

## Eerherstel voor de rede

Maarten van Doorn  
*Radboud Universiteit*  
martinus.vandoorn@ru.nl

De rede is niet krachtig genoeg om ons moreel te motiveren. Althans, dat leerden we van Hume. En dat is een probleem. Want we worden tegenwoordig verwacht moreel rekening te houden met mede-aardebewoners die ver van ons af staan. Dat wil zeggen, verschillen in identiteit (we moeten geven om iedereen, niet alleen onze ‘ingroup’), ruimte (problemen die heel ver weg zijn horen net zo goed tot onze verantwoordelijkheid) en tijd (ook dat iets zich pas ver in de toekomst laat gelden is geen morele excuserende factor meer) zijn niet moreel relevant. Alleen, onze onderbuik reageert vooral op wat tastbaar is: op wij (in plaats van iedereen), hier (in plaats van overal) en nu (in plaats van altijd). En ‘de rede is niet bij machte om die intuïties op te voeden of te overrulen’. Helaas is het daarom nauwelijks meer relevant dat we rationeel concluderen dat vreemdelingen en hun ongeboren nakomelingen gelijke rechten hebben: ‘Velen horen het inclusieve appel wel, maar worden er in de praktijk niet of nauwelijks door geactiveerd doordat ons morele gevoel minder ver reikt dan onze rationele morele overwegingen... Vooronderstelt de universalistische ethiek geen onthechte en zuiver rationele heiligen, in plaats van mensen van vlees en bloed die leven binnen lokale relationele verbanden?’. En is het dan niet psychologisch onrealistisch om te verwachten dat mensen zich door zo’n ethiek laten leiden?

Dat is het centrale probleem in het interessante artikel van Tjalling Swierstra en Evelien Tonkens. Zij beargumenteren dat twee dominante oplossingen voor deze puzzel niet zullen werken. Volgens optie één, het ‘evo-progressivisme’ van Pinker, Singer, Sunstein, Buchanan en Powell, beseffen we dat de rechten van onbekende anderen een appel op ons doen, en kunnen we daarnaar handelen, ‘omdat we inmiddels voldoende redelijk aan het worden/geworden zijn’. Swierstra en Tonkens lijken het met Haidt eens te zijn dat zo’n ‘inclusieve ethiek bij velen niet aanslaat’, omdat ‘een appel op de rede [geen] zoden aan de dijk [zet] om mensen te bewegen afstanden in identiteit, ruimte en tijd te overbruggen’. Al menen zij dan

weer dat *zijn* visie – optie twee – geen ruimte laat voor morele progressie. Dus pleiten ze – optie drie – voor een ‘brede conceptie van de rede, van Systeem 2, die naast theoretische ook mimetische en mythisch-narratieve elementen bevat’. Waarbij je, als ik het goed begrijp, de ‘circle of concern’ uitbreidt door te extrapoleren vanuit concrete betrokkenheid, in plaats abstracte consistentie.

Swierstra en Tonkens voeren een indrukwekkend scala denkers op om hun zienswijze te illustreren. Zo ‘narratief-emotionele vorm van reflectie’ zal ongetwijfeld ook wel haar plaats hebben in een compleet verhaal over morele deliberatie en verbeelding. Toch wil in dit commentaar wat redenen geven om te denken dat Tonkens en Swierstra de relevantie van klassiek – wat zij noemen ‘theoretisch’ – moreel redeneren onderschatten. Er zijn goede empirische redenen om te denken dat die niet zo motivationeel inert is als zij schetsen. En dus dat de oplossing van ‘evo-progressieven’ meer potentieel heeft dan zij erkennen.

## 1 Dual-Process Theory en rationaliteit

De kwestie is dus of we door te reflecteren op onze morele intuïties hun normatieve en motivationele beperkingen kunnen overstijgen. Swierstra en Tonkens benaderen dit vanuit de bekende Dual-Process Theory (DPT) (Kahneman 2011). Ik wil hun schets van die theorie uitbreiden. Enkele voor hun argument cruciale empirische aanvullingen zijn op zijn plaats, omdat die erop wijzen dat de rede krachtiger is dan aangenomen.

Hoe zij ‘Systeem 1 biases’ uitleggen aan de hand van de beweerde bias ‘We geloven wat we willen geloven’ is op dit gebied kenmerkend. Hoe liever we willen dat iets waar is, hoe waarschijnlijker we het achten, is de suggestie. Waarheid of waarschijnlijkheid inschatten aan de hand van wenselijkheid, is blijkbaar een vuistregel van ons associatieve denksysteem. Net zoals dat we bijvoorbeeld de afstand van een object schatten we aan de hand van haar waargenomen helderheid. Hoe helderder, des te dichterbij.

In werkelijkheid heeft onderzoek naar zulk ‘wishful thinking’ nauwelijks bewijs gevonden voor een directe invloed van wenselijkheid op waarschijnlijkheidsoordelen, zoals die er *wel* is van waargenomen helderheid op afstandsoordelen (Bar-Hillel & Budescu, 1995; Harris, Corner, & Hahn, 2009). Het lijkt er eerder op dat mensen *alleen* geloven wat ze willen geloven in de mate dat ze die conclusie met geldige redenen kunnen rechtvaardigen ten overstaan van anderen (Kunda, 1990; Hahn & Harris, 2014). Die rol van sociale rechtvaardigingen onderschrijven Swierstra en Tonkens wel, maar

gaat, zo zal ik in § 2 beargumenteren, verder dan zij erkennen. Eerst de empirische aanvullingen over de correctieve kracht van de rede.

Swierstra en Tonkens schrijven: '[E]volutie heeft ons toegerust met groepsgerichte morele intuïties en affecten, en dus met denken in wij versus zij, kortom met exclusieve moraal'. Onze morele intuïties wijken dus af van het normatief relevante criterium, want alle grote ethische theorieën, zoals het utilitarisme, zijn *universalistisch* van aard. Helaas slaagt Systeem 2 er niet betrouwbaar in om (die) fouten van Systeem 1 (1) te herkennen en (2) te verbeteren. En dus laten cognitieve biases en tribale morele standpunten zich 'niet uitroeien'. Maar die redeneerstap ziet twee goed onderzochte correctiemogelijkheden van onze biases over het hoofd.

De eerste daarvan bespreek ik nu. De tweede heeft te maken met de zojuist genoemde sociale rechtvaardigen, en bewaar ik dus ook voor sectie twee.

Ten eerste is het, als we niet op Systeem 2 moeten rekenen, mogelijk om biases weg te werken door *Systeem 1* te trainen. Want hoewel ons denken vaak wordt beïnvloed door verkeerde heuristieken (de auteurs noemen naast het zojuist besproken wishful thinking effect ook nog de 'availability bias' en de 'anchoring bias'), suggereren recente debiasing-onderzoeken dat onze rationaliteit aanzienlijk kan worden opgeschroefd door korte trainingsinterventies waarin de juiste antwoorden op redeneerproblemen worden uitgelegd (Claidière, Trouche, & Mercier, 2017; Hoover & Healy, 2017; Morewedge et al., 2015; Purcell, Wastell, & Sweller, 2020). Je zou denken dat de verbetering die dat oplevert, komt doordat die uitleg de proefpersonen in staat stelt om via Systeem-2 denkwerk hun foutieve intuïtieve respons te corrigeren. Maar dat is niet hoe het werkt. Na de training is die correctieve deliberatie niet eens meer nodig, want proefpersonen ontwikkelen een intuïtief Systeem-1 gevoel voor de juiste oplossing (Boissin et al., 2021). Eenzelfde wisselwerking kunnen we verwachten bij morele intuïties. Precies zoals de auteurs schrijven dat Pinker (2011) betoogt: 'Wanneer [correcties] succesvol zijn, [kunnen] deze wel degelijk langzamerhand deel worden van de pre-reflectieve routines van Systeem 1'.

Dat is al één route naar rationele intuïties, lijkt het. Swierstra en Tonkens beweren echter dat onze morele intuïties zó 'off-track' zijn, dat dit slechts een druppel op de hete plaat is. Om twee redenen die ik nu uiteen zal doen, is dit een betwifelbare suggestie.

Ten eerste betogen moreel realisten in de context van het Evolutionary Debunking Argument (EDA; Street, 2006) al lang dat het gat tussen (a) evolutionair adaptieve en (b) normatief acceptabele morele oordelen overschat wordt. Parfit (2011, vol. II, 533) noemt bijvoorbeeld de overtuigingen

dat pijn slecht is, overleven en gezondheid goed zijn en we verplichtingen hebben naar onze kinderen. Vier van de vele opvattingen die breed gedeeld worden, moreel juist én evolutionair voordelig zijn. We moeten ons door het tribale element in onze morele psychologie dus niet laten verleiden om te denken dat evolutie ons op moreel gebied totaal de verkeerde kant op heeft gestuurd. Daarnaast is er empirisch bewijs dat universalistisch-ethische redenen wel degelijk aanslaan op onze intuïties (Baron, 2017). Die dus niet zo ongevoelig zijn voor universalistische overwegingen als Swierstra en Tonkens aannemen. Proefpersonen die bijvoorbeeld utilistische reacties geven op morele dilemma's wanneer ze hun Systeem 2 mogen gebruiken, geven doorgaans hetzelfde antwoord wanneer het experimentele ontwerp hen dwingt om puur intuïtief te antwoorden. Ze hadden dus al de universalistische intuïtie. Conclusie: 'System 2 processing does not necessarily lead to correct decisions and utilitarian moral judgments on sacrificial moral dilemmas do not necessarily require System 2 ... Although the dual process framework has great merits, it should not be used to make inferences about the normativity of moral judgments' (De Neys, 2020, p. 94).

Ondanks het plausibele evolutionaire achtergrondverhaal plaatsen deze overwegingen twijfels bij het idee dat onze morele intuïties in principe ontoereikend en oncorrigeerbaar zijn. Dat zijn al goede redenen waarom klassieke rationaliteit meer potentie heeft voor het accommoderen van een inclusieve ethiek dan Swierstra en Tonkens suggereren.

## 2 Samen beter worden

Tegelijkertijd lijkt het meeste van wat ik zojuist gezegd wellicht koren op de molen van Swierstra en Tonkens. Zij betogen immers met Haidt dat de rede nauwelijks correctieve kracht heeft, maar vrij machteloos staat tegenover onze intuïties, die ze slechts kan rechtvaardigen maar niet verbeteren. Net zoals dat rede weinig voorkwam in mijn verhaal over morele en trainbare intuïties! De tweede correctiemogelijkheid die Swierstra en Tonkens over hoofd zien, toont echter waarom het voorgaande toch consistent is met het voorstaan van een grotere rol voor klassiek redeneren.

Swierstra en Tonkens onderschrijven Singers claim dat mensen zich in een discussie naar elkaar moeten rechtvaardigen. Dat vereist dat zij *argumenten* produceren ter rechtvaardiging van hun intuïtie. Dat is de rol die de auteurs, met Haidt, aan de rede toekennen.

Hier zijn een paar dingen interessant aan. Onze sterke drang om argumenten te vinden voor onze intuïtieve oordelen, duidt ten eerste op een basisrespect voor andermans oordelen over onze standpunten. Anders zouden we die moeite niet doen. Bovendien geven we door het geven van argumenten toe dat ‘rational considerations, valid by criteria which are more than biases arising from private interests or from unexamined and unverifiable preconceptions, are relevant to the issue under discussion’ (Lovejoy, 1961, p. 62). Dit is belangrijk, want zodra we onze argumenten uitspreken, kunnen anderen die rechtvaardigingen aan de hand van zulke criteria staven en aanvallen als ze tekortschieten. Als we er dan niet in slagen ons standpunt adequaat te rechtvaardigen – omdat onze rede, ons Systeem 2, geen adequate argumenten weet te vinden – moeten we van gedachten veranderen op straffe van reputatieschade of inconsequentie (Sperber & Baumard 2012).<sup>1</sup> Om dat te voorkomen, is het aannemelijk dat we van gedachten veranderen, als we beseffen dat we geen rationele rechtvaardiging kunnen geven voor een bepaald standpunt: ‘Rationalizations serve reputation management goals, [and] affect our behaviors because we are held accountable by our peers’ (Altay & Mercier, 2020, p.e29).

Zo is er een indirecte maar belangrijke rol voor klassiek moreel redeneren in het vormen van onze morele standpunten. Die bovendien een verhaal heeft over morele motivatie: mensen zijn gemotiveerd om van morele positie te veranderen als ze geen goede argumenten hebben, teneinde reputatieschade te voorkomen.

Kortom: een inclusieve ethiek is psychologisch realistischer dan Swierstra en Tonkens schetsen, omdat zij enkele cruciale feiten over menselijk redeneren over het hoofd zien. Onze intuïties zijn wel degelijk trainbaar en helemaal niet zo anti-universalistisch. En in een sociale context maakt het zeker uit of er rationele (door Systeem 2 geproduceerde en geëvalueerde) argumenten zijn voor een bepaald moreel standpunt. Want de afwezigheid daarvan is een sterke reputatieprikkel en motivator – intuïtie of niet – voor morele meningsverandering.

<sup>1</sup> Dit belang van een ‘division of cognitive labor’ laat trouwens ook zien dat ‘Systeem 1 óf Systeem 2 moet de waarheid opleveren’ een vals dilemma is. Het checken van onze intuïties kan gebeuren via de rationaliteit *van anderen*. De bevinding dat Systeem 2 in individuele denkers niet altijd in staat is om onjuiste intuïties te corrigeren betekent dan ook niet, zoals de auteurs suggereren, dat ‘de hoop vervlogen [is] dat we via de rede, via ethische reflectie dus, onze automatische, parochiale, morele intuïties en affecten kunnen corrigeren’. Die reflectie kan namelijk ook in een groepscontext plaatsvinden, en dan spelen (door de rede geproduceerde) argumenten een cruciale rol.

## Bibliografie

- Altay, S, & Mercier, H. (2020). Rationalizations Primarily Serve Reputation Management, Not Decision Making, *Behavioral and Brain Sciences* 43, pp. e29. DOI: 10.1017/S0140525X19002115.
- Bar-Hillel, M., & Budescu, D. (1995) The Elusive Wishful Thinking Effect, *Thinking and Reasoning* 1(1), pp. 71–103. DOI:10.1080/13546789508256906
- Baron, J. (2017) Utilitarian vs. Deontological Reasoning: Method, Results, and Theory, in: J.-F. Bonnefon, B. Trémolière (eds.), *Moral Inferences*, Hove: Psychology Press, pp. 137–151.
- Boissin, E., Caparos, S., Raelison, M., & De Neys, W. (2021) From Bias to Sound Intuiting: Boosting Correct Intuitive Reasoning, *Cognition* online prepub. DOI: 10.1016/j.cognition.2021.104645
- Claidière, N., Trouche, E., & Mercier, H. (2017) Argumentation and the Diffusion of Counter-intuitive Beliefs. *Journal of Experimental Psychology* 146(7), pp. 1052–1066. DOI: 10.1037/xge0000323.
- De Neys, W. (2020) Morality, Normativity, and the Good System 2 Fallacy, *Diametros* 17(64), pp. 90-95 DOI: 10.33392/diam.1447.
- Haidt, J. (2001) The Emotional Dog and Its Rational Tail, *Psychological Review* 108(4), pp. 814-34. DOI: 10.1037/0033-295x.108.4.814.
- Hahn, U, & Harris, A. J. L. (2014). What Does It Mean to Be Biased, *Psychology of Learning and Motivation* 61, pp. 41–102. DOI:10.1016/B978-0-12-800283-4.00002-2.
- Harris, A. J. L., Corner, A., & Hahn, U. (2009) Estimating the Probability of Negative Events, *Cognition* 110(1), pp. 51–64. DOI: 10.1016/j.cognition.2008.10.006.
- Hoover, J. D., & Healy, A. F. (2017) Algebraic Reasoning and Bat-and-Ball Problem Variants: Solving Isomorphic Algebra First Facilitates Problem Solving Later, *Psychonomic Bulletin & Review* 24(6), pp. 1922-28. DOI: 10.3758/s13423-017-1241-8.
- Kahneman, D. (2011) *Thinking, Fast and Slow*. Allen Lane.
- Kunda, Z. (1990) The Case for Motivated Reasoning, *Psychological Bulletin* 108(3), pp. 480-98
- Lovejoy, A. (1961) *Reflections on Human Nature*. Baltimore: The Johns Hopkins Press.
- Morewedge, C. K., Yoon, H., Scopelliti, I., Symborski, C. W., Korris, J. H., & Kassam, K. S. (2015) Debiasing Decisions: Improved Decision Making with A Single Training Intervention, *Policy Insights From the Behavioral and Brain Sciences* 2(1), pp. 129–140. DOI: 10.1177/2372732215600886.
- Parfit, D. (2011) *On What Matters, Volume 2*, Oxford: Oxford University Press.
- Pinker, S. (2011) *The Better Angels of Our Nature: The Decline of Violence in History and Its Causes*, London: Penguin.

- Purcell, Z. A., Wastell, C. A., & Sweller, N. (2020) Domain-Specific Experience And Dual-Process Thinking, *Thinking & Reasoning*, pp. 1–29. DOI: 10.1080/13546783.2020.1793813
- Sperber, D. & Baumard, N. (2012) Moral Reputation: An Evolutionary and Cognitive Perspective, *Mind & Language* 27(5), pp. 495–518. DOI: 10.1111/mila.12000
- Street, S. (2006) A Darwinian Dilemma for Realist Theories of Value, *Philosophical Studies* 127(1), pp. 109-66. DOI: 10.1007/s11098-005-1726-6

## Over de auteur

**Maarten van Doorn** verdedigt over enkele maanden zijn proefschrift in de meta-ethiek aan de Central European University (CEU). Hij heeft ook een podcast en blog (<https://maartenvandoorn.com>) over cognitieve psychologie. En publiceert in 2022 het boek *Wij zijn rationeel* bij Noordboek Uitgeverij.