

Mireille Hildebrandt & Antoinette Rouvroy (eds.), *Law, Human Agency, and Autonomic Computing*

Mark Coeckelbergh

Mireille Hildebrandt & Antoinette Rouvroy (eds.), *Law, Human Agency, and Autonomic Computing* (Abingdon/New York: Routledge, 2011), 248 p.

Moderne rechtssystemen veronderstellen ideeën over menselijk handelen, identiteit, subjectiviteit, intentionaliteit en de persoon die steeds ter discussie gesteld worden door filosofen die zich daarvoor – al dan niet terecht – beroepen op de wetenschap. Zo stellen sommigen dat de neurowetenschappen de idee dat mensen een vrije wil hebben, problematiseren. Maar het is niet alleen wetenschappelijke kennis die de rechtsfilosofie uitdaagt: er worden ook *dingen* gemaakt, *artefacten* die traditionele ideeën over mens en menselijk handelen op de proef stellen. Het gaat dan om kunstmatige entiteiten zoals robots en, meer algemeen, om kunstmatige systemen die steeds meer ‘autonoom’ worden: door middel van artificiële intelligentie kunnen ze zichzelf besturen en worden ze steeds zelfstandiger, dat wil zeggen ze hebben steeds minder nood aan direct menselijk ingrijpen. Maar, zo vragen steeds meer filosofen zich af, betekent dit niet dat de afstand tussen mens en ding, en tussen handelend subject en be-handeld object kleiner wordt? Wat betekenen de traditionele filosofische begrippen nog waarmee we dachten de mens te definiëren? Wat betekenen deze ontwikkelingen voor de menselijke verantwoordelijkheid?

Een recent gepubliceerde bundel van Hildebrandt en Rouvroy biedt een interessante en significante bijdrage aan deze discussie door techniekfilosofen en rechtsfilosofen te laten reflecteren over wat ‘autonomic computing’ en ‘ambient intelligence’ betekenen voor centrale rechtsbegrippen zoals subjectiviteit, mens en persoon. Daarbij komen filosofen van allerlei pluimage aan het woord, met een nadruk op de (post)fenomenologische en poststructuralistische traditie. Gezien de dominantie van de analytische filosofie in dit soort discussies verdient de bundel daarom alleen al aandacht.

Het boek begint in de Amsterdamse urinoirs, waar gepoogd wordt menselijk gedrag te beïnvloeden. De Amerikaanse filosoof Don Ihde ontwaart er een dubbele hermeneutiek: we worden aangezet tot een bepaald gedrag en de technologie probeert dus ons gedrag te voorspellen en te beïnvloeden, maar tegelijkertijd proberen wij uit te vogelen wat de technologie van ons ‘denkt’ en van ons ‘wil’. Daarmee is de toon gezet voor de volgende techniekfilosofische bijdragen over ‘smart’ technologie en mens-techniekrelaties. Verbeek denkt na over een ethiek die rekening zou moeten houden met de verwevenheid van mens en techniek, en De Mul en Van den Berg beweren dat menselijk handelen altijd al tot op zekere hoogte

‘remote controlled’ is geweest en dus van buitenaf gestuurd. De conclusie van deze eerste bijdragen lijkt te zijn dat we ons misschien wel niet al te veel zorgen moeten maken over bedreigingen van ‘agency’ door techniek: enerzijds omdat, zoals Ihde argumenteert, er nog helemaal niet zoiets bestaat als ‘autonomic computing’ en we kritisch moeten zijn over deze ‘hype’ en ‘fantasie’, zeker in het licht van de discussie over artificiële intelligentie (p. 17); anderzijds omdat we de mens niet kunnen definiëren los van techniek, zoals Verbeek, De Mul, en Van den Berg laten zien.

Brownsword daarentegen stelt voor om eerst onze ideeën over autonomie en verantwoordelijkheid scherp te stellen en verdedigt de idee van een gemeenschap van rechten. Modern denken wordt echter in de volgende hoofdstukken weer bekritiseerd: Durante, Kang en Rouvroy argumenteren tegen het cartesiaanse subject en voor het erkennen van ‘embodiment’, en brengen de ‘virtualiteit’ van agency ter sprake. Rouvroy brengt ook terecht de *politieke* dimensie van het probleem ter sprake. Het recht komt echter pas echt op de voorgrond in het derde deel, waar Hildebrandt de relatie bespreekt tussen de voorwaarden van ‘criminal accountability’ en ‘autonomic computing’. Erodeert de laatste onze capaciteit om autonoom te handelen, bijvoorbeeld omdat we ons niet bewust zijn van wat er gaande is? Kallinikos laat zien dat het antwoord op deze vraag ambigu is: zo kan gegevensverwerking volgens hem net zo goed die capaciteit vergroten. Rodotà heeft het over de constructie van identiteiten door ‘autonomic computing’ en toont de uitdagingen voor wetgever en gerecht. Ten slotte bespreekt Mathias de impact van ‘autonomic computing’ op het lichaam, de persoon, en identiteit, en volgt nog een epiloog van Rouvroy.

Dit boek moet verwelkomd worden: het moedigt aan om op een filosofisch originele, empirisch geïnformeerde en historisch-kritische manier naar techniek te kijken, het geeft een goed overzicht van mogelijke problemen die ‘autonomic computing’ oproept voor enkele voor het recht belangrijke filosofische begrippen, het betreft verschillende filosofische tradities bij dit project, en probeert twee (sub)disciplines te overbruggen die nog te weinig met elkaar praten. Zoals Hildebrandt aangeeft in haar inleiding, kan dit veel opleveren voor beide.

De vraag is wel of dit hier gelukt is: de nadruk ligt in deze bundel toch meer op de techniekfilosofie en op het denken over agency op zich; de band met recht en de (dialoog met de) rechtsfilosofie komt wat minder goed uit de verf. Bovendien komen zoals gesuggereerd het merendeel van de bijdragen uit een welbepaalde (techniek)filosofische traditie en samen stellen ze de moderne rechtsbegrippen ter discussie. Waarom echter deze keuze gemaakt is (als het een keuze was) en waarom met dit thema niet meer gedaan is om de bijdragen in deze bundel te integreren en meer focus te geven, daar heeft de lezer het raden naar. Verder stelt zich na de eerste bijdragen de vraag wat er nu eigenlijk zo dringend aan is om na te denken over ‘autonomic computing’: als de problemen die besproken worden, ook dikwijls bij andere technieken en artefacten aan de orde komen (ook bij low-tech zoals de afbeelding van de vlieg op het urinoir) en als ‘autonomic computing’

Mark Coeckelbergh

slechts de natte droom is van een (ouderwetse?) systeembeheerder, waarom moeten we het er dan überhaupt over hebben?

Ten slotte moet de kanttekening gemaakt worden dat wat betreft het filosofische begrippenkader het bestreken gebied wel erg groot is: het gaat over identiteit, subjectiviteit, agency, persoon, intentionaliteit, de mens, enzovoort. Iets meer focus had wellicht de door de samenstellers beoogde integratie ten goede gekomen. Met het begrip 'human agency' wordt gepoogd alles aan elkaar te rijgen, maar deze term zelf behoort al tot een bepaald discours over de mens (toch voornamelijk in de Engelstalige analytische filosofie) en kan dus niet helemaal de rol van neutrale noemer waarmaken. Dit wordt het duidelijkst als het gaat over hybridisatie en als de notie van het liberale subject wordt bekritiseerd: hier wordt niet alleen *een bepaald soort* begrip van agency op de korrel genomen maar komt ook het begrip 'agency' zelf onder druk. Sommige bijdragen kunnen dan ook nog moeilijk bestempeld worden als bijdragen over agency. Anderzijds maakt deze confrontatie van verschillende filosofische tradities de bundel nu net spannend: veel 'dialogen' beperken zich gewoonlijk tot de eigen kring. Het boek onderscheidt zich positief door een waaier van perspectieven aan bod te laten komen, ook al blijkt dat uiteindelijk de diversiteit toch beperkt is en dat de nagestreefde dialoog en integratie tussen écht verschillende benaderingen en tussen subgebieden van de wijsbegeerte toch nog meer wensdroom dan realiteit zijn.