

Het dilemma van de trolley en de bredere context

De kracht van systeemdenken in IT

Voor IT'ers die willen reflecteren op ethische aspecten van hun werk, heeft systeemdenken veel voordelen. Ze kunnen er soepel mee heen en weer bewegen tussen technische details en bredere context, ziet Marc Steen. Zo kunnen ze waar nodig veel effectiever bijsturen.

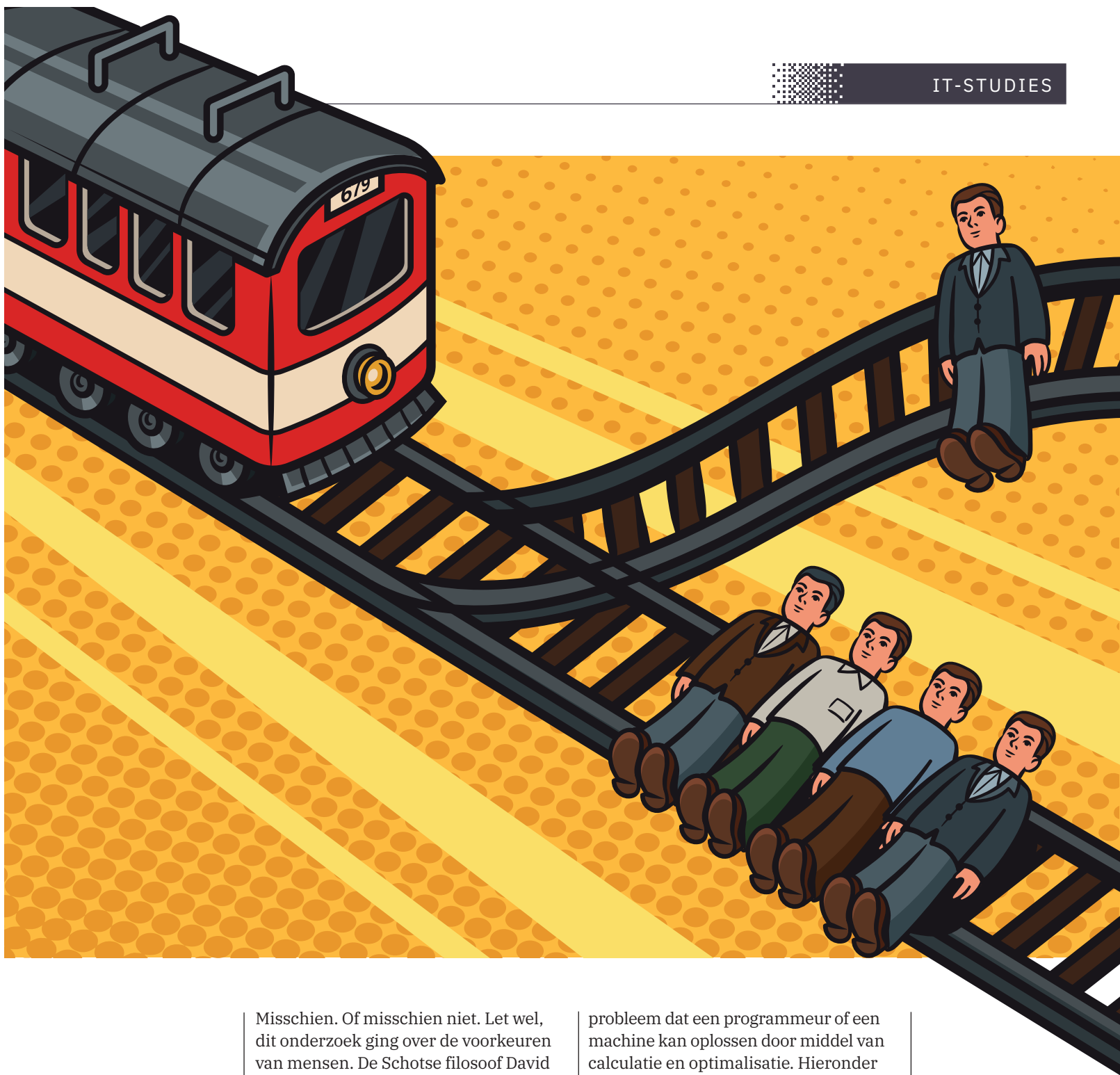
IN 1967 SCHREEF DE ENGELSE FILOSOOF PHILIPPA FOOT OVER DE 'DOCTRINE VAN HET DUBBELE EFFECT'; als je handelt vanuit goede bedoelingen, kun je ook, per ongeluk, bijkomende schade veroorzaken. Ze gaf het volgende voorbeeld: stel je voor dat jij de bestuurder bent van een op hol geslagen trolley: vóór je op het spoor werken vijf personen aan het spoor; je hebt de mogelijkheid om de wagon te sturen naar een ander spoor, waarop één persoon staat. Wat doe je?

In de jaren daarna zijn allerlei variaties bedacht op dit thema. Je staat naast het spoor, met een hendel om de wissel te bedienen: je kunt één persoon opofferen om de dood van vijf personen te voorkomen. Of de fat person-variant; je staat op een loopbrug boven het spoor, samen met een persoon met een fors gewicht. Gooi je die persoon van de brug, om de trolley tot stilstand te brengen? Veel mensen deinzen daarvoor terug; het voelt alsof je die persoon vermoordt. Dat voelt anders dan aan de hendel trekken en het moorden te delegeren aan de trolley. In de afgelopen

jaren werd het *Trolley Problem* een terugkerend thema in technologie en ethiek.

Het probleem

Er zijn ook psychologische experimenten uitgevoerd, zoals de Moral Machine van MIT. Dat was een online enquête waarin mensen moesten beslissen wat een zelfrijdende auto zou moeten doen bij een dreigend ongeval. Miljoenen mensen deden mee. Ze kregen plaatjes van verkeerssituaties, met verschillende soorten en aantallen voetgangers en passagiers. Voor elk paar plaatjes moesten ze kiezen tussen doorrijden en voetgangers doden, of tegen een obstakel botsen en passagiers doden. Op basis van 40 miljoen reacties van mensen uit meer dan tweehonderd landen ontdekten de onderzoekers enkele algemene voorkeuren, zoals het sparen van mensen boven dieren. Ze vonden ook verschillen tussen culturen, zoals dat mensen uit landen met collectivistische culturen oudere mensen sparen boven jongere. Wat kunnen we leren van zo'n experiment? Helpt het bij het programmeren van zelfrijdende auto's?



Misschien. Of misschien niet. Let wel, dit onderzoek ging over de voorkeuren van mensen. De Schotse filosoof David Hume herinnert ons eraan dat we 'is' en 'ought' niet moeten verwarren. Op basis van bevindingen uit empirisch onderzoek kun je geen morele uitspraken afleiden. Als mensen X doen of zeggen dat ze X zouden doen, volgt daaruit niet dat X moreel aanvaardbaar of wenselijk is.

Het probleem met het vraagstuk met de op hol geslagen treinwagon is dat mensen kunnen denken dat ethiek een probleem is dat we kunnen oplossen; een

probleem dat een programmeur of een machine kan oplossen door middel van calculatie en optimalisatie. Hieronder zal ik een andere invalshoek presenteren, meer in lijn met de oorspronkelijke functie van het gedachte-experiment, namelijk: het bevorderen van ethische reflectie en deliberatie.

Bredere context

Ten eerste kunnen we de trolley, de mensen op het spoor, en de hoofdpersoon van het gedachte-experiment, de *moral agent*, in een bredere context

We kunnen systeemdenken inzetten om een complexer systeem te bevragen

plaatsen. De originele omschrijving is met opzet heel beknopt. Maar de wereld is veel complexer. We kunnen systeemdenken (*systems thinking*) inzetten om ons een grotere, complexer systeem voor te stellen en daarover vragen te stellen.

Waar zijn de remmen?

Waarom zit er bijvoorbeeld geen claxon op de trolley om mensen op het spoor te waarschuwen? En waarom werkten de remmen niet goed? Een fout in het ontwerp van het remsysteem? Was er voldoende budget voor onderhoud?

Of zijn de onderhoudsexperts onlangs ontslagen, als bezuiniging? Trouwens, hoe besteedt het trolleybedrijf zijn budget? Biedt het eerlijke salarissen? Kunnen medewerkers hun vak bijhouden?

Worden ze beloond om fouten of ongevallen te melden? Krijgen bestuurders voldoende rust tussen hun diensten door? Welke *health and safety policies* zijn er voor het werken aan het spoor? We kunnen verder uitzoemen en vragen stellen over regelgeving en governance. Heeft de staat de vervoersbedrijven geprivatiseerd? Hoe wordt concurrentie tussen vervoerders geregeld? Zijn er prikkels om samen te werken op het gebied van veiligheid? We kunnen inzoomen, tot op het niveau van

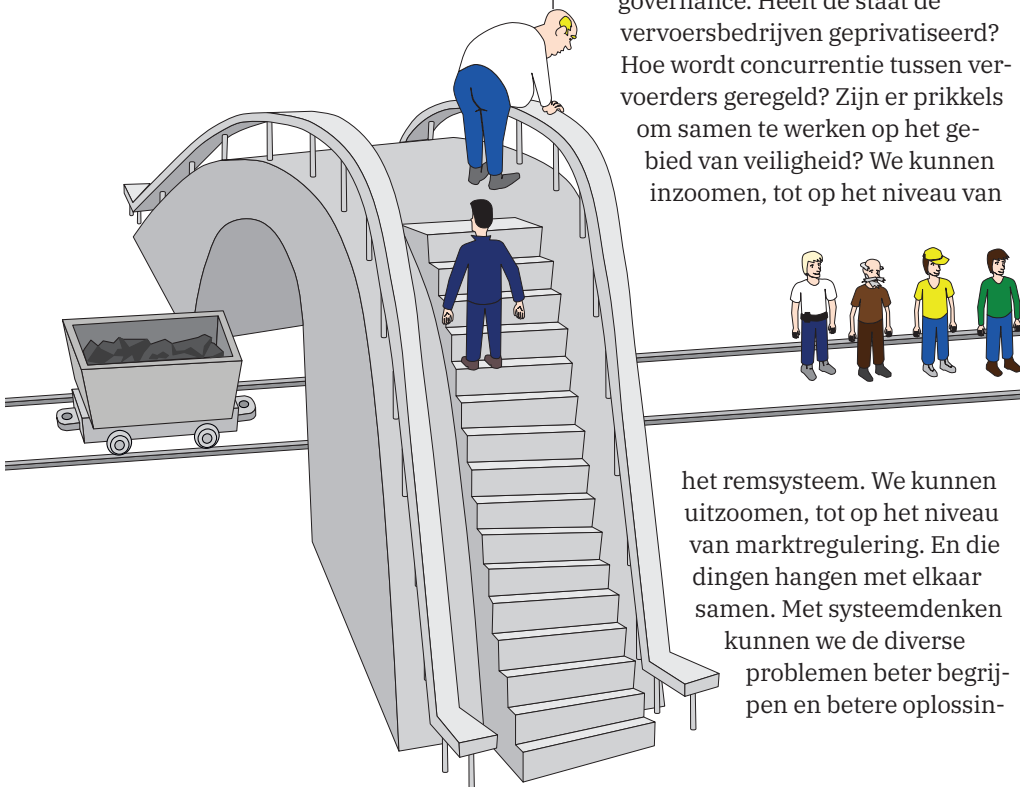
gen ontwikkelen. Ten tweede kunnen wij - als professionals, betrokken bij het ontwikkelen en toepassen van technologie - nadenken over onze *agency*. Welke mogelijkheden hebben wij om te handelen? We kunnen meer doen dan aan de kant staan en aan een hendel trekken. We kunnen het hebben over de samenwerking tussen *agency* en structure; hoe de context, de organisatie waarin je werkt, het project waarin je werkt, van invloed is op jouw mogelijkheden, en, omgekeerd, hoe jij die context kunt beïnvloeden. Wat kun jij doen, als ontwikkelaar, ontwerper, programmeur, manager?

In het trolleyprobleem heb je slechts twee handelingsopties: wel of niet aan de hendel trekken; over de voetgangers rijden of de auto laten botsen. Zo'n beperkte visie op *agency* is echter totaal onbruikbaar als we willen reflecteren op ethische aspecten in ons werk, in ons projecten. Je kunt nieuwsgierigheid gebruiken om de context beter te begrijpen, zoals we net zijn begonnen te doen. Je kunt creativiteit gebruiken om betere oplossingen te bedenken. Simpele oplossingen, zoals, in deze casus, een toeter om mensen te waarschuwen. Of complexe oplossingen, zoals een proces waarin de mensen worden aangemoedigd om te leren van ongevallen en maatregelen kunnen nemen om ongelukken te voorkomen.

Orde scheppen

Bovenstaande voorbeelden kunnen overkomen als een ratjetoe. We kunnen systeemdenken inzetten om wat orde te scheppen. We kunnen dat op verschillende niveaus doen. Op het niveau van de samenleving; dan kunnen we kijken naar de positieve en negatieve effecten van een bepaalde technologie of toepassing voor verschillende groepen mensen. Bijvoorbeeld voor eigenaren van zelfrijdende auto's versus voetgangers. Op het niveau van één specifieke organisatie; dat kan ons helpen om te begrijpen hoe mensen en machines interacteren, hoe ze informatie uit de

het remsysteem. We kunnen uitzoemen, tot op het niveau van marktregulering. En die dingen hangen met elkaar samen. Met systeemdenken kunnen we de diverse problemen beter begrijpen en betere oplossin-



omgeving gebruiken en op basis daarvan handelen. Hierbij kunnen we ons wenden tot vakgebieden zoals science and technology studies en cybernetica. Of we kijken naar het niveau van één specifiek project, bijvoorbeeld een project waarin we technologie ontwikkelen of inzetten; op dat niveau kan systems thinking ons helpen om de agency en structuur van de projectteamleden te begrijpen, en bijvoorbeeld om een proces van ethische reflectie en deliberatie te organiseren. Systeemdenken gaat over het zien van zowel de delen als het geheel; zowel de bomen als het bos.

We kunnen kijken naar de bomen, de takken, de bladeren en de bloemen. Naar de insecten die helpen bij het verspreiden van het stuifmeel van de bloemen, zodat vruchten zich kunnen ontwikkelen. Naar de bodem en de micro-organismen en schimmels daarin die vitaal zijn voor de bomen. En naar de kleine en grote dieren en vogels die in het bos leven, en, belangrijker nog, naar de ingewikkelde relaties en interacties tussen al deze verschillende organismen.

Uitzoomen en inzoomen

In de praktijk pas ik systeemdenken toe door heen en weer te bewegen tussen abstractieniveaus, afhankelijk van de situatie. In een meeting waarin mensen in abstracte termen praten, bijvoorbeeld over rechtvaardigheid, zoom ik in en stel ik een vraag over een praktisch detail, bijvoorbeeld hoe het zit met de bias in de trainingsdata van een algoritme. En als ze over allerlei technische details praten, bijvoorbeeld over de effectiviteit van een algoritme, dan zoom ik uit en stel ik een bredere vraag, bijvoorbeeld over de schade die die *false positives* veroorzaken en hoe we dat kunnen voorkomen.

Juist het heen-en-weer bewegen tussen technische details en de bredere maatschappelijke context kan enorm productief werken en kan nieuwsgierigheid en creativiteit stimuleren. Feedbackloops kunnen werken als

een virtueuze cirkel; ze dragen dan bij aan het bereiken van betere, gewenste resultaten. Of als een vicieuze cirkel; het systeem ontspoord dan en kan veel schade veroorzaken. Donella Meadows, een pionier van systems thinking, stelde dat we een systeem het meest effectief kunnen (bij) sturen door te sleutelen aan de onderliggende doelen en dieperliggende aannames (paradigma's) van een systeem. Ze noemt dat een hefboom, een leverage; in plaats van te focussen op allerlei losse elementen aan de oppervlakte van dat systeem (2008: 145-165).

Om dit verhaal af te ronden keren we terug naar dat beeld van die op hol geslagen trolley en die hendel. Jij kunt veel meer dan aan de kant staan en aan een hendel trekken. Jij kunt aan de slag met ethische reflectie en deliberatie. En met systems thinking; je kunt inzoomen en uitzoomen, de bredere context begrijpen van het systeem waaraan je werkt. Jij kunt op zoek naar een hefboom (leverage) om het systeem waaraan je werkt effectief bij te sturen. 🌐

Referenties

Foot, P., *The Problem of Abortion and the Doctrine of the Double Effect*. Oxford Review, 1967.

Meadows, D.H., *Thinking in Systems: A Primer*. 2008, White River Junction, Vermont: Chelsea Publishing.

O'Neil, C., *Weapons of Math Destruction*. 2016, London: Penguin.



Marc Steen

werkt als senior research scientist bij TNO, als expert in responsible innovation en toegepaste ethiek. Hij behaalde MSc, PDEng- en PhD-graden in Industrial Design Engineering aan de TU Delft. Onlangs verscheen zijn boek: *Ethics for people who work in tech* (Routledge).