
Het dilemma van de gevangenen

ERIC VAN DAMME

Als ieder in zijn eigenbelang handelt, resulteert er voor elk, en dus voor de maatschappij als geheel, een slechte uitkomst. Dat is de kern van het *prisoners' dilemma*. Individuele rationaliteit conflicteert met sociale wenselijkheid; de onzichtbare hand van Adam Smith werkt niet. Voor de inrichting van de samenleving maakt het nogal wat uit welk idee men als uitgangspunt neemt. Wanneer en hoe het individuele gedrag te reguleren om sociaal optimale uitkomsten te bewerkstelligen?

De film *A Beautiful Mind* vertelt het verhaal van de geniale, krankzinnige, John Nash, een van de grondleggers van de speltheorie. Om Nash' basisideeën te illustreren, bedacht zijn promotor in 1950 de oervorm van het dilemma. Twee mannen worden verdacht van het gezamenlijk plegen van een misdrijf, maar er is geen sluitend bewijs. Ze zijn afzonderlijk opgesloten, zodat ze hun gedrag niet kunnen afstemmen. Elke verdachte weet dat als een van hen bekent, het bewijs sluitend is, de

klikspaan vrijuit gaat, terwijl de ander dan tien jaar moet brommen. Als beiden bekennen, krijgt elk strafvermindering en vijf jaar cel. Als beiden zwijgen, is er geen bewijs en gaat elk voor een lichter vergrijp één jaar de cel in. Voor het collectief is ontkennen duidelijk het beste. Als een gevangene in zijn eigenbelang handelt, zal hij echter bekennen: wat de partner ook doet, bekennen leidt een kortere straf. Als beiden deze logica volgen, gaat elk echter vijf jaar de cel in.

Het *prisoners' dilemma* is een voorbeeld van een sociaal dilemma. Elk individu moet beslissen hoe te handelen. Elke actie levert individueel nut op, maar ook kosten en baten voor de overige deelnemers in de samenleving. Bij zelfzuchtig gedrag laat ieder zich alleen door het eigen nut leiden, en worden de externe effecten die dat handelen voor anderen heeft niet in beschouwing genomen. Individuele optimalisering leidt dan al snel tot sociaal inferieure uitkomsten.

Bij een brand in een disco haast iedereen zich

meteen naar de uitgang, met opstopping, chaos en onnodige doden als gevolg. Bij een popconcert is er de neiging om op de tenen te gaan staan om beter te kunnen zien. Als iedereen dat doet, ziet ieder even goed (of slecht) maar geniet iedereen minder. Als we alle tegelijk de weg opgaan, zijn files het gevolg. Als aan de andere kant van de weg een ongeluk gebeurd is, ontstaat aan deze kant een langere kijkfile. Als mijn concurrenten adverteren, ben ik gedwongen dat ook te doen, ook als adverteren op zich niet tot extra verkopen leidt. Als schaarse goederen op 'eerst komt eerst maalt' basis verdeeld worden, moet ieder die iets wil hebben vroeg in de rij gaan staan. Inwoners van een dorp profiteren van een lokale supermarkt, maar als iedereen voor de reguliere boodschappen naar de discounter in de stad gaat, kan de dorpssuper niet overleven. Landen en steden zijn geneigd om met excessieve subsidies economische bedrijvigheid aan te trekken. De opwarming van de aarde wordt veroorzaakt doordat landen wel kijken naar de individuele baten en lasten, maar de negatieve externe effecten onvoldoende meenemen. De cle-

Zelfselectie ondermijnt representativiteit.

mentieprogramma's die mededingingsautoriteiten gebruiken om kartels op te sporen zijn gebaseerd op het gevangenendilemma: de eerste melder krijgt geen straf.

Adam Smith' *invisible hand* werkt alleen als er geen externe effecten zijn, of als deze geïnternali-

seerd kunnen worden. Er zijn diverse mogelijkheden om aan het dilemma te ontsnappen: toedelen van eigendomsrechten, maken van bindende afspraken, wet- en regelgeving die het gewenste gedrag afdwingt, herhaalde interactie, en morele normen, of formele of informele instituties die sociaal gedrag stimuleren. Elinor Ostrom, Nobelprijswinnaar Economie in 2009, heeft voor een groot aantal concrete dilemmasituaties onderzocht onder welke voorwaarden dergelijke mechanismen effectief zijn.

Contracten en wetgeving werken volgens het principe dat iemand die afwijkt van wat gewenst is een boete moet betalen. Slim gekozen boetes, of belastingen, leiden er toe dat eigenbelang en sociale wenselijkheid parallel gaan lopen. Rekeningrijden werkt volgens dit idee. Europese Staatssteunregels verbieden EU-lidstaten om aan bedrijven oneigenlijke subsidies te geven en voorkomen daarmee verkwisting van belastinggeld. De milieuproblematiek laat zien dat het niet altijd even eenvoudig is om tot bindende afspraken te komen. Ook gedrag in een brandende disco is op deze manier niet te reguleren: de transactiekosten zijn te hoog.

Als het sociale dilemma herhaald optreedt, er voldoende informatie over het verleden is en de deelnemers de toekomst voldoende belangrijk achten, is een overheid niet nodig om maatschappelijk betere uitkomsten te bereiken: 'oog om oog, tand om tand'-strategieën kunnen gebruikt worden om samenwerking te stabiliseren. Als gewenst gedrag beloond wordt en opportunisme bestraft, weegt de korte termijn winst die zelfzuchtig handelen behaald kan worden niet op tegen de lange termijn kosten daarvan, zodat het lange-termijn eigenbelang het beste gediend wordt door sociaal te handelen. De speltheorie heeft laten zien dat stabiele

samenwerking in veel gevallen kan ontstaan, maar dat dit niet het enige mogelijke (Nash) evenwicht is. Laboratoriumexperimenten en het veldwerk van Ostrom geven inzicht in de voorwaarden waaronder goede evenwichten geselecteerd worden, maar ons begrip hiervan is nog verre van volledig.

De (traditionele) economische literatuur gaat uit van eigenbelang. Niet alle mensen zijn echter zelfzuchtig. Veel mensen zijn bereid samen te werken als anderen dat ook doen, terwijl ze egoïstisch handelen afkeuren en soms afstraffen. Als anderen zo handelen is het ook voor een zelfzuchtig iemand optimaal zich coöperatief op te stellen. Een robuust inzicht, dat niet verenigbaar is met de aanname van *homo economicus*, is dat (onverbindende) communicatie tot meer coöperatie leidt. Emoties en schuldgevoel leveren een mogelijke verklaring. Als iemand zegt dat hij zal samenwerken, dan verwacht de ander dat de eerste woord zal houden, en wel omdat de spreker bij afwijken de verwachting beschaamt en zich schuldig voelt. Omdat schuldgevoelens pijnlijk zijn, is een aankondiging om samen te gaan werken niet zuiver loze praat, maar een signaal met een zekere betrouwbaarheid. Experimenten met het *prisoners' dilemma* verwerpen het traditionele economische model. In de moderne gedragseconomie wordt gewerkt met een meer realistisch mensbeeld, en verwacht mag worden dat dit meer inzicht zal opleveren in de vraag in welke situaties mensen in staat zijn samen te werken.

De vraag: 'Wat maakt sommige samenlevingen welvarend en andere arm?' fascineerde Adam Smith en staat ook nu nog op de agenda. Het vermogen van mensen om samen te werken en sociale dilemma's op te lossen levert een sleutel tot het antwoord. Het *prisoners' dilemma* laat zien dat samenwerking geen vanzelfsprekendheid is, maar dat

bepaalde instituties effectieve samenwerking kunnen bewerkstelligen. Een bevredigend antwoord op de vraag: wat werkt en wat niet, kan vermoedelijk echter alleen door effectieve samenwerking tussen de verschillende sociale wetenschappen gevonden worden.