

TECHNOLOGIE TWIJFELT OOK

Leesexemplaar

TECHNO- LOGIE TWIJFELT OOK

Dertien dilemma's
over innovatie
en ethiek

Filip Van den Abeele

P E L C K M A N S

© 2023, Filip Van den Abeele en Pelckmans Uitgevers nv
pelckmans.be
Brasschaatsteenweg 308, 2920 Kalmthout, België

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand of openbaar gemaakt, op welke wijze ook, zonder de uitdrukkelijke voorafgaande en schriftelijke toestemming van de uitgever, behalve in geval van wettelijke uitzondering. Informatie over kopieerrechten en de wetgeving met betrekking tot de reproductie vindt u op www.reprobel.be.

All rights reserved. No part of this book may be reproduced, stored or made public by any means whatsoever, whether electronic or mechanical, without prior permission in writing from the publisher.

Omslagontwerp: Studio Gestaag
Vormgeving: www.intertext.be

D/2023/0055/307
ISBN 978 94 6401 443 3
NUR 740
THEMA PDR, TB

Ook verkrijgbaar als:
E-book: ISBN 978 94 6401 488 4

pelckmans.be

 facebook.com/pelckmans.be
 twitter.com/Pelckmans_be
 instagram.com/pelckmans.be

Inhoud

Ten geleide	7
Inleiding: boeiende tijden	9
1. De zelfrijdende auto: wie stuurt het geweten?	17
2. Big data of big brother?	47
3. Artificiële intelligentie: het einde van het mensdom?	75
4. DNA: de maakbaarheid van de wens?	111
5. E-health: 'an app a day keeps the doctor away'	143
6. Rise of the robots: wie laat wie nu uit?	185
7. Zorgrobots: zorgzaam of bot?	219
8. Schuld en boete: rechter-muisknop?	251
9. E-ducatie: machine leert mens	283
10. Klimaatverandering: de tragedie van ons onvermogen?	321
11. Energietransitie: een kans om te herbronnen?	371
12. Mens of machine? Olympia wordt Olymp-AI	409
13. Smart farming: hongeren naar meer?	465
Conclusie: bijzonder boeiende tijden	531
Noten	541

Opgedragen aan Stanislav Petrov,
die begreep dat technologie twijfelt
en zo een wereld borgde
waarin dit boek kon worden geschreven

Leesexemplaar

Ten geleide

De wereld is in volle transitie, en het lijkt er steeds meer op dat een nieuw tijdperk in de wereldgeschiedenis in aantocht is. De klimaatopwarming dwingt ons tot een transitie van fossiele naar hernieuwbare energie. De demografische ontwikkeling noopt ons tot digitalisering om verdere productiviteitsgroei te ondersteunen. De globalisering stoot op haar grenzen en doet ons beslissen tot reshoring van de productie. De beleving van werken, leven en wonen verandert van het platteland naar de steden. De democratie en de westerse waarden staan steeds meer onder druk.

En dan rijst de vraag: wie kan de wereld redden? De overheden, de bedrijven, de wetenschap, de technologie, of de cultuur...? Mijn antwoord luidt: we zullen *samen* de wereld moeten redden en beter maken. Het is overduidelijk dat technologie een essentieel deel van de oplossing is – en zal blijven – om de uitdagingen van de komende decennia aan te pakken. Hernieuwbare energie, big data, artificiële intelligentie, het internet der dingen, robots, drones... We zullen meer dan ooit alle technologie nodig hebben. Daarom is investeren in onderzoek en ontwikkeling zo belangrijk.

De technologie zal voor grote veranderingen zorgen, zowel voor de overheden, de bedrijven als de burgers. En iedere verandering brengt angst en twijfel, roept weerstand en tegenstand op, lokt actie en reactie uit. En ja: er zullen vele vragen en antwoorden,

opmerkingen en bemerkingen, oprispingen en bekommernissen komen. Wat zal de technologie brengen? Gaat de technologie niet te snel? Wat zullen de nevenwerkingen van nieuwe technologie zijn? En wat met de gevolgen, vandaag bekend en nog niet bekend? En wat zal dat uiteindelijk betekenen voor de mens als persoon, als burger, als homo economicus? Wat betekent die technologie voor mij persoonlijk?

Technologie geeft vandaag hoop, maar schrikt ons ook af. Twijfel overvalt ons soms in deze uitdagende tijden. Filip Van den Abeele is tegelijkertijd ingenieur en technoloog, maar ook auteur en kritische wetenschapscommunicator. In dit boek *Technologie twijfelt ook* legt hij op onnavolgbare wijze de scheidingslijn bloot tussen de mogelijkheden van technologische innovatie, en de beperkingen ervan. Op die scheidingslijn ontstaat de twijfel, zelfs bij vooruitgangsoptimisten die de technologische evolutie en revolutie omarmen.

Laten we bij het lezen van dit boek dat dilemma tussen innovatie en ethiek in het achterhoofd houden. Met de stellige overtuiging dat technologie de wereld zal kunnen redden en beter maken. Maar ook met een gezonde dosis twijfel. In het besef dat twijfel het begin is van waarheid en wijsheid. De toekomst is aan ons.

Hans Maertens
Gedelegeerd bestuurder VOKA

Inleiding: boeiende tijden

‘**M**ay you live in interesting times!’ Dat waren de gevleugelde woorden waarmee Robert Kennedy in 1966 de laatstejaarsstudenten van de universiteit van Kaapstad een boeiend leven toewenste¹. En mijn hemel, wat leven wij in boeiende tijden!

Wetenschappelijke vooruitgang en technologische innovaties scheppen steeds meer welzijn en welvaart. Slimme stofzuigers en zelflerende grasmaaiers maken ons leven lichter en leuker. Onze digital natives groeien op in een wereld waarin computers intelligenter zijn en dieper denken dan de grootste schaakmeesters. Recente ontwikkelingen in artificiële intelligentie plaatsen een hefboom op het vermogen van ons beperkte brein. We mogen hardop dromen van kwantumsprongen in de geneeskunde en de gezondheidszorg. Fitbits en wearables meten al onze fysieke conditie, en intelligente pillen kunnen onze levensverwachting gevoelig verhogen. En misschien gidst nog tijdens dit ‘interessante leven’ een slimme machine ons naar een onverhoopte oplossing voor het klimaatvraagstuk of een revolutionair medicijn tegen kanker.

We leven inderdaad in boeiende tijden, waarin armoede, hongersnood en kindersterfte nooit lager lagen. Wereldwijde welvaart, comfort en levensverwachting bereiken dan weer ongeken-

de hoogten. En dat allemaal dankzij technologische vooruitgang². We zijn in een rotvaart geëvolueerd van griffel en lei naar een *connected society*, waarin onze smartphone ons de weg wijst naar

**Ideeën zijn de
grondstoffen voor
innovatie, en innovatie
is de hefboom voor
vooruitgang.**

het beste restaurant, de kortste route en het spannendste boek. We stoten ook allemaal steeds meer data uit. We hebben supercomputers opgeleid om die data te leren destilleren tot heilzame inzichten. Dankzij bigdatamining liggen verbluffende doorbraken in onder meer landbouw, voedselvoorziening en

de optimalisatie van verkeersstromen binnen handbereik. Door een combinatie van slimme sensoren, artificiële intelligentie en de nieuwste generatie van gps-technologie wordt dat verkeer steeds vaker een lust in plaats van een last. Binnenkort ontzorgt de zelfrijdende auto ons immers van stuurloosheid, filefrustraties en de tergende zoektocht naar parkeerplaatsen.

Kortom: we leven in boeiende en benijdenswaardige tijden. We waren nooit met meer mensen op deze planeet, en zelden zo intens met elkaar verbonden. Dankzij het internet der dingen kunnen we (honderd)duizenden ideeën met lichtsnelheid met elkaar delen. Ideeën zijn de grondstoffen voor innovatie, en innovatie is de hefboom voor vooruitgang. Het vergt een behoorlijke zelfbeheersing om niet razend enthousiast te worden van de mogelijkheden die voor het grijpen liggen.

Het was dan ook met grote geestdrift dat ik de voorbije jaren het land mocht rondtrekken, als handelsreiziger in onverbetelijk vooruitgangsoptimisme. Ik sprak vol vuur over de vierde industriële revolutie: een ecosysteem van slimme sensoren en intelligente machines, waarin onbegrensde personalisatie mogelijk wordt. Ik begeleidde grootbanken, energieleveranciers en zelfs

staalproducenten en glasfabrikanten in hun digitale transformatie. In bibliotheken en aan universiteiten gaf ik lezingen over de op handen zijnde energietransitie, die ons van fossiele brandstoffen naar hernieuwbare vormen van energie zal leiden. En in culturele centra probeerde ik telkens opnieuw een publiek van diverse pluimage te overtuigen van het potentieel van big data.

Enthousiasme werkt aanstekelijk, want tegen authentieke geestdrift is weinig kruid gewassen. Mijn #filipinwonderlandtournee werkte dan ook zichzelf een beetje in de hand, en bracht me naar alle uithoeken van ons vlakke Vlaanderen, van Alveringem over Keerbergen tot in Zutendaal. Mijn rondreizende wetenschaps-circus stelde me bloot aan een hoogst wetenswaardig amalgaam van vragen en voorbehoud. Want wat gebeurt er als een zelfrijdende auto moet kiezen tussen het beschermen van de inzitende of het aanrijden van een klein kind? Mag een emotioneel intelligente robot op mijn (klein)kinderen passen? Zijn al die zelflerende machines een dankbaar instrument van efficiëntie, of eerder een bedreiging op de arbeidsmarkt? En waar ligt de grens tussen big data en big brother?

Onder het voorwendsel dat mijn laptop nog bij de techniek zat, deed ik wel vaker het licht uit in de bar na een voordracht. Tussen pot en pint worden de indiscreties onverbloemder, en de opmerkingen stekeliger. Een geschiedenisleerkracht op rust vertelde, met de nostalgie die eigen is aan het lege glas, hoe hij vroeger met zijn kameraden cowboy en indiaan speelde. 'Wij werkten ons nog in het zweet om kampen te bouwen in de bossen', sprak hij weemoedig, 'waar we dan stiekem onze eigen omeletten bakten. Onze herinneringen zijn gekristalliseerd in geuren. Mijn kleinzoon van zeven, daarentegen, is gekluisterd aan een scherm, waarop hij urenlang geestdodende games speelt. En in schuttingtaal schreeuwt naar een klasgenoot die twee dorpen verderop

ook als een junkie aan een tablet hangt. En jij noemt dat vooruitgang?’

Een andere (naar ik aanneem: goede) huisvader maakte zich zorgen over zijn factuur voor gas en elektriciteit. Hij overwoog zich een dieselgenerator aan te schaffen, en informeerde voorzichtig of en hoe je uit bliksem batterijen kunt opladen. Een vrouw van middelbare leeftijd, die na de lezing geen vragen durfde te stellen, werd na twee glazen wijn plots minder schroomvallig. Haar man was net ontslagen in de fabriek, en vervangen door ‘één van die robots waar gij met zoveel vuur over spreekt’. En dat het als vijftiger zonder diploma geen sinecure was om opnieuw aan de bak te komen. In een *menage* met twee inwonende zonen die hogere studies volgden, nota bene. Ze sprak het uit op een toon die duidelijk maakte dat haar opmerking als een verwijt wilde worden gelezen.

Een pientere jongeman die mij een Duvel aanbod³, won advies in over zijn studiekeuze. Hij wilde een ingenieursopleiding volgen, maar vroeg zich af of de maakindustrie in Vlaanderen nog een toekomst had. ‘Elektromechanica draagt mijn voorkeur weg,’ sprak hij beslist, ‘maar ik betwijfel of we binnen afzienbare tijd hier nog machines zullen bouwen. Zou ik, in het licht van big data en het internet der dingen, niet beter computerwetenschappen volgen?’ Nog voor ik kon antwoorden, werd ik omsingeld door een jong koppel, weliswaar met lovende feedback. Ze hadden genoten van mijn betoog over drones, bitcoins en continue innovatie die welvaart democratisch maakt. Maar ze wezen er fijntjes op dat mijn optimisme nogal hol klonk nu zij ‘met scharrelen en schrappen’ een eigen woonst probeerden te verwerven.

Ik wentelde vele vragen af door te wijzen op de creativiteit van de mensheid. Dankzij de wetenschap hebben we een oplossing

gevonden voor het probleem van zure regen en het gat in de ozonlaag. De intrede van achtereenvolgens het mechanische weefgetouw, de lopende band en de robot heeft onze productiviteit enorm verhoogd en werk zinvoller gemaakt. Niemand rouwt nog om de automatisatie van de textielindustrie of de teeloorgang van zware fabrieksarbeid. Het internet stelt ons in staat om te videobellen met familieleden aan de andere kant van de oceaan, actuele informatie uit grenzeloze encyclopedieën te raadplegen en in een handomdraai kennis en ideeën te delen. En zijn wetenschappers er ten slotte niet in geslaagd om op minder dan twaalf maanden tijd een doeltreffend vaccin te ontwikkelen voor een virus dat de volledige wereldbevolking bedreigde? Als we de intelligentie van het mensdom oordeelkundig kanaliseren, dan kunnen we ook andere mondiale uitdagingen zoals klimaatverandering, bevolkingstoename of opstandige robots ongetwijfeld het hoofd bieden.

Ansichtkaartjes om in te kaderen

Toen het schrijven van dit boek een aanvang nam, kende ik nauwelijks het verschil tussen plichtethiek en deugdethiek, laat staan dat ik de verschillende strekkingen binnen het utilitarisme uit het blote hoofd kon opnoemen. Ik had nog nooit van Feng-Hsiung Hsu gehoord, en zelfs Nick Bostrom was voor mij een nobele onbekende. Een ‘trolley’ was nog een vorm van handbagage in plaats van een probleem, en ‘Scanadu’ klonk als een overzeese vakantiebestemming. Ik ging er stilzwijgend van uit dat *ethereum* de Latijnse benaming was voor een vloeistof om vetvlekken te verwijderen, en dacht bij ‘regels voor robots’ onwillekeurig aan menstruerende machines.

Kortom: ik had nog veel (bij) te leren, om het nog zacht uit te drukken. Tijdens het schrijven van *Technologie twijfelt ook* las ik een honderdtal boeken en een veelvoud aan wetenschappelijke publicaties, rapporten, artikels en blogs. Die vele documenten verlichtten mijn pad, verdiepten mijn inzichten en daagden mijn denken uit. En ze wezen de weg naar nóg meer razend interessante lectuur. Ik heb mij de voorbije jaren dan ook vaak een ontdekkingsreiziger gevoeld. Ik verliet de platgetreden paden van mijn eigen begrensde kennis om verafgelegen *terra incognita* in kaart te brengen. Ik groef door de recente geschiedenis van wetenschap en technologie, en ploegde al eens een fossiel van een vergeten beschaving naar boven. Onderweg ontdekte ik nieuwe werelddelen, onbekende talen en verloren gewaande bosbloemen.

Ik trof een schat aan kennis en wijsheid aan, die als grondstof fungeerde om mijn verhaal vastere vorm te geven. Maar een schat wordt pas echte rijkdom wanneer je die kunt delen. Net als een ontdekkingsreiziger voel ik vaak een onweerstaanbare drang om nieuw verworven inzichten aan een breed publiek te verkondigen. Want wist jij dat je van selder een levensbedreigende allergieaanval kunt krijgen? Dat Garri Kasparov zijn eerste wedstrijd tegen schaakcomputer Deep Blue heeft gewonnen? Dat het risico op verdrinkingsdood lijkt toe te nemen wanneer in de bioscoop een film met Nicolas Cage draait? Of welke militaire doeleinden de eerste prototypes van zelfrijdende auto's dienden?

Zulke onverwachte inzichten zijn vaak anekdotisch: ze schragen het verhaal dat ik wil brengen niet, maar ze zijn ook te mooi om níét te melden. Het zijn het soort weetjes dat je deelt terwijl je aanschuift bij de bakker. En waarvan ik heimelijk hoop dat je ze morgen verder vertelt. Ik heb ervoor gekozen om mijn

INLEIDING

meest verrassende ontdekkingen als lichtvoetige kaderstukjes op te nemen in dit boek. Je kunt ze lezen als de ansichtkaarten die een pelgrim opstuurt naar het thuisfront. Dat betekent meteen dat je die kadertjes ook kunt overslaan, zonder de kern van het betoog te moeten missen.

En toch. De vervelende vragen bleven steken. Tijdens de lange terugritten uit het verre Veurne of de Hasseltse Grenslandhallen nestelden die opmerkingen zich als weerhaakjes in mijn hoofd. De volgende ochtend werd ik wakker met de jeuk van wondjes die niet willen genezen. Jeuk is de meest milde vorm van pijn, maar het ergert niettemin. Hoe hoort die zelfrijdende auto te reageren wanneer een kind plots de straat oversteekt? Wat zou ik zelf doen, en hoe kleurt mijn beslissing wanneer het mijn zoon is die achter een bal aanholt? Of indien mijn vrouw in de auto zit?

Zo is dit boek ontstaan. Als een remedie tegen de jeuk van die vervelende weerhaakjes. Een zoektocht naar zalf om de hardnekkige wondjes in mijn hoofd te helen. In *Technologie twijfelt ook* breng ik graag verslag uit van mijn meanderende odyssee, op zoek naar antwoorden. Daarvoor ging ik gulzig te rade in boeken⁴ en bibliotheken, las ik wetenschappelijke rapporten, woonde lezingen bij en sprak met diverse deskundigen. Het werd voor mij een leerrijk traject van voortschrijdend inzicht, pendelend tussen begrip en betekenis. Al was het vaak ook een Echternachprocessie, toen bleek dat mijn referentiekader nog maar eens vierkant draaide.

Dit boek is een zoektocht naar zalf om de hardnekkige wondjes in mijn hoofd te helen.

Dit boek benoemt dertien dilemma's over innovatie en ethiek. Het volgt een organische opbouw, maar je kunt de dertien hoofdstukken die volgen ook afzonderlijk en in willekeurige volgorde lezen. Het boek stelt de netelige vragen onverbloemd, en probeert die met open vizier te beantwoorden. Ik heb dit boek geschreven als geïnteresseerde wereldburger en ingenieur. Maar ik ben allerminst onderlegd als ethicus. Moraalfilosofen, zo leerde ik onderweg, spreken een andere taal. Ze hanteren cryptische woorden en referenties aan oude wijsgeren in een poging de complexe kwesties die onze boeiende tijden stellen te duiden.

Ingenieurs hanteren een meer pragmatische aanpak. Zij zien geen problemen, maar eerder een ruggengraat voor mogelijke antwoorden. Ik heb, naar godsvrucht en vermogen, geprobeerd om die antwoorden op schrift te stellen. Al stelde ik vast dat oplossingen een spectrum vormen dat veel breder oogt dan simpelweg 'goed' of 'kwaad'. De antwoorden krijgen gestalte in een spanningsveld dat nuance vergt, en ruimte laat voor individuele levensvisie en persoonlijke inborst. En soms meer ambiguïteit duldt dan waar een ingenieur zich comfortabel bij voelt. Verwacht dus niet louter antwoorden met 'ja' of 'nee', maar aanvaard af en toe een 'misschien'. Per slot van rekening mocht ik onderweg ontdekken dat ook technologie soms twijfelt.

HOOFDSTUK 1

**De zelfrijdende auto:
wie stuurt het geweten?**

Leesvoorbeeldjaar

Mag een zelfrijdende
auto moedwillig een
kind aanrijden of
moet hij de passagier
beschermen?

En wie hoort die
keuze te maken?

‘Over driehonderd meter rechts afslaan. Daarna: houd links aan.’ Er klinkt geen zweem van twijfel in het marsorder, dus ik vertraag volgzzaam en laat de richtingaanwijzer knipperen. Zelfs wanneer ik alleen onderweg ben, blijf ik gehoorzamen aan de vrouwenstem. ‘Over anderhalve kilometer: ga rechtdoor over de rotonde.’ Voor een matige chauffeur als ik is de gps een zegen. Met ‘verkeersinformatie’ als drogreden ga ik zelden of nooit zonder Waze, TomTom of Google Maps op pad. Ik betrap mezelf erop dat ik nog nauwelijks op eigen kracht de weg terugvind tussen schoolpoort en garagepoort. ‘Over vierhonderd meter, rechts aanhouden’, vervolgt mijn virtuele copiloot met kordate stem die laat vermoeden dat daar niet over te onderhandelen valt. Ik mijmer, en denk terug aan die dagen waarin je analoge atlassen moest raadplegen om zélf een heuristische knooppuntenroute samen te harken. Dat vertaalde zich doorgaans in een bos veelkleurige post-its op het dashboard, met een sequentie aan grote steden (Rijsel – Reims – Dijon – Lyon), een salvo aan alfanumerieke codes (E17! A26! A7!) én de even permanente als heilige vrees om bij een verkeerde afslag plots onderweg te blijken naar Parijs. Toen computers hun intrede deden, ontleenden wij ons traject aan de intelligentie van de vroegste routeplanners. Aan de binnenkant van de voorruit plakte toen een printje van Mappy.be met een statische route, die bij de eerste wegenwerken, wegomleiding of – meestal – verstrooidheid volledig betekenisloos werd. En onvermijdelijk uitmondde in verhitte discussies met die andere vrouwenstem.

Met een schamper ‘Probeer om te keren’ haalt de elektronische meesteres me opnieuw bij de les. Terwijl ik begin te beseffen dat

ik de juiste afslag heb gemist, wordt de route feilloos voor mij herberekend. Met een moeilijk mis te verstante 'Neem de afslag links' word ik opnieuw op het juiste pad gebracht, en dwalen mijn gedachten andermaal af. Die eerste gps, herinner ik me grijnzend, had nog de omvang van een vissenkomp. Met duimdikke kabels en een grotesk zware zuignap nam dat toestel een prominente plaats in op de voorruit. Het eiste vaak meer aandacht op dan het verkeer. Het was bovendien een behoorlijk dom ding, dat weliswaar een route naar je bestemming voor ogen had, maar niet of nauwelijks kon anticiperen op files en flitspalen. En je deed stoppen aan de grens om de cd-rom te verwisselen.

Bzzz. Bzzz. Ditmaal is het de Driver Alert die me terughaalt van *memory lane* naar de Zwijnaardsesteenweg. Een slim stuur dat mijn aandacht tijdig kanaliseert is een noodzaak voor mijn meanderende geest, en een zegen voor de nabije weggebruikers. Ook de Parking Assist is voor mij van goudwaarde, omdat ik sinds jaar en dag geboekstaafd sta als een hardnekkige recidivist in blikshade. De intelligente instrumenten in moderne wagens leggen steeds meer de nadruk op de eerste twee lettergrepen van 'autorijden'. Ze zorgen er zo voor dat ik verkeersdeelname niet langer hoeft te ervaren als een straf. 'Over twee kilometer, ga rechtdoor', krijg ik te horen, terwijl het begint te motregen. Ik overweeg heel even, uit pure balorigheid, de instructie te negeren. Zou het mogelijk zijn je gps op stang te jagen? Is er iemand bij Garmin die het toestel programmeert om, na zevenmaal moedwillig misrijden, te antwoorden: 'Bij het volgende kruispunt: zoek het zelf maar uit'? De avond valt, en het vooruitzicht van een woordentwist tussen mens en machine doet me monkellachen.

Crash! De crash van metaal tegen metaal klinkt verrassend dof. Mijn hoofd wordt vooruit geslagen, veert vervolgens terug tegen de zetel, en landt ten slotte in de airbag. Alles wordt dof, waar-

door het lijkt alsof ik me onder water bevind. Ik duizel als een Zwitser, heb een wee gevoel in de maag en voel een scherpe pijn in de onderbenen. Dit is hoe een aanrijding aanvoelt, hoor ik mezelf denken, om er meteen aan toe te voegen: je ziet het dus écht niet aankomen... Een aantal minuten later kan ik de auto op eigen kracht verlaten. Mankend, weliswaar, want het linkerbeen is gehavend, en ook mijn bekken betreurt de brutale manier waarop ik tot stilstand kwam. Mijn tegenligger blijkt een dwarsligger: het voertuig staat haaks op de weg, aan het einde van een onheilspellend oliespoor. De voorruit is gebarsten, het kofferdekseel is een kreukelzone, en de linkervoordeur hangt uit de hengsels. Maar het blijft vooral onrustwekkend stil. De schrik slaat me om het hart. Heb ik iemand verwond? Of erger? Was ik in mijn recht, of verzuimde ik voorrang te verlenen?

Met klamme handen en wild kloppend hart nader ik de auto. Op de achterbank ligt een oudere vrouw zonder zichtbare verwondingen, maar duidelijk buiten bewustzijn. De chauffeur blijkt spoorloos. Is hij uit het voertuig geslingerd? Onder mijn auto terechtgekomen? Of heeft die simpelweg de benen genomen...? Ik stel vast dat, ondanks de harde klap van de aanrijding, de airbag zich niet heeft opgeblazen. Dan pas merk ik op dat de wagen geen stuur heeft. Het onvermijdelijke inzicht raakt me als een nieuwe frontale klap: ik ben zonet gecrasht tegen een zelfrijdende auto! Waar kwam die in vredesnaam vandaan? Hoe kan zo'n computer op wielen in een ongeval betrokken raken? En kan een zelfrijdende auto überhaupt *uit* zijn recht zijn?!

De baan ligt breed open voor de zelfrijdende auto

Het voorgaande crashscenario is – voor alle duidelijkheid – virtueel, al kan de omschrijving van mijn krimpande aandachts-spanne achter het stuur wel degelijk sporen van waarheid bevatten. Nochtans is een toekomstbeeld waarin zelfrijdende auto's binnenkort deelnemen aan het verkeer niet langer een technologische utopie. Wij hebben als mens telkens opnieuw de neiging om de ogenblikkelijke impact van nieuwe technologie te overschatten. Maar we onderschatten schromelijk de invloed op langere termijn. Sciencefictionfilms uit de jaren tachtig tonen zonder uitzondering visoenen over vliegende auto's, teleportatie en uurwerken aangedreven door microscopische kernfusiereactoren. Maar wanneer de hoofdpersonages elkaar opbellen, doen ze dat nog steeds via vreemde bakelieten hoorns waaruit een don-

Wij hebben als mens telkens opnieuw de neiging om de ogenblikkelijke impact van nieuwe technologie te overschatten. Maar we onderschatten schromelijk de invloed op langere termijn.

kere draad krult. Waarmee ik wil zeggen: de zelfrijdende auto is nog niet voor morgen, maar de digital natives die vandaag geboren worden, zullen misschien⁵ nooit meer een rijbewijs nodig hebben. Volgens de meer optimistische scenario's zal die 'Generation Z' kunnen gebruikmaken van een vloot intelligente, geconnecteerde Google Cars of zelfrijdende Tesla's om hen van hot naar her te voeren. Je spreekt hier bovendien over een cohort

van mondige wereldburgers met een hoge affiniteit voor technologie, die niet langer ondanks alles een auto willen bezitten. Ze swipen eenvoudigweg de beschikbare Google Car die zich – hier en nu – het dichtst in hun buurt bevindt. Ze huren die 'dienstwagen' vervolgens in om hun doel te bereiken. Zo geven zij gestalte aan een volledig nieuwe vorm van transport, waarin auto-

bezit de baan moet ruimen voor *mobility as a service*. Het is weliswaar toekomstmuziek, maar ze klinkt bijzonder veelbelovend. Zeker als je even doordenkt over de repercussies op het vlak van verkeersdruk, ochtendfiles en de tergende zoektocht naar parkeerplaatsen.

De eerste octaven van die toekomstmuziek kun je vandaag al horen. De zelfrijdende auto kwam breed onder de aandacht dankzij het eerste prototype van de Google Car, een sympathiek ogende tweezitter die het midden hield tussen een helikopter en een eierkoker. Maar de initiële blauwdrukken voor autonome voertuigen dateren van veel vroeger. Zo schetste General Motors al op de Wereldtentoonstelling van 1939 een toekomstvisie met Futurama, een gedurfd voorstel met zelfrijdende auto's en volledig geautomatiseerde snelwegen. De verkeersstromen in Futurama leken verrassend veel op het mobiliteitsconcept van Generatie Z hierboven. De auto's ontleenden hun zelfrijdend vermogen aan een elektrisch circuit in het wegdek, net zoals speelgoedautootjes op een racebaan. Met de Firebird III uit 1962 heeft GM ook zo'n prototype gebouwd, al bleef het bij een conceptcar die enkel in het laboratorium en op testcircuits zelfstandig mocht rijden. Begin jaren tachtig bouwde uitvinder, pionier⁶ en professor Ernst Dickmanns een Mercedesbestelbus om tot een zelfrijdende robot. Hij slaagde erin het 5 ton zware gevaarte met een gemiddelde snelheid van 50 kilometer per uur zelfstandig te laten rondtuffen... in een verkeersvrije straat.

De verdere ontwikkeling van zelfrijdende auto's werd geschraagd door grootschalige onderzoeksprogramma's zoals het Prometheusproject⁷. Dat R&D-programma met een budget van maar liefst 750 miljoen euro bracht toonaangevende universiteiten en automobielabrikanten samen om uitdagingen rond artificiële intelligentie, communicatie tussen verschillende voertuigen en het

vermijden van aanrijdingen het hoofd te bieden. Met vereende krachten werden twee demonstratieauto's (met de poëtische namen Vamp en Vita) gebouwd om nieuwe technologieën te testen. Aan het einde van de rit (in 1995) bleken beide auto's in staat om, op een autosnelweg in de buurt van Parijs, volledig zelfstandig door druk verkeer te bewegen. Hun arsenaal aan stuurmanskunsten omvatte inhaalmanoeuvres, automatisch voorsorteren, en deelnemen aan het verkeer met snelheden tot 130 kilometer per uur. De Vamp werd naderhand ingezet om professor Dickmanns van München naar Kopenhagen en terug te rijden. Een tocht van bijna 2000 kilometer met amper menselijke interventie en, nota bene, zónder gps aan boord.

Een tocht door de woestijn... op steunwiel-tjes

Ook de DARPA-uitdagingen fungeerden als katalysator bij het verder perfectioneren van de capaciteiten van zelfrijdende auto's. De competitie is genoemd naar haar broodheer DARPA. Dat Defense Advanced Research Projects Agency is zo'n beetje de Willy Wortelafdeling van het Amerikaanse ministerie van Defensie. De uitdaging bestaat uit een jaarlijkse competitie, met een prijzengeld van 1 miljoen dollar, waarbij autonome voertuigen een vast traject moeten afleggen in de Mojavewoestijn tussen California en Nevada. De eerste editie in 2004 was, naar alle mogelijke maatstaven, een ontzettende ontzuivering. Slechts zeven van de meer dan honderd ingeschreven teams slaagden in de kwalificatieronde: een (brokken)parcours van nauwelijks 1,5 kilometer. Acht andere robots werden door de jury opgevestigd en 'goed genoeg' geacht om ook aan de echte wedstrijd deel te nemen. De race bood een traject van 150 mijl door onherbergzaam gebied, langs onver-

DE ZELFRIJDENDE AUTO

harde wegen en op onbekend terrein. Van de vijftien auto's haalde geen enkele de finish. Twee auto's haalden niet eens de startlijn; een derde reed zich al na 50 meter te pletter tegen een betonnen paaltje. Een robotbuggy die uit de film *Mad Max* leek verdwaald, raakte ook in de wedstrijd het spoor bijster en reed hopeloos verloren. Een zelfrijdende pick-uptruck met technische problemen bleef wanhopig rondjes rijden, als een hond die in zijn eigen staart probeert te bijten. Na drie uur van gedeukt koetswerk, gescheurde riemen en gebroken dromen reed de Humvee van de Carnegie Mellonuniversiteit zich vast in een duin. Hij bleef in het mulle zand amechtig de banden spinnen tot ze in brand vlogen. Met een schamele 11 kilometer op de teller strandde de Humvee als laatste, met in zijn zog een waar autokerkhof van 'zelf-rijtende' auto's.

Ondanks de groteske mislukking van de eerste editie werd de DARPA Grand Challenge een jaar later opnieuw georganiseerd, en werd het prijzengeld verdubbeld. De drieëntwintig voertuigen die in 2005 aan de start kwamen, en de technologie die ze aan boord hadden, waren nauwelijks herkenbaar. Alle deelnemers – op eentje⁸ na – raakten vrijwel moeiteloos verder dan die 11 kilometer uit 2004. Vijf auto's haalden de finish zonder de minste menselijke interventie. Stanley, een verbouwde Volkswagen Touareg op steroïden en artificiële intelligentie, legde het traject van 212 kilometer af in iets minder dan zeven uur, dus met een gemiddelde snelheid van ruim 30 kilometer per uur. Een halfuur later hadden ook Sandstorm, Highlander en Kat-5 foutloos de finish bereikt. Net zoals de ruimtewedloop in de jaren zestig fungeerden de DARPA-uitdagingen als een hefboom op innovatie en een brandversneller voor aanstekelijke ideeën.

In 2007 legde DARPA de lat nóg hoger met een Urban Challenge, waarbij de zelfrijdende auto's op een gesloten circuit moesten deelnemen aan verkeerssituaties. De elf deelnemers reden tussen dertig andere voertuigen, en konden zo hun kennis en kunde van inhalen, voorsorteren en voorrang verlenen demonstreren. Zes teams haalden zonder blutsen en builen de finish, een overtuigende aanwijzing dat zelfrijdende auto's in stadsverkeer binnen bereik kwamen te liggen. DARPA kan met recht en reden claimen dat hun uitdagingen mee het pad hebben geëffend voor de komst van autonome voertuigen.

Onhandige auto's voor hersenloze chauffeurs

De technologie is inmiddels tot volle wasdom gekomen, en zelfrijdende auto's staan te popelen om zich op onze wegen te begeven. Sterker zelfs: in ons alledaagse verkeer rijden al gecamouflerde robots rond, zij het met verschillende gradaties van autonomie. Het Amerikaanse instituut voor automobiëdeskundigen (SAE)⁹ onderscheidt vijf sporten op de ladder naar volledig autonome voertuigen. Het eerste niveau is de zogeheten 'voetloze' auto, waar je dankzij de cruisecontrol het gaspedaal kunt lossen. Het automatische remsysteem herkent voorliggers en vertraagt wanneer jij dat nalaat. Ook parkeerhulp en technologie die je binnen de lijntjes van je rijvak doet kleuren, behoren tot die vroege verschijningsvormen van geautomatiseerd rijden. Zowat alle hedendaagse auto's hebben die snuffes al aan boord.

De volgende stap in autonomie is de 'handloze' auto, waar het voertuig instaat voor versnellen, vertragen en sturen. De inzittende hoeft enkel nog een oogje in het zeil te houden. Tesla was met zijn 'Autopilot' een van de eersten om die mate van zelfstandig-

heid te introduceren. Het is een ontwikkeling die bij uitstek dienstig is bij voorspelbare verkeersstromen en voertuigen die in konvooi rijden. Het ligt dan ook in de lijn der verwachtingen dat dit type technologie brede ingang zal vinden bij onder meer vrachtverkeer, vuilniswagens en shuttles op bijvoorbeeld luchthavens. Hoewel dat soort voertuigen autonoom kunnen rijden, zijn hun intelligentie en beslissingsvermogen onvoldoende robuust om er blind op te kunnen vertrouwen. De bestuurder moet nog steeds permanent paraat staan. Het bracht Alex Davies¹⁰ tot de smalende opmerking dat dit type autonomie 'even nuttig is als een peuter die je helpt bij de afwas'.

Zelfrijdende auto's van SAE-niveau 3 laten je wél toe om de blik af te wenden, en zelfs een boek te lezen achter het stuur. Ze worden daarom ook 'oogloos' genoemd. Ze rijden nog niet in groten getale rond op onze Europese snelwegen, maar constructeurs als Volvo, Audi en Honda hebben al beloftevolle prototypes gebouwd. De stap naar volledige autonomie wordt niveau 4 of de 'hersenloze' wagen. Mijn dagelijkse pendeltocht tussen Gent en Brussel suggereert dat er al verrassend veel van dit type auto's in omloop zijn. Niettemin blijft dat nog toekomstmuziek, al claimt Tesla dat in recente modellen alle hardware al voorhanden is. Die auto's zouden – na een software-update – binnenkort volledig autonoom kunnen worden. Een zelfrijdende wagen van niveau 4 heeft geen stuur of pedalen meer nodig, en laat toe dat de passagier letterlijk in slaap valt. Dat niveau impliceert dan ook een verschuiving in verantwoordelijkheid van mens naar machine. Waymo – het bedrijf boven de vroegere Google Car – heeft al meer dan 10 miljoen onbemande kilometers op Amerikaanse (snel)wegen achter de kiezen. Het ligt momenteel aan de leiding om de eerste commerciële zelfrijdende auto van niveau 4 op de markt te brengen. Ook Aurora experimenteert met technologie die taxi's zonder chauffeur op korte termijn mogelijk moet maken.

Het ultieme walhalla voor de zelfrijdende auto is niveau 5. Dat is een auto die in iedere verkeerssituatie en onder alle (weers)omstandigheden perfect autonoom kan rijden. Het wordt, met andere woorden, een machine die de wegcode altijd respecteert, nooit drank of drugs gebruikt, en niet vermoeid raakt of in slaap kan vallen. Een droom van een auto, zeg maar, die aanrijdingen degradeert tot een verre herinnering, en het aantal verkeersslachtoffers drastisch kan doen dalen¹¹. Maar meteen ook een machine die aanleiding geeft tot vervelende vragen en stekelige keuzes.

Zelfrijdende auto botst met stuurloze tram

Een zelfrijdende auto ontwerpen is – op papier, althans – helemaal niet zo moeilijk. Je moet een robot bouwen die exact weet waar hij zich bevindt, en bovendien begrijpt waar hij heen wil. Vervolgens dwing je die robot zich te onderwerpen aan de wegcode. Dat is een even coherente als consistente set van ondubbeltzinnige regels en afspraken, en meteen het soort draaiboek waar machines dol op zijn. Vandaag weet al iedere auto, en ook elke gsm, waar hij zich bevindt dankzij het Global Positioning System (gps)¹². Door radiogolven uit te wisselen met satellieten die op 20.000 kilometer hoogte om de aarde cirkelen, kun je je positie op aarde bij benadering berekenen. De zelfrijdende auto heeft weliswaar militaire gps-precisie nodig. Een nauwkeurigheid van ‘slechts’ een meter kan immers het verschil maken tussen het linkerrijvak en spookrijden. De auto geeft zich ook rekenschap van zijn omgeving door het gebruik van camera’s. Dankzij patroonherkenning en artificiële intelligentie kun je het voertuig zijn omliggers leren ‘zien’. Maar ook camera’s zijn niet onfeilbaar. De beelden kunnen misleidend¹³ zijn, en ze bieden bovendien geen dieptezicht. Zelfrijdende auto’s ontlenen hun stereoscopie aan light detection and ranging (lidar). De auto strooit laserpulsen in

het rond (meestal vanaf die discobal boven op het dak) en meet hoelang het duurt om de gereflecteerde signalen te ontvangen. Het is hetzelfde principe als de radar, maar met lichtsignalen in plaats van radiogolven. De meeste zelfrijdende wagens hebben trouwens beide technologieën aan boord: Lidar werkt goed voor korte afstanden, radar laat dan weer toe om (veel) verder te kijken.

En zo begint de zelfrijdende auto op de cockpit van een vliegtuig te lijken, waar een veelheid aan metingen en signalen binnenkomen. De automatische piloot moet zinvolle en daadkrachtige beslissingen nemen om op koers te blijven. De zondvloed aan binnenkomende signalen en het brede spectrum aan technologie bieden het voordeel van vergevingsgezindheid¹⁴. Veel ruis wordt gefilterd, en je kunt je een aantal foute metingen of slechte sensoren permitteren. Als iets eruitziet als een paard, klinkt als een paard en ruikt als een paard... dan staat er hoogstwaarschijnlijk een viervoeter naast je wagen. De automatische piloot die betekenis moet ontleen aan de data op het dashboard, en vervolgens de koers uitzet, is een zelflerende machine die getraind wordt aan de hand van neurale netwerken. Dat betekent meteen dat een zelfrijdende auto, net zoals iedere chauffeur, kilometers moet maken om voldoende ervaring te krijgen. Voor de gemiddelde (jong) volwassene is dat naar verluidt¹⁵ minstens 10.000 kilometer. Zelfrijdende auto's zullen wellicht 1 miljoen kilometer trainingsritten nodig hebben, met nog een veelvoud aan simulatiedata daarbovenop. Maar zodra de technologie geperfectioneerd is, biedt de nuchtere, onvermoeibare en (artificieel) intelligente zelfrijdende auto uitzicht op een veel verkeersveiliger toekomst.

De verwachting is dat de komst van de zelfrijdende auto het aantal verkeersslachtoffers, maar ook het aantal overtredingen en aanrijdingen, sterk zal doen dalen. Maar we kunnen de kans op

een botsing niet volledig uitsluiten, want ook machines zijn niet onfeilbaar. De kwetsbaarheid van zelfrijdende auto's werd brutaal aangetoond op 7 mei 2016, toen Joshua Brown zijn Tesla Model S op autopiloot liet rijden. Ter hoogte van een kruispunt sloeg een tegenligger – een uit de kluiten gewassen vrachtwagen op achttien wielen – linksaf, waardoor hij dwars op de weg kwam te staan. De Tesla autopiloot slaagde er niet in de witte flank van de vrachtwagen te onderscheiden van de omgeving, en reed zich zonder te vertragen te pletter. Joshua Brown stierf ter plaatse, en het ongeval veroorzaakte heel wat beroering¹⁶. Twee jaar later overleed Elaine Herzberg toen ze na valavond, naast haar met boodschappen beladen fiets, een viervaksbaan probeerde over te steken. Halverwege haar traject werd ze aangerezen door een prototype van een zelfrijdende Uber, die haar niet (h)erkende als zwakke weggebruiker. De Uber vertraagde niet, en Elaine Herzberg liep bij de impact fatale verwondingen op. Die dodelijke ongevallen betekenden ook een zware slag voor het vertrouwen in de capaciteiten van zelfrijdende auto's. En ze legden meteen hun achilleshiel bloot: autonome voertuigen blijken het meest kwetsbaar in hun interactie met menselijke vrachtwagenchauffeurs, of voetgangers die in het donker plots oversteken. Het toont aan dat, in het bijzonder bij de introductie van de eerste prototypes in 'gemengd' verkeer, het risico op botsingen niet denkbeeldig is. En dat leidt tot vervelende vragen, en zelfs ethische dilemma's. Mag een zelfrijdende auto moedwillig een kind aanrijden om mijn veiligheid te beschermen? Wat gebeurt er als die wagen moet kiezen tussen een frontale impact met een schoolbus of een val in een diep ravijn? En moet hij, bij een onvermijdelijke aanrijding, mikken op een jonge vrouw met kinderwagen of op een bejaard koppel? Het zijn terechte en pertinente vragen die vergen dat we robots programmeren met een vermogen om ethische beslissingen te nemen. Hoe die computercode er zal uitzien, en wie de gewichtsfactoren in de vergelij-

kingen moet bepalen, is nog hoogst onduidelijk. Maar de morele dilemma's rond zelfrijdende auto's brachten Hannah Fry¹⁷ er alvast toe te besluiten dat 'de essentie van deze nieuwe technologie vragen oproept over macht, verwachting, controle en het delegeren van verantwoordelijkheid'.

Wie rijdt jij aan: het kleine kind of de kalende opa?

Neem dus vooral zélf even plaats in een zelfrijdende auto terwijl je dit boek aan het lezen bent. Het is een stralende lentedag, en je rijdt 58 kilometer per uur in een zone waar je 60 mag. Je gedachten worden in beslag genomen door de netelige kwesties die ik net heb aangehaald, wanneer plots een kind op een fiets vlak voor je auto opduikt. Het kind draagt geen helm, en zal bij een botsing onvermijdelijk zware verwondingen oplopen. De enige mogelijke uitweg is vliegensvlug naar rechts zwenken, waardoor je inrijdt op een nietsvermoedende voorbijganger. Wat doe je? Of, juister: wat hoort je zelfrijdende auto te doen? Daarbij komt al een eerste subtiele – maar niettemin heel belangrijke – nuance dagzomen. Als chauffeur zou je het stuur in pure paniek kunnen omgooien, waardoor je per ongeluk en zonder voorbedachten rade de voetganger in het vizier krijgt. Maar paniek laat zich niet programmeren. Dus het gedrag van een zelfrijdende auto heeft niet langer het karakter van een reactie, maar dat van een beslissing.

Paniek laat zich niet programmeren. Dus het gedrag van een zelfrijdende auto heeft niet langer het karakter van een reactie, maar dat van een beslissing.

En de motieven voor die beslissing moeten, wellicht jaren voordien, door mensenhanden worden geprogrammeerd. Opnieuw: welke optie krijgt bij het algoritme de bovenhand? Het kind op de fiets aanrijden, of de voorbijganger van het voetpad maaien? En wat indien ik je nog een derde mogelijkheid bied: de auto kan

naar links zwenken, waar een betonnen muur wacht. En het risico op levensgevaar voor de inzittende.

Bij dit soort vragen ligt de zelfrijdende auto op ramkoers met een stuurloze tram. Inderdaad, de aard van die ethische vragen doet denken aan het fameuze trolleyprobleem, een gedachte-experiment waarover al heel wat inkt⁸ is gevloeid. De originele formulering bestaat uit een op hol geslagen tram (trolley) die dreigt in te rijden op vijf spoorarbeiders. Bij de impact zullen die allen onvermijdelijk sterven. Je kunt als toeschouwer echter een hendel activeren, waardoor de tram op een ander spoor terechtkomt. Op dat spoor werkt één onfortuinlijke arbeider, die onherroepelijk moet sterven om de vijf anderen te redden. Dat gedachte-experiment is een soort lakmoesproef die peilt naar onze moraliteit. Sommigen vinden het deontologisch onverantwoord om wetens en willens te doden, en laten de hendel dus ongemoeid. Anderen – onder wie de auteur van dit boek – voelen dan weer de morele plicht om in te grijpen, en dus een leven op te offeren om vijf andere te redden. Een alternatieve formulering van het probleem luidt dat je een corpulente man van een brug op de sporen duwt. De dikkerd sterft, maar hij brengt de tram tot stilstand. Die handeling lijkt voor velen een brug te ver, hoewel je hier opnieuw vijf mensen redt door één leven op te offeren. Het is een subtiel gewijzigd scenario, maar het appelleert aan het verschillende morele gewicht dat wij toekennen aan ‘doden’ of ‘laten sterven’.

Vijf tinten grijs

Het vraagstuk van de losgeslagen tram vertoont gelijkenissen met de ethische dilemma's die een zelfrijdende auto bij crash-scenario's wachten, al vinden velen⁹ het trolleyprobleem te zwart-wit geformuleerd. Het is uiteraard correct dat een levensechte verkeerssituatie veel genuanceerder is dan een geïdealiseerd ge-

dachte-experiment. Er vallen op zijn minst vijf tinten grijs²⁰ te onderscheiden. Eerst en vooral is de 'morele agent' bij een aanrijding met een zelfrijdende auto geen toeschouwer. De beslissing komt uit de boordcomputer van een voertuig dat zelf deelneemt aan de crash, en dus ook de gevolgen ervan ondergaat. Bovendien verschuift zowel de morele verantwoordelijkheid als de wettelijke aansprakelijkheid van de toeschouwer (bij de tram) naar de programmeur (van de zelfrijdende auto). Een derde – vaak gehoord en ook gewichtig – argument luidt dat verkeerssituaties veel onzekerheid, waarschijnlijkheid en toeval inhouden, terwijl het trolleyprobleem in een heel strak keurslijf is gegoten. Dat argument houdt in elk geval steek: een zelfrijdende auto kan onmogelijk het gewicht, de snelheid en het traject van alle andere weggebruikers perfect berekenen. Het valt te betwijfelen of camera's en gezichtsherkenning toelaten om de leeftijd van voetgangers te begroten. Ook de gevolgen van een impact (licht- of zwaarwonden? aantal doden?) zijn nooit vooraf bekend. Een vierde manco in de analogie luidt dat het trolleyprobleem statisch is, als in steen geschreven. Een botsing daarentegen is bijna altijd een hoogst dynamische gebeurtenis, waarbij op heel korte termijn verschillende betrokken partijen moeten onderhandelen over blikshade en lijfsbehoud. Bovendien is een zelfrijdende auto een lerende machine, die door verkeersdeelname voortschrijdend inzicht verwerft. Het gedrag en de reactie van het voertuig veranderen met andere woorden ook in de loop van de tijd. Een laatste vermeldenswaardige kritiek op het trolleyprobleem is dat sommige filosofen het concept van een gedachte-experiment verwerpen als leidraad voor realistische crashscenario's met verre gaande gevolgen. Patrick Lin, professor ethiek en expert op het vlak van autonome voertuigen, hanteert een pragmatischer visie²¹. Hij wijst erop dat gedachte-experimenten, net zoals fysische experimenten in een laboratorium, ontworpen zijn als stresstesten om bepaalde variabelen te isoleren, en onze intuïtie daarover te onderzoeken.

Ik sluit me aan bij die gedachtegang, en wil een aantal variaties op het trolleyprobleem formuleren om de ethische dilemma's van zelfrijdende auto's scherp te stellen, en op zoek te gaan naar mogelijke antwoorden.

Navigeren op een moreel kompas

Het voert me terug naar mijn eigen crashscenario: een kind op een fiets aanrijden, een toevallige voorbijganger van het voetpad maaien, of zelf tegen een betonnen muur crashen? Het hangt ervan af aan wie je het vraagt, maar de ethiek onderscheidt ruwweg twee²² stromingen die het antwoord richting kunnen geven. De deontologie of plichtethiek gaat uit van absolute gedragsregels die vaak als normen worden gesteld. Een beetje zoals de stenen tafelen van Mozes, zeg maar. Aangezien het vijfde gebod luidt: 'Gij zult niet doden', zal een deontoloog die daarin gelooft de voorkeur geven aan een gewisse dood tegen de betonnen muur om de fietser en de voetganger te sparen. Diezelfde deontoloog blijft trouwens ook notoir afzijdig in de meeste trolleyscenario's. Het utilisme, een tak van de gevolgenethiek, biedt een andere benadering. De morele waarde van een handeling wordt afgemeten aan de (positieve en/of negatieve) gevolgen ervan. Een utilist is met andere woorden geneigd om de hendel over te halen in het trolleyprobleem. Maar hij ziet zich als programmeur van een zelfrijdende auto geconfronteerd met een nóg moeilijker probleem: hoe weeg je de waarde van het leven van de voetganger af ten opzichte van dat van het kind – of van jezelf?

Op grond van zijn leeftijd zou je het kind kunnen sparen, net zoals je zou kunnen overwegen om een bejaard koppel aan te rijden in plaats van een jonge vrouw met een kindervagen. Velen zullen die – weliswaar hartverscheurende – keuze gerechtvaar-

digd vinden, en misschien is ook het bejaarde koppel bereid hun leven te ruilen voor dat van een jonge moeder met haar baby. Maar valt het ethisch te verantwoorden dat we zelfrijdende auto's trainen om actief te discrimineren op leeftijd? Die vraag wordt nog vervelender wanneer we dat dilemma door een andere lens bekijken. Het onstuimige en onvoorzichtige kind op de fiets, en ook de overstekende moeder met de kinderwagen, nemen actief deel aan het verkeer. Ze aanvaarden dus impliciet blootstelling aan de risico's die inherent zijn aan verkeerdeelname. Valt het dan te verantwoorden dat we ervoor kiezen om in te rijden op omstanders of toevallige passanten die aanvankelijk geen deel uitmaken van de verkeerssituatie? Of hebben we net de morele plicht om buitenstaanders te ontzien²³, en dus als het echt moet in te rijden op een tegenligger in plaats van te zwenken in de richting van een terras? Het zijn vele vervelende vragen die zich vandaag al stellen, en waarop de boordcomputer van een zelfrijdende wagen eerder vroeg dan laat een antwoord moet bieden.

Het goede nieuws is dat de computercode achter zo'n beslissing zich verrassend eenvoudig²⁴ laat programmeren. Ze vormt zelfs de grondslag voor vele controle- en optimalisatiealgoritmes. Ingenieurs en programmeurs zijn heel vertrouwd met controlestrategieën op basis van kostenfuncties. Zo'n kostenfunctie is in principe niet veel meer dan de gewogen som van variabelen zoals positie, snelheid, brandstofverbruik of omgevingstemperatuur. Het is een wiskundige manier om een reële situatie (zoals het traject van een voertuig) te vertalen naar cijfers. Sinds jaar en dag bestaan er krachtige algoritmes²⁵ die, aan de hand van feedbacklussen, toelaten om kostenfuncties te optimaliseren. Door het verschil tussen de gemeten baan en het gewenste traject van een object zo klein mogelijk te maken, slagen ingenieurs er bijvoorbeeld in om de marsrobot Perseverance met bijna chirurgische precisie te laten landen in het midden van de Jezerokrater. En dat

na een astronomische tocht van bijna 500 miljoen kilometer. Dezelfde algoritmes bepalen ook de meest efficiënte bezorggronde van je postbode, en herberekenen de snelste route wanneer je gps file of vertraging op je traject ontdekt.

Deontologische keuzes laten zich nóg gemakkelijker implementeren. Het is vrij eenvoudig om een robot te leren dat hij nooit door een rood licht mag rijden, en dat het verboden is om over een ononderbroken witte lijn te rijden. Je kunt dat soort verbodsregels als ‘harde randvoorwaarden’ of ‘no-gozones’ toevoegen bij het formuleren van het wiskundige probleem, al is dat niet altijd de meest verstandige aanpak. Soms is het wenselijk – zo niet noodzakelijk – om een witte lijn te overschrijden: bijvoorbeeld om een ziekenwagen door te laten, of om de confrontatie met een spookrijder uit de weg te gaan. In dat soort conflictsituaties kun je de harde verbodsregels afzwakken tot een extra term in de kostenfunctie, weliswaar met een heel hoog gewicht. Een zelfrijdende auto kan dan bereid zijn om door het rood te rijden, maar enkel om groter onheil te vermijden.

Mag je een mensenleven met wiskunde wegen?

De ethische dilemma’s die zelfrijdende auto’s wachten bij een botsing, laten zich dus al met bestaande en bewezen algoritmes programmeren als een kostenfunctie. Maar het precaire in het debat ligt in het oogmerk van die kostenfunctie. Je wilt immers geen brandstofverbruik beperken, de productiviteit van een machine maximaliseren of de doorlooptijd van een proces verkorten. Bij crashscenario’s voor zelfrijdende auto’s wil je, letterlijk, de schade beperken. Het kiezen van de wegingsfactoren in je vergelijking komt dus in feite neer op het verdelen van leed, en het onderhandelen over de ernst ervan. En dat is allerminst eenvoudig in cijfers te vatten. Is het leven van een kind meer waard dan

dat van een bejaarde, en zo ja: hoeveel? Mag je twee (of drie?) volwassenen verwonden om een aanrijding met een adolescent te vermijden? En wil je zelf dat je auto de betonnen muur opzoekt omdat hij de sticker 'Baby on Board' heeft gelezen op de achterraut van je voorligger?

Is het leven van een kind meer waard dan dat van een bejaarde, en zo ja: hoeveel?

Het zijn moeilijke vragen waar ethici en filosofen zich vandaag over buigen. En de antwoorden worden niet bepaald eenvoudiger zodra je dieper doordenkt. Neem nu een scenario waarin je je voorligger nog nauwelijks kunt vermijden, en het risico loopt op zeer zware verwondingen. Je kunt zwenken naar links, waardoor je een motorrijder raakt die een helm draagt. Of je kunt naar rechts uitwijken, waardoor je botst met een motorist zónder helm. Een algoritme dat getraind is om de schade te beperken, zal de voorkeur geven aan impact met gehelmden weggebruikers, omdat die nu eenmaal een hogere kans hebben om de aanrijding te overleven. Het is een pervers effect van de optimalisatie-routine, omdat plichtsbewuste burgers in het vizier komen, en manifeste verkeersovertreders buiten schot blijken te blijven. Een vergelijkbaar probleem stelt zich wanneer je beide motards vervangt door auto's: eentje met een hoge crashbestendigheid, en een andere met een povere veiligheidsscore. Om het leed te minimaliseren, zal je zelfrijdende auto willen botsen met de meest veilige omligger. Maar het zorgt er meteen voor dat goed beveiligde wagens het preferentiële doelwit worden bij crashscenario's...

De discussie wordt zo mogelijk nog moeilijker wanneer je je eigen lijf en leden in rekening moet brengen. Je bent op een vroege ochtend in de herfst onderweg naar je werk: het is donker, het regent pijpenstelen, en je rijdt op een brug over een diep ravijn.

Plots wijkt je tegenligger van zijn baanvak af, en komt met hoge snelheid naar jou toe gereden. De spookrijder blijkt een schoolbus met twintig kinderen aan boord en een chauffeur die in slaap is gevallen. Een aanrijding betekent (op zijn minst) zware verwondingen voor jezelf en een tiental kinderen. Hoort je zelfrijdende auto zich in zo'n scenario dan in het ravijn te rijden, je gewisse dood tegemoet? Je hoeft die confronterende vraag niet meteen te beantwoorden, want de buschauffeur ontwaakt nog n et op tijd om een dantesk drama te vermijden. De adrenaline pompt nog door je aderen, en je hartslag zoekt nog steeds een evenwicht wanneer je een kwartier later, in de dorpskom van de buurgemeente, remt voor een zebrapad. Je verleent voorrang aan een tiental kinderen die ter hoogte van de schoolpoort de weg oversteken²⁶. Je voelt stiekem afgunst om hun onschuld en na iviteit, maar je hoort jezelf denken: hoe mooi en onbezoedeld is de jeugd. Zie, ze glimmen van hoop. Net op dat moment ontwaart je zelfrijdende auto een onrustwekkende achterligger: een vrachtwagen komt steeds dichterbij, en lijkt geen aanstalten te maken om te remmen. Je kunt een impact – en het risico op zware verwondingen – vermijden door veilig de berm in te rijden. Dat manoeuvre maakt weliswaar de baan vrij voor een botsing tussen een tien-tonner en een tiental schoolkinderen, van wie velen het leven zullen laten. Er wacht je een bijzonder wrange keuze tussen gebroken botten of schuldig verzuim²⁷. Wat nu gezongen...?

Dat zijn opnieuw gedachte-experimenten die onze ethische intu tie bevragen (en zelfs uitdagen) over het verschil tussen doden en laten sterven, waarbij de beslissing ditmaal fatale gevolgen kan hebben voor onszelf. Jean-Fran ois Bonnefon merkt op²⁸ dat er een breed maatschappelijk draagvlak bestaat voor zelfrijdende auto's die geprogrammeerd zijn op basis van utilitaire principes, en dus schade beperken bij impact. Maar hij stelt ook fijntjes vast dat vrijwel niemand bereid blijkt om een wagen te kopen die

zichzelf daarvoor zou opofferen. Ook Patrick Lin wijst op het belang van een draagvlak: je kunt sommige wegingsfactoren nu eenmaal niet op wetenschappelijke gronden afleiden of als ingenieur berekenen. Zo'n afweging vraagt eerder om een maatschappelijke consensus over wat wenselijk is. Het vereist dat je een vergelijk vindt tussen alle mogelijke betrokken partijen, dus vaak tussen mensen met soms sterk verschillende ideologische strekkingen. Je zou de netelige vragen dus bij wijze van spreken kunnen opwerpen tijdens een soort consensusconferentie²⁹, en de grootste gemene deler hanteren bij het bepalen van delicate wegingsfactoren. Per slot van rekening is ook de huidige wegcode een bijzonder brede consensus. Het is een set van regels die (althans een diagonale doorsnede van) onze maatschappij aannemelijk acht, en die dankzij een wetgevend kader afdwingbaar is.

Vrijwel niemand blijkt bereid om een wagen te kopen die zichzelf zou opofferen.

The Moral Machine: Magda of Dany?

Over de ethiek van zelfrijdende auto's is al een consensusconferentie georganiseerd. En wel een omvangrijke, met meer dan 40 miljoen deelnemers uit 233 landen. Die indrukwekkende cijfers verraden dat het om een online conferentie gaat: The Moral Machine is een webplatform³⁰, ontwikkeld door het Massachusetts Institute of Technology (MIT), om de brede bevolking te bevragen over morele beslissingen van intelligente machines in het algemeen, en zelfrijdende auto's in het bijzonder. Het platform lijkt op een jukebox met crashscenario's en trolleyproblemen, waarbij achtereenvolgens dieren, kinderen, zwangere vrouwen en zelfs Nobelprijswinnaars de straat

oversteken. Het dwingt je te kiezen tussen daklozen die een stoplicht negeren of Olympische atletes die dat niet doen. Het daagt je morele intuïtie uit door katten te wegen met kinderen, of twee jongedames met drie oudere heren. De bevraging wordt gaandeweg confronterender. Na verloop van tijd doet ook een betonblok zijn intrede, waardoor je zelfrijdende wagen de existentiële vragen over zelfopoffering niet langer uit de weg kan gaan. Is je eigen leven meer waard dan het geluk van een koppel van middelbare leeftijd? En verandert je mening indien dat koppel de verkeersregels overtreedt? Ben je bereid je eigen leven en dat van je zoon op te offeren om een kroostrijk gezin te redden? En hoe kies je in vredesnaam tussen een priester en een zwangere vrouw, of een jong meisje met haar dikke vader?

The Moral Machine is een sterk staaltje empirische ethiek, en peilt op lepe wijze naar voorkeuren in geslacht, gezondheid, leeftijd en sociale status. Het legt ook inzichten bloot over het belang dat je hecht aan rechtschapenheid, betrokkenheid, zelfbescherming en – opnieuw – het verschil tussen doden en laten sterven. Je mag The Moral Machine dus gerust lezen als een wereldwijd referendum over de morele principes voor zelfrijdende auto's. Een volksbevraging die meteen de delicate wegingsfactoren voor onze ethische kostenfuncties kan bepalen aan de hand van crowdsourcing. De resultaten van het onderzoek werden gepubliceerd in het gezaghebbende tijdschrift *Nature*³¹, en zijn even verrassend als ontluisterend. Er blijkt immers helemaal geen universele ethiek te bestaan!

De resultaten toonden sterk verschillende voorkeuren volgens nationaliteit, maar ook correlaties met culturele afkomst en ideologische overtuiging. Fransen, bijvoorbeeld, hadden een uitgesproken voorkeur om kinderen te sparen, terwijl deelnemers

uit China, Japan en Taiwan veel meer geneigd waren de ouderen te beschermen. Westerse landen toonden dan weer minder begrip voor verkeersovertreders dan ontwikkelingslanden. Voor individualistische culturen (zoals bij de Amerikanen en de Britten) weegt het aantal levens dat op het spel staat zwaar door in de beslissing. En het zal misschien niet verbazen dat Japanners met voorsprong de meest gulle inborst hebben rond zelfopoffering, terwijl Chinezen net tot elke prijs een zelfrijdende auto willen die de inzittende beschermt. In hun helikopterzicht konden de onderzoekers drie clusters kleuren op de wereldkaart. Westerse culturen (zoals in Noord-Amerika en de meeste Europese landen) dragen een sterke invloed van het christendom, en lijken daar hun morele waarden over respect, verantwoordelijkheid en onbaatzuchtigheid aan te ontlenen. De leer van Confucius schemert nog steeds door in de oosterse cluster (met landen als Japan, Indonesië en Pakistan), waar de antwoorden geïnspireerd worden door deugdelijkheid, rechtvaardigheid en oog voor het groter belang. En deelnemers uit de zuiderse cluster (met voornamelijk landen in Centraal- en Zuid-Amerika) gaven dan weer een statistisch significante³² voorkeur aan het sparen van jongeren, vrouwen en welgestelden.

Universele ethiek bestaat niet!

Een brede bevraging zoals The Moral Machine leert ons dat – voor zover dat al wenselijk³³ zou zijn – het onmogelijk is om de ethische code van een zelfrijdende auto te programmeren aan de hand van de stem van het volk. Het antwoord klinkt immers meerstemmig, en dreigt aanleiding te geven tot babylonische spraakverwarring vanaf het eerste zebra-pad. Je zou natuurlijk kunnen overwegen om regionale dialecten in de programmeer-

taal toe te laten: per slot van rekening rijdt men in Groot-Brittannië en Malta ook aan de ‘verkeerde’ kant van de weg. Maar het dreigt de logistiek er niet op te vereenvoudigen, en je durft als Duitser wellicht nooit meer rond te rijden in Japan. De zelfrijdende auto moet dus tot nader order, ondanks een bibliotheek aan filosofische essays, navigeren op een troebel moreel kompas.

Niettemin tekenen enkele krijtlijnen zich af, dus sta mij toe mijn voortschrijdend inzicht over de ethiek van zelfrijdende auto’s samen te vatten. Een Google Car programmeren aan de hand van louter deontologische regels – dus met de stenen tafelen van Mozes als wegcode – lijkt onwerkbaar, omdat dit (te) strakke keurslijf heel wat onheil in de hand kan werken. In conflictsituaties moet je toelaten dat ‘harde’ randvoorwaarden kunnen worden omgeboogen in ‘buigzame’ voorschriften. Je kunt zulke regels dan overtreden, zolang dat bijdraagt tot het hogere doel: de schade beperken. Het verdelen van leed wordt dan wiskundig gereduceerd tot het minimaliseren van een kostenfunctie, wat voor ingenieurs en programmeurs een eitje is. Het bepalen van de wegingsfactoren bij die wiskunde blijkt echter een hardgekookt probleem. Wereldwijd bestaan verschillende strekkingen, die terug te voeren zijn tot een verschil in afkomst, cultuur en ideologie. Maar een aantal principes geniet brede overeenstemming. Zo is vrijwel iedereen het erover eens dat het moreel meer verwerpelijk is te doden dan te laten sterven. Ook in afgezwakte en toegepaste vorm³⁴ vindt dat idee een brede consensus: negatieve of ‘koude’ rechten (‘Ik mag niet omvergereden worden’) wegen moreel zwaarder door dan positieve of ‘warme’ rechten (‘Ik moet geholpen worden’). Een derde aspect dat velen beamen is het belang van betrokkenheid. Je tegenligger, of de weggebruiker op het zebrapad, neemt willens en wetens deel aan het inherent risicovolle verkeer. Hij heeft bij een crashscenario dan ook een cynisch streepje voor op de toerist op het terras. Je zou die mate van betrokkenheid zelfs

kunnen becijferen, en weggebruikers toelaten te onderhandelen over de aanspraak die zij maken op blikshade. Er blijven nog vele vragen, maar dat zijn alvast enkele leidraden die ons kunnen gidsen wanneer we de blauwdruk van het geweten van zelfrijdende auto's uittekenen.

Nog meer obstakels op het traject

Tot hiertoe heb ik vooral gefocust op de ethiek van zelfrijdende auto's bij crashscenario's, maar er rijzen nog veel meer ethische dilemma's rond autonome voertuigen. Een eerste is alvast die focus op botsingen: omdat ze onze ethische intuïtie zo sterk interpellieren, hebben vraagstukken met overstekende oma's en aanstormende schoolbussen veel aandacht gekregen. Buitensporig veel aandacht, volgens sommigen³⁵, en dat vormt op zich een ethisch bezwaar. Zelfrijdende auto's zijn immers lerende machines, die getraind worden aan de hand van artificiële intelligentie (AI) of neurale netwerken op steroïden. Hoewel artificiële intelligentie en (deep) machinelearning vandaag buzzwords zijn, kun je het voortschrijdend inzicht van zelfrijdende auto's nog het best vergelijken met de manier waarop je honden traint. Een hond is allerminst een talig wezen, maar kan na verloop van tijd wel de vraag om een 'pootje' onderscheiden van het commando 'zit!'. Je hond leert dat dankzij een zelfversterkend systeem van positieve leerervaringen. Weglopen, blaffen en zelfs zitten bij de klanken 'pootje' worden bestraft, terwijl het opheffen en plooiën van de juiste voorpoot om onduidelijke redenen met lekkers wordt beloond. Na voldoende veel herhalingen en het nodige engelengeduld bij het baasje, moet het mogelijk zijn om iedere hond het onderscheid aan te leren tussen zitten en pootjes uitdelen. Zeldzame rassen kun je dankzij die beproefde techniek zelfs het zwijgen opleggen. De zelfrijdende auto wordt op een vergelijkbare manier geconditioneerd.

Door voldoende blootstelling aan snelheidslimieten, stopborden, verkeerslichten en andere weggebruikers kun je de pavloviaanse reflexen kweken die vereist zijn om een baanvaste robot te worden. Maar je bereikt die stuurvaardigheid pas na enorme hoeveelheden trainingsdata en simulaties. En daar ligt net het gevaar van een disproportionele focus op crashscenario's. Zo dreig je een auto op te leiden die nauwelijks kan sturen, maar wel met verve weet te botsen. Het is wellicht wenselijker om het leeuwendeel van de opleiding te besteden aan stuurmanskunsten en verkeersvaardigheid. Zo ontwerp je immers een door de wol geverfde artificiële chauffeur, die misschien nog nauwelijks hoeft te botsen. Dus ook de keuze van het trainingsprogramma voor zelfrijdende auto's vergt beslissingen met een ethische inslag.

Nog een hele vloot vervelende vragen

De opkomst van zelfrijdende auto's werpt de schaduw van nog andere maatschappelijke en morele problemen voor zich uit. De boordcomputer moet oordeelkundig worden getraind, maar wij dreigen als menselijke chauffeur het autorijden te verliezen. Dat kan, zeker in de vroege fase van 'hybride' verkeer, gevaarlijke situaties scheppen. Ook de interactie tussen mens en machine zal op de proef worden gesteld. Zelfrijdende auto's willen schade beperken, en zullen zich dus conflictvermijnd gedragen. Een diagonale doorsnede van het hedendaagse verkeer leert dat zo'n gedrag uitnodigt tot allerhande herrie. Fietsers of voetgangers zouden van dat defensieve karakter misbruik kunnen maken. Door lepe manoeuvres (of zelfs ostentatief treiteren) kunnen ze de auto degraderen tot zwakke weggebruiker.

De sociale repercussies vormen een voor de hand liggende bezorgdheid. Indien zelfrijdende voertuigen brede ingang vinden, dreigen miljoenen busbestuurders, taxi- en vrachtwagenchauf-

feurs werkloos te worden. Andere maatschappelijke gevolgen zijn subtieler. Het ligt in de lijn der verwachtingen dat autonome bestelbusjes zullen worden ingezet om boodschappen aan huis te bezorgen. Dat kan de doodsteek betekenen voor de kleine kruidenier, maar ook de vorming van ‘voedselwoestijnen’³⁶ in de hand werken. Door zijn belofte van veilige wegkapitein zou de zelfrijdende auto ook (overmatig) alcoholgebruik kunnen aanmoedigen.

Ook de technologie is voer voor discussie. Om de verkeersstroom te optimaliseren, zullen zelfrijdende auto’s heel veel data met elkaar moeten uitwisselen. Zijn wij bereid te allen tijde onze positie – maar ook ons traject, onze bestemming en wellicht nog andere metadata – continu met onze omgeving te delen? Ons voertuig kan binnenkort een aanslag worden op onze privacy. Bugs of kwetsbaarheden in de software kunnen er dan weer toe leiden dat carjacking straks gelijkstaat aan *carhacking*. En als je auto virtueel wordt gekaapt: hoe kun je die dan nog tot stilstand brengen...?

Ons voertuig kan binnenkort een aanslag worden op onze privacy.

Nog een vaak gestelde vraag is die van aansprakelijkheid. Bij een botsing verschuift de verantwoordelijkheid van de inzittende naar de constructeur – of programmeur – van de auto. Dat zal nieuwe wet- en regelgeving vergen. Het betekent meteen ook een aardverschuiving voor de verzekeringsmaatschappijen. Sommigen opperen dat, indien de technologie op punt staat, autoverzekeraars zullen uitsterven. Ook de budgetten van lokale besturen kunnen krimpen, omdat het aantal verkeersboetes spectaculair zal dalen. En als meer morbide variant op datzelfde thema: bij vrijwel perfecte zelfrijdende auto’s zullen er nog amper orgaandonoren zijn.

Dat laatste mag ons er uiteraard niet van weerhouden de technologie verder te optimaliseren. Zelfrijdende auto's bieden namelijk het benijdenswaardige vooruitzicht op filevrije snelwegen, en de mogelijkheid te shoppen zonder de gebruikelijke herrie rond parkeerplaatsen en -boetes. Bovendien zijn zelfrijdende auto's intelligent en robuust: ze zijn onvermoeibaar, en vrij van drank of drugs. Autonome voertuigen zullen het aantal aanrijdingen drastisch doen dalen. Waymo gelooft dat het aantal verkeersslachtoffers (met momenteel een zeer zware tol van meer dan 1,3 miljoen doden per jaar) bij volledig autonoom verkeer met minstens 90 procent zal dalen. Alleen al daarom is de snelle en grootschalige introductie van zelfrijdende auto's in ons verkeer moreel meer dan wenselijk.

HOOFDSTUK 2

Big data of big brother?

Leesexemplaar

Worden wij reeds
geschaduwd door
duistere algoritmes,
terwijl onze uitstoot
van data ons net
doorzichtig maakt?