde.
En toch moet hij ergens dat stukje nog hebben, ik voel het nog kloppen. Heel erg ver weg.

Er is een muts, hij heeft drie felle kleuren. Ik heb hem ooit van een meisje gekregen.
Het meisje had piercings een buik en twee benen. Ze had de allergrootste tranen die je ooit hebt gezien.
Haar vader was lief, hij sprak over schrijvers. We snapten het half en zo was het prima.
Haar moeder was degene die de muts heeft gebreid. Als deze kapot gaat dan moet ik haar bellen.
Het meisje met benen heeft een stukje van mij, maar niet in haar buik. Het zit op haar schouder.
En als ik haar zie springt het zo naar de mijne, zo door de ruit. Dat kost een fortuin.
Het stukje is warm, te warm voor mijn schouder. Het is ook hoekig en groot, het zit echt heel naar.
Het stukje moest vaak, echt heel erg vaak, huilen. Maar durfde dat niet waardoor het nog meer moest huilen.
Het stukje vond allemaal moeilijke dingen. Over te dit, en te dat. Wat er moet en niet mag.
Dan stak het zijn been voor wat ik ook probeerde. Ik kon alleen nog maar struikelen als in een heel vreemde dans.
Soms struikelde zij mee en dansten we samen. Ik vond het verschrikkelijk, het voelde zo slecht.
In lelijke ritmes zo in te kleine kamers, vallend en troostend precies naast de maat.
Het stukje van mij at zo stukjes van haar, en stukjes van mij dus als ik te dichtbij kwam.
Vaak moet ik dan denken aan een heel groot geweer met een heel groot vizier en een heel erg klein stukje.
In mijn hoofd ligt het daar dan, dood op de tegels. Bloed op de stoep en hastalapasta.
Soms hoor ik het stukje nog zeuren en zagen. Ergens waar zij is, veel te dichtbij.

Illustratie door Mette van Liempd

Iemand schreef ooit dat vlammen één keertje branden. Dat voelde toen waar en werd een stukje van mij.
Een meisje met staartjes houdt haar hand op een deurknop. Ik vond dat zo lief dat het een stuk werd van mij.
Een rossige jongen, hij was mooier op foto's, ging over vijf dagen weg en vroeg wat ik dan wilde.
Een heel groot paars boek op de bovenste plank en een lelijke fles in 'gunmetal black'.
En nog heel erg veel stukjes waar ik niet meer naar zoek omdat ik te bang ben ze niet meer te vinden.
Er zijn ook stukjes van mij die ik maar heel erg soms zie, als ik heel erg lang wacht en heel erg diep adem.
Er is zo veel van mij dat het toch niet zou passen. Mijn buik is te klein en ik heb erg smalle handen.
Dus plak ik stukjes van mij op ruggen van vrienden, laat ze achter in struiken en voer ze aan eenden. ■

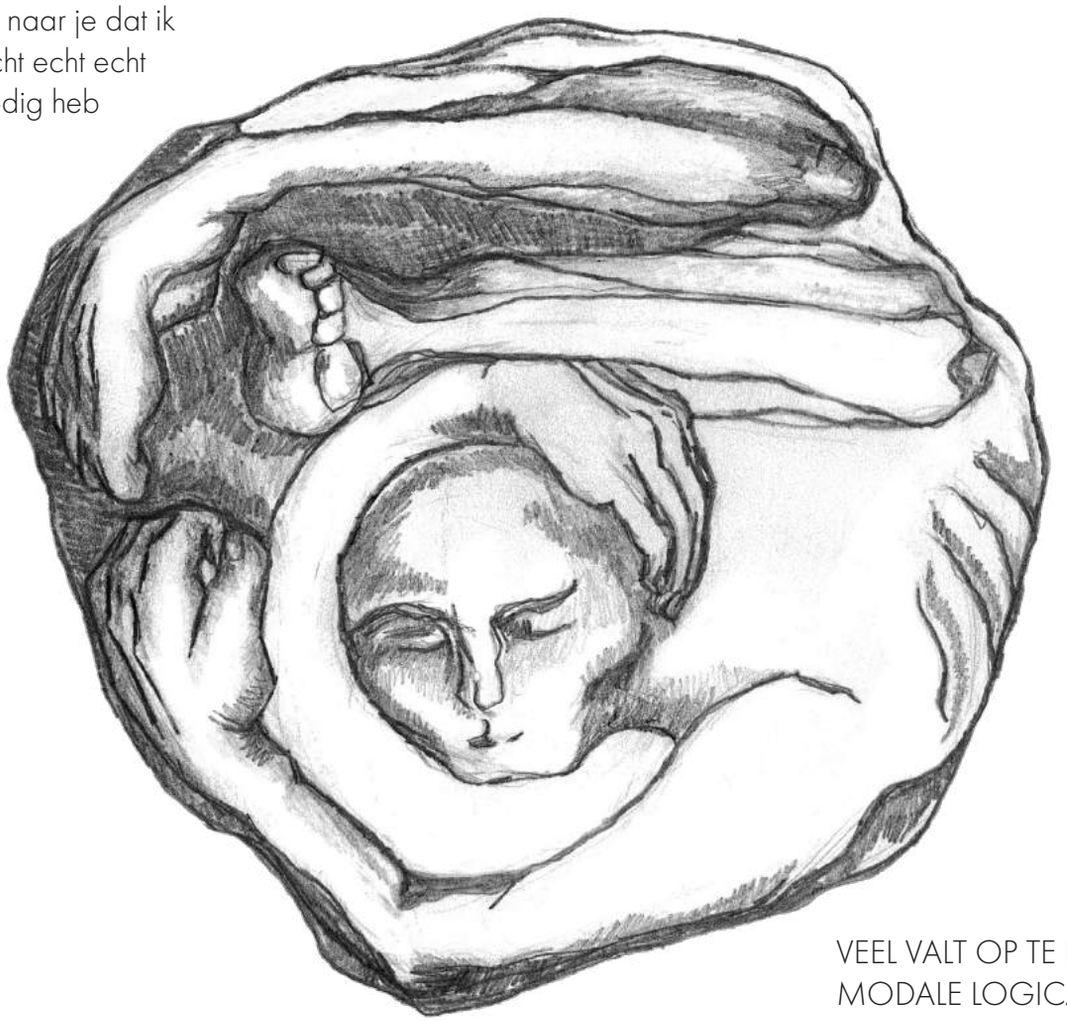
Cas van de Laar is filosoof en schrijft ter lering ende vermaak.

LOOPS III, IV

ZELFVOORZIENEND

Ik adem in de kom van mijn handen
Ik drink de borstvoeding uit mijn tepels
Ik lik het bloed van mijn wonden
Ik eet de plak van mijn tanden
Mijn plas staat in een tupperware in de koelkast

In een trui geweven van mijn zelf afgeknipte haar
sta ik bovenop een hoop huidschilfers,
schreeuw ik naar je dat ik
echt echt echt echt echt
niemand nodig heb



Illustratie door Willemijn Debets

VEEL VALT OP TE LOSSEN MET
MODALE LOGICA

In het eindeloze weefsel van
mogelijkheden
is er een wereld denkbaar waarin
ik pas

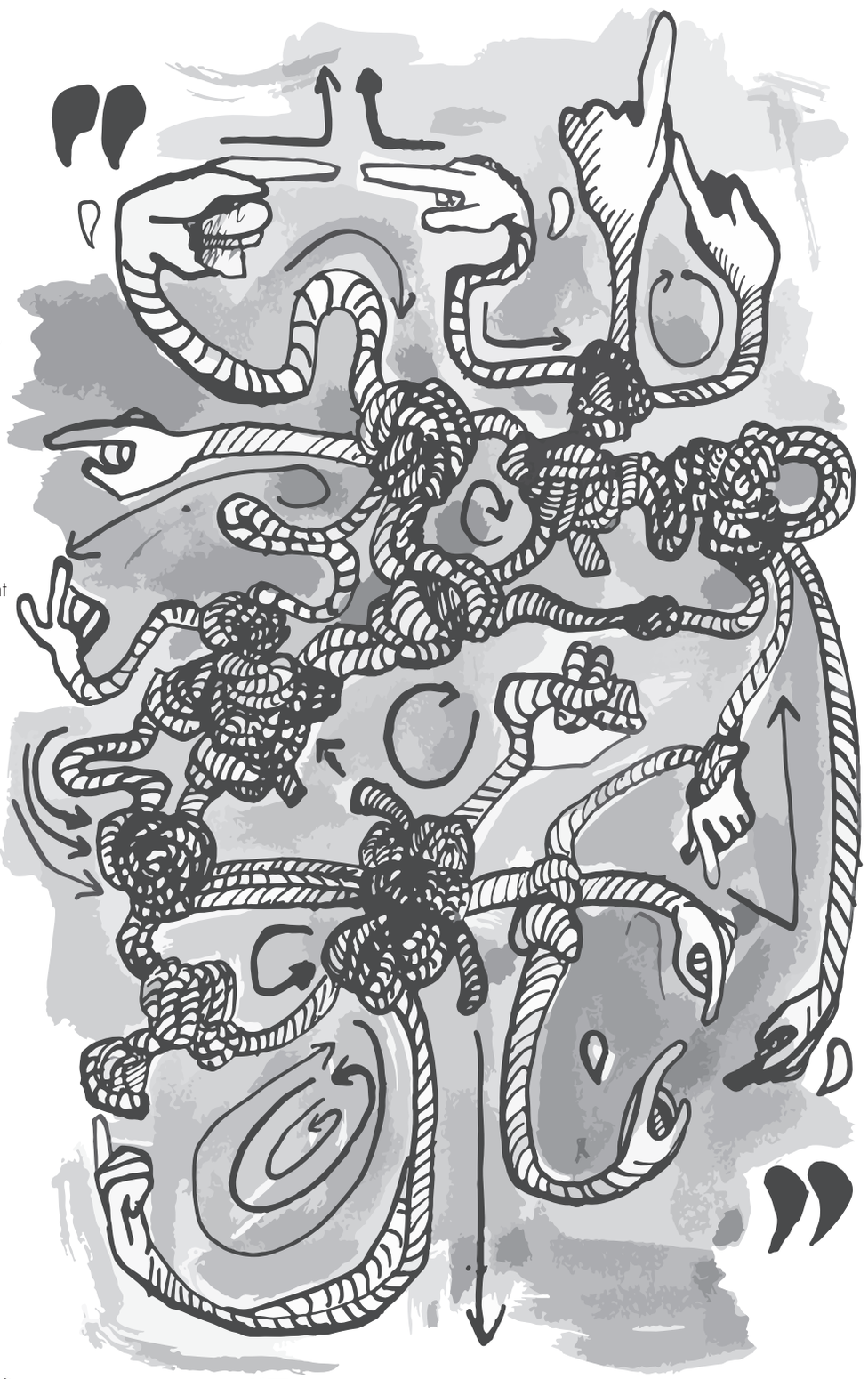
DEZE COLUMN GAAT OVER ZELF-REFERENTIE

Waarin **Victor** zich in zelfreferentiële knopen schrijft.

Deze column gaat over zelfreferentie en refereert hiermee aan deze uitgave die over zelfreferentie gaat en waar deze column zelf dus deel van is. Dit thema grijpt dan weer terug op zichzelf, omdat het deze uitgave, zelfreferentie, is, die zelf verwijst aan de aan zichzelf refererende column die over zelfreferentie gaat.
Dit is dus, in essentie, een naar zichzelf verwijzende column over zelfreferentie. Het is deze referentie aan zichzelf die deze column, als deel uitmakend van een groter geheel dat aan zichzelf refereert, zijn inhoud verschaft, omdat het aan zichzelf refereren hetgeen is waar dit thema naar refereert, namelijk zelfreferentie. Deze zin maakt dat duidelijk.

Nu het duidelijk is waar deze column over gaat (zelfreferentie) zal deze column echter een kritische noot bij zichzelf moeten plaatsen. In dit aan zichzelf refereren zelf komt men de meest problematische paradoxen tegen, die niet vanzelf weggaan en zelfs van zulke ernstige aard zijn dat de zelf-referentie zelf op losse schroeven komt te staan! Men zal zichzelf afvragen of men ooit op zichzelf de zelf-referentie moet toepassen. Door hieraan te refereren refereert deze column aan zelf-referentie en hiermee aan zichzelf, maar ik geloof dat u mijn punt nu wel begrijpt (zelfs als u deze referentie zelf niet snapt).

Dit klassieke voorbeeld licht zichzelf het beste toe: Een man – een zelf voor de lezers die genderneutraal taalgebruik prefereren – zegt over zichzelf: "Ik spreek nooit de waarheid" en daar staan wij, de toeschouwer, dan met ons goede gedrag. Het feit dat de taalfilosofie nog altijd niet heeft weten op te lossen dat leugenaars niet over zichzelf de waarheid kunnen spreken daargelaten, kan u, beste lezer, uzelf inbeelden hoeveel groter het probleem wordt als deze man daarna nog zou spreken "Ik refereer altijd aan mijzelf"! Want als deze zelf inderdaad altijd aan zichzelf refereert en deze uitspraak dus met betrekking tot zichzelf correct is, dan is dat zelf incoherent met de zelfreferentiële incoherente zelfreferentie waar in de eerste uitspraak aan gerefereerd >>



Illustratie door Mette van Liempd

wordt. Deze zou op zijn beurt de incoherentie van de tweede zelf-referentie afdwingen, wat tot een zelfreferentiële incoherentie zou leiden. Dit is echter triviaal.

Ik spreek hier natuurlijk de taal van Bertrand Russel, die met zijn eigen zelfreferentiële paradox de Duitse filosoof Gottlob Frege (die overigens graag aan zichzelf refereerde) een decennium lang de mond snoerde. Bevat de verzameling van alle verzamelingen die niet zichzelf bevatten zichzelf? Of, beter gezegd, refereert de referentie van alle niet-zelfreferenties aan zichzelf? Als u hier zelf niet uitkomt refereer ik u graag door naar de onderwijsstaf van theoretische filosofie, die vast een paar verhelderende referenties voor u heeft.

Refereert de referentie van alle niet-zelf-referenties aan zichzelf?

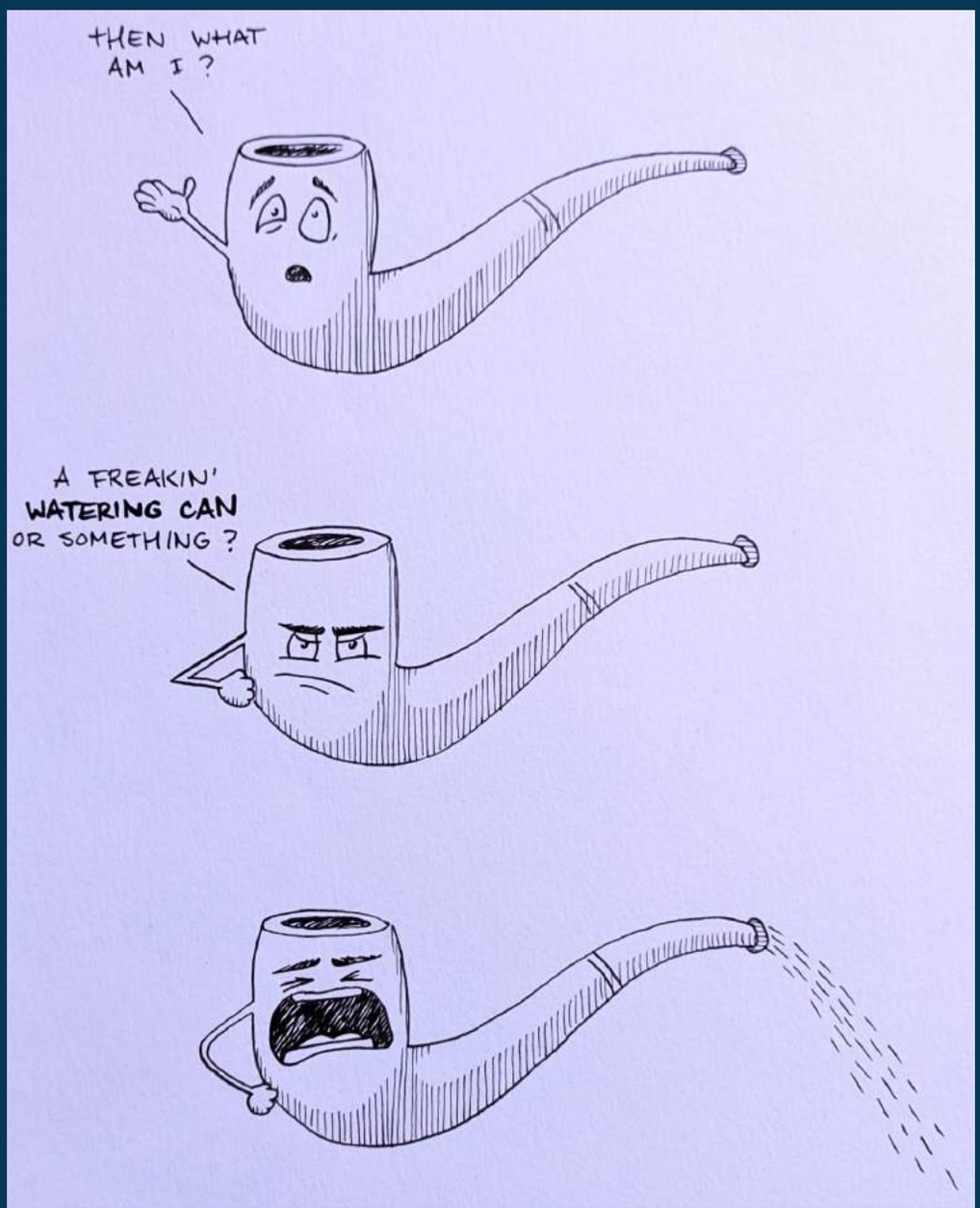
Als we dit alles zichzelf in acht laten nemen, dan wordt duidelijk dat zelfs de zelfreferentie zelf een precair onderwerp is. Dit is allesbehalve intuïtief. Ikzelf heb altijd gedacht dat de zelfreferentie nooit zelf de bron van zoveel problematiek kon zijn omdat het zelf refereert aan dat wat het zelf al is, namelijk zichzelf. Of het probleem dan bij het zelf of bij het zelf ligt is lastig te bepalen en i kzelf heb hier altijd de fout gemaakt om aan te nemen dat het inderdaad het zelf is waar het probleem ligt. Het zelf, of preciezer zelf-1, dat refereert aan het zelf, zelf-2, is echter in het refereren aan het zelf, zelf-3 (?) het probleem zelve, omdat de referentie hier aan zichzelf refereert, en niet aan het zelf. Dit is het probleem van zelfreferentie.

"Zelfreferentie" is dan ook een term die met de uiterste zorgvuldigheid toegepast dient te worden en niet zomaar willekeurig. Dit zou tot ernstige problemen en bovendien zeer beperkte leesbaarheid leiden. Zelf probeer ik mijn gebruik van zelfreferenties dan ook zeer te beperken. Als u na dit alles uw interesse voor de theoretische filosofie nog altijd niet verloren heeft, zal ik u voor mijn laatste referentie dan ook niet aan mijzelf, maar aan het UWV refereren. Daar zitten ze vast op meer theoretische filosofen te wachten. ■

Victor Smit studeert zowel Liberal Arts and Sciences als Filosofie in Utrecht en is columnist bij *De Filosoof*

STRIP

Then What Am I?



Hannah Waayers is grafisch ontwerper, schilder en illustrator. Ze heeft haar bachelordiploma filosofie op zak en is nu druk bezig met uiteenlopende kunstprojecten. Ben je benieuwd naar haar werk? Neem dan een kijkje op @hnnh_art op Instagram.

AUTEUR **HANNAH WAAYERS**

De Filosoof

www.facebook.com/defilosoof.uu

Volg ons op