

Een commentaar bij het artikel 'Vrije wil, verantwoordelijkheid en broncontrole' van Pieter Gilles

Jan Verplaetse en Delphine De Smet¹

Het manipulatie-argument (MA) is één van de krachtigste wapens waarover een incompatibilist beschikt in het debat over vrije wil en verantwoordelijkheid (Verplaetse, 2011; Pereboom, 2001). Incompatibilisten geloven niet dat een causaal-deterministisch wereldbeeld verzoenbaar is met onze intuïtieve noties van verantwoordelijkheid, verdienste en schuld. Deze morele noties steunen op het bestaan van een echte, libertaristische vrije wil die je de mogelijkheid geeft om anders te handelen, bepaalde acties of beslissingen te vermijden en bijgevolg niet te doen waartoe je in een causaal-deterministische wereld geprogrammeerd bent. Wanneer we een misdadiger verachten om zijn misdaad of een laureaat prijzen omwille van zijn verdienste, dan zijn onze blaam en lof gebaseerd op de vooronderstelling dat beide de mogelijkheid hadden om anders te handelen. Die vrijheid maakt hen verantwoordelijk. In een wereld waarin die vrijheid ontbreekt, een causaal-deterministisch universum, verdwijnt die verantwoordelijkheid. Let wel, de incompatibilist zegt niet dat onze kosmos causaal en deterministisch in elkaar zit. Sommige incompatibilisten ontkennen dit namelijk. Zij denken dat het bestaan van een echte vrij wil niet uitgesloten is. Incompatibilisten geloven alleen dat *indien* het universum zo in elkaar steekt, dit het einde betekent voor deze fundamentele morele noties. Is er sluitend bewijs dat alles causaal en deterministisch werkt, dan moeten we van deze concepten afscheid nemen.

Zo ver zijn we nog niet. Compatibilisten zoals Pieter Gilles, maar in Vlaanderen ook Etienne Vermeersch en Maarten Boudry en in Nederland Marc Slors en Maureen Sie verwerpen deze radicale gevolgtrekking. Verantwoordelijkheid, schuld en verdienste blijven overeind ook al zou blijken dat ons universum – en bijgevolg ook ons brein – causaal-deterministisch in elkaar zit. Compatibilisme is een comfortabele positie omdat ze wetenschap verzoent met traditionele morele noties. Een wetenschappelijk wereldbeeld botst dan niet langer met onze ethische concepten. Je kunt criminelen blijven verwijten en verdienstelijke mensen blijven loven, al waren ze niet in staat om hun slechte of goede daden te vermijden.

1. Vakgroep Metajuridica, Privaatrecht en Ondernemingsrecht – Rechtsgeleerdheid UGent.

Morele concepten steunen niet op dubieuze of zelfs achterhaalde filosofische denkbeelden, zoals de idee dat er zaken zijn die niet veroorzaakt worden of geen materiële of fysische oorzaken hebben (acausaliteit). Je hoeft dus geen dualist, van welke subtiele soort ook, te zijn om te geloven in verantwoordelijkheid. Sommige compatibilisten vinden de omgekeerde bewering zelfs zo bizar dat ze de incompatibilistische zienswijze niet begrijpen. In een discussie met neurofilosofe Patricia Churchland vorige herfst in Nederland kregen incompatibilistische opposanten onophoudelijk te horen dat er niets bestaat dat ontsnapt aan de causale orde. Hoe naïef, ouderwets en filosofisch problematisch is het niet om te denken dat moderne wetenschap ruimte laat voor dualisme. Aan één David Chalmers hebben we genoeg. Al zijn we het wat dit betreft met Churchland eens, ze begrijpt de incompatibilistische zienswijze verkeerd. Incompatibilisten zeggen niet dat ons brein acausaal en indeterministisch werkt, maar wel dat noties zoals verantwoordelijkheid en schuld zulke werking veronderstellen. Niet de wereld is dualistisch maar onze morele basisconcepten stammen uit een dualistische erfenis.

Alleen: is dat wel zo? Compatibilisten die wel dit onderscheid begrijpen, zijn niet overtuigd. Centraal in het meningsverschil staat het manipulatie-argument (MA) (Verplaetse 2011; Mele 2008; Pereboom 2001; Kane 1996; Taylor 1974). Het vervelende – sommigen zullen zeggen: het boeiende – aan dit argument is dat het slechts een gedachte-experiment is. Of verantwoordelijkheid bestaat of niet kan je niet waarnemen met een microscoop of met een hersenscanner. Verantwoordelijkheid is een *intuïtie*. Je vindt een persoon X of Y verantwoordelijk voor wat X of Y gedaan heeft. Maar wanneer vinden we dat? En wanneer verdwijnt dit gevoel of deze intuïtie. Met behulp van MA trachten incompatibilisten aan te tonen dat deze intuïtie verdwijnt in een causaal-deterministisch universum. Echter, omdat we ons zulk universum moeilijk kunnen voorstellen – onze morele basisnoties zouden wel eens dualistisch kunnen zijn – maakt MA gebruik van een vereenvoudiging. Niets vat het beeld van zulk universum sterker dan een manipulatie. Als je iemand manipuleert is er altijd een oorzaak die buiten de gemanipuleerde ligt (causaliteit) en is er slechts één gevolg, met name de uitkomst die de manipulator beoogt (determinisme). Opdat de gemanipuleerde niet zou protesteren tegen de manipulatie weet hij niet van de manipulatie af. MA heeft iets van de film *The Manchurian Candidate*, oorspronkelijk gebaseerd op een roman van Richard Condon (1959) maar in 1962 en 2004 verfilmd door Hollywood, waarin een zoon gehersenspoeld wordt door zijn eigen moeder. De gebruikte manipulatietechniek volgt trends. In de jaren '60 was hypnose de meest geloofwaardige *brainwash*-methode, terwijl dat in het nieuwe millennium hersenimplantaten zijn geworden. Eenzelfde schommeling in populariteit zie je ook in de verschillende versies van het manipulatie-argument. Hersenmodulatie is nu hipper dan psychotherapie.

Hoezeer de techniek ook verschilt, de essentie blijft dezelfde. X voert een immorele daad uit die afkeuring oproept en in normale omstandigheden aan X wordt toegerekend. De toehoorders of lezers van MA vinden X aanvankelijk verantwoordelijk. Maar krijgen die te horen dat X dusdanig gemanipuleerd werd dat niet zozeer X maar de manipulator besliste om die immorele daad uit te voeren, waarbij X slechts een uitvoerend instrument is, dan vervalt die toerekening van verantwoordelijkheid. Vraag je aan je toehoorders of X verantwoordelijk is, dan zeggen zij doorgaans 'neen'. We zeggen uitdrukkelijk 'doorgaans' omdat er wel altijd een kleine minderheid is die vindt dat X verantwoordelijk blijft. De redenen zijn divers. Sommige toehoorders zijn van oordeel dat elke dader meteen verantwoordelijk is, al heeft hij geen morele schuld. Of het nu baby's, honden of volwassen mensen zijn die iets verkeerd doen, ze zijn verantwoordelijk. Anderen vinden X verantwoordelijk omdat ze bang zijn voor de maatschappelijke gevolgen als ze X onverantwoordelijk achten. Moeten we X vrijlaten? X interneren? Weer anderen vinden X misschien niet verantwoordelijk voor de immorele daad, maar wel voor de weg die naar het uitvoeren van die daad heeft geleid. Om het gedachte-experiment wat realistischer te maken, valt de immorele daad van X meestal niet uit de lucht, maar krijgt de uiteindelijke misdaad maar geleidelijk aan gestalte. Voor die voorafgaande keuzes achten de toehoorders X wel verantwoordelijk. Nooit heb je een zaal toehoorders die unaniem van oordeel is dat X in dit manipulatieverhaal niet verantwoordelijk is. Er is altijd wel wat verdeeldheid. Overigens is MA tot op heden niet uitgebreid wetenschappelijk getoetst. De enkele empirische vertalingen die experimentele filosofen ondernamen, kampen met methodologische en conceptuele tekorten (Sripada 2012; Feltz 2013; Nahmias, Shepard & Reuter 2014; Nahmias 2015; voor kritiek, zie Gorin 2013). Het wordt tijd om dit aan de hand van meer systematisch empirisch onderzoek beter te doen.

MA heeft een stapsgewijze opbouw. In een eerste verhaal heeft de manipulatie een beperkt bereik. De manipulator stuurt slechts één vermogen van X gedurende een korte tijd, maar niettemin voldoende om de misdaad te begaan. Bijvoorbeeld, X duwt op een (verkeerde) knop waardoor hij iemand doodt. De motoriek van X werd slechts even in de war gestuurd, maar genoeg om iemand te vermoorden. In de volgende verhalen neemt die manipulatie toe. Meerdere vermogens worden gedurende langere periodes gemanipuleerd, zonder dat X hiervan afweet. Telkens wordt aan de toehoorders of lezers gevraagd of X nog wel verantwoordelijk is, als je weet dat de manipulatie het werk deed. Niet alleen de externe vermogens van X (motoriek) maar ook zijn interne vermogens zijn voorwerp van manipulatie. Zijn verlangens, wilsuitingen, motieven, overwegingen, afwegingen, gedachten: dit alles kan worden gestuurd. In een causaal-deterministische wereld kan niets van manipulatie uitgesloten worden. De meest interne drijfveer heeft een mate-

riële oorsprong en kan dus gestuurd worden. De kracht van MA steunt uitiem op de allerlaatste stap die je neemt. Eenmaal je duidelijk gemaakt hebt dat het niet uitgesloten is dat X zijn hele leven al gemanipuleerd werd en dat die manipulatie ook betrekking heeft op alle mentale ervaringen die hij had, vraag je in een laatste verhaal om de manipulatie volledig weg te denken en je een wereld voor te stellen waarin alles causaal-deterministisch werkt, net zoals in een universum met manipulatie maar dan alleen met normale oorzaken en gevolgen die zich op verschillende niveaus bevinden, van biochemische wetmatigheden tot sociologische patronen. Op dat moment begrijpen de toeschouwers of lezers dat een universum met manipulatie niet van een universum zonder manipulatie verschilt. Een universum met manipulatie maakt het je gemakkelijker om je een causaal-deterministische kosmos voor te stellen. Natuurlijk is zo'n kosmos met een manipulator simplistisch, onrealistisch, slechte *science fiction*, maar dat is het punt niet. Wat de elementaire basismechanisme betreft is er tussen beide geen verschil. Wat nog belangrijker is: wie denkt dat in een universum met manipulatie 'X niet verantwoordelijk is', moet logischerwijze ook concluderen dat in een universum zonder manipulatie er geen verantwoordelijke mensen kunnen bestaan, waarmee aange-toond is dat causaal-determinisme verantwoordelijkheid, schuld en verdienste uitsluit.

MA steunt op twee premissen. Vooreerst is er de logische veralgemening. Als je vindt dat een kleine en kortstondige manipulatie iemand onverantwoordelijk maakt, dan moet je ook vinden dat een grote en langdurige sturing iemand zijn verantwoordelijkheid afneemt. De manipulatie is alleen toegenomen. Een kwantitatief verschil kan geen verschil in beoordeling rechtvaardigen. Wie dus al in het allereerste verhaal toegeeft dat X niet verantwoordelijk is, zit aan deze conclusie vast tot het voorlaatste verhaal. Ten tweede is er de gelijkstelling van een universum zonder manipulatie met een universum met manipulatie, zoals die tot uiting komt in het laatste verhaal. MA tracht aannemelijk te maken dat een gemanipuleerde X gelijk is aan een niet-gemanipuleerde X. Of preciezer gesteld: dat in een causaal-deterministische wereld X altijd gemanipuleerd is, weliswaar niet door een demonische hersenchirurg maar door natuur, cultuur, omgeving, etc. Elke menselijke beslissing is de uitkomst van subtiele manipulatie waarover de beslisser geen recht van hoger beroep heeft of dat hoger beroep is zelf weer het gevolg van causatie. Door alle veroorzaking als manipulatie te begrijpen illustreert MA wat het betekent om in een causaal-deterministische wereld te leven. Voor incompatibilisten is dat logisch en intuïtief een wereld zonder vrije wil en verantwoordelijkheid.

MA kun je op verschillende manieren bekritisieren (King 2013; Kearns 2012; McKenna 2008). Veel kritiek heeft betrekking op de tweede premisse. Ergens zou er toch een conceptueel verschil blijven bestaan tussen de gewone wereld zonder

sturing en het gemanipuleerde universum, tussen normale X en manipulatie-X, waardoor we gerechtvaardigd zijn om de verantwoordelijkheid van beide personen anders te beoordelen. De kritiek van Pieter Gilles mikt evenwel op de eerste premisse, de logische veralgemening. De reden waarom veel toehoorders X in het eerste verhaal niet verantwoordelijk houden, gaat terug op een manipulatie van buitenaf. Wat X wil, is bijvoorbeeld niet wat X eigenlijk wil, maar datgene wat de hersenwetenschappers willen dat X doet of wil. Zelfs al zorgen die wetenschappers ervoor dat X toch wil wat hij wil, al is het maar voor even, dan blijft die wil een opgedrongen wil en dat maakt X net niet verantwoordelijk. Zolang de verhalen van MA deze vorm hebben, zo speculeert Gilles want op empirische gronden weten we het eigenlijk niet, dan zullen mensen geneigd zijn om X van elke verantwoordelijkheid te ontslaan. Maar van zodra de manipulatie verder reikt en langer duurt, en X zich werkelijk identificeert met die opgedrongen wil omdat een al evenzeer opgedrongen hogere wil instemt met die lagere wil, zouden mensen de neiging hebben om X wel verantwoordelijk te achten. Er is een punt in de sturing van X waarop je intuïties overgaan van 'X is onverantwoordelijk' omwille van een beperkte, tijdelijk en als vreemd ervaren manipulatie naar 'X is verantwoordelijk' wegens een onbeperkte, langdurige of altijd al actieve en als eigen ervaren manipulatie. Hoewel de kritiek van Gilles ons niet altijd zo duidelijk is, vermoeden we dat hij zoiets bedoelt. Hij zou bij het derde verhaal van Verplaetse's versie van MA Jan verantwoordelijk achten voor het doden van Paul (Verplaetse 2011). De ene manipulatie is de andere niet. Hoe rijker, complexer en langduriger de manipulatie werkt, hoe meer we geneigd zullen zijn om zelfs een gemanipuleerde X verantwoordelijk te achten. Ook al kan X niet anders handelen, hij blijft gemanipuleerd, er is grote kans dat hij als verantwoordelijk wordt beschouwd. Zijn verlangens zijn stabiel, zijn wil is consistent, zijn beslissingen gemotiveerd, en zijn gedrag in overeenstemming met hetgeen hij altijd al wilde. Bouwen hersenchirurgen een mensenbrein met een evenwichtige persoonlijkheid, dan zal die fabricatie onze intuïtie van verantwoordelijkheid niet beïnvloeden. Om die reden, zo zouden we hieraan kunnen toevoegen, vinden we psychopaten ondanks alle ontdekkingen over hun aparte genetica, breinen, en omgevingsfactoren nog altijd toerekeningsvatbaar en dus verantwoordelijk. Hun persoonlijkheidsstoornis is niet bruusk genoeg om onze tegengestelde intuïtie te activeren.

Drijven onze intuïties van verantwoordelijkheid, verdienste, schuld niet zozeer op impliciet dualisme of veronderstelde acausaliteit, maar op afwezigheid van vreemde, storende en verwarrende invloeden die tijdelijk en beperkt zijn, dan heeft Gilles een krachtige kritiek op MA in het bijzonder en op de incompatibilistische positie in het algemeen. Het diepere twistpunt is het volgende: onderscheiden we 'X is verantwoordelijkheid' van 'X is onverantwoordelijkheid' op grond van een metafysische aanname (*acausaliteit versus causaliteit*) of op basis van

een psychologisch vertrekpunt (*normaliteit versus abnormaliteit*)? Wie van beide gelijk heeft, dat weten we momenteel niet. Elke stap in de richting van een wetenschappelijk antwoord moeten we aanmoedigen. Wel weten we nu al welke argumenten niet welkom zijn. Zo is het erg verleidelijk om op deze feitelijke vraag een normatief antwoord te geven. Zo is het niet denkbeeldig dat incompatibilisten uiteindelijk wel erkennen dat veel gewone mensen gebruikmaken van een psychologisch criterium, maar dat ze vinden dat ze eigenlijk gebruik horen te maken van een metafysische toetssteen. Wie 'X is onverantwoordelijk' zegt bij een kleine manipulatie moet volgens de strenge wetten van logica ook 'X is onverantwoordelijk' zeggen bij een grotere manipulatie, en zeker bij een volledige manipulatie. Alleen verzekert niets ons dat onze intuïties ook logische wetten volgen. Ook garandeert niets ons dat er maar één intuïtie is. Misschien zijn gewone mensen verdeelt tussen "metafysici" en "psychologen" of verschilt de toetssteen van situatie tot situatie. Dat is de tol die filosofen moeten betalen voor het verlaten van hun leunstoel. Maak je grote filosofische disputen afhankelijk van wat de man in de straat voelt en denkt, dan moet je niet verrast zijn dat je ongemakkelijk resultaten krijgt. Van foutieve intuïties kun je dan langer spreken. Maar nogmaals: we hebben die resultaten voorlopig niet.

Ten slotte dit: mocht blijken dat het MA de empirische toetsing niet overleeft, dan is dit niet het einde van de incompatibilistische positie. Er zijn nog andere argumenten die deze positie filosofische legitimiteit geven, zoals het consequentie-argument en het basisargument. Toch zou het verlies van MA wel een grote tegenslag betekenen. MA bezit een aparte schoonheid en aantrekkingskracht. Indien succesvol laat dit gedachte-experiment zien dat onze dagelijkse intuïties innig verweven zijn met diepe metafysische vraagstukken over causaliteit, gedetermineerdheid, wetmatigheden en het mind-body probleem. Is die verwevenheid evenwel een illusie, dan verliest het vrije wildebat aan filosofische reikwijdte. Metafysica wordt voor een stuk minder relevanter. Compatibilisten hebben die irrelevantie altijd wel al verdedigd. Vrije wil en verantwoordelijkheid kon je perfect wetenschappelijk begrijpen. Maar voor wie geloofde dat onze diepere intuïties uliem steunden op filosofische denkbeelden, komt dit mogelijke verlies toch hard aan. Paradoxaal genoeg hebben de incompatibilisten wetenschap nodig om hun punt te maken en het belang van filosofie veilig te stellen.

Bibliografie

- Feltz, A. 2013. "Pereboom and Premises. Asking the right questions in the experimental philosophy of free will." *Consciousness and Cognition* 22: 53-63.
- Gorin, Moti. 2013. "What Makes an Intuition a Compatibilist Intuition? A response to Sripada." *Philosophia* 41: 1205-1215.

- Kane, Robert. 1996. *The Significance of Free Will*. Oxford: Oxford University Press.
- Kearns, Stephen. 2012. "Aborting the Zygote Argument." *Philosophical Studies* 160: 379-389.
- King, Matt. 2013. "The Problem with Manipulation." *Ethics* 124, 1: 65-83.
- McKenna, Michael. 2008. "A Hard-line Reply to Pereboom's Four-case Manipulation Argument." *Philosophy and Phenomenological Research* 77, 1: 142-159.
- Mele, Alfred. 2008. "Manipulation, Compatibilism, and Moral Responsibility." *The Journal of Ethics* 12, 3-4: 263-286.
- Nahmias, E., Shepard, J. & Reuter, S. 2014. "It's OK if 'My brain made me do it': People's intuitions about free will and neuroscientific prediction." *Cognition* 133: 502-516.
- Pereboom, D. 2001. *Living Without Free Will*. Cambridge: Cambridge University Press.
- Sripada, Chandra. 2011. "What Makes a Manipulated Agent Unfree?" *Philosophy and Phenomenological Research* 85, 3: 1-31.
- Verplaetse, J. 2011. *Zonder vrije wil. Een essay over verantwoordelijkheid*. Amsterdam: Nieuwezijds.