

Hoe ontstaat polarisatie in de samenleving? Waarom veranderen mensen van attitude of opinie? Om deze vragen te beantwoorden presenteren Han van der Maas en collega's een cascade-model van attitudeverandering, dat zowel de interacties tussen gevoelens, opvattingen en gedrag binnen het individu als de interacties tussen individuen omvat. 'Een manipulatie van progressieven in conservatieve richting leidt tot een massale opinieverandering in de richting van het progressieve standpunt.'

EEN FORMELE BENADERING

POLARISATIE ALS COMPLEX VERSCHIJNSEL

Waarom valt iemand opeens van zijn geloof? Hoe verspreiden complottheorieën zich? Wat doen we tegen de toenemende polarisatie in de samenleving? Antwoorden op deze belangrijke vragen worden gezocht in de psychologie, de sociale wetenschappen maar ook in de natuurkunde en informatica. In de laatste gebieden wordt opvallend veel gepubliceerd over de verspreiding van opinies in sociale netwerken.

Dit artikel bespreekt onze benadering van attitude- of opinieverandering, waarin we ideeën uit deze verschillende vakgebieden integreren. Ons werk hierover is formeel van aard. Dat wil zeggen dat we wiskundige modellen voor deze processen voorstellen, zoals dat al veel langer in de natuurwetenschappen gebruikelijk is. Dat is ongewoon in de psychologie, maar heeft als voordeel dat het ons dwingt tot grote nauwkeurigheid in onze modelkeuzes. Een nadeel is dat veel psychologen en sociologen onbekend zijn met wiskundige modellen. We houden daarmee rekening in deze introductie en vermijden formules. We zullen ook spaarzaam zijn met verwijzingen. Alles wat we bespreken is gepubliceerd in de wetenschappelijke literatuur en in de publicaties die we noemen zijn verdere referenties te vinden.

Onze hoofdvraag is hoe polarisatie in de samenleving

ontstaat. Polarisation kan hierbij het best omschreven worden door een toename in variantie, waarbij de verdeling van opinies uiteindelijk bimodaal (twee toppig) wordt. We willen weten welke rol psychologische processen *tussen* en *binnen* mensen hierbij spelen. Het wiskundige model dat we zullen introduceren, kent daarom twee beschrijvingsniveaus: ten eerste modelleren we de attitude van één persoon. Dit deel richt zich op één individu en kan worden beschreven als een psychologisch netwerkmodel (van der Maas et al., 2006). Hiermee modelleren wij de attitude, of opinie, van de persoon als een netwerk van gevoelens, opvattingen en gedragingen ten opzichte van het attitudeobject (Dalege et al., 2016). Deze gevoelens, opvattingen en gedragingen zijn de knopen van het netwerk (zie figuur 1, linker deel). Het kan zijn dat deze knopen elkaar beïnvloeden; dan zijn deze knopen verbonden. Een specifiek gevoel kan bijvoorbeeld bepaalde gedragingen veroorzaken.

Als de knopen van het attitude-netwerk sterk verbonden zijn, hebben we een stabiele en sterke attitude. Een overtuigd vegetariër vindt de slacht van dieren bijvoorbeeld naar (gevoelens), de bio-industrie slecht voor het klimaat (een overtuiging), en hij eet geen vlees (gedrag). Deze gevoelens, opvattingen en gedragingen zijn sterk met elkaar verbonden.

Een andere mogelijkheid is dat verbonden knopen

tegenstelde waarden hebben. De meeste rokers weten dat roken ongezond is maar roken toch. Deze inconsistentie maakt het netwerk minder stabiel. Als mensen zich bewust zijn van die inconsistentie, dan spreken we wel van cognitieve dissonantie.

We weten niet precies hoe attitudenetwerken in elkaar zitten, ook al omdat dit per individu kan variëren. Maar we zullen zien dat we met dit algemene netwerkmodel van de attitude, op basis van heel zwakke aannames over de organisatie van het netwerk, al interessante verklaringen kunnen opstellen.

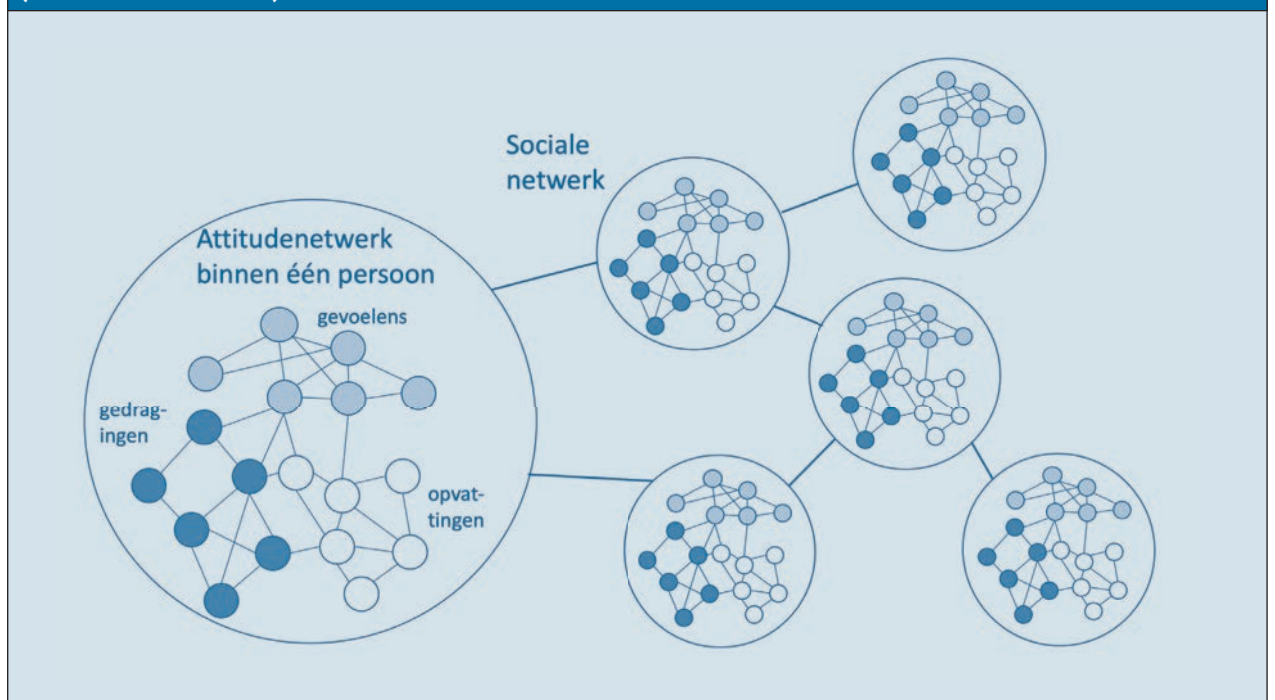
Individen hebben echter niet alleen een netwerk van gedragingen, overtuigingen en gevoelens op hun individuele niveau; ze zijn ook onderdeel van sociale netwerken. Daarom heeft ons model een tweede beschrijvingsniveau, waarin onderlinge beïnvloeding tussen individuen wordt gemodelleerd (zie figuur 1, rechter deel). Sociale netwerken zijn uitgebreid bestudeerd in de sociale wetenschappen maar ook in een deelgebied van de natuurkunde, de statistische fysica. Modellen ontleend aan de statistische fysica worden gebruikt om de verspreiding van attitudes te verklaren. Dit onderzoeksgebied wordt ook wel sociofysica genoemd. Ook

hierin staan netwerken centraal, alleen is in deze netwerkrepresentatie de persoon zelf een knoop, en de sociale verbanden (bijvoorbeeld interacties tussen mensen) worden gerepresenteerd als een verbinding tussen de knopen in een netwerk. Ook dit niveau kan worden gezien als een netwerkmodel.

Als je deze twee netwerkrepresentaties van individuele en sociale netwerken combineert, dan krijg je een netwerk van netwerken: ieder persoon heeft a) zijn eigen attitudenetwerk en is b) als individu een knoop binnen een sociaal netwerk. Dit wordt al snel heel ingewikkeld. Er is evenwel een manier om dit complexe modelsysteem te vereenvoudigen. Het gedrag van het netwerk van vele variabelen van één individu kan namelijk samengevat worden met een transitie-model met slechts drie variabelen (attitude, informatie en aandacht) in één vergelijking.

Transities zijn plotselinge veranderingen in gedrag (attitude) naar aanleiding van kleine veranderingen in een onafhankelijke variabele (bijvoorbeeld informatie) – daarover straks meer. Deze individuele transities zijn verbonden in een sociaal netwerkmodel waar deze transities elkaar beïnvloeden en zich soms snel verspreiden. Een metafoor

FIGUUR 1. EEN SOCIAAL NETWERK (TUSSEN PERSONEN) VAN PSYCHOLOGISCHE ATTITUDENETWERKEN (BINNEN EEN PERSOON).



hiervoor zijn de omvallende dominosteentjes. In de technische literatuur over onder andere klimaatverandering wordt dit wel een ‘cascading transition’-model genoemd (Brummitt et al., 2015). We noemen ons model daarom een Cascade-model van attitudeverandering. In dit model is het voor het eerst mogelijk om tegelijk de interacties tussen gevoelens, opvattingen en gedrag binnen het individu te modelleren, én de interacties tussen individuen.

We zullen zien dat dit model, met zijn twee niveaus, verklaart waarom attitudes soms zeer moeilijk te beïnvloeden en soms juist zeer instabiel zijn en snel veranderen. Het model biedt ook een geheel nieuwe verklaring voor polarisatie in de samenleving en opties om aan deze polarisatie te ontsnappen.

HET ISING-MODEL VAN INDIVIDUELE ATTITUDES

We gaan nu wat dieper in op het netwerkmodel van individuele attitudes. Met het idee van attitudes als netwerken van relevante gevoelens, opvattingen en gedragingen kunnen we namelijk nog alle kanten op. Netwerken bestaan in allerlei soorten en maten. We kiezen echter een zo eenvoudig mogelijk netwerkmodel op basis van de volgende aannames (Dalege et al., 2018):

- Knopen zijn binair. Men vindt vlees lekker of niet, gelooft wel of niet in de klimaatgevolgen van vleesproductie en heeft leren schoenen of niet. Soms is een continue knoop misschien een betere keuze, maar voor nu sluiten we die uit om het model niet te ingewikkeld te maken.
- Knopen hebben los van hun burens een voorkeur voor één van de twee toestanden. Deze disposities of voorkeuren kunnen specifiek zijn voor een knoop (‘vlees is nou eenmaal lekker’), maar ook extern als een sociale druk van de omgeving (‘iedereen om mij heen is vegetariër’). Ook een informatiecampaagne van de overheid speelt zo’n externe rol
- Knopen zijn in een netwerk verbonden via oorzakelijke verbanden en neigen naar aanpassing aan hun burens. Negatieve emoties over de bio-industrie kunnen het eetgedrag beïnvloeden, terwijl stoppen met het eten van vlees deze emoties weer kunnen versterken. Naast wederzijdse beïnvloeding kan de relatie ook eenzijdig zijn, als in een klassiek oorzakelijk verband. Kennis over de gevolgen van roken zullen het gedrag beïnvloeden maar niet andersom. Als de beïnvloeding sterk is zijn mensen over het algemeen consistent.
- Die consistentie is wel afhankelijk van de sterkte van de

Een overtuigd vegetariër vindt de slacht van dieren bijvoorbeeld naar (gevoelens), de bio-industrie slecht voor het klimaat (een overtuiging), en hij eet geen vlees (gedrag)

aandacht die men besteedt aan de attitude, oftewel hoe betrokken men is bij (of nadenkt over) het onderwerp. Sterke aandacht of betrokkenheid leidt tot een hoge mate aan consistentie. Als de aandacht laag is, veranderen de knoop waarden voortdurend van waarde. Aandacht is dus een cruciale variabele in ons cascade-model.

Het voordeel van deze lijst met versimpelende aannames is dat we zo een model verkrijgen dat zeer goed onderzocht is in de statistische fysica: het Ising-model voor magnetisme. In dit model stellen we de magneet voor als een tweedimensionaal rooster van atomaire minikompaasnaaldjes die met hun pool omhoog óf omlaag staan. Ieder naaldje voelt de invloed van zijn directe burens en neigt zich aan te passen aan de burens. Bij een lage temperatuur gaat deze aanpassing heel goed, bij hoge temperatuur trilt het naaldje veel meer en is de invloed van de burens geringer. Tot slot kan er nog een extern magnetisch veld bestaan, waardoor de naaldjes een voorkeur hebben voor één van de twee richtingen. Zo’n extern effect kan ook knoopspecifiek zijn, bijvoorbeeld een verbod op een gedraging als roken op de werkplek (dan spreken we van een random veld Ising-model).

Er is een parallel te trekken tussen de beschrijving van de magneet en de attitude. De attitude elementen (gevoelens, opvattingen en gedragingen) zijn de naaldjes die onder invloed van hun burens en een extern veld naar één van de twee toestanden springen. Het externe veld is bijvoorbeeld een informatiecampaagne. Aandacht of betrokkenheid is het equivalent van temperatuur, of eigenlijk het omgekeerde van temperatuur (Dalege et al., 2018). Een zeer betrokken attitude komt overeen met een koude magneet waarin de knopen sterk samenhangen en nauwelijks fluctueren. Deze analogie impliceert een concreet wiskundig model voor attitudes.

Maar wat betekent dat nu precies en wat hebben we eraan? De eerste vraag is zeker niet makkelijk te beantwoorden (Finnemann et al., 2021). In de eerste plaats gebruiken fysici het Ising-model om heel precieze kwantitatieve voorspellingen af te leiden. Onze meetinstrumenten voor attitudes laten zo'n niveau van kwantificatie en verificatie niet toe. Gelukkig maakt het Ising-model ook interessante kwalitatieve voorspellingen, die we later zullen bespreken. In de tweede plaats loopt de analogie niet altijd lekker. In ons dagelijks taalgebruik associëren we oplopende temperatuur ('een verhit debat') met polarisatie, terwijl polarisatie in de magneet juist optreedt bij lage temperatuur. Ook is het energiebegrip dat centraal staat in het Ising-model (deeltjes minimaliseren energie door gelijk te gaan staan met hun buren) enigszins problematisch voor het attitude-model. Wat nu het equivalent is van energie als het gaat om attitudes, is niet helemaal duidelijk. Inconsistenties in een belangrijke geachte attitude kosten mentale energie, maar wat mentale energie nu precies is blijft onduidelijk.

Wat hebben we eraan? In de eerste plaats dwingt de formalisering van individuele attitudes als Ising-model ons om zeer precies en concreet te zijn – iets wat lang niet altijd het geval is in de psychologie (Borsboom et al., 2021). In de tweede plaats kunnen we beter begrijpen waarom attitudes soms instabiel zijn en op een ander moment juist weer zeer stabiel zijn.

Aan de hand van dit model hebben we een nieuwe verklaring opgesteld voor de verschillen tussen impliciete en expliciete metingen van attitudes (Dalege & van der Maas, 2020). Impliciete metingen beogen een attitude te meten zonder de aandacht op de attitude te vestigen – bijvoorbeeld door te kijken hoe snel mensen positieve woorden als 'zomer' en negatieve woorden als 'kakkerlak' aan een onderwerp verbinden. Bij expliciete metingen ('Wat vindt u van X') is die aandacht inherent verbonden aan de manier van meten. Uit die verklaring volgt dat impliciete metingen echt anders kunnen zijn dan expliciete metingen maar dat zij wel inherent onbetrouwbaar zijn.

Een weinig betrokken persoon is als een windvaan: hij waait met alle winden mee en zonder wind gaat hij hangen

Twee andere implicaties van het model hebben te maken met weerstand tegen contra-informatie en polarisatie. Maar daarvoor moeten we iets meer uitleggen over het Ising-model.

FASETRANSITIES

Zoals natuurkundigen geïnteresseerd zijn in het gedrag van de hele magneet, zo zijn wij dat ook in het gedrag van het hele attitudesysteem van het individu. We willen dus niet zozeer weten hoe elk naaldje staat maar wat het gemiddelde gedrag is van de naaldjes. Met de formules van het Ising-model en slimme rekentrucs valt daar veel over te zeggen.

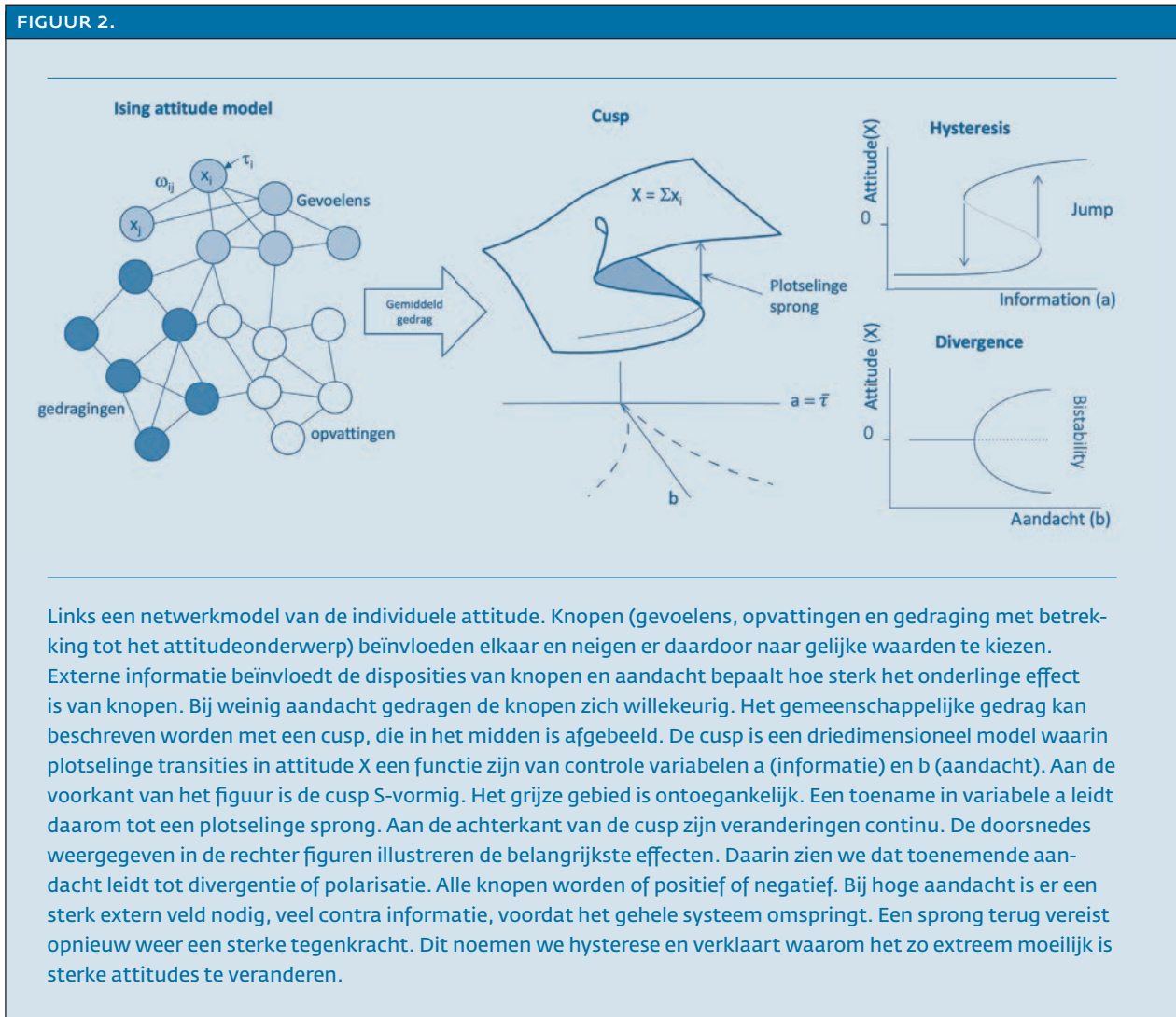
Een interessant geval krijgen we als een hete magneet zonder extern veld, waarin de naaldjes dus wild heen en weer trillen, wordt afgekoeld. In dat geval is er een kritische temperatuur waarop de naaldjes allemaal in het gelid gaan staan in één van de twee oriëntaties. Welke oriëntatie dat zal zijn, is onvoorspelbaar maar plotseling ontstaat er orde in de wanorde. Vertaald naar attitudes zou dit betekenen dat als het belang van een attitude flink toeneemt, mensen snel naar één kant zullen neigen. Dit kan een verklaring zijn voor de gepolariseerde meningen over COVID-vaccins. Dit noemen we ook wel divergentie of polarisatie.

Een ander geval is dat we een magneet bij een lage temperatuur in een toestand ('noord') plaatsen in een extern veld met de omgekeerde oriëntatie ('zuid'). Als dit externe veld in sterkte toeneemt vindt er een plotselinge transitie plaats waarin alle naaldjes omspringen naar 'zuid'. Als we dan weer de andere kant op willen springen, dan is weer een extra groot omgekeerd veld nodig. Deze asymmetrie wordt hysteresis genoemd. In de attitude-literatuur staat dit wel bekend als weerstand tegen contra-informatie (Eagly & Chaiken, 1993).

Er zijn dus drie variabelen in het spel, (a) het gemiddelde gedrag van de naaldjes, (b) de temperatuur en (c) extern veld. Voor de attitude zijn dit de som of gemiddelde van de attitude-elementen en de twee controlevariabelen, aandacht en informatie. De gecombineerde effecten van de twee controlevariabelen kunnen worden weergegeven in een cusp-model, waarin divergentie en hysteresis gecombineerd zijn (Abe et al., 2017). In Figuur 2 wordt dit geïllustreerd.

Dit model genereert belangrijke nieuwe inzichten, die nog wel verdere toetsing behoeven. Willen we bijvoorbeeld iemand overtuigen van een bepaald standpunt of zover krijgen ander gedrag te vertonen, dan hebben we last van een aandachtsparadox. Is de aandacht of betrokkenheid bij het

FIGUUR 2.



onderwerp hoog, dan helpt contra-informatie nauwelijks vanwege het hysteresis-effect. Zijn aandacht en betrokkenheid laag, dan kunnen we iemand tijdelijk beïnvloeden, maar deze invloed beklijft dan niet. Een weinig betrokken persoon is als een windvaan: hij waait met alle winden mee en zonder wind gaat hij hangen. Willen we iemand overtuigen door eerst het belang van de kwestie te bedrukken en dan met nieuwe informatie te komen, dan maken we volgens het model niet veel kans. Bij verhoogde aandacht is de kans op overtuiging immers kleiner. Het is beter eerst informatie te geven en pas dan het belang te benadrukken.

Het model zegt ook iets over conflictbeheersing. Als beide partijen zeer betrokken zijn, is uitwisseling van standpunten

niet vruchtbaar. Eerst moet men kalmeren en pas dan is beweging mogelijk. Vermoedelijk is dit ook het effect van urenlange nachtelijke onderhandelingen. Op een bepaald moment is iedereen moe en hongerig en zijn de hysteresis-effecten verkleind. Pas dan is een akkoord mogelijk.

In eerder onderzoek (van der Maas et al., 2003) hebben we bewijs gevonden voor dit divergentie- of polarisatie-effect (zie figuur 2). Het is ook mogelijk met behulp van bijvoorbeeld vragenlijst-data attitude-netwerken in kaart te brengen (Dale et al., 2017).

MODELLEN VAN OPINIEVERSPREIDING

We gaan nu de stap maken naar het tweede beschrijvings-

niveau, dat van sociale netwerken bestaande uit individuen met attitudes zoals hierboven beschreven.

Als het gaat om de verspreiding van attitudes over populaties – men spreekt dan meestal over opinies –, komen we namelijk opnieuw uit op netwerkmodellen. Mensen maken immers onderdeel uit van sociale netwerken. Mensen zijn verbonden met elkaar en in die verbindingen is er sprake van clusters zoals familie, verenigingen en collega's op werk. Ook hier geldt dat sommige verbindingen sterk zijn en andere verbindingen zwakker.

Ook hier zijn modellen uit de statistische fysica relevant. Er is zelf een heel vakgebied ontstaan, soms sociofysica genoemd, waarin honderden modellen voor opinieverspreiding zijn voorgesteld en onderzocht, vaak met computersimulaties (Castellano et al., 2009).

Zulke modellen zijn uiteraard versimpelingen van onze werkelijkheid waarin opinieverspreiding verloopt via allerlei kanalen. De eerste versimpeling betreft de definitie van opinie op individueel niveau, met name of deze continu (van negatief naar positief) of discreet (-1,1) is. Afhankelijk van deze keuze komt men in geheel andere modeltradities terecht.

Kiest men voor discreet, dan komen varianten van het Ising-model voor magnetisme weer op tafel (Galam, 2008). De knopen zijn nu mensen die in een sociaal netwerk interacteren. In dit model kunnen clusters van positieve en negatieve opinies ontstaan. Ook in continue opiniemodellen kan polarisatie optreden (Hegselmann & Krause, 2002). Men neemt doorgaans aan dat interacties tussen personen leiden tot een bepaalde mate van convergentie in opinie. Met elke interactie komen we wat nader tot elkaar. Toch kan in deze modellen ook polarisatie optreden. De hoofdverklaring hiervoor is dat netwerken vaak onvoldoende verbonden zijn. Het blijkt dat er extreme polarisatie kan ontstaan als men veronderstelt dat personen die te veel verschillen in opinie gewoonweg niet (goed) communiceren.

Een andere versimpeling betreft de structuur van het sociale netwerk zelf. Echte sociale netwerken zijn zeer complex, maar worden in simulaties sterk vereenvoudigd. Zo kunnen we iedereen op een cirkel plaatsen waarbij iedereen uitsluitend met twee burens interacteert, of we kunnen iedereen in een rooster plaatsen waar men met vier of acht burens praat. Er zijn nog veel meer mogelijkheden, zoals de schalingsvrije netwerken waarin slechts enkele knopen (hubs) extreem veel verbindingen hebben (Barabási et al., 1999).

De derde versimpeling gaat over wat er gebeurt tijdens

interacties die in werkelijkheid ook weer zeer divers en complex zijn en ook media en sociale media omvatten. Vaak kiest men ervoor in elke stap van de simulatie steeds een willekeurige knoop in het netwerk te laten interacteren. Tijdens deze interactie wordt de opinie van de burens op een of andere wijze meegewogen.

Het is onmogelijk een compleet beeld van alle modelopties te geven, maar we hopen dat de lezer nu een beeld heeft van de diversiteit van opiniemodellen.

HET CASCADE-MODEL VAN ATTITUDES

Ons model van opinieverspreiding (van der Maas et al., 2020) wijkt af van eerdere modellen omdat we de attitude van een individu niet opvatten als een simpele binaire of continue variabele maar als een niet-lineaire variabele. Zoals gezegd beschouwen we de individuele attitude als een Ising-netwerk en zijn individuen weer onderdeel van een sociaal netwerk. We hebben dus eigenlijk een model van netwerken van netwerken.

Zo'n model is echter zeer complex en nauwelijks te bestuderen. We maken daarom gebruik van de versimpeling in de vorm van een cusp die in figuur 2 is weergegeven. Het gemiddelde gedrag van het individuele attitudenetwerk kan beschreven worden als een cusp. De cusp lijkt, zeker als je die voor het eerst ziet, ook geen simpel model maar laat zich wel vangen in slechts één (niet-lineaire) wiskundige vergelijking. Ons model is daarmee een netwerk van cusps. Omdat de cusp zich soms continu (als aandacht laag is) en soms discontinu (als aandacht hoog is) gedraagt, is het model een generalisatie van de bestaande discrete en continue modellen.

Het idee van een netwerk van cusps is niet nieuw. De wiskundige eigenschappen van dit model zijn decennia geleden al enigszins in kaart gebracht (Abraham et al., 1991). De toepassing is onderzocht voor neurale netwerken, immuunsystemen en klimaatsystemen. Onze toepassing op opinieverspreiding wijkt wel af van deze toepassingen, met name omdat we beide controlevariabelen laten veranderen.

In het cascade-model van attitudeverandering wordt de attitude van een persoon beschreven met de cusp als afgebeeld in figuur 2. De toestand van één persoon is dus gedefinieerd aan de hand van drie variabelen. De eerste is de opinie zelf, die kan zowel positief als negatief zijn, maar ook neutraal. Dat geldt ook voor de tweede variabele, de informatie waarover de persoon beschikt. Informatie is een vergaarbak van variabelen die met de pro/contra-opinie te maken hebben. Als het gaat om bijvoorbeeld de opinie over het eten

De psychologische processen die een rol spelen in individuele radicalisering kunnen niet los gezien worden van sociologische, politieke en economische processen

van vlees, kan het gaan om medische informatie, maar ook over culturele achtergrond. De derde variabele is aandacht voor of betrokkenheid bij het onderwerp. Deze variabele is dus onafhankelijk van de informatiewaarde. Bij neutrale informatie en hoge aandacht kan de opinie extreem positief of negatief zijn, maar niet neutraal (ontoegankelijke grijze gebied in de cusp in figuur 2). Zo raken opinies dus veelal gepolariseerd bij een hoge aandacht voor het onderwerp.

Het cascade-model van attitudes is gebaseerd op een drietal assumpties over de effecten van interacties tussen individuen:

- Interacties in het netwerk worden eerder aangegaan door individuen met een hoge aandacht. Wie in het geheel niet geïnteresseerd is in bijvoorbeeld het abortusvraagstuk, zal daar bij de koffie ook niet gauw een discussie over starten.
- Als mensen interacteren wisselen ze informatie uit en bewegen ze enigszins naar elkaar toe. We veronderstellen dat deze beweging afhangt van de betrokkenheid. Iemand die laag betrokken is, zal geneigd zijn meer op te schuiven dan iemand die zeer betrokken is.
- Aandacht voor of betrokkenheid bij een onderwerp neemt slechts langzaam af, maar wordt door interacties weer verhoogd. De kernenergiediscussie is in Nederland al langere tijd verstomd, maar de deelnemers aan die discussie verliezen maar heel langzaam betrokkenheid bij dat debat. Nieuwe discussies hierover kunnen de betrokkenheid weer snel verhogen.

Het is mogelijk deze aannames in de vorm van vergelijkingen of computercode op te schrijven. Met computersimulaties kunnen we vervolgens het gedrag van dit model onderzoeken. Dat werkt eigenlijk heel eenvoudig. Eerst kiezen we een sociaal netwerk, bijvoorbeeld een 'random network' waarin personen met een vaste uniforme kans verbonden zijn. We kiezen dan initiële waarden voor informatie en aandacht. Met de cusp-formule berekenen we dan de bijbehorende opinie. Dan kiezen we steeds random een persoon gewogen met de aandacht van de personen, dus

waarschijnlijk iemand met hoge aandacht. We kiezen vervolgens een willekeurige buur van deze persoon en wisselen informatie uit. Dat wil zeggen dat we beide informatieposities iets naar elkaar opschuiven afhankelijk van de aandacht van beide personen. Van deze twee personen verhogen we de aandacht, van alle andere personen verlagen we de aandacht een klein beetje. Tot slot berekenen we opnieuw van alle personen de opinie met behulp van de cusp-vergelijking. En deze stappen herhalen we keer op keer, een vast aantal keren of totdat er niets meer echt verandert.

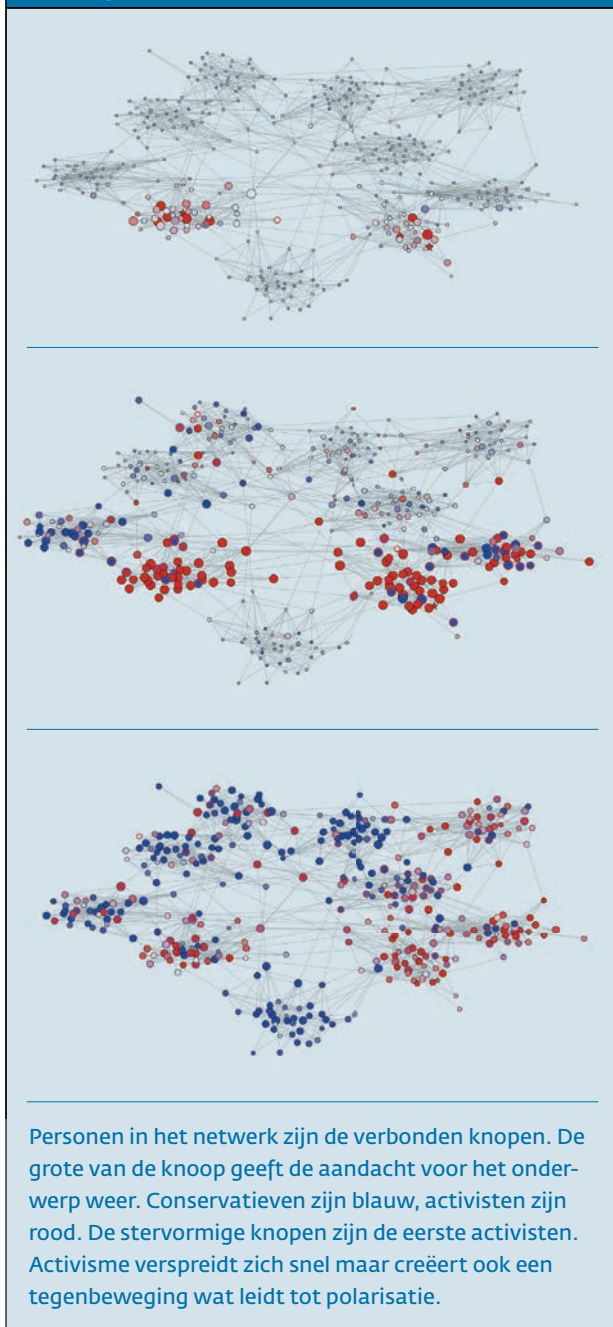
DE ZWARTE PIET-SIMULATIE

Om te illustreren hoe op onverwachte wijze polarisatie ontstaat in dit model hebben we het Zwarte Piet-scenario bedacht. Eerst kiezen we een sociaal netwerk, in dit geval een stochastisch blokmodel. Hierin bestaan clusters van personen met relatief veel verbindingen, terwijl verbindingen tussen clusters schaarser zijn (zie figuur 3). De initiële waarden kiezen we analoog aan de situatie in Nederland voor het Zwarte Piet-debat op gang kwam. Dat wil zeggen dat we aan de overgrote meerderheid, de conservatieven, een lage aandacht voor het onderwerp toekennen en een opinie en informatiewaarde licht positief ten opzichte van Zwarte Piet. Maar een klein aantal personen, activisten, hebben juist erg veel aandacht en negatieve informatie voor het onderwerp en zijn sterk gekant tegen Zwarte Piet. Wat zal er nu gebeuren in de simulatie?

We verwachten dat de activisten interacties initiëren (assumptie a) en de discussie winnen (assumptie c). De conservatief heeft immers minder aandacht voor het onderwerp en zal dus meer opschuiven. Op basis van deze twee assumpties verwachten we dus een snelle verspreiding van de activistische opinie. Maar het effect van assumptie b verandert de zaak. Doordat de interacties toenemen, neemt ook de aandacht toe. In een deel van de conservatieven verandert de informatie en opinie snel en neemt dan de aandacht langzaam toe. Zij veranderen van standpunt. Maar in andere conservatieven verandert de aandacht te snel. Divergentie treedt op (figuur 2) en deze conservatieven

radicaliseren de andere kant op, ze worden activist pro Zwarte piet. Deze divergentie treedt niet noodzakelijk op door een toename aan positieve informatie – soms is het voldoende dat aandacht toeneemt en is de globale opinie extremer dan de onderliggende informatie.

FIGUUR 3. EEN POLARISEREND NETWERK.



Figuur 3 laat de uitkomst van de simulatie zien. Eerst winnen de activisten terrein (eerste figuur), dan komt anti-activisme op (tweede figuur) en we eindigen met een gepolariseerd netwerk (derde figuur).

Uiteraard hangt de uitkomst af van allerlei parameterinstellingen. Als aandacht maar langzaam toeneemt dan winnen de activisten vaker. We hebben een website gemaakt met een uitleg van het model en de mogelijkheid zelf eenvoudig simulaties uit te voeren (<https://lukovdm.github.io/hiom-sim/>).

EEN TEGEN-INTUÏTIEVE VOORSPELLING

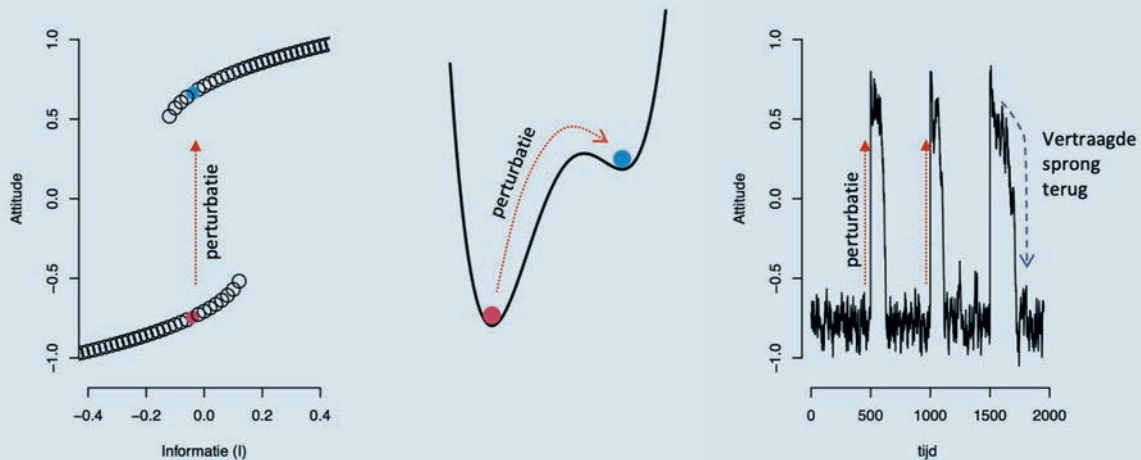
In de modellen waarin opinie als een continue variabele wordt opgevat, ontstaat polarisatie door gebrek aan contact. Als er contact is, vindt toenadering plaats. Een belangrijke rol speelt het concept van 'bounded confidence', wat we hier als beperkt vertrouwen zullen vertalen. Het idee is dat als mensen te veel van opinie verschillen, ze de interactie zullen weigeren waardoor er geen toenadering kan optreden. Het is uitgebreid aangetoond dat dit kan leiden tot polarisatie in twee of meer groepen (Lorenz, 2007).

Als we dit concept inbouwen in ons model, dan zien we inderdaad dat activisten en conservatieven niet meer met elkaar interacteren en polarisatie optreedt. Ons model suggereert echter een uitweg. Het is mogelijk om de activist tijdelijk een conservatieve opinie te laten aannemen ondanks de progressieve informatietoestand. Dit kan omdat de cusp bistabiel is (twee stabiele punten heeft) bij voldoende aandacht. Dit is geïllustreerd in figuur 4. Bij een informatie-waarde van -1 (progressief) kan een verstoring van de opinie naar conservatief enige tijd stabiel zijn. We zien wel dat na enige tijd de opinie terugspringt naar de meer stabiele progressieve waarde.

In de tijdelijk conservatieve toestand is deze persoon opeens wel gesprekspartner voor de 'echte' conservatieven omdat de opinies nu dicht bij elkaar liggen en er dus voldoende vertrouwen is. De informatie die de getransformeerde conservatief uitwisselt, is alleen wel progressief van aard. En als deze ook meer aandacht heeft voor het onderwerp, zal de echte conservatief mogelijk flink opschuiven naar progressief.

Figuur 5 illustreert het effect van deze manipulatie. We beginnen met een situatie van weinig betrokken conservatieven en een minderheid van activisten. De bovenste figuren laten zien wat er gebeurt zonder manipulatie. De activisten overtuigen weinig conservatieven die wel meer

FIGUUR 4. EEN POLARISEREND NETWERK.



Een progressieve persoon wordt een conservatieve opinie opgelegd (een perturbatie, figuur links). Deze toestand is ondanks de progressieve informatietoestand stabiel (figuur midden) maar zal door verstoringen (op tijdstippen 500, 1000 en 1500) niet lang standhouden (figuur rechts).

aandacht voor het onderwerp ontwikkelen. Resultaat is sterke polarisatie. De onderste figuren laten het effect van de manipulatie zien. Om de zoveel tijd duwen we een aantal progressieven even in de conservatieve opinie. Vanuit dezelfde startpositie zien we dat deze activisten wel anderen overtuigen zonder dat extreme polarisatie ontstaat.

Dit is een opvallende predictie van het model: een manipulatie van progressieven in *conservatieve* richting leidt tot een massale opinieverandering in de richting van het *progressieve* standpunt. Tot nu toe hebben we deze predictie nog niet getest in een experiment. Er bestaat wel aanverwant onderzoek naar duale identiteiten (González & Brown, 2006) two experiments were conducted involving groups of different size and/or status that worked together on a cooperative task. Three categorization strategies (decategorization, recategorization, and dual identity en de rol van bruggenbouwers tussen geradicaliseerde groepen (Hegselmann & Krause, 2015) waarin soortgelijke effecten gevonden zijn.

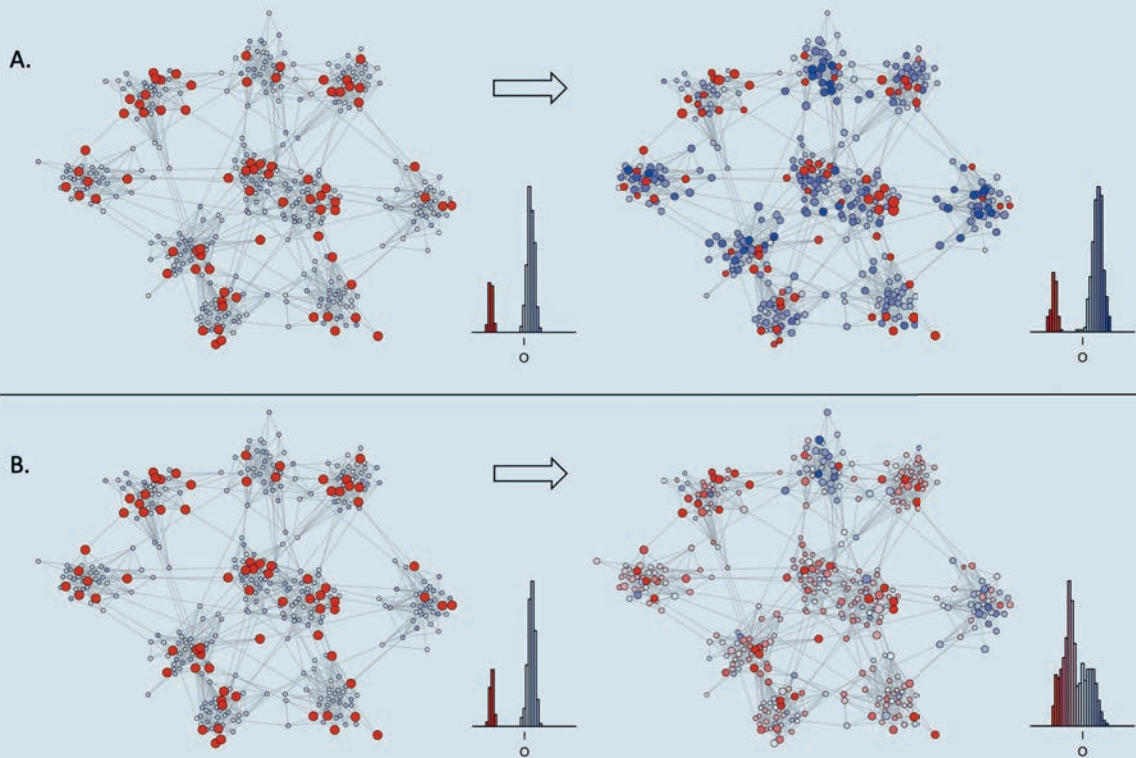
DISCUSSIE

We hebben polarisatie verklaard uit de interactie binnen personen en tussen persoonsprocessen. Binnen personen speelt aandacht en daarmee de behoefte aan consistentie een cruciale rol. Op het sociale niveau, tussen personen, leidt dat

uiteindelijk tot clustering van opinies. De psychologische processen die een rol spelen in individuele radicalisering kunnen dus niet los gezien worden van sociologische, politieke en economische processen. Radicalisering op individueel niveau en polarisatie op groepsniveau zijn gekoppelde complexe verschijnselen. In de studie van complexe systemen zoeken we naar een formele beschrijving van de essentie van de verschijnselen. Het bekendste voorbeeld is de modelering van files waarbij zeer ingewikkelde auto's bestuurd door zeer ingewikkelde mensen in simulaties versimpeld worden tot autootjes die harder gaan als ze de ruimte hebben en afremmen als ze een andere auto naderen. Het blijkt dat we files met dit eenvoudige model goed kunnen begrijpen en de effecten van interventies (bijvoorbeeld een wegverbreding) kunnen voorspellen (Nagatani, 2002).

Wij volgen een soortgelijke benadering van de ontwikkeling van attitudes in een sociale context. Ons netwerkattitudemodell is een basale beschrijving van de essentie van de attitude, die overigens opvallend consistent is met verschillende klassieke attitudetheorieën en een veelheid van empirische gegevens. Zo leidt aandacht voor de attitude tot hogere consistentie tussen de knopen van de attitude. Er is veel bewijs voor dit fenomeen van dissonantiereductie (McGrath, 2017). En ander voorbeeld is het 'mere thought'-effect, waarbij alleen het proces van nadenken over de

FIGUUR 5.



In geval A ontstaat sterke polarisatie omdat conservatieven (blauw) interacties met progressieven (rood) weigeren. In geval B worden een deel van de progressieven een conservatieve opinie gegeven terwijl hun onderliggende informatiewaarde progressief blijft. Door het hysteresis-effect duurt het enige tijd voor deze personen terugschieten in hun progressieve opinie. In dit tijd zijn ze wel gesprekspartner voor conservatieven die daardoor steeds progressiever worden. Manipulatie richting conservatief leidt dus uiteindelijk tot een progressieve populatie.

attitude al leidt tot polarisatie (Tesser, 1978). De conceptualisatie van psychologische systemen als anetwerken van cognitieve functies, klinische symptomen en hier dus attitude-elementen wordt momenteel breed omarmd in de psychologie.

De stap om deze psychologische netwerken op hun beurt weer als onderdeel te zien van sociale netwerken is nieuw. Simulatiemodellen voor de verspreiding van opinies bestaan in allerlei variëteiten. Het is opvallend hoe populair dit onderzoek is in de bètawetenschappen. Een bekend overzichtspaper (Castellano et al., 2009) met de titel 'Statistical physics of social dynamics', is bijna 4000 keer geciteerd. Ons model heeft vele overeenkomsten met de standaardmodellen uit dit veld, bijvoorbeeld als het gaat om de uitwisseling van

informatie. Een groot verschil is de modelering van het individu. In de meeste modellen kiest men zonder veel onderbouwing voor een zeer vergaande reductie van de psychologische component van het model.

Wij hebben ervoor gekozen het netwerkmodel van attitudes als uitgangspunt te nemen. Ook wij kiezen voor een reductie (de cusp als samenvatting van de dynamica het psychologisch netwerk), maar die is wel rijker dan die van de standaardmodellen. Ten eerste introduceren we aandacht als nieuwe variabele die, zoals we hebben kunnen zien, een cruciale rol kan spelen. Ten tweede is de opinie in ons model een continue of discrete variabele afhankelijk van de aandacht voor het onderwerp. Ten derde treedt er hysteresis op in de opinie op individueel niveau, met een reeks van

Veel meer complexe psychologische verschijnselen die gekenmerkt worden door plotselinge transities zijn ingebed in sociale verschijnselen

gevolgen, waaronder weerstand tegen contra-informatie, polarisatie (figuur 3) maar ook nieuwe opties voor de bestrijding van polarisatie (figuur 5).

Hoe dit model zich precies verhoudt tot andere modellen vereist nog verder onderzoek. De assumpties van het model vereisen ook nog meer onderbouwing en empirische toetsing van de predicties moet om die reden op de agenda. Wij zien dit model als startpunt van een nieuwe benadering van psychosociale systemen. Veel meer complexe psychologische verschijnselen die gekenmerkt worden door plotselinge transities zijn ingebed in sociale verschijnselen. Een goed voorbeeld is verslaving. De dynamica van verslaving is complex en bijvoorbeeld stoppen maar ook terugval vertonen de typische kenmerken van fase transities (Piazza & Deroche-Gamonet, 2013). Tegelijkertijd is verslaving een sociaal verschijnsel, middelengebruik is besmettelijk, denk bijvoorbeeld aan de verspreiding van roken onder tieners. Ook andere psychologische stoornissen zoals angst zijn in bepaalde mate sociaal besmettelijk. Een ander positief voorbeeld is leren. Het aanleren van bijvoorbeeld het oplossen van Rubik's kubus is een complex leerproces dat zeer snel kan verlopen als kinderen elkaars leerproces stimuleren.

Terug nog naar onze oorspronkelijke vragen. Waarom valt iemand opeens van zijn geloof? In het netwerkmodel van attitudes zijn hiervoor twee routes. De eerste is geleidelijk. Als iemand minder aandacht heeft voor het onderwerp, dus minder betrokken raakt, dan is deze persoon gevoeliger voor contra-informatie en kan van mening veranderen. In onderzoek naar de-radicalisering is dit een bekend verschijnsel met als nadeel dat dit een langzaam proces is. De andere route is dat inconsistenties in het netwerk, door contra-informatie, de bestaande attitude destabiliseren, waarna een plotselinge transitie kan optreden. Dit kan een rol spelen bij stoppen met roken, wat bijna altijd een abrupt proces is.

Het in dit paper naar voren gebrachte perspectief biedt ook inzicht in hoe complottheorieën zich verspreiden. Het is evident dat radicalisering besmettelijk is en dat complottheorieën hierin een grote rol spelen. Een complottheorie beschermt een radicale attitude tegen destabiliserende

inconsistentie door contra-informatie verdacht te maken.

Hiernaar is meer theoretisch en empirisch onderzoek vereist.

En wat te doen tegen de toenemende polarisatie in de samenleving? Dat is helaas geen makkelijk te beantwoorden vraag. Ons model geeft wel beter begrip van de processen onderliggend aan radicalisering en polarisatie maar biedt niet direct een oplossing. De reden is dat aandacht voor en betrokkenheid bij een onderwerp snel kunnen toenemen maar slechts langzaam lijken af te nemen. De voortdurende aandacht voor het vaccinatievraagstuk, in de media maar ook in directe discussies tussen mensen, maakt depolariseren bijvoorbeeld lastig. In een pandemie kunnen we moeilijk afspreken het er maar een jaartje niet meer over te hebben. Het huidige model biedt niet veel andere aanknopingspunten anders dan contact blijven houden. Zodra de communicatie tussen groepen verdwijnt, is polarisatie volgens elk model onvermijdelijk.

We hopen met dit werk niet alleen meer inzicht te hebben geboden in de polarisatie van attitudes in de samenleving maar ook het nut van het modelleren in de psychologie te hebben aangetoond. De formele aanpak van theorievorming dwingt ons tot meer precisie in definities. Nieuwe predicties volgen controleerbaar uit de computersimulaties van het model. Deze wetenschappelijke ontwikkeling is natuurlijk al langer gangbaar in de natuurwetenschappen maar biedt ook kansen aan de psychologie.

OVER DE AUTEURS

Han L. J. van der Maas¹ is hoogleraar Algemene Methodologie aan de Universiteit van Amsterdam (UvA). Dr. Jonas Dalege is postdoctoral fellow aan het Sante Fe Institute in de Verenigde Staten. Frenk van Harreveld en Denny Borsboom zijn respectievelijk hoogleraar Sociale Psychologie en hoogleraar Grondslagen van de Psychometrie en de Psychologie aan de UvA. Dr. Lourens Waldorp is universitair hoofddocent Methodenleer aan de UvA.

Correspondentie aangaande dit artikel via Han van der Maas:

H.L.J.vanderMaas@uva.nl.

¹ Han van der Maas ontving in april dit jaar een ERC Advanced Grant van de Europese Unie voor verder onderzoek op dit terrein.

Summary

POLARIZATION AS COMPLEX PHENOMENON

H. L. J. VAN DER MAAS, J. DALEGE,
F. VAN HARREVELD, D. BORSBOOM &
L. WALDORP

How does polarization arise in society? Why do people change attitudes or opinions? To

answer these questions, we present a cascading model of attitude change, integrating ideas from different scientific fields. The model includes both the interactions between feelings, opinions, and behaviors within the individual and the interactions between individuals. We will see that this

model, with its two levels, explains why attitudes are sometimes very difficult to influence and sometimes very unstable and change rapidly. The model also offers an entirely new explanation for polarization in society and options for escaping this polarization.

Literatuur

- Abe, Y., Ishida, M., Nozawa, E., Ootsuka, T., & Yahagi, R. (2017). Cusp singularity in mean field Ising model. *European Journal of Physics*, 38(6), 065102. <https://doi.org/10.1088/1361-6404/aa82fc>
- Abraham, R., Keith, A., Koebe, M., & Mayer-Kress, G. (1991). Computational unfolding of double-cusp models of opinion formation. *International Journal of Bifurcation and Chaos*, 01(02), 417–430. <https://doi.org/10.1142/S0218127491000324>
- Barabási, A.-L., Albert, R., & Jeong, H. (1999). Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and Its Applications*, 272(1), 173–187. [https://doi.org/10.1016/S0378-4371\(99\)00291-5](https://doi.org/10.1016/S0378-4371(99)00291-5)
- Borsboom, D., van der Maas, H. L. J., Dalege, J., Kievit, R. A., & Haig, B. D. (2021). Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. *Perspectives on Psychological Science*, 1745691620969647. <https://doi.org/10.1177/1745691620969647>
- Brummitt, C. D., Barnett, G., & D'Souza, R. M. (2015). Coupled catastrophes: Sudden shifts cascade and hop among interdependent systems. *Journal of The Royal Society Interface*, 12(112), 20150712. <https://doi.org/10.1098/rsif.2015.0712>
- Castellano, C., Fortunato, S., & Loreto, V. (2009). Statistical physics of social dynamics. *Reviews of Modern Physics*, 81(2), 591–646. <https://doi.org/10.1103/RevModPhys.81.591>
- Dalege, J., Borsboom, D., Harrevel, F. van, & Maas, H. L. J. van der. (2018). The Attitudinal Entropy (AE) Framework as a General Theory of Individual Attitudes. *Psychological Inquiry*, 29(4), 175–193. <https://doi.org/10.1080/1047840X.2018.1537246>
- Dalege, J., Borsboom, D., van Harrevel, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, 123(1), 2–22. <https://doi.org/10.1037/a0039802>
- Dalege, J., Borsboom, D., van Harrevel, F., & van der Maas, H. L. (2017). Network analysis on attitudes: A brief tutorial. *Social psychological and personality science*, 8(5), 528–537.
- Dalege, J., & van der Maas, H. L. J. (2020). Accurate by Being Noisy: A Formal Network Model of Implicit Measures of Attitudes. *Social Cognition*, 38(Supplement), s26–s41. <https://doi.org/10.1521/soco.2020.38.supp.s26>
- Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.
- Finnemann, A., Borsboom, D., Epskamp, S., & van der Maas, H. L. J. (2021). The Theoretical and Statistical Ising Model: A Practical Guide in *R*. *Psych*, 3(4), 593–617. <https://doi.org/10/gm4fvt>
- Galam, S. (2008). Sociophysics: A review of galam models. *International Journal of Modern Physics C*, 19(03), 409–440. <https://doi.org/10.1142/S0129183108012297>
- González, R., & Brown, R. (2006). Dual identities in intergroup contact: Group status and size moderate the generalization of positive attitude change. *Journal of Experimental Social Psychology*, 42(6), 753–767. <https://doi.org/10.1016/j.jesp.2005.11.008>
- Hegselmann, R., & Krause, U. (2002). Opinion dynamics and bounded confidence: Models, analysis and simulation. *J. Artificial Societies and Social Simulation*, 5.
- Hegselmann, R., & Krause, U. (2015). Opinion dynamics under the influence of radical groups, charismatic leaders, and other constant signals: A simple unifying model. *Networks & Heterogeneous Media*, 10(3), 477. <https://doi.org/10.3934/nhm.2015.10.477>
- Lorenz, J. (2007). Continuous opinion dynamics under bounded confidence: A survey. *International Journal of Modern Physics C*, 18(12), 1819–1838. <https://doi.org/10.1142/S0129183107011789>
- McGrath, A. (2017). Dealing with dissonance: A review of cognitive dissonance reduction. *Social and Personality Psychology Compass*, 11(12), e12362. <https://doi.org/10.1111/spc3.12362>
- Nagatani, T. (2002). The physics of traffic jams. *Reports on Progress in Physics*, 65(9), 1331–1386. <https://doi.org/10/dkvhk5>
- Piazza, P. V., & Deroche-Gamonet, V. (2013). A multistep general theory of transition to addiction. *Psychopharmacology*, 229(3), 387–413. <https://doi.org/10/f49rfv>
- Tesser, A. (1978). Self-Generated Attitude. In L. Berkowitz (Red.), *Advances in Experimental Social Psychology* (Vol. 11, pp. 289–338). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60010-6](https://doi.org/10.1016/S0065-2601(08)60010-6)
- van der Maas, H. L. J., Dalege, J., & Waldorp, L. (2020). The polarization within and across individuals: The hierarchical Ising opinion model. *Journal of Complex Networks*, 8(cnaa010). <https://doi.org/10.1093/com-net/cnaa010>
- van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, 113(4), 842–861. <https://doi.org/10.1037/0033-295X.113.4.842>
- van der Maas, H. L. J., Kolstein, R., & van der Pligt, J. (2003). Sudden Transitions in Attitudes. *Sociological Methods & Research*, 32(2), 125–152. <https://doi.org/10.1177/0049124103253773>