# Inhoud | Contents

## Artikelen | Features

### Wiskunde en de Sociale Wetenschappen



**Onderwijs | Education**



**Onderzoek | Research**



**Evenement | Event**



**In Memoriam | Obituary**



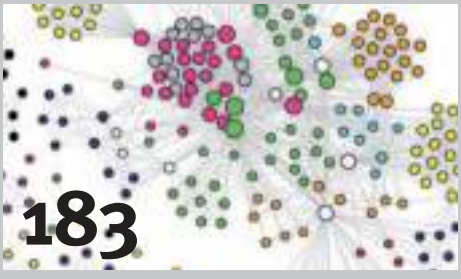
**In Memoriam | Obituary**



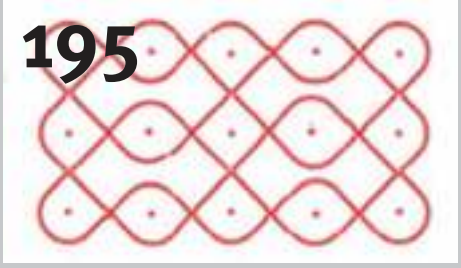
## Rubrieken | Departments



**Uitgelicht**




Denny Borsboom en Maarten Marsman bespreken netwerkbenaderingen van de psychometrie.




Het tekenen van koloms is een eeuwenoude traditie in Zuid-India. S. Naranan, T.V. Suresh en Swarna Srinivasan beschrijven de wiskunde achter deze volkskunst.

## Colofon

Het Nieuw Archief voor Wiskunde is een uitgave van het Koninklijk Wiskundig Genootschap en verschijnt vier maal per jaar. Het tijdschrift richt zich op eenieder die zich beroepsmatig met wiskunde bezighoudt, als academisch of industrieel onderzoeker, student, leraar, journalist of beleidsmaker. Het stelt zich ten doel te berichten over ontwikkelingen in de wiskunde in het algemeen en in de Nederlandse wiskunde in het bijzonder.

ISSN 0028-9825

**Adres**
Nieuw Archief voor Wiskunde
Centrum Wiskunde & Informatica
Postbus 94079, 1090 GB Amsterdam
tel. 020-5924288
redactie@nieuwarchief.nl
www.nieuwarchief.nl



**Druk**
Veldhuis Media

**Advertenties**
advertenties@nieuwarchief.nl

---

**Koninklijk Wiskundig Genootschap**
www.wiskgenoot.nl

**Platform Wiskunde Nederland**
www.platformwiskunde.nl

**European Mathematical Society**
www.euro-math-soc.eu

**International Mathematical Union**
www.mathunion.org

**Koninklijk Wiskundig Genootschap**
Het Koninklijk Wiskundig Genootschap (KWG) is een landelijke vereniging van beoefenaren van de wiskunde. In 1778 opgericht onder de zinspreuk: 'Een onvermoeide arbeid komt alles te boven' is het 's werelds oudste nationale wiskundegenootschap. Leden ontvangen het Nieuw Archief voor Wiskunde als onderdeel van hun lidmaatschap.

**Bestuur**
Jan Wiegerinck (UvA, voorzitter), Marije Elkenbracht (ABN AMRO, penningmeester), Marie-Colette van Lieshout (CWI, secretaris), Danny Beckers (VU), Theo van den Bogaart (HU), Sonja Cox (UvA), Barry Koren (TU/e), Michael Müger (RU), Wioletta Ruszel (TUD).

**Secretariaat**
Koninklijk Wiskundig Genootschap, CWI
Postbus 94079, 1090 GB Amsterdam
wiskgenoot@wiskgenoot.nl
www.wiskgenoot.nl

**Ledenadministratie KWG / Abonnementen**
admin@wiskgenoot.nl, tel. 020-5924008
Wijzigingen of opzegging kunt u ook doorgeven door in te loggen op www.wiskgenoot.nl.

**Contributie**
Tarieven per jaar (bij betaling per automatische incasso krijgt u € 2,50 korting):
– gewoon lid: € 92,50
– reciprociteitslid: € 72,50
– gepensioneerden en werklozen: € 47,50
– aio's, oio's, studenten: € 37,50 (max. 4 jaar)
– lid zonder Nieuw Archief: € 52,50
Studenten en pas afgestudeerden kunnen een jaar lang gratis lid worden.
Voor verzending van het Nieuw Archief naar het buitenland wordt een portotoeslag van € 10 in rekening gebracht.
Instituten in Nederland en buitenland kunnen zich abonneren op het Nieuw Archief voor Wiskunde voor € 100 (Nederland), € 110 (Europa) of € 112,50 (buiten Europa).

**Members of foreign societies**
Members of *SIAM*, the *Société Mathématique de France*, the *Gesellschaft für Angewandte Mathematik und Mechanik*, the *Deutsche Mathematiker-Vereinigung*, and of the *American*, *Australian*, *Belgian*, *Indian*, *London*, *Spanish*, and *South-African Mathematical Societies* living outside of the Netherlands pay € 72,50 instead of the full membership fee of € 92,50 a year. Conversely, these foreign societies, as well as the VvS+OR and the NVORWO, have reduced membership fees for KWG members. A postage charge of € 10 is added for members living abroad but in Europe, and a charge of € 12,50 for members living outside of Europe.

**Exchange subscriptions**
Koninklijk Wiskundig Genootschap Library
Centrum Wiskunde & Informatica
P.O. Box 94079, NL-1090 GB Amsterdam
KWG-exchange@cwi.nl

## Informatie voor auteurs

**Kopij**
Het Nieuw Archief voor Wiskunde is een gerediger tijdschrift. Kopij dient per e-mail te worden aangeboden aan kopij@nieuwarchief.nl. Uitgebreide informatie en auteursinstructies kunt u vinden op www.nieuwarchief.nl.

**Volgende nummers**

| *nummer* | *maand verschijnen* | *deadline kopij* |
|---|---|---|
| 20(4) | december 2019 | verstreken |
| 21(1) | maart 2020 | 01-11-2019 |
| 21(2) | juni 2020 | 01-02-2020 |
| 21(3) | september 2020 | 01-05-2020 |

## Instituutsleden

De publicaties van het Koninklijk Wiskundig Genootschap worden mede mogelijk gemaakt door de bijdragen van de volgende instituten.

- Centrum Wiskunde & Informatica
- Rijksuniversiteit Groningen
- Universiteit van Amsterdam
- Universiteit Leiden
- Vrije Universiteit
- Universiteit Utrecht
- Technische Universiteit Eindhoven
- Eurandom
- Technische Universiteit Delft
- Universiteit Twente
- Radboud Universiteit Nijmegen
- Freudenthal Instituut

## Op het omslag

Toetsen en examens hebben veel invloed op schoolcarrières en kansen op een baan. Het is dan ook van groot belang dat de uitslagen ervan betrouwbaar zijn. Lees op pagina 161 het artikel van Bernard Veldkamp over de tak van wiskunde die gebruikt wordt bij toetsen en examens en bekend staat als psychometrie.

# Wiskunde en de Sociale Wetenschappen

Mijn vrouw kent alle mensen in de wijk bij naam en toenaam. Niet alleen de namen en gezichten, maar ook waar ze met vakantie naartoe gaan of welke hobby's ze hebben. Als we gaan winkelen wordt ze voortdurend gestopt door deze of gene en ben ik de zwijgende getuige van een lang gesprek. "Wie was dat?" vraag ik dan na afloop. "Onze buurman/buurvrouw", is steevast haar antwoord. Ik ben mij van mijn gebrek aan sociale vaardigheden welbewust en daarom schrijf ik met gepaste trots de inleiding van dit themanummer over Wiskunde en de Sociale Wetenschappen.

Wie door dit themanummer bladert, ziet al snel dat grafen een voorname rol spelen in de sociale wetenschappen, maar onder een andere naam. Een graaf heet een netwerk en eigenlijk is dat beter, want bij graaf denk je al gauw aan een Haagse edelman en zijn zoontje Jantje. Overigens was een *graaf* oorspronkelijk helemaal geen edelman, maar gewoon een hoge ambtenaar die kon schrijven. Graaf, als in *graveren*. In Engeland en Frankrijk was schrijven minder belangrijk dan tellen en daarom heette zo'n man daar een *count*, of een *comte*, maar dit terzijde.

Vroeger waren mensen die konden rekenen en schrijven zo bijzonder dat ze vanzelf boven kwamen drijven in de maatschappij. In onze tijd komen mensen bovendrijven als ze de wildgroei aan informatie en communicatie kunnen behappen. Algoritme is het nieuwe toverwoord, maar waar blijven de wiskundigen? Alexander Nix, de voormalig directeur van Cambridge Analytica, is een kunsthistoricus. Aleksandr Kogan, de man achter het Facebook-datalek, een psycholoog. Nate Silver, de statisticus die de Amerikaanse sport en politiek in de gaten houdt, een econoom. Hans Rosling, een medicus. Wiskundigen hebben nu eenmaal weinig met de wereld om ons heen. "Wat doet je man eigenlijk?" vragen ze soms aan mijn vrouw. "Hij is een wiskundige", zegt ze dan. Daarna wordt er begrijpend geknikt.

Een echte wiskundige verdiept zich alleen in de echte wereld als het echt moet. Toen Kurt Gödel in 1947 het Amerikaanse staatsburgerschap aanvroeg, moest hij slagen voor een inburgeringstest die werd afgenomen door een rechter. Ter voorbereiding van de test had Gödel twee weken gestudeerd op de staatsinrichting. Oskar Morgenstern was aanwezig als getuige en tekende de volgende woordenwisseling op:

*Judge:* Now Mr. Gödel, where do you come from?
*Gödel:* Where I come from? Austria.
*Judge:* What kind of government did you have in Austria?
*Gödel:* It was a republic but the constitution was such that it finally was changed into a dictatorship.
*Judge:* Oh! This is very bad. This could not happen in this country.
*Gödel:* Oh yes, I can prove it!
*Judge:* Oh God, let's not get into this.

En zo werd het bewijs van Gödel afgekapt door een rechter, net zoals het bewijs van Fermat ooit werd afgekapt door een kantlijn. Zullen we ooit weten wat Gödel in gedachten had? Gödels Oostenrijk kende een sterk verdeelde samenleving, met een progressieve hoofdstad en een conservatief platteland. Het klimaat werd uiteindelijk zo giftig dat de eenheid in het land alleen kon worden hersteld door een dictator. Dat was lang geleden en hopelijk gebeurt het niet meer, maar je weet het nooit. Het huidige klimaat in de Amerikaanse maatschappij is ook nogal verdeeld en giftig. Wie weet komt er binnenkort een begaafde politicus, die het verloren bewijs van Gödel in volle glorie weet te herstellen, net zoals het verloren bewijs van Fermat.

Het verdelen van een maatschappij in twee delen — hoeken en kabeljauwen, conservatieven en liberalen, katholieken en protestanten — komt aan bod onder het kopje *sociale balans* in de bijdrage van Vincent Buskens en Arnout van de Rijt over wiskunde en sociologie. De sociaal psycholoog Doc Cartwright en de wiskundige Frank Harary toonden aan dat het "vijand van mijn vijand is mijn vriend"-principe de maatschappij verdeelt in twee kampen. Het mechanisme achter een giftig sociaal klimaat is dus bekend. Nu nog een tegengif. Ach, waren er maar meer mensen die elke wijkbewoner kennen bij naam en toenaam.

**Robbert Fokkink**, hoofdredacteur
*Delft Institute of Applied Mathematics, TU Delft*

# Agenda

| Upcoming Events

1 januari – 31 december 2019

❑ **Spelen met wiskunde**

Doe-tentoonstelling. Maak zelf een draaikolk en bouw je eigen brug. Ontdek hoe geluid eruitziet en speel als een kunstenaar met perspectief. Wiskunde is écht overal.

***plaats*** *Rijksmuseum Boerhaave, Leiden*

***info*** *rijksmuseumboerhaave.nl*

## september 2019

11 september 2019

❑ **Networks Workshop on Random Graphs, Counting and Sampling**

Workshop georganiseerd door Viresh Patel and Leen Stougie.

***plaats*** *Amsterdam*

***info*** *cwi.nl*

13 september 2019

❑ **Finale Nederlandse Wiskunde Olympiade**

Nationale finale van de wiskundeolympiade voor middelbare scholieren.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *wiskundeolympiade.nl*

16–20 september 2019

❑ **Machine Learning in Heliophysics**

Workshop over hoe machine- en deep learning-technieken, informatietheorie en Bayesiaanse analyse worden gebruikt in allerlei problemen in het ruimteonderzoek.

***plaats*** *Koninklijk Instituut vd Tropen, Amsterdam*

***info*** *event.cwi.nl/ml-helio-2019*

17–18 september 2019

❑ **The Third CWI-Inria Workshop**

Workshop ter ere van de samenwerking tussen het CWI en het Inria instituut.

***plaats*** *Amsterdam*

***info*** *project.inria.fr/inriacwi/workshop-2019/*

20 september 2019

❑ **Wiskundetoernooi**

Internationaal wiskundetoernooi waaraan leerlingen uit 4/5 vwo in teams kunnen deelnemen. De wedstrijd vind tegelijkertijd plaats in Nijmegen, Bonn en Leuven.

***plaats*** *Radboud Universiteit, Nijmegen*

***info*** *ru.nl/wiskundetoernooi*

23–27 september 2019

❑ **Statistics to the Stars and Back: Convincing Inference Across the Interdisciplinary Gap**

Workshop georganiseerd door Elena Sellentin en Harry van Zanten over statistiek in theoretische natuurkunde.

***plaats*** *Lorentz Center@Snellius, Leiden*

***info*** *lorentzcenter.nl*

28 september 2019

❑ **Junior Wiskunde Olympiade**

Jaarlijks landelijke wiskundewedstrijd voor leerlingen van havo en vwo, georganiseerd door de VU in samenwerking met de Nederlandse Wiskunde Olympiade en W4Kangoeroe.

***plaats*** *Vrije Universiteit Amsterdam*

***info*** *beta.vu.nl/nl/scholieren*

28 september 2019

❑ **Denken met Dingen, 25e symposium van de werkgroep geschiedenis van de NVvW**

Een symposium over hoe en waarom er in het verleden in de wiskundeles gebruik werd gemaakt van instrumenten en modellen.

***plaats*** *Nationaal Onderwijsmuseum, Dordrecht*

***info*** *nvvw.nl/denken-met-dingen*

30 september – 4 oktober 2019

❑ **ENUMATH 2019**

The European Conference on Numerical Mathematics and Advanced Applications (ENUMATH). Conferentie over numerieke wiskunde.

***plaats*** *Hotel Zuiderduin, Egmond aan Zee*

***info*** *enumath2019.eu*

## oktober 2019

9–11 oktober 2019

❑ **44rd Woudschoten Conference 2019**

Een jaarlijkse conferentie van de Nederlandse en Vlaamse Numerieke Analyse-groepen, georganiseerd door de Werkgemeenschap Scientific Computing (WSC).

***plaats*** *Woudschoten, Zeist*

***info*** *wsc.project.cwi.nl/events*

16–18 oktober 2019

❑ **Workshop YEQT XIII: Data-Driven Analytics and Optimization for Stochastic Systems**

Workshop over het combineren van technieken uit theoretische stochastische modellen en optimalisering met moderne statistische technieken.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *eurandom.nl*

*Suggesties voor deze agenda zijn welkom.*
*Redacteur: Margriet Oomen*
*agenda@nieuwarchief.nl*

## november 2019

2 november 2019

❑ **Jaarvergadering NVvW**

Jaarlijkse vergadering en studiedag van de Nederlandse Vereniging van Wiskundeleraren.

***plaats*** *Ichthus College, Veenendaal*

***info*** *nvvw.nl/jaarvergadering*

8 november 2019

❑ **Prijsuitreiking Nederlandse Wiskunde Olympiade 2019**

De prijsuitreiking van de finale van de Nederlandse Wiskunde Olympiade.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *wiskundeolympiade.nl*

11–13 november 2019

❑ **Stochastics Meeting Lunteren 2019**

Bijeenkomst van stochastiek- en statistiek- groepen uit heel Nederland.

***plaats*** *Congrescentrum De Werelt, Lunteren*

***info*** *homepages.cwi.nl/~colette/lunteren/2019.html*

22 november 2019

❑ **Voorronde Wiskunde A-lympiade 2019**

Voorronde van de jaarlijkse competitie voor leerlingen uit de bovenbouw havo/vwo met wiskunde A in hun pakket.

***plaats*** *middelbare scholen in Nederland*

***info*** *wiskundeinteams.sites.uu.nl*

22 november 2019

❑ **Wiskunde B-dag**

De Wiskunde B-dag is een 1-daagse opdracht voor op school voor teams uit 5 havo en 5/6 vwo met wiskunde B in hun profiel.

***plaats*** *middelbare scholen in Nederland*

***info*** *wiskundeinteams.sites.uu.nl*

25–29 november 2019

❑ **Enabling Data-Driven Methods for Inverse Problems in 3D Imaging: Uniting Data, People and Algorithms**

Workshop van Joost Batenburg, Christoph Brune, Maureen van Eijnatten en Tristan van Leeuwen.

***plaats*** *Lorentz Center@Oort, Leiden*

***info*** *lorentzcenter.nl*

## December 2019

9–13 december 2019

❑ **Workshop Heavy Tails**

Workshop voor statistici en stochastici alsmede natuurkundige, besliskundigen, financiële wiskundigen en informatici, om nieuwste technieken en theorie over 'heavy tails' uit te wisselen.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *eurandom.nl*

## Januari 2020

9 januari 2020

❑ **Panama-conferentie 2020**

Conferentie met thema 'Rekenen-wiskunde van … tot …' biedt expliciet de ruimte om verschillende visies, doelen en didactiek van het rekenwiskundeonderwijs te delen.

***plaats*** *Woudschoten, Zeist*

***info*** *panamaconferentie.sites.uu.nl/*

20–30 januari 2020

❑ **Eerste ronde Nederlandse Wiskunde Olympiade**

Wiskundewedstrijd voor middelbare scholieren in Nederland.

***plaats*** *middelbare scholen in Nederland*

***info*** *wiskundeolympiade.nl*

20–22 januari 2020

❑ **The NETWORKS Match Making Event 2020: Social Sciences Meets Mathematics**

Doel van de workshop is het bij elkaar brengen van sociologen, economen, wiskundigen en informatici die onderzoek doen naar netwerken.

***plaats*** *Doorn*

***info*** *thenetworkcenter.nl*

31 januari – 1 februari 2020

❑ **Nationale Wiskunde Dagen**

Landelijke tweedaagse bijeenkomst voor wiskundeleraren.

***plaats*** *De Leeuwenhorst, Noordwijkerhout*

***info*** *uu.nl//nationale-wiskunde-dagen*

## Maart 2020

13 maart 2020

❑ **Tweede ronde Nederlandse Wiskunde Olympiade**

Wiskundewedstrijd voor middelbare scholieren.

***plaats*** *twaalf universiteiten in Nederland*

***info*** *wiskundeolympiade.nl*

13 maart 2020

❑ **Nationale Rekencoördinator Dag**

Een dag voor rekencoördinatoren en leerkrachten, waarop het rekenonderwijs van de basisschool centraal staat.

***plaats*** *Domstad, Utrecht*

***info*** *nrcd.sites.uu.nl/*

25 maart 2020

❑ **Grote Rekendag**

Dag voor basisschoolleerlingen die geheel in het teken van rekenen staat.

***plaats*** *basisscholen in Nederland en Belgie*

***info*** *uu.nl*

## April 2020

6–9 april 2020

❑ **Graph Limits**

Workshop mede georganiseerd door Remco van der Hofstad, Frank den Hollander en Viresh Patel.

***plaats*** *Technische Universiteit Eindhoven*

***info*** *eurandom.nl*

14–15 april 2020

❑ **Nederlands Mathematisch Congres**

Jaarlijkse bijeenkomst van wiskundig Nederland georganiseerd door het KWG, met dit jaar onder andere de uitreiking van de Brouwerprijs.

***plaats*** *Van der Valk Hotel Utrecht*

***info*** *wiskgenoot.nl*

# Nieuws

| News

**Erelid Cor Baayen overleden**

Op 22 mei jongstleden is professor Cor Baayen overleden. Baayen was lid van het bestuur van het Wiskundig Genootschap van 1965 tot 1981, waarvan drie jaar als voorzitter. Vanwege zijn langjarige inzet en grote verdiensten voor zowel het genootschap als de wiskunde in het algemeen werd hij in 2004 benoemd tot erelid.

Pieter Cornelis Baayen werd geboren op 10 maart 1934 in Klaten op Java. Hij promoveerde in 1964 bij Jan de Groot aan de Universiteit van Amsterdam, waarna hij in 1965 werd benoemd tot hoogleraar aan zijn alma mater, de Vrije Universiteit. Tot aan zijn pensionering zou Baayen verbonden blijven aan de VU.

Ook was Baayen van 1959 tot 1995 verbonden aan het Mathematisch Centrum, later het CWI. Vanaf 1980 was hij er wetenschappelijk directeur. Hij slaagde erin om in een tijd van grote bezuinigingen het instituut te laten groeien door de missie te verruimen naar de informatica en door nieuwe vakgebieden zoals discrete wiskunde, computer algebra, cryptografie en beeldanalyse te stimuleren.

In Europees perspectief zal Baayen vooral herinnerd worden als mede-oprichter en eerste directeur van ERCIM, het European Research Consortium for Informatics and Mathematics. Bij zijn afscheid als directeur in 1994 werd hij benoemd tot President d'Honneur. Ook draagt de jaarlijks uitgereikte Young Researcher Award voor de meest veelbelovende onderzoeker in de toegepaste wiskunde of informatica zijn naam. *wiskgenoot.nl*

Foto: cwi.nl

*Deze rubriek is een kroniek van wiskundige activiteiten in Nederland. Toekomstige activiteiten worden aangekondigd en van voorbije activiteiten wordt verslag gedaan. Wilt u uw aankondiging of verslag in deze rubriek geplaatst zien? Stuur ons dan uw bijdrage, zo mogelijk met illustratie. De redactie behoudt zich het recht voor berichten te weigeren of in te korten.*

*Redacteur: Margriet Oomen*

*nieuws@nieuwarchief.nl*

**Vijf medailles op de Internationale Wiskunde Olympiade**

Op de Internationale Wiskunde Olympiade in Bath heeft Nederland vijf medailles gehaald. Matthijs van der Poel (18) uit IJsselstein behaalde voor het derde jaar op rij een zilveren medaille. Jesse Fitié, Richard Wols, Jovan Gerbscheid en Jippe Hoogeveen sleepten alle vier een bronzen medaille in de wacht. Het zesde lid van het Nederlandse team, Szabi Buzogány, verdiende een eervolle vermelding, omdat hij één van de zes opgaven van de wedstrijd foutloos oploste. Met deze score eindigde het Nederlandse team op de 37ste plek in het landenklassement. In totaal deden er 112 landen mee. Het Nederlandse team werd begeleid door Birgit van Dalen (Universiteit Leiden en ISW Hoogeland Naaldwijk), Quintijn Puite (Technische Universiteit Eindhoven en Alberdingk Thijm College Hilversum) en Jeroen Huijben (Universiteit van Amsterdam). *wiskundeolympiade.nl*

Foto: wiskundeolympiade.nl

**Wiskundige bewijst het dertig jaar oude 'sensitivity conjecture'**

Sensitiviteit is een begrip uit de theoretische informatica dat gebruikt wordt om Boolean functies te karakteriseren. Een Boolean functie is een functie van $\{0,1\}^k \to \{0,1\}$, waarbij $k \in \mathbb{N}$. De sensitiviteit beschrijft wat de invloed van $1$ bit is op de uitkomst van een Boolean functie. Op deze manier is de sensitiviteit een maat voor de complexiteit van een Boolean functie. Sensitiviteit is niet de enige maat voor de complexiteit van Boolean functies, een andere maat is bijvoorbeeld de query complexity. De query complexity van een Boolean functie geeft aan hoeveel input-bits je moet weten om de uitkomst van de functie te berekenen. Van alle maten voor complexiteit behalve de sensitiviteit is bekend dat ze aan elkaar gerelateerd zijn. Als je bijvoorbeeld *query complexity* weet, kun je inschatten wat de andere complexiteitsmaten zijn. De sensitivity conjecture stelt dat de sensitiviteit ook een inschatting zou moeten geven voor de andere complexiteitsmaten. Echter het lukte informatici dertig jaar lang niet om dit te bewijzen.

In juli publiceerde de wiskundige Hao Huang, assistant professor aan Emory University, een artikel waarin hij het sensitiviteitsvermoeden bewijst. Tot grote verbijstering van zijn collega's beslaat het artikel slechts vijf pagina's en is het bewijs bijzonder elegant. Voor de meeste combinatoristen is Huangs bewijs zelfs in enkele minuten te begrijpen. Voor zijn bewijs koppelde Huang Boolean functies aan $n$-dimensionale eenheidskubussen. Huang kleurde de hoekpunten aan de hand van de uitkomst van de Boolean functie. Vervolgens beschreef Huang $n$-dimensionale eenheidskubussen als een netwerk en stelde voor dit netwerk een zogeheten *adjencency matrix* op. Door het 200 jaar oude Cauchy interlacing theorema te gebruiken, kon Huang het sensitiviteitsvermoeden bewijzen met behulp van de eigenwaarden van de matrix. *quantamagazine.org*

**Portret Alan Turing op het nieuwe 50 pondbiljet**

The Bank of England heeft de wiskundige Alan Turing (1912–1954) gekozen als het nieuwe gezicht voor het 50 pondbiljet. Turings werk vormde de basis voor de huidige theoretische informatica. Tijdens de Tweede Wereldoorlog ontwierp Turing codekrakende machines die een belangrijke rol speelden in het verloop van de oorlog. Voor zijn werk tijdens de oorlog kreeg Turing in 1946 de onderscheiding 'Officer of the Order of the British Empire' en in 1951 werd Turing benoemd tot Fellow of the Royal Society (FRS). Een vervolging wegens homoseksualiteit overschaduwde echter de erkenning van Turings werk. Nadat Turing zich verplicht chemisch moest laten castreren, overleed hij in 1954. Na zijn dood kreeg Turing vele postume onderscheidingen.

Het nieuwe 50 pondbiljet toont op de ene zijde het portret van Turing en op de andere zijde het gezicht van Queen Elizabeth. Daarnaast zal het biljet de volgende quote van Turing tonen uit een interview in 1949 in *The Times* over zijn zelfgebouwde computers: "This is only a foretaste of what is to come, and only the shadow of what is going to be." *bankofengland.co.uk*


Conceptontwerp van het 50 pondbiljet met het portret van Alan Turing

**ELA-presentatie award voor Bernard Zweers**

Bernard Zweers, promovendus van de CWI Stochastics-groep, heeft de prijs voor de beste presentatie gewonnen op de 24ste workshop voor PhD-studenten van de European Logistics Association (ELA). Op de workshop die plaatsvond in Edinburgh gaf Zweers de presentatie 'Optimizing inland container terminal operations'. De presentatie ging over algoritmes die ontworpen zijn om containers zo efficiënt mogelijk te verplaatsen in een containerterminal. Voor dit onderzoek werkte Zweers samen met zijn begeleiders Sandjai Bhulai en Rob van der Mei. *cwi.nl*

**Piet Hemker overleden**

Op 27 mei is prof. dr. Piet Hemker op 77 jarige leeftijd overleden. Piet Hemker trad in 1970 in dienst bij het CWI, als jong wetenschappelijk medewerker bij de rekenafdeling. Vanuit het CWI heeft hij de Nederlandse numerieke wiskunde gedurende vele jaren mede vormgegeven en een hoog internationaal aanzien bezorgd. Van 1989 tot aan zijn pensionering in 2006 was hij tevens deeltijdhoogleraar in de industriële wiskunde aan de Universiteit van Amsterdam. Piet Hemker is bij zijn pensionering benoemd tot Ridder in de Orde van de Nederlandse Leeuw. Hij kreeg deze onderscheiding onder andere vanwege zijn verdiensten als bruggenbouwer tussen de wiskunde en de industrie en tussen wiskundigen in West- en Oost-Europa. *wiskgenoot.nl*

Foto: cwi.nl

**Statisticus Aad van der Vaart krijgt koninklijke onderscheiding**

Aad van der Vaart, statisticus aan de Universiteit Leiden, is benoemd tot Ridder in de Orde van de Nederlandse Leeuw. Tijdens een symposium ter gelegenheid van zijn zestigste verjaardag kreeg Aad van der Vaart zijn lintje opgespeld door de Leidse burgemeester Henri Lenferink. Van der Vaart krijgt het lintje niet alleen voor zijn baanbrekende werk in de statistiek, maar ook voor de vele bestuurlijke functies die hij vervulde. *universiteitleiden.nl*

**Leids statistiekcentrum LUXs geopend**

Op 19 juni is het Leids statistiekcentrum Leiden University Center for Statistical Science (LUXs) geopend. De opening vond plaats tijdens een internationale statistiekconferentie ter ere van de zestigste verjaardag van Aad van der Vaart. Ter gelegenheid van de opening van het LUXs gaf de internationaal bekende 'rockster' van de statistiek, Bradley Efron, een lezing. LUXs is het eerste universiteitsbrede statistiekcentrum van Nederland. Het doel is om statistici, die over verschillende vakgroepen binnen de universiteit verspreid zitten, bij elkaar te brengen en samenwerking tussen verschillende statistici te bevorderen. *universiteitleiden.nl*

**Humies Silver Award voor Peter Bosman en collega's**

Op de Genetic and Evolutionary Computation Conference (GECCO 2019) hebben Peter Bosman (CWI) en Hoang Luong (CWI) samen met collega's van het Amsterdam UMC de Humies Silver Award in ontvangst mogen nemen. Ze kregen de prijs voor het door hen ontworpen algoritme dat een automatisch behandelplan met brachytherapy genereert voor kankerpatiënten. De Humies Awards zijn prijzen voor toepasbaar wetenschappelijk onderzoek dat problemen oplost met behulp van technieken uit de *evolutionary computation*. Daarnaast won Stef Maree, promovendus van Peter Bosman, de Competition on Niching Methods for Multimodal Optimization. *cwi.nl*

**Teun Koetsier Honorary Member IFToMM en winnaar ASME Award**

Op het 15de Wereldcongres van IFToMM (International Federation for the Promotion of Mechanism and Machine Science) in juli van dit jaar in Krakau is Teun Koetsier (VU) benoemd tot 'Honorary Member' van IFToMM voor zijn werk binnen IFToMM en zijn onderzoek op het gebied van de geschiedenis van machines. Tevens zal hij in november van dit jaar in Salt Lake City de Engineer-Historian Award of the American Society of Mechanical Engineers (ASME) ontvangen. Hij krijgt de award voor zijn boek getiteld *The Ascent of GIM, the Global Intelligent Machine – A History of Production and Information Machines*, Springer, 2019. *vu.nl*

## Koninklijk Wiskundig Genootschap

❑ **NMC 2020**

Het Nederlands Mathematisch Congres 2020 zal plaatsvinden op 14 en 15 april in het Van der Valk Hotel Utrecht (Winthontlaan 4, Utrecht). Het hotel is makkelijk bereikbaar met auto of openbaar vervoer.

❑ **Brouwermedaille 2020**

Op het NMC 2020 zal de Brouwermedaille worden uitgereikt aan professor David John Aldous (Berkeley). De Brouwermedaille is een driejaarlijkse prijs van het KWG. Als gebruikelijk heeft het bestuur het onderwerp vastgesteld waarbinnen de laureaat moet werken. Het onderwerp is dit keer waarschijnlijkheidsrekening. Vervolgens is een selectiecommissie voor de laureaat samengesteld bestaande uit Remco van der Hofstad (TU/e), Frank den Hollander (UL, voorzitter), Frank Redig (TUD) en Bert Zwart (CWI). Deze commissie is na ampele overweging unaniem tot de conclusie gekomen dat David John Aldous de beste kandidaat voor de Brouwermedaille 2020 is. Het bestuur heeft de aanbeveling overgenomen. Inmiddels heeft professor Aldous toegezegd naar het congres te komen, de Brouwerlezing te geven en de medaille in ontvangst te nemen.

❑ **Milieuvriendelijke folie voor NAW en Pythagoras**

De folie die tot nu toe gebruikt werd om NAW en Pythagoras te sealen is gemaakt van aardolie. Hoewel deze folie 100% recyclebaar is, kan het beter. Vandaar dat onze drukker nu overschakelt op biobased sealfolie. Ook deze folie is 100% recyclebaar maar is gemaakt uit suikerriet en heeft een biobased keurmerk van certificeringsinstelling TÜV Austria.

Bernard P. Veldkamp
*Faculty of Behavioural, Management and Social Sciences, Universiteit Twente, Enschede*
*en Research Center voor Examinering en Certificering, Vaassen*
*b.p.veldkamp@utwente.nl*

# Het wiskundige fundament van toetsen en examens

**Toetsen en examens hebben veel invloed op schoolcarrières en kansen op een baan, maar hoe worden cijfers toegekend? In dit artikel gaat Bernard Veldkamp in op het wiskundige fundament van onderwijskundig meten. De twee meestgebruikte modellen uit de psychometrie worden geïntroduceerd en hun belangrijkste begrippen, formules en toepassingen worden beschreven. Klassieke testtheorie bestaat het langst en is het meest bekend vanwege de scoringsregel, die het cijfer bepaalt door de punten van de individuele opgaves op te tellen. Item-responstheorie is minder bekend, maar wordt met name gebruikt bij grootschalige toetsen en examens zoals eindexamens en inburgeringsexamens. De voor- en nadelen van beide modellen worden benoemd. Tot slot wordt kort ingegaan op actuele uitdagingen voor het vakgebied van de psychometrie.**

De centrale eindexamens zijn weer achter de rug. Tienduizenden scholieren zwoegden op examens die van grote invloed zijn op het vervolg van hun carrière. De examens vormen niet alleen de afronding van het voortgezet onderwijs, maar het diploma dat ermee wordt verkregen geeft toegang tot vervolgonderwijs, het wordt gebruikt om studenten te selecteren voor specifieke programma's en er wordt bij sollicitaties vaak gebruik van gemaakt om een kandidaat te beoordelen op geschiktheid. De behaalde cijfers worden bovendien niet alleen gebruikt om de prestaties van leerlingen te beoordelen. De gemiddelde scores die leerlingen behalen op toetsen en examens worden in de praktijk gebruikt om docenten en scholen te beoordelen. Schoolleiders gebruiken gemiddelde cijfers bij jaargesprekken en de Inspectie van het Onderwijs onderzoekt of de cijfers die behaald worden op het centraal schriftelijk examen niet te veel afwijken van de cijfers op de schoolexamens. Ten slotte gebruiken bestuurders en politici de cijfers om een indruk te krijgen van het algemene onderwijsniveau en van de prestaties van Nederlandse leerlingen ten opzichte van leerlingen uit andere landen. Vanwege dit grote civiele effect is het belangrijk dat de cijfers die leerlingen behalen op examens een betrouwbare weergave zijn van werkelijke vaardigheden van de leerlingen.

Bij het geven van cijfers voor toetsen of examens spelen verschillende aspecten een rol. Allereerst moet beoordeeld worden of en in hoeverre een leerling de individuele vragen op de juiste manier heeft beantwoord. Bij multiplechoicevragen kan deze beoordeling geautomatiseerd plaatsvinden. Bij open vragen gebeurt deze beoordeling door een of meer beoordelaars aan de hand van een beoordelingsvoorschrift. Vervolgens worden de scores op de individuele vragen, al dan niet gewogen, bij elkaar opgeteld. De resulterende score wordt vergeleken met een standaard of cesuur, die aangeeft hoeveel punten behaald moeten worden voor een voldoende beoordeling, en op basis van deze vergelijking wordt een cijfer toegekend. Zowel bij het wegen van de verschillende vragen, bij het optellen van de (gewogen) scores, als bij het vergelijken van de behaalde score met de cesuur, speelt wiskunde een belangrijke rol.

De tak van wiskunde die gebruikt wordt bij toetsen en examens staat bekend als *psychometrie*. Formeel gesproken is psychometrie de wetenschap die zich bezighoudt met de technieken van het meten van psychologische fenomenen zoals kennis, vaardigheden, attituden, eigenschappen en persoonskenmerken. Voor het meten van kennis en vaardigheden wordt gebruikgemaakt van toetsen en examens, terwijl attitude, eigenschappen en persoonskenmerken gemeten worden met psychologische testen.

In dit artikel wordt een beschrijving gegeven van de geschiedenis van de psychometrie. Vervolgens wordt ingegaan op de twee bekendste en meest toegepaste families van modellen, te weten de klassie-

ke testtheorie en de item-responstheorie. Het artikel eindigt met het beschrijven van de meest recente ontwikkelingen binnen de psychometrie.

**Klassieke testtheorie**

Het meestgebruikte model om cijfers toe te kennen aan toets- en examenresultaten is gebaseerd op de som-correctscore, ook wel totaalscore genoemd. Als dit model wordt toegepast, dan worden bij het nakijken punten gegeven voor elke individuele vraag. Vragen in een toets worden ook wel items genoemd. Vervolgens worden de punten van alle items opgeteld en omgezet in een cijfer. Dit model wordt toegepast bij de meeste toetsen en proefwerken in het reguliere onderwijs. Het model volgt uit de klassieke testtheorie.

Klassieke testtheorie is gebouwd op drie aannames:

$$X = T + E,$$
$$E(E) = 0,$$
$$\rho_{TE} = 0.$$

De eerste aanname geeft aan dat een geobserveerde score $X$ opgebouwd is uit een werkelijke score $T$ (true score) en een ruiscomponent $E$ (error). De geobserveerde score is de som-correctscore, zoals die hierboven is geïntroduceerd. De werkelijke score is het echte niveau van de kandidaat. Dit kun je vergelijken met de gemiddelde score die een kandidaat zou halen als hij de toets een groot aantal keer zou maken, waarbij zijn hersens gespoeld zouden worden na elke poging om te voorkomen dat hij de vragen onthoudt. Hierbij wordt ervan uitgegaan dat alle items even goed bijdragen aan de te meten vaardigheid. Deze werkelijke score is een latente variabele die niet direct geobserveerd kan worden, maar die wordt afgeleid uit de antwoorden. De ruiscomponent is een meetfout die veroorzaakt kan worden door allerlei toevallige omstandigheden en die een correcte meting verstoren. Bij toetsen en examens zijn we er in geïnteresseerd om de werkelijke score $T$ zo nauwkeurig mogelijk te meten.

De tweede aanname geeft aan dat de verwachting van de ruiscomponent gelijk is aan nul. Bij toetsen kunnen er verschillende oorzaken zijn voor ruis. Ruis kan veroorzaakt worden door eigenschappen van de persoon, van de toets, of van de situatie. Vermoeidheid kan ervoor zorgen dat de prestaties van de persoon verminderen of iemand kan een moeilijk item per ongeluk correct beantwoorden. Slechte items kunnen zorgen voor ruis ten gevolge van het meetinstrument en een surveillant die verkouden is en veel moet niesen, is een omgevingsfactor die ruis kan veroorzaken. De consequentie van de tweede aanname is dat de verwachting van de geobserveerde score gelijk is aan de verwachting van de werkelijke score. Dit geeft aan dat, op het moment dat de klassieke testtheorie geldt, we verwachten dat de geobserveerde score een goede weergave is van de werkelijke score.

De derde aanname geeft aan dat er geen correlatie is tussen de ruiscomponent en de werkelijke score van een persoon. Dat wil zeggen dat het niet uitmaakt of een kandidaat juist een hoge vaardigheid of een lage vaardigheid heeft, er is geen relatie met de gemaakte meetfout. Deze drie aannames kunnen gebruikt worden om een aantal afleidingen te maken.

*Betrouwbaarheid in klassieke testtheorie*

Allereerst kunnen we iets zeggen over de betrouwbaarheid $\rho^2_{XT}$ van een examen. De betrouwbaarheid is gedefinieerd als dat deel van de variantie van de geobserveerde score dat verklaard wordt door de werkelijke score. De betrouwbaarheid kan waardes aannemen van $0$ tot en met $1$, mits de variantie in de geobserveerde score groter is dan $0$. Een betrouwbaarheid van $\rho^2_{XT} = 0$ geeft aan dat er geen relatie is tussen de geobserveerde score en de werkelijke score waar we in geïnteresseerd zijn, oftewel, de score op de toets is volledig gebaseerd op de ruiscomponent. Een betrouwbaarheid van $\rho^2_{XT} = 1$ daarentegen, geeft aan dat de geobserveerde score volledig wordt verklaard door de werkelijke vaardigheid van de persoon, oftewel, de toets heeft geen meetfout. Deze betrouwbaarheid kan geformuleerd worden als:

$$\rho^2_{XT} = \frac{\sigma^2_T}{\sigma^2_X}.$$

Omdat $X = T + E$ en $\rho_{TE} = 0$ kunnen we afleiden dat $\sigma^2_X = \sigma^2_T + \sigma^2_E$. Daarom geldt voor de betrouwbaarheid:

$$\rho^2_{XT} = \frac{\sigma^2_T}{\sigma^2_T + \sigma^2_E}.$$

Deze laatste formule laat zien hoe de ruiscomponent van invloed is op de betrouwbaarheid. Hoe kleiner $\sigma^2_E$ des te hoger de betrouwbaarheid. Bij het construeren van toetsen en examens wordt er daarom naar gestreefd om de betrouwbaarheid van toetsen en examens te maximaliseren en de ruiscomponent te minimaliseren. In de praktijk is het alleen niet mogelijk om de betrouwbaarheid $\rho^2_{XT}$ uit te rekenen omdat we de variantie van de werkelijke score $\sigma^2_T$ niet kunnen meten. Daarom zijn er verschillende manieren bedacht om de betrouwbaarheid te kunnen schatten. De drie bekendste methodes zijn test-hertestbetrouwbaarheid, de split-halfmethode en Cronbachs alfa. Bij test-hertestbetrouwbaarheid wordt een toets twee keer afgenomen bij dezelfde populatie en wordt de correlatie tussen de beide afnames genomen als schatting van de betrouwbaarheid. Een mogelijke verstoring hierbij is wel dat kandidaten vragen kunnen onthouden, waardoor deze schatting niet optimaal is. Bij de split-halfmethode wordt de toets verdeeld in twee helften en wordt de correlatie tussen de beiden helften uitgerekend als schatting voor de betrouwbaarheid. Deze methode heeft als nadeel dat de betrouwbaarheid effectief slechts berekend wordt op basis van een halve toetslengte. De derde manier is door gebruik te maken van Cronbachs alfa:

$$\alpha = \frac{n}{n-1}\left(1 - \frac{\sum_i \sigma^2_i}{\sigma^2_X}\right),$$

waarbij $n$ het aantal items is, $\sigma^2_X$ de variantie van de totaalscore is en $\sigma^2_i$ de variantie van de score van item $i$ is. Cronbachs alfa is nog steeds een ondergrens van de betrouwbaarheid, maar het grote voordeel is dat Cronbachs alfa berekend kan worden op basis van bekende varianties.

*De meetfout*

Naast betrouwbaarheid wordt er vaak gesproken over de meetfout van een toets. Met deze meetfout wordt de wortel van de ruisvariantie bedoeld. Met behulp van de formules voor de betrouwbaarheid, kan de meetfout worden berekend als:

$$\sigma_E = \sigma_X \sqrt{(1 - \rho_{XT^2})}.$$

Hoe kleiner deze meetfout, hoe nauwkeuriger de toets of het examen meet. Een opvallend kenmerk van klassieke testtheorie is dat de meetfout onafhankelijk is van het werkelijke niveau van de kandidaat. Uit de drie aannames van klassieke testtheorie volgt dat de meetfout een eigenschap is

van de toets. Deze meetfout is constant en voor de berekening maakt het niet uit hoeveel kandidaten de toets maken.

*De invloed van toetslengte*

Wat wel uitmaakt is het aantal items waaruit de toets bestaat. Eggen en Sanders [2] laten zien hoe Spearman en Brown afgeleid hebben dat je op basis van de formule voor de betrouwbaarheid kunt aantonen dat de betrouwbaarheid van een nieuwe toets die $k$ keer zo lang is als de oude toets kunt berekenen met de formule

$$\rho^2_{XT_{\text{nieuw}}} = \frac{k\rho^2_{XT_{\text{oud}}}}{1+(k-1)\,\rho^2_{XT_{\text{oud}}}}.$$

De voorwaarde hierbij is dat de items in de nieuwe toets allemaal van dezelfde kwaliteit zijn als de items uit de oude toets. Om de betrouwbaarheid van de toets te verhogen en de meetfout te verkleinen wordt er in de praktijk daarom vaak voor gekozen om extra items af te nemen. De Cotan (COmmissie Test Aangelegenheden Nederland) heeft een aantal richtlijnen opgesteld voor de betrouwbaarheid. Zo geldt dat de betrouwbaarheid hoger moet zijn dan $0{,}90$ voor examens en voor andere toetsen die gebruikt worden voor belangrijke beslissingen. Voor toetsen die gebruikt worden bij minder belangrijke beslissingen, zoals voortgangstoetsen, geldt dat de betrouwbaarheid hoger moet zijn dan $0{,}80$. Als er slechts op groepsniveau wat gedaan wordt met de uitkomsten van de toetsen geldt dat de betrouwbaarheid $0{,}70$ moet zijn. Door de lengte van de toets te variëren kan geprobeerd worden om aan deze richtlijnen te voldoen.

Door de toetslengte te variëren, kan de betrouwbaarheid beïnvloed worden. In de vorige paragraaf is al opgemerkt dat een van de uitgangspunten daarbij is dat de kwaliteit van de oude en nieuwe items vergelijkbaar is. In de praktijk blijkt dit zelden het geval te zijn. De strategie die dan meestal gehanteerd wordt, is dat kwalitatief mindere items uit de toets verwijderd worden om op die manier de betrouwbaarheid te verhogen. Daarbij wordt gekeken naar de moeilijkheid van de items en naar het onderscheidend vermogen. De moeilijkheid van een item wordt berekend met het percentage correcte antwoorden op een vraag, ook wel $p$-waarde genoemd. Voor het onderscheidend vermogen van een item wordt als maat vaak de $r_{\text{it}}$-waarde gebruikt, dat is de correlatie tussen de score op het item en de toets in zijn geheel. Voor een hoge betrouwbaarheid is het van belang dat er voldoende spreiding zit in de moeilijkheid van de items en dat de toets bestaat uit items met een hoog onderscheidend vermogen ($r_{\text{it}} \geq 0{,}40$).

*Validiteit*

De betrouwbaarheid zegt iets over de meetkwaliteit van de toets of het examen. Bij een hoge betrouwbaarheid, meet de toets consistent. Maar meet de toets ook wat hij moet meten? Kun je testscores gebruiken om conclusies te trekken over de vaardigheid waarvoor de toets ontworpen en afgenomen is? Om die vraag te beantwoorden moet gekeken worden naar de validiteit. Bij validiteit spelen verschillende aspecten een rol. Messick [8] beschrijft dat traditioneel gezien er vooral gekeken wordt naar inhoudsvaliditeit, criteriumvaliditeit en begripsvaliditeit. Inhoudsvaliditeit zegt iets over de formulering van de items: is de inhoud duidelijk verankerd in de leerdoelen en is er een evenwichtige verdeling van de vragen of de leerdoelen? Criteriumvaliditeit zegt iets over hoe goed de toetsscore voorspellend is voor een extern criterium, zoals toekomstige prestaties of een tweede onafhankelijke meting met een ander instrument. Begripsvaliditeit, ten slotte, geeft aan hoe goed de verschillende items een representatie zijn van het onderliggende construct dat je eigenlijk wilt meten met de toets. De traditionele manier van validiteitsonderzoek vergeet alleen mee te nemen hoe de toetsscore in de praktijk gebruikt wordt, aldus Messick [8]. In recenter onderzoek naar validiteit [13], wordt validiteit dan ook veel meer gekoppeld aan geschiktheid voor een specifiek doel. Door de validiteit van een toets of examen te onderzoeken, kunnen ten slotte uitspraken gedaan worden over hoe de toetsscore gebruikt kan worden om uitspraken te doen over de kandidaten.

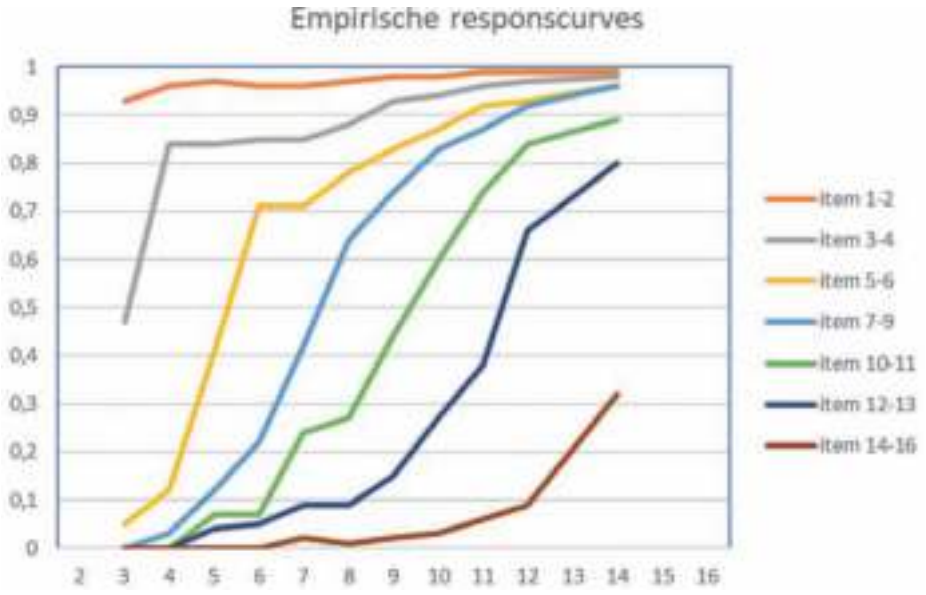
*Beperkingen van de klassieke testtheorie*

De klassieke testtheorie, zoals die hierboven kort is beschreven, vormt het fundament onder het gebruik van een totaalscore of som-correctscore bij het beoordelen van de resultaten van leerlingen bij toetsen en examens. Dit is alleen niet het hele verhaal. Bij de meeste belangrijke toetsen en examens, zoals de eindexamens voortgezet onderwijs, bij de inburgeringsexamens of bij de eindtoetsen basisonderwijs wordt geen gebruikgemaakt van deze methode. De reden hiervoor is dat de klassieke testtheorie een aantal grote nadelen [5] heeft. Alle indices en scores die ermee berekend worden zijn gekoppeld aan de toets en de populatie waarmee ze zijn berekend. De betrouwbaarheid van een toets wordt bijvoorbeeld berekend voor een specifieke populatie. Op het moment dat de toets bij een andere populatie, zoals een andere klas, een andere school, of leerlingen uit een ander leerjaar wordt afgenomen, moet de betrouwbaarheid opnieuw worden berekend. Een tweede voorbeeld betreft de score op de toets. Als twee leerlingen verschillende toetsen maken, dan kunnen hun scores niet onderling worden vergeleken. Om met dit soort bezwaren om te kunnen gaan is er veel onderzoek gedaan naar mogelijkheden om de indices en score te kunnen generaliseren. Door te werken met parallelle toetsen [2], kunnen de resultaten van examens en herexamens, bijvoorbeeld, worden vergeleken. Ondanks al het onderzoek naar deze mogelijkheden, bleef er veel kritiek op de klassieke testtheorie. Daarom is item-responstheorie ontwikkeld.

**Item-responstheorie**

Item-responstheorie ontstond gelijktijdig in de Verenigde Staten en in Europa in een poging om tot betere modellen te komen om de antwoorden van de kandidaten aan eigenschappen van toetsen en van individuele items te koppelen. Daarvoor zijn een aantal modellen ontwikkeld, die de kans dat een kandidaat een item correct beantwoordt modelleren als functie van een latente vaardigheid en een of meer eigenschappen van de items. Die vaardigheid wordt latent verondersteld omdat hij niet direct te observeren is, maar geschat moet worden uit de antwoorden die de kandidaat geeft.

Lord [6] omschreef, voor het eerst, het concept van een item-karakteristieke curve, een grafiek die de relatie tussen de vaardigheid van een kandidaat en de kans op een correct antwoord weergeeft. In Europa werkte de Deense wiskundige Georg Rasch onafhankelijk aan hetzelfde idee. Rasch (1960) bestudeerde grafieken waarin hij de kandidaten opdeelde in categorieën op basis van hun totaalscore. Hij ordende de vragen op basis van hun moeilijkheid en keek wat voor elk van de categorieën de kans was op een correct antwoord. Een




**Figuur 1** Empirische item-responscurves Rasch.

voorbeeld van een dergelijke grafiek wordt gegeven in Figuur 1.

In Figuur 1 zijn items gerangschikt van makkelijk naar moeilijk en ingedeeld in zeven moeilijkheidscategorieën. Voor deze toets van zestien items hebben de kandidaten totaalscores gehaald die variëren van $2$ tot $14$. Het is duidelijk te zien dat Rasch bij het ontwikkelen van zijn modellen in eerste instantie nog werkte binnen het raamwerk van de klassieke testtheorie, omdat hij nog werkte met totaalscores. De grafiek laat zien wat de kans was dat leerlingen met een bepaalde totaalscore items binnen een bepaalde groep correct beantwoorden. Zo kan uit de grafiek afgelezen worden dat leerlingen met een totaalscore van $8$ een kans hadden van $0.09$ om bijvoorbeeld item 12 correct te beantwoorden.

*Verschillende item-responstheoriemodellen*

*Rasch-model.* Bij het bestuderen van deze grafieken, maakte Rasch onderscheid tussen de vaardigheden van kandidaten en de moeilijkheden van items, als twee onafhankelijke grootheden die van invloed waren op de kans dat een kandidaat een item correct beantwoordde. Met betrekking tot de curves die ontstonden, constateerde Rasch drie dingen:

1. De curves zijn niet-lineair.
2. De curves snijden elkaar niet.
3. De curves voor de moeilijke items stijgen minder snel dan de curves voor de makkelijke items.

Op basis van deze constateringen ging Rasch op zoek naar een model dat uitgaande van het niveau van de kandidaat en de moeilijkheid van de items een accurate voorspelling doet van de kans dat de betreffende kandidaat de vraag correct beantwoordt. Hij deed daarbij twee aannames. De aanname van unidimensionaliteit houdt in dat de kans om het item correct te beantwoorden slechts beïnvloed wordt door de vaardigheid die de toets beoogt te meten. In de praktijk zullen meestal ook andere vaardigheden, persoonlijkheid en omstandigheden waaronder de toets wordt afgenomen van invloed zijn op deze kans, maar die worden in het model niet meegenomen. De tweede aanname van lokale onafhankelijkheid houdt in dat de kans dat een kandidaat een item correct beantwoordt niet afhankelijk is van het responsgedrag op andere items in de toets. Oftewel, twee items zijn alleen aan elkaar gerelateerd via de invloed van de onderliggende vaardigheid die ze meten. Als het ene item bijvoorbeeld aanwijzingen bevat voor het correct beantwoorden van andere items, dan wordt de aanname van lokale onafhankelijkheid geschonden. Een ander voorbeeld van een schending is dat meerdere items gekoppeld kunnen zijn aan een tekst of grafiek. In die gevallen wordt de afhankelijkheid vaak expliciet mee gemodelleerd om aan de tweede aanname te blijven voldoen.

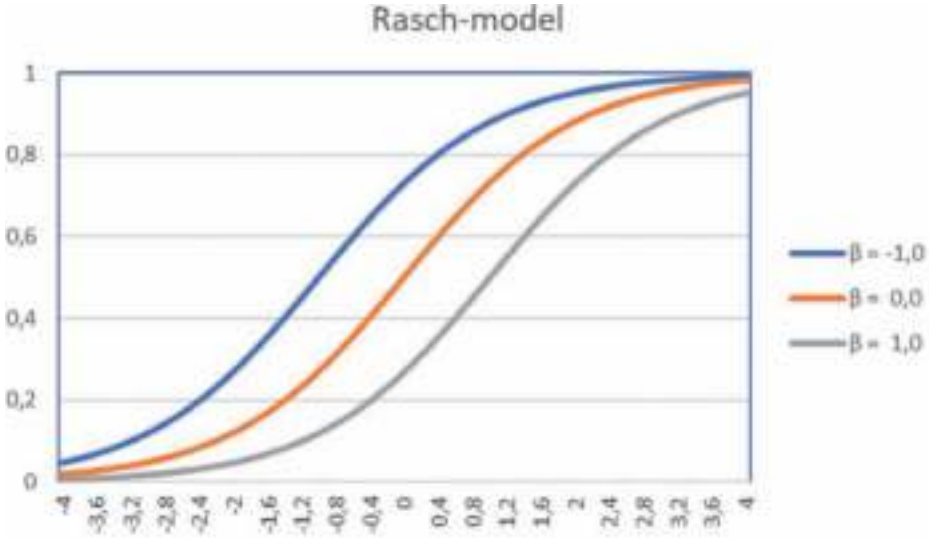
Om het niet-lineaire verband te modelleren gebruikte Rasch een logistische linkfunctie, waarbij de relatie tussen twee variabelen gemodelleerd wordt als:

$$Y = \frac{e^X}{1+e^X}$$

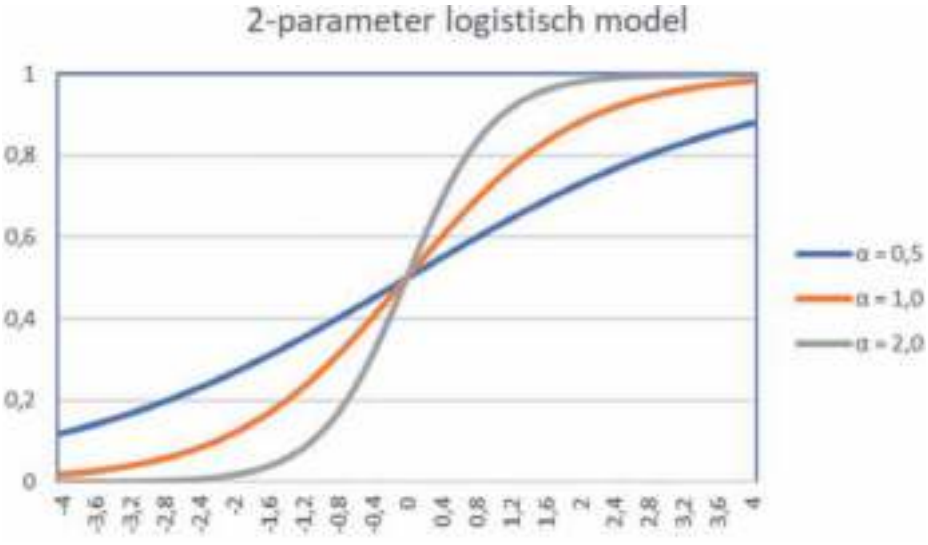
Rasch modelleerde de kans dat een kandidaat een correct antwoord gaf op een vraag ($X = 1$), afhankelijk van de moeilijkheid ($\beta$) van item $i$ en de vaardigheid ($\theta$) van kandidaat $j$, als:

$$P((X=1) \mid \theta, \beta) = \frac{e^{(\theta-\beta)}}{1+e^{(\theta-\beta)}}.$$

De vaardigheid is daarbij gedefinieerd als een latente variabele die het beheersingsniveau van de kandidaat weergeeft, waarbij $-\infty < \theta < \infty$. De moeilijkheid is gedefinieerd als een locatieparameter die aangeeft bij welke vaardigheid de kandidaat een kans heeft van $p = 0{,}50$ om de vraag correct te beantwoorden ($-\infty < \beta < \infty$).




**Figuur 2** Item-karakteristieke curves Rasch-model.




**Figuur 3** Item-karakteristieke curves voor het 2-parameter logistisch model.

Is de vaardigheid van de kandidaat hoger dan de moeilijkheid, dan is de kans op een correct antwoord groter dan $0{,}50$, als de vaardigheid lager is dan de moeilijkheid, dan is de kans kleiner dan $0{,}50$. Dit Rasch-model wordt ook wel het $1$-parameter logistisch model genoemd, omdat de eigenschappen van het item gemodelleerd worden met één parameter, namelijk de moeilijkheid. Figuur 2 geeft de item-karakteristieke curves weer voor items met moeilijkheden $\beta_1 = -1$, $\beta_2 = 0$, $\beta_3 = 1$. Voor een kandidaat met een vaardigheid gelijk aan $\theta = 1{,}5$ is de kans op een correct antwoord op het item gelijk aan respectievelijk $0{,}92$; $0{,}82$; $0{,}62$.

Kenmerkend voor het Rasch-model is dat de item-karakteristieke curves allemaal dezelfde vorm hebben en dat ze alleen verschillen in locatie. Dit is een vrij strakke eis waaraan in de praktijk lang niet altijd wordt voldaan.

*Het 2-parameter logistisch model.* Om wat meer flexibiliteit aan te brengen in item-responstheoriemodellen, werd het $2$-parameter logistisch model ontwikkeld [7]. Naast de moeilijkheidparameter ($\beta$) kent dit model ook een discriminatieparameter ($\alpha$). Deze parameter geeft aan hoe goed een item onderscheid kan maken tussen kandidaten die een vaardigheid hebben lager dan de moeilijkheid en de kandidaten die een vaardigheid hebben hoger dan de moeilijkheid. Het $2$-parameter logistisch model kan geformuleerd worden als:

$$P((X=1) \mid \theta, \alpha, \beta) = \frac{e^{\alpha(\theta-\beta)}}{1+e^{\alpha(\theta-\beta)}}.$$

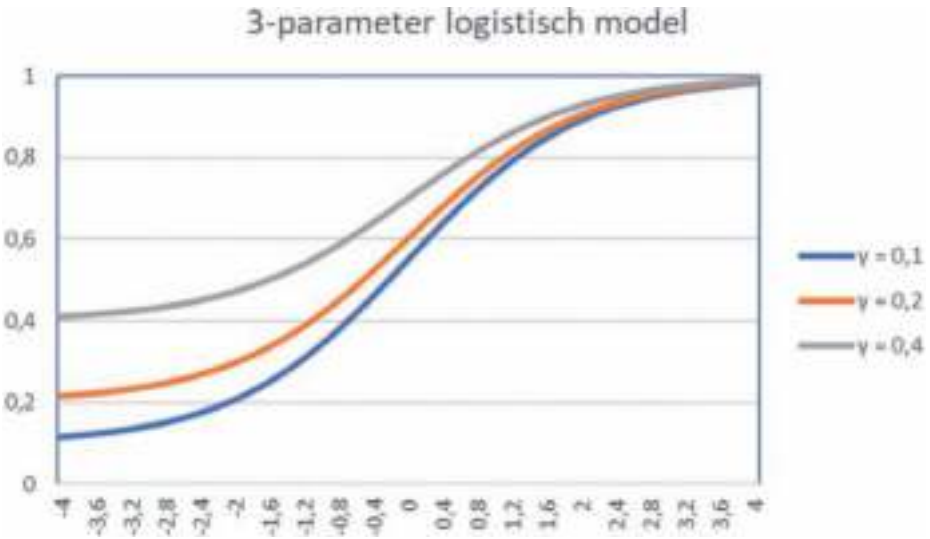
In Figuur 3 worden de karakteristieke curves weergegeven van drie items met discriminatieparameters $0{,}5$; $1{,}0$; $2{,}0$ waarbij de moeilijkheid van alle drie de items gelijk is aan $\beta = 0{,}0$.

In Figuur 3 is te zien dat het verschil in kans op een goed antwoord tussen kandidaten met een vaardigheid $\theta = -0{,}4$ en kandidaten met een vaardigheid $\theta = 0{,}4$ voor items met een laag onderscheidend vermogen ($\alpha = 0{,}5$) slechts gelijk is aan $0{,}10$, terwijl het voor items met een hoog onderscheidend vermogen ($\alpha = 2{,}0$) gelijk is aan $0{,}39$. Items met een hoog onderscheidend vermogen kunnen dus veel beter onderscheid maken tussen kandidaten met een vaardigheid lager dan de moeilijkheid en kandidaten met een vaardigheid hoger dan de moeilijkheid. Deze items leveren daarom meer informatie over de kandidaten.

*Het 3-parameter logistisch model.* Een tweede uitbreiding heeft te maken met de mogelijkheden voor een kandidaat om het correcte antwoord te raden. Als een toets bestaat uit multiplechoicevragen, dan heeft een kandidaat een kans groter dan nul om goed te gokken, op het moment dat hij of zij een van de antwoorden aanvinkt. Om hiervoor te corrigeren is het $2$-parameter logistisch model uitgebreid met een pseudo-gokparameter ($\gamma$). Deze parameter geeft aan dat iedere kandidaat een kans heeft gelijk aan $\gamma$ om het item correct te beantwoorden, ongeacht zijn of haar vaardigheid. Het $3$-parameter logistisch model wordt geformuleerd als:

$$P((X=1) \mid \theta, \alpha, \beta, \gamma) = \gamma + (1-\gamma)\frac{e^{\alpha(\theta-\beta)}}{1+e^{\alpha(\theta-\beta)}}.$$

Op het moment dat een multiplechoice-item vier antwoordcategorieën heeft, zou je verwachten dat de pseudo-gokparameter de waarde aanneemt $\gamma = 0{,}25$. Deze schatting houdt er alleen geen rekening mee dat de verschillende antwoordcategorieën niet allemaal even waarschijnlijk zijn. Het gevolg hiervan is dat de pseudo-gokparameter verschillende waardes aan kan nemen, zelfs al is het aantal antwoordcategorieën vergelijkbaar. Figuur 4 laat de item-karakteristieke curves zien van items van een pseudo-gokparameter gelijk aan $\gamma_1 = 0{,}1$; $\gamma_2 = 0{,}2$; $\gamma_3 = 0{,}4$, terwijl de moeilijkheidsparameters en discriminatieparameters gelijk zijn aan $\beta = 0{,}0$ en $\alpha = 1{,}0$.




**Figuur 4** Item-karakteristieke curves voor het 3-parameter logistisch model.

In Figuur 4 is goed te zien hoe de pseudogokparameter fungeert als een onderasymptoot voor de kans op een correct antwoord. Het Rasch-model, het 2-parameter logistisch model en het 3-parameter logistisch model zijn alle drie gemodelleerd met een logit-linkfunctie. Al vanaf de beginjaren van item-responstheorie is er ook gebruikgemaakt van equivalente modellen die gebaseerd waren op een probit-linkfunctie.

*Polytome item-responstheoriemodellen.* Beide soorten modellen kunnen gebruikt worden om de kans op een correct antwoord te berekenen voor een gegeven vaardigheidsniveau en bekende itemparameters. Hierbij moet wel opgemerkt worden dat deze modellen ervan uitgaan dat een item correct of incorrect wordt beantwoord. Dit worden ook wel dichotome items genoemd. Als een item ook gedeeltelijk correct beantwoord kan worden, of als een kandidaat meerdere punten kan behalen voor een item, dan is er sprake van een polytoom item en kan er gebruikgemaakt worden van polytome item-responstheoriemodellen [9]. Een voorbeeld hierbij is de vraag:

$$\sqrt{\frac{7{,}5}{0{,}3}-16}=\ldots?$$

Om tot het correcte antwoord te komen, moet een kandidaat zowel de deling, het aftrekken, als het worteltrekken correct uitvoeren. Als kandidaten ook punten krijgen voor het correct oplossen van deelstappen, is er sprake van een polytoom item. Daarnaast is het mogelijk dat er meerdere vaardigheden nodig zijn om een vraag correct te beantwoorden. Een voorbeeld hierbij zijn wiskundeopgaven waarbij kandidaten de benodigde informatie uit een tekst moeten halen. Naast de vaardigheid wiskunde is ook begrijpend lezen nodig om tot een goed antwoord te komen. Een overzicht van multi-dimensionale item-responstheoriemodellen en hun toepassingen wordt gegeven in Reckase [11].

*Informatiefuncties en meetfout*

Binnen de item-responstheorie wordt betrouwbaarheid vervangen door het concept informatie. Hoe meer informatie een item geeft over de vaardigheid van een kandidaat, des te nauwkeuriger de vaardigheid geschat kan worden. Om uit te rekenen hoeveel elk item bijdraagt, wordt gebruikgemaakt van statistische informatietheorie [1]. Fishers informatiefunctie is een van de manieren om te berekenen hoeveel informatie een geobserveerde variabele (het antwoord van de kandidaat) geeft over een latente variabele (de vaardigheid), als de kans op deze geobserveerde variabele afhangt van deze latente variabele. Voor de verschillende modellen kan de informatiefunctie $I_i(\theta)$, die laat zien hoe de informatie van item $i$ afhankelijk is van de vaardigheid $\theta$, uitgerekend worden als:

$$I_i(\theta)=\frac{\left[\frac{\partial P_i(\theta)}{\partial\theta}\right]^2}{P_i(\theta)\left(1-P_i(\theta)\right)},$$

waarbij $P_i(\theta)$ staat voor de item-responsefunctie $P((X=1)\mid\theta,\beta)$ van item $i$. Iteminformatiefuncties voor de verschillende item-responstheoriemodellen kunnen gevonden worden in bijvoorbeeld Embretson en Reise [3]. Deze informatiefuncties hebben twee nuttige eigenschappen. Allereerst kan de informatie voor een complete toets, de toetsinformatiefunctie (TIF), uitgerekend worden door de informatie van de verschillende items op te tellen:

$$TIF(\theta)=\sum_{i=1}^{N}I_i(\theta).$$

Daarnaast kan de meetfout uitgerekend worden met:

$$SE(\theta)=\frac{1}{\sqrt{TIF(\theta)}}.$$

*Schatten vaardigheid en itemparameters*

Er zijn verschillende schattingsmethoden ontwikkeld om de vaardigheid ($\theta$) en de itemparameters ($\alpha$,$\beta$,$\gamma$) te schatten. Al deze methoden gaan uit van de likelihood, oftewel de waarschijnlijkheid van de geobserveerde antwoordpatronen. Voor een toets van $N$ items die beantwoord is door $J$ personen, waarbij de antwoorden van persoon $j$ op item $i$ genoteerd wordt als $u_{ij}$ ($u_{ij}=1$ voor een correct antwoord en $u_{ij}=0$ voor een incorrect antwoord), en de kans dat persoon $j$ een correct antwoord geeft op item $i$ als $P_{ij}$ kun je afleiden dat:

$$P_{ij}^{u_{ij}}(1-P_{ij})^{1-u_{ij}}=\begin{cases}P_{ij} & \text{als } (u_{ij}=1),\\ 1-P_{ij} & \text{als } (u_{ij}=0).\end{cases}$$

Voor $u_{ij}=1$ geldt immers dat $P_{ij}^{u_{ij}}=P_{ij}^{1}=P_{ij}$ en voor $u_{ij}=0$ geldt dat $P_{ij}^{u_{ij}}=P_{ij}^{0}=1$. Daarmee kan de likelihood geformuleerd worden als:

$$L(u_{11},u_{12},\ldots,u_{NJ}\mid\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma})=\prod_{i=1}^{N}\prod_{j=1}^{J}P_{ij}^{u_{ij}}(1-P_{ij})^{1-u_{ij}},$$

waarbij ($\boldsymbol{\theta}$) de vector met vaardigheidsparameters weergeeft en ($\boldsymbol{\alpha},\boldsymbol{\beta},\boldsymbol{\gamma}$) de matrix met itemparameters. Deze likelihood is de vermenigvuldiging van de kansen op correcte ($P_{ij}$) of incorrecte ($1-P_{ij}$) antwoorden voor alle items in de toets en alle kandidaten die meegedaan hebben. Als bijvoorbeeld de vragen 1 tot en met 5 op de volgende manier beantwoord worden door kandidaat 3: (correct, correct, incorrect, incorrect, correct) en $P_{ij}$ op dezelfde manier gedefinieerd is als hiervoor, ziet de likelihood, hier genoteerd als $L(\cdot)$ er uit als:

$$L(\cdot)=P_{13}P_{23}(1-P_{33})(1-P_{43})P_{53}.$$

Bij het schatten van de vaardigheids- en de itemparameters wordt gezocht naar parameterwaardes die deze likelihood optimaliseren. Joint Maximum Likelihood-schatters, die proberen om tegelijkertijd zowel de vaardigheden als de itemparameters te schatten, leveren helaas inconsistente schatters op. Alternatieve methodes hiervoor zijn Marginal Maximum Likelihood-schatters en Conditional Maximum Likelihood-schatters. De Marginal Maximum Likelihood-schatters schatten eerst de itemparameters door de vaardigheidsparameters uit de likelihoodvergelijking te integreren, waarbij ervan uitgegaan wordt dat de vaardigheden van de hele populatie standaardnormaal verdeeld zijn. Na het schatten van itemparameters kunnen de vaardigheidsparameters geschat worden met een Joint Maximum Likelihood-schatting. De Conditional Maximum Likelihood-schatters gebruiken een vergelijkbare aanpak waarbij eerst de itemparameters geschat worden, conditioneel op de vaardigheid. Vervolgens worden de vaardigheidsparameters van de individuele kandidaten apart geschat. Voor een uitgebreide beschrijving van deze schattingsmethodes, zie Eggen en Sanders (1993). Naast deze beide frequentistische methodes, kunnen de vaardigheden- en de itemparameters ook geschat worden met een Bayesiaans algoritme, waarbij een priorverdeling voor zowel de vaardigheden als de itemparameters wordt aangenomen. Voor een overzicht van Bayesiaanse IRT-methodes wordt verwezen naar Fox [4]. Zowel voor de maximum likelihood-methodes als voor de Bayesiaanse methode kan

gebruikgemaakt worden van standaard opensource-softwarepakketten.

Omdat de vaardigheid en de itemparameters los van elkaar geschat worden, werken de bovenstaande schattingsmethodes ook bij missing data, dat wil zeggen dat ze ook werken als de kandidaten maar een deel van de items hebben beantwoord. Voorwaarde is wel dat deze missing data niet afhangt van de vaardigheid van de kandidaten. Het mag dus niet komen omdat de kandidaat vragen die hij of zij te moeilijk vindt over heeft geslagen. Deze eigenschap wordt gebruikt om verschillende toetsen of examens onderling vergelijkbaar te maken. Beide toetsen kunnen op dezelfde schaal gebracht worden door ze te equivaleren. Hiervoor is nodig dat een deel van de kandidaten, zowel de vragen van de eerste als van de tweede toets maken. Op die manier krijg je een gelinkt design. Twee manieren om dit te organiseren worden weergegeven in Figuur 5, waarbij de rijen verschillende groepen kandidaten weergeven en de kolommen de verschillende items. De blauwe gedeeltes geven de items aan die de verschillende groepen beantwoorden, de grijze gedeeltes staan voor items die niet aan de betreffende groep worden aangeboden.

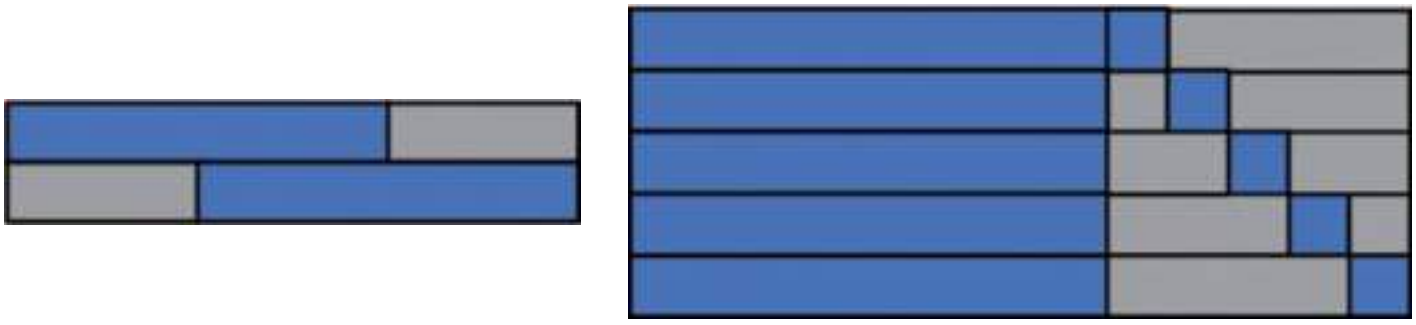
In het design van Figuur 5(a) worden twee toetsen aan elkaar gelinkt doordat twee groepen kandidaten een deel van de items gezamenlijk hebben. Allereerst worden de itemparameters van de eerste toets geschat op basis van de eerste groep kandidaten. Daarmee wordt de schaal vastgelegd. De itemparameters van de items die beide groepen gemeenschappelijk hebben hoeven niet opnieuw geschat te worden voor de tweede groep. De ontbrekende itemparameters van de tweede toets worden vervolgens op dezelfde schaal geschat als de items waarvan de parameters al bekend zijn. Dit gebeurt door deze itemparameters te fixeren. De parameters van de eerste toets en de tweede toets liggen nu op dezelfde schaal waardoor de toetsscores vergelijkbaar zijn. Bij het design in Figuur 5(b) maken vijf verschillende groepen de hele eerste toets. Daarnaast maken ze allemaal een deel van de tweede toets. De itemparameters van de eerste toets zijn identiek voor alle vijf de groepen. De vaardigheidsverdeling van de vijf groepen liggen daarmee op dezelfde schaal. Die schaal wordt vervolgens gebruikt om de itemparameters van de tweede toets op dezelfde schaal te schatten. Hiermee worden de cijfers op de tweede toets vergelijkbaar met die op de eerste toets.

Op deze manier kan er bijvoorbeeld voor gezorgd worden dat de eindexamens van verschillende jaren onderling vergelijkbaar zijn. Voorafgaand aan de examenperiode worden daarvoor items van het nieuwe examen samen met een aantal items van het oude examen uitgeprobeerd bij een kleine groep leerlingen. De informatie die dit oplevert, wordt gebruikt om de nieuwe items te kalibreren op dezelfde schaal. Voor het nieuwe eindexamen worden items gekozen die voldoen aan de randvoorwaarden voor de inhoud, maar die gezamenlijk ook een testinformatiefunctie hebben, die vergelijkbaar is met de examens van de jaren ervoor. Op deze manier worden beide examens geëquivaleerd.

*Voor- en nadelen van item-responstheorie*

Item-responstheorie heeft een groot aantal voordelen. Allereerst wordt de vaardigheid van de kandidaten nauwkeuriger geschat, waarbij rekening gehouden wordt met verschil in moeilijkheid en onderscheidend vermogen van de items. De schatting van de vaardigheid is onafhankelijk van de moeilijkheid van de toets en van de gebruikte items. Bovendien kan de informatiefunctie gebruikt worden om een individuele schatting van de meetfout te berekenen. Ook met betrekking tot de itemparameters kent item-responstheorie grote voordelen. Door deze parameters separaat te schatten krijgen we veel meer inzicht in de kwaliteit van de individuele items. De itemparameters kunnen bovendien onafhankelijk van de steekproef van kandidaten geschat worden. Daarnaast liggen de itemmoeilijkheid en de vaardigheid van de kandidaat op dezelfde schaal, wat veel inzicht geeft. Het gebruik van item-responstheorie kan daarom leiden tot eerlijkere en accuratere toetsing. Ten slotte maakt item-responstheorie het mogelijk om toetsen en examens adaptief af te nemen. Dat houdt in dat tijdens de afname van een toets of examen al een inschatting van de vaardigheid van de kandidaat gemaakt wordt en dat de moeilijkheid van de items afgestemd wordt op het niveau van de kandidaat. Dit leidt tot kortere toetsen en voorkomt frustratie vanwege het moeten beantwoorden van veel te makkelijke of veel te moeilijke items.

Naast deze voordelen heeft item-responstheorie ook een aantal nadelen. De vaardigheid wordt geschat op een standaardnormaal verdeelde schaal. De schattingen lopen daarmee van $-4{,}0$ tot $4{,}0$ met een gemiddelde van $0{,}0$. In tegenstelling tot wat we in Nederland gewend zijn met de klassieke som-correctscore, is deze schaal bovendien niet lineair. Voor de meeste leerkrachten en leerlingen is daarom een vertaling nodig naar een schaalscore. Dit gebeurt door middel van een standaard-setting procedure. In deze procedure werken inhoudsdeskundigen en psychometrici samen om te bepalen bij welke vaardigheidsscore een kandidaat voldoende score voor de test. Deze vaardigheidsscore krijgt de waarde $5{,}5$ op de scoreschaal. Vervolgens wordt een tabel ontwikkeld die de te behalen vaardigheidsscores ver-



**Figuur 5** Voorbeelden van gelinkte designs.

taalt naar een schaalscore die loopt van 0 tot 10. Een positieve uitzondering hierbij is het Rasch-model, waarvan aangetoond is dat je daarbij wel de som-correctscore kunt gebruiken [2]. Een tweede nadeel is dat de steekproef voldoende groot (minstens een paar honderd kandidaten) moet zijn om de itemparameters nauwkeurig genoeg te kunnen schatten. Daardoor is item-responstheorie alleen toepasbaar bij grotere toetsen en examens. Een derde nadeel, ten slotte, is dat de item-responsmodellen wel een goede fit moeten laten zien bij de responsdata. Als de geobserveerde scores erg afwijken van de item-karakteristieke curves, dan is er sprake van misfit en zijn item-responsmodellen niet toepasbaar.

**Conclusie**

Toetsen en examens nemen een belangrijke plaats in binnen het Nederlandse onderwijs. Geschat wordt dat leerkrachten en leerlingen gemiddeld 30% van hun tijd hieraan besteden. De cijfers die de leerlingen krijgen, hebben daarnaast veel invloed op hun schoolcarrière en hun kansen op een baan. Het is daarom van groot belang dat de cijfers een betrouwbaar beeld geven van de vaardigheden van de kandidaten. In dit artikel is een korte inleiding gegeven in de psychometrie. Deze relatief onbekende tak van wiskunde is ontwikkeld om een fundament te leggen onder het meten in het onderwijs. Psychometrie heeft zich in de afgelopen honderd jaar ontwikkeld tot een rijk vakgebied en omdat er steeds meer opensource-software beschikbaar komt, is het ook toepasbaar voor een groot publiek.

Psychometrie is overigens niet exclusief ontwikkeld voor het onderwijs, al heeft de grootschalige toepassing binnen de examinering er wel een enorme boost aan gegeven. Een tweede belangrijke toepassing ligt in het analyseren van vragenlijstdata. Binnen de psychologie en de gezondheidszorg wordt bijvoorbeeld op grote schaal gebruikgemaakt van Likert-items, waarbij respondenten op een schaal van één tot drie, één tot vijf of één tot zeven alternatieven, aan moeten geven in hoeverre ze het met een uitspraak eens zijn. Concrete voorbeelden van toepassingen in de psychometrie zijn vragenlijsten die depressie meten of intelligentietesten. Met name de klassieke testtheorie wordt hierbij veel gebruikt, terwijl sinds de eeuwwisseling item-responstheorie ook binnen de psychologie en gezondheidszorg steeds vaker wordt toegepast.

Een van de uitdagingen waar de psychometrie op dit moment voor staat is dat er steeds meer data en anderssoortige data beschikbaar komt over kandidaten. Terwijl de modellen in de psychometrie met name gericht zijn op het analyseren van antwoorden, komt er steeds meer informatie beschikbaar over het leerproces. Online leersystemen leggen in logfiles exact vast wat de verschillende stappen zijn die de kandidaat doorlopen heeft om tot een antwoord te komen. Ook kan de tijd die de kandidaat nodig gehad heeft bruikbare informatie geven over zijn of haar vaardigheid. Van der Linden [12] stelde al voor hoe item-responstheorie modellen uitgebreid kunnen worden om responstijden mee te kunnen modelleren. Maar de vraag hoe logfilegegevens gemodelleerd moeten worden en de vraag hoe je (wiskundig) kunt verantwoorden dat deze data gebruikt wordt vaardigheden of groei te schatten, liggen nog open.

**Referenties**


1 T.M. Cover en J.A. Thomas, *Elements of Information Theory,* John Wiley & Sons, 2012.

2 T.J.H.M. Eggen en P.F. Sanders, *Psychometrie in de praktijk*, Cito Instituut voor Toetsontwikkeling, 1993.

3 S.E. Embretson en S.P. Reise, *Item Response Theory for Psychologists*, Lawrence Earlbaum Associates, 2000.

4 J.P. Fox, *Bayesian Item Response Modeling: Theory and Applications,* Springer Science & Business Media, 2010.

5 R.K. Hambleton, H. Swaminathan en H.J. Rogers, *Fundamentals of Item Response Theory*, Vol. 2, Sage, 1991.

6 F.M. Lord, A theory of test scores, *Psychometric Monographs* (1952).

7 F.M. Lord en M.R. Novick, *Statistical Theories of Mental Test Scores*, Addison-Wesley, 1968.

8 S. Messick, (Meaning and values in test validation: The science and ethics of assessment, *Educational Researcher* 18(2) (1989), 5–11.

9 R. Ostini en M.L. Nering, *Polytomous Item Response Theory Models*, No. 144, Sage, 2006.

10 G. Rasch, *Studies in Mathematical Psychology: I. Probabilistic Models for Some Intelligence and Attainment Tests*, 1960.

11 M.D. Reckase, *Multidimensional Item Response Theory*, Springer, 2009.

12 W.J. van der Linden, A hierarchical framework for modeling speed and accuracy on test items, *Psychometrika* 72(3) (2007), 287.

13 S. Wools, T.J. Eggen en A.A. Béguin, Constructing validity arguments for test combinations, *Studies in Educational Evaluation* 48 (2016), 10–18.

Arnout van de Rijt
*Faculteit Sociale Wetenschappen*
*Universiteit Utrecht*
*a.vanderijt@uu.nl*

Vincent Buskens
*Faculteit Sociale Wetenschappen*
*Universiteit Utrecht*
*v.buskens@uu.nl*

# Wiskunde in de sociologie

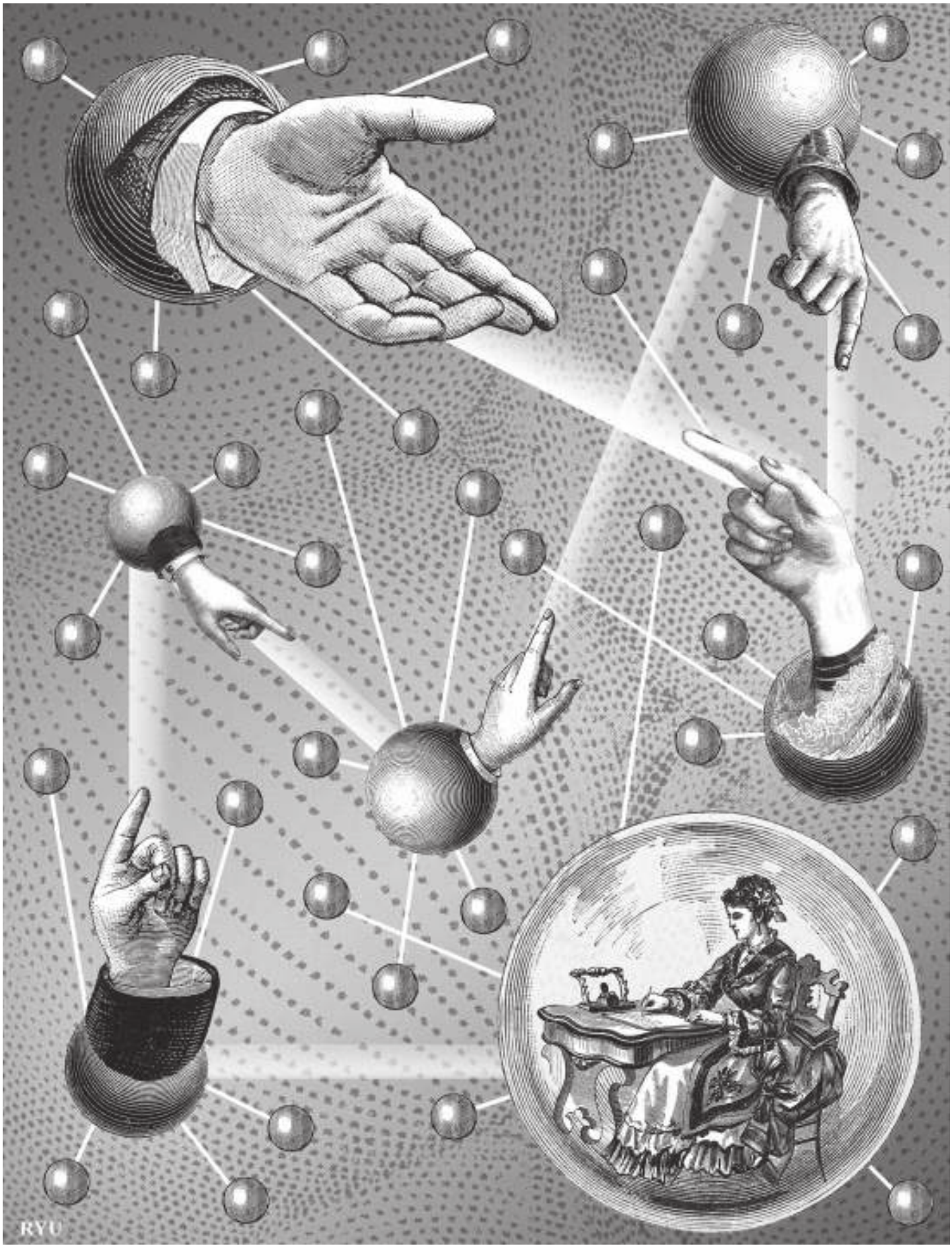
**De wiskundige sociologie ontstond na de Tweede Wereldoorlog. Zij onderscheidt zich van de reguliere sociologie, waarin theorieën meestal niet geformaliseerd worden. Mathematisering van theorieën blijft een uitzonderingssituatie in de hedendaagse sociologie. Wiskundige sociologie is in Nederland relatief oververtegenwoordigd in vergelijking met het buitenland. In Nederland zijn de wiskundig sociologen vooral te vinden aan de Universiteit Utrecht, de Universiteit van Amsterdam, de Radboud Universiteit en de Rijksuniversiteit Groningen. Arnout van de Rijt en Vincent Buskens van de Universiteit Utrecht laten aan de hand van voorbeelden zien wat wiskundige sociologie inhoudt.**

Er zijn leuke wiskundige resultaten mogelijk in de sociologie die verrassende inzichten geven. Voor de geïnteresseerde lezer citeren we een aantal inleidingen in de wiskundige sociologie [7, 12, 14, 15, 23, 32]. De grondleggers van de wiskundige sociologie waren erop uit om door middel van formele modellering van sociale processen het logische beargumenteren van gevolgen van assumpties over menselijk gedrag voor uitkomsten over de samenleving te verbeteren. Het blijkt namelijk dat bijvoorbeeld vanwege domino-effecten uitkomsten op het niveau van de samenleving vaak op niet-triviale manieren afhangen van wat individuen willen en doen. De uitkomsten van dit soort domino-effecten zijn vaak via informele argumentaties moeilijk te anticiperen. Zij kunnen echter inzichtelijk gemaakt worden door de formalisering van aannames en bijbehorende processen. In dit stuk bespreken we vier voorbeelden uit onze onderzoekspraktijk.

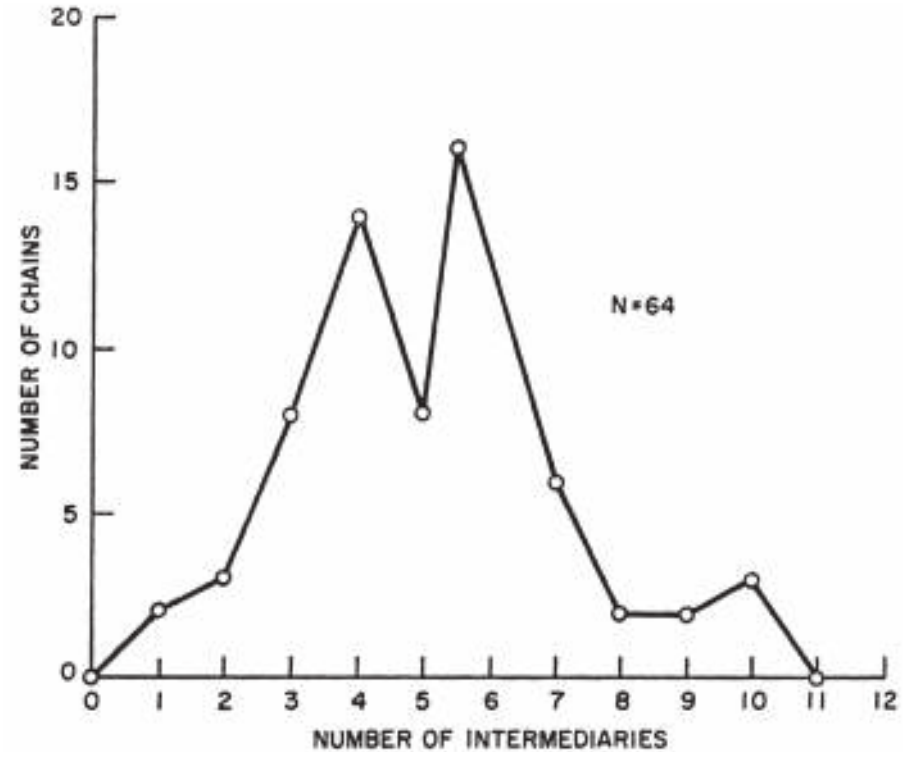
## Small worlds

Eerst een puzzel. In 1967 rapporteerde Stanley Milgram de resultaten van een experiment [33]. Zo'n driehonderd inwoners van Kansas en Nebraska werd een envelop met bescheiden gegeven. De uiteindelijke bestemming, de 'Target Person', was iemand in Massachusetts. De inwoners werd gevraagd de envelop naar iemand te sturen die de betreffende persoon mogelijk zou kennen. Als ze zo'n persoon niet kenden, werd ze gevraagd de envelop te sturen naar iemand die zo iemand zou kennen, et cetera. Verrassend genoeg vond Milgram dat de brief gemiddeld in zes stapjes zijn bestemming bereikte. Dit komt overeen met vijf tussenpersonen, of 'intermediaries'. De originele grafiek van Milgram met de small-worlds-bevinding is weergegeven in Figuur 1.

Dit empirische resultaat is gemakkelijk te verklaren door je een wereld voor te stellen waarin iedereen vriendjes is met, zeg, vijftig willekeurig gekozen anderen. In dat geval kun je via de vrienden van je vrienden van je vrienden van je vrienden van je vrienden al in zes stappen $50^6$




Illustratie: Ryu Tajiri




**Figuur 1** Originele grafiek van Stanley Milgram over 'small worlds'.

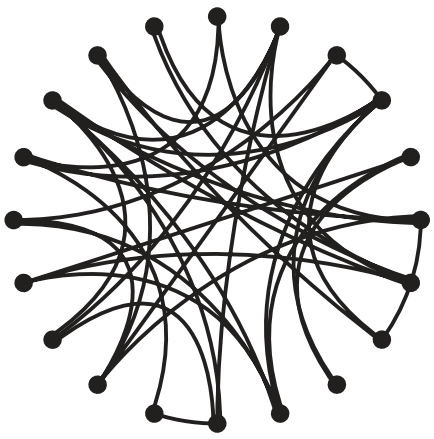
mensen bereiken, dus iedereen op aarde. Een miniatuurversie van zo'n wereld is afgebeeld in Figuur 2. Dit model is een 'random network' en is opgesteld door Erdős en Rényi. (In dit model is het Erdősgetal $3$ van de tweede auteur niks speciaals!)

Maar in zo'n wereld leven we niet. We kennen vooral mensen die heel dichtbij wonen en die een soortgelijk inkomen, dezelfde hobby, en dezelfde cultuur hebben als wijzelf. Dat creëert kliekjes. De vrienden van mijn vrienden zijn veelal ook mijn vrienden. De wereld ziet er meer uit als de 'regular ring lattice' in Figuur 3. En daarin zit hem de puzzel [22]. In de ring lattice kun je alleen lokaal wandelen. Vertices kennen alleen hun buren en overburen. Je hebt veel meer stapjes nodig om elders uit te komen. Hoe kan de wereld er nou enerzijds grotendeels uitzien als een regular ring lattice (Figuur 3), en anderzijds toch de eigenschap delen met het random netwerk (Figuur 2) dat je slechts een paar vriendschappen van iedereen verwijderd bent?

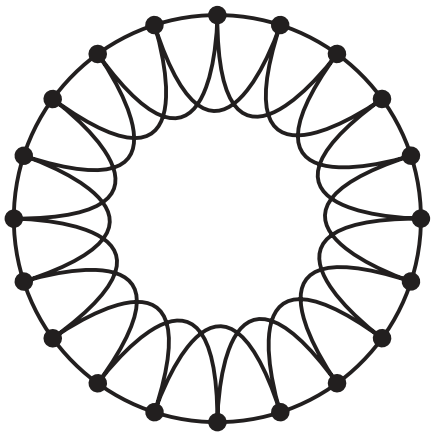
Deze sociologische puzzel werd wiskundig opgelost door Duncan Watts en Steve Strogatz. De oplossing staat in een beroemd artikel dat in 1998 gepubliceerd werd in *Nature* [35]. Watts en Strogatz namen de regular ring lattice als uitgangspunt maar voegden daar een beetje randomness aan toe. Dit deden ze door zijden uit de graaf te verwijderen en aan willekeurige andere punten weer vast te maken (zie Figuur 4). Daarmee creëerden ze een hybride netwerk dat $99\%$ overeenkomt met een ring lattice en $1\%$ met een random netwerk. Ze lieten vervolgens zien dat de gemiddelde kortste route tussen twee personen in dit netwerk bijna gelijk is aan die van een random netwerk: $\ln(n)/\ln(k)$. Hier is $n$ het aantal personen en $k$ het aantal vriendschappen per persoon. In onze wereld met $n \approx 7.000.000.000$ levert een gemiddeld aantal vriendschappen van $k = 50$ een kortste route van ongeveer zes stappen op. Precies die 'six degrees of separation' die Milgram 31 jaar eerder proefondervindelijk ontdekte.

Een vraag die onbeantwoord blijft is echter: Hoe wisten Milgrams proefpersonen nou in hemelsnaam naar wie ze de envelop moesten doorsturen? Ze hadden tenslotte geen wereldkaart van het mondiale sociale netwerk waarop je de kortste route zou kunnen opzoeken. Toch wisten ze heel korte routes te vinden. Kennelijk was het genoeg dat de proefpersonen slechts een paar dingen van de bestemmingspersoon afwisten zoals beroep, opleiding en etniciteit. Op basis daarvan stuurden ze het naar vrienden en kennissen die langs die paar dimensies het meest op de bestemmingspersoon leken. Maar is dit een plausibele verklaring?

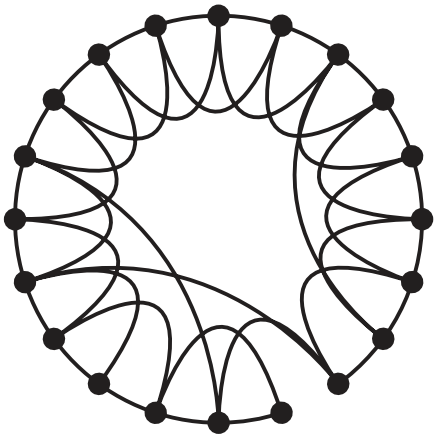
Om dit beter te begrijpen voerde de eerste auteur met Xiaomeng Ban en Jie Gao in een artikel in 2010 de volgende exercitie uit [4]. Zij namen vijf bestaande sociale netwerken en bedden elk in in een laagdimensionale ruimte, waarbij vrienden dicht bij elkaar werden geplaatst. In si-



**Figuur 2**

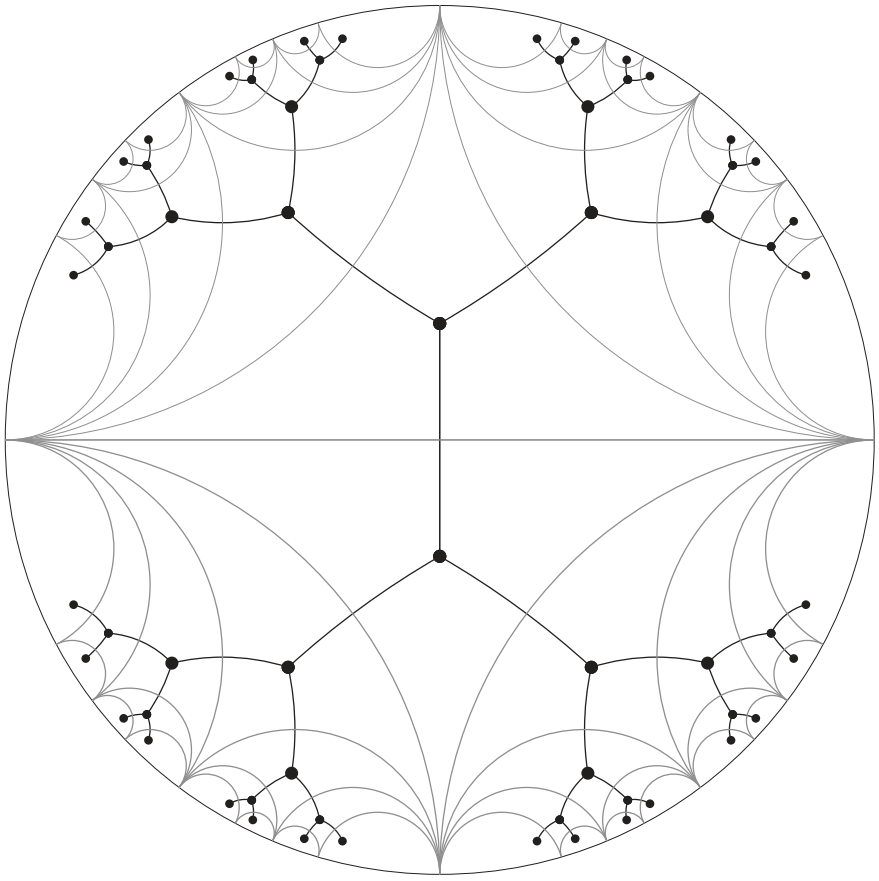


**Figuur 3**



**Figuur 4**

Illustraties: D.J. Watts en S.H. Strogatz [35]



Illustratie: R. Kleinberg [21]

**Figuur 5**

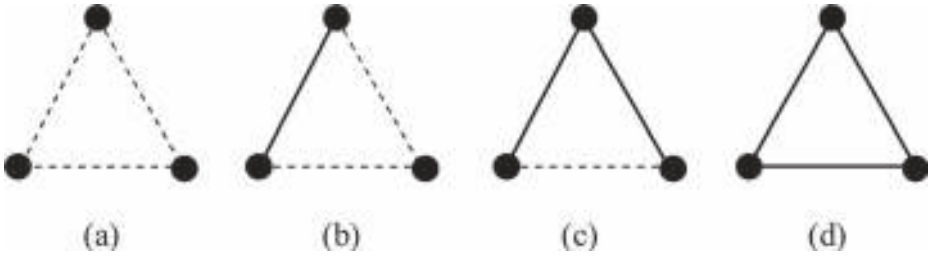
mulaties werd synthetische proefpersonen informatie en een simpele beslissingsregel gegeven. Zij kregen de coördinaten in die ruimte vervolgens toegewezen. Zij kregen ook de taak een envelop richting een andere proefpersoon te sturen waarvan coordinaten werden gegeven, net als in Milgrams small-world-experiment. Een 'greedy algorithm' werd toegepast waarbij bots de envelop doorstuurden naar een bevriende bot die op de kleinste afstand van de bestemming zat. Dit greedy algorithm bootst dus een proefpersoon van Milgram na die de envelop doorstuurt naar een kennis die het meest lijkt op de bestemming. Dit bleek inderdaad te werken! In alle vijf de netwerken waren de bots in staat hun bestemming in hooguit zes stappen te bereiken. Dit werkte het beste als de netwerken werden ingebed in het hyperbolische vlak (zie Figuur 5). Computerwetenschapper Robert Kleinberg [21] had eerder al bewezen dat 'greedy delivery' hierin nooit vastloopt. Onze resultaten laten zien dat levering in alle netwerken niet alleen gegarandeerd is, maar dat het ook in slechts een handjevol stappen kan. Zo kunnen we dus begrijpen dat Milgrams proefpersonen middels een simpele heuristiek korte paden zonder schatkaart met succes wisten te vinden.

## Sociale balans

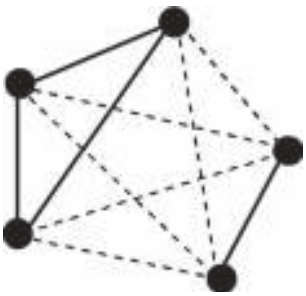
In de oploop naar zowel de Eerste als Tweede Wereldoorlog organiseerden landen zich in twee allianties [2]. In het verloop van de oorlogen wisselden sommige landen van loyaliteit, maar altijd waren ze óf "with us" óf "against us". Ook in de bilaterale relaties tussen landen in het Midden Oosten zien we zulke macro-patronen [24]. Gangs vormen vaak tweezijdige allianties bij het voeren van strijd over territorium in steden [25]. En in schoolklassen zien we ook vaak twee groepen die tegenover elkaar staan [28]. Hoe verklaren we de neiging van mensen, organisaties, en landen om zich zo dualistisch te gedragen, om twee vijandige clubjes te vormen?

In 1956 deden Dorwin Cartwright en Frank Harary een prachtige wiskundige ontdekking [11]. Ze herformuleerden eerst een gedragsprincipe van Frits Heider betreffende menselijke relaties in termen van netwerken met positieve en negatieve relaties: de vriend van een vriend is een vriend, de vijand van een vriend is een vijand, de vriend van een vijand is een vijand, en de vijand van een vijand is een vriend [20]. Er zijn vier soorten driehoeksrelaties, weergegeven in Figuur 6. Solide lijnen stellen positieve relaties voor, onderbroken lijnen negatieve. Twee van de vier driehoeksrelaties, (b) en (d), voldoen aan het balans-theoretische principe. De andere twee, (a) en (c), schenden het.

Cartwright en Harary lieten zien dat als alle driehoekjes in een netwerk van type (b) of (d) zijn, dan bestaat het netwerk noodzakelijkerwijs uit precies twee vijandige groepen wanneer er conflict is! Een voorbeeld is weergegeven in Figuur 7. Drie actoren zitten in de ene groep, de overige twee in de andere. Binnen groepen zijn er alleen positieve relaties, tussen groepen slechts negatieve. Dit resultaat laat dus zien dat in situaties waar individuen, organisaties of landen Heiders gedragsprincipe




**Figuur 6**



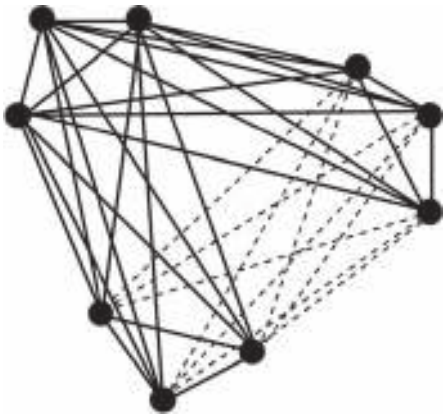
Illustraties: A. van de Rijt [30]

**Figuur 7**

strikt volgen, zij bij conflict noodzakelijkerwijs in een tweekampensituatie terechtkomen. Iets anders is onmogelijk zonder schending van het gedragsprincipe.

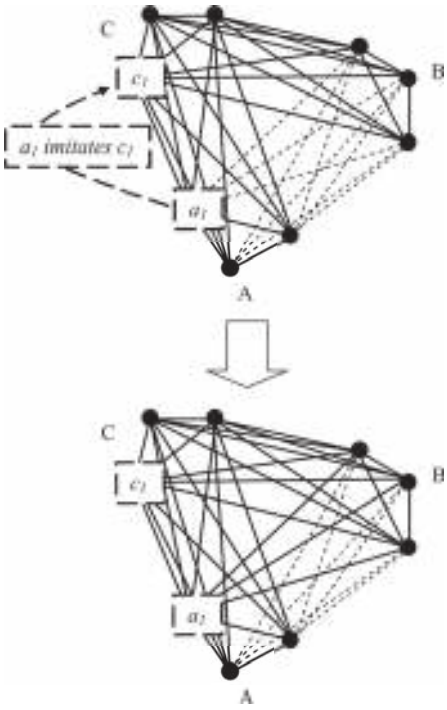
Het resultaat van Cartwright en Harary zegt echter niets over óf zo'n gebalanceerde graaf bereikt zal worden vanuit een situatie waarin niet elke driehoekje gebalanceerd is. Stel dat we in een situatie zitten waarin er één of meerdere ongebalanceerde driehoekjes in het netwerk aanwezig zijn. Is het dan altijd mogelijk voor een actor om het aantal ongebalanceerde driehoekjes waarvan zij deel uitmaakt te verminderen door een vriendschap in een ruzie om te zetten of vice versa, van een vijand een vriend te maken? Deze vraag stelden wiskundigen Tibor Antal, Pavel Krapivsky en Richard Redner in 2005 [2]. Zij ontdekten zodoende het bestaan van 'jammed configurations', waarmee het antwoord "nee!" werd gevonden. In het netwerk in Figuur 8 bijvoorbeeld, zitten tal van ongebalanceerde driewegsrelaties, maar elke verandering maakt de situatie alleen maar erger. Het betreft een dynamisch evenwicht. Het is dus mogelijk dat actoren vast komen te zitten in complexere vijand-vriendpatronen dan de simpele Axis-Allies-configuraties van Cartwright en Harary.

Hoe verklaren we dan de empirische tendens voor het vormen van tweekamppatronen? De eerste auteur onderzocht in 2011 wat er zou gebeuren met de dynamica van balanstheorie als individuen soms een fout maken door een vriend in een vijand te veranderen, of andersom, wanneer dit het aantal ongebalanceerde driehoekjes verder verhoogt [30]. Kan de dynamiek hiermee worden losgeschud uit een jammed state? Het geïmpliceerde spel heeft een potentiaal die alleen gemaximaliseerd wordt



Illustratie: A. van de Rijt [30]

**Figuur 8**




Illustratie: A. van de Rijt [30]

**Figuur 9**

in een Cartwright–Harary-netwerk. Alleen evenwichten die een maximale potentiaal hebben zijn 'stochastisch stabiel'. Met stochastisch stabiel wordt bedoeld dat als de kans op fouten richting de limiet van nul gaat, de kans dat de dynamiek bij het evenwicht in de buurt is niet ook naar nul gaat [16]. Verder bekeek de eerste auteur ook het scenario waarin actoren meer dan één verandering tegelijk kunnen aanbrengen in hun relaties. Het bleek te bewijzen dat bij meerdere veranderingen tegelijk in de relaties van een persoon altijd een verlaging van het aantal gebalanceerde driehoekjes mogelijk is. Dit kan een actor altijd doen door de relaties van een andere actor te kopiëren. Dit is geïllustreerd in Figuur 9: $a_1$ imiteert $c_1$. Daarmee verlaagt $a_1$ het aantal ongebalanceerde driehoekjes. Hieruit volgt ook weer dat er dan uiteindelijk altijd een Cartwright–Harary-netwerk bereikt wordt.

### Coöperatie in netwerken

Een andere puzzel. Veel mensen op deze wereld vertrouwen elkaar of werken met elkaar samen. Dit terwijl ze weten dat anderen goede redenen kunnen hebben om vertrouwen te misbruiken of te profiteren van de inspanningen van anderen. Speltheorie is een wiskundig instrumentarium voor sociale wetenschappers zoals sociologen en economen. Het helpt om gedrag in situaties waarin mensen van elkaar afhankelijk zijn beter te begrijpen. Het gevangenendilemma is een traditionele representatie van zo een probleem. Twee gevangenen kunnen kiezen tussen een ander aangeven of zwijgen. Hun gevangenisstraf hangt af van wat ze zelf doen en wat de ander doet. Het dilemma is zo geframed dat wat de ander ook doet, je altijd minder jaar de gevangenis in hoeft als je de ander aangeeft. Je bent het allerbeste af als de ander zwijgt en je zelf de ander aangeeft. Zie Tabel 1. Daarom voorspelt speltheorie ook dat beide gevangenen de ander zullen aangeven. Dit is het zogenaamde Nash-evenwicht van het spel [26]: een uitkomst waarbij beide spelers hun eigen uitkomst optimaliseren als we het gedrag van de ander als gegeven beschouwen. Het probleem is, dat beide gevangenen langer de gevangenis in gaan als ze beide de ander aangeven dan wanneer ze beide blijven zwijgen. De keuze zwijgen staat hier dus voor samenwerken met de ander om beide minder straf te krijgen. Maar hoe kan het dan dat we veel samenwerking om ons heen zien?

Speltheorie leert ons ook dat als dit spel niet eenmalig gespeeld wordt de analyse verandert. Stel dat twee spelers na elk spel met een vaste kans $p$ het spel opnieuw spelen en met de complementaire kans de serie spellen eindigt. Deze kans $p$ heeft meestal te maken met de situatie waarin coöperatie plaatsvindt. Vaak is het alleen van belang of een $p$ groter of kleiner is als we twee situaties vergelijken. Bijvoorbeeld in een buurt met veel mobiliteit is de kans kleiner dat mensen heel vaak elkaar tegenkomen om elkaar te helpen dan in een buurt met weinig mobiliteit.

Laten we even een getallenvoorbeeld bekijken, bijvoorbeeld een Gevangendilemma waarbij wederzijdse samenwerking $3$ oplevert, $1$ als beide niet samenwerken. De combinatie van samenwerken en niet samenwerken levert $0$ op voor degene die samenwerkt en $4$ voor de ander. Stel dat dit spel zich nu met een kans $p = 0{,}5$ herhaalt en iemand gebruikt als strategie om samen te werken zolang de partner dat ook doet. Zodra de partner één keer niet samenwerkt, werkt hij nooit meer samen. Als deze persoon iemand met dezelfde strategie tegenkomt, is hun verwachte uitbetaling: $3 + 1{,}5 + 0{,}75 + \cdots = 6$. Merk op dat we hier ook aannemen dat de uitbetaling $0$ is als het spel ophoudt. Dus de verwachte uitkomst van het tweede spel

| Gevangenendilemma | | Gevangene 2 | |
|---|---|---|---|
| | | Ander aangeven | Zwijgen |
| Gevangene 1 | Ander aangeven | 3 jaar, 3 jaar | 0 jaar, 5 jaar |
| | Zwijgen | 5 jaar, 0 jaar | 1 jaar, 1 jaar |

**Tabel 1** Matrixweergave gevangenendilemma.

is $0{,}5 \cdot 0 + 0{,}5 \cdot 3 = 1{,}5$. Als deze persoon iemand tegenkomst die nooit samenwerkt is zijn uitbetaling $0 + 0{,}5 + 0{,}25 + \cdots = 1$, en voor de ander $4 + 0{,}5 + 0{,}25 + \cdots = 5$. Twee personen die nooit samenwerken hebben een verwachte uitbetaling van $1 + 0{,}5 + 0{,}25 + \cdots = 2$. Als we de proefpersonen zouden verplichten uit deze twee strategieën te kiezen, kunnen we het spel weer in een matrix als in Tabel 2 weergeven met in de cellen de verwachte uitkomsten. Uit dit voorbeeld blijkt dat als beide kiezen om samen te werken totdat ze teleurgesteld worden doordat de ander dat niet doet, dit ook een Nash-evenwicht is, waarbij samenwerking dus tot stand komt. Merk op dat nooit samenwerken ook een evenwicht blijft.

Standaard resultaten laten zien dat als de kans om te stoppen maar klein genoeg is (voor de limiet voor $p$ naar $0$), er Nash-evenwichten bestaan waarin de spelers altijd samenwerken. Overigens zijn deze resultaten veel algemener. Er zijn namelijk Nash-evenwichten te construeren voor elke mogelijk gemiddelde uitkomst over de spellen zolang beide spelers maar ten minste de uitbetaling krijgen die ze krijgen in het Nash-evenwicht van het eenmalige spel. Deze uitkomst kan gezien worden als een soort dreigpunt. Het is ook vrij makkelijk een strategie te beschrijven om dit voor elkaar te krijgen. Stel twee spelers zijn het eens het spel te spelen zodat ze gemiddeld ieder meer krijgen dan ze in het dreigpunt krijgen. En stel dat als iemand afwijkt van de afspraak de andere speler daarna dan altijd de Nash-evenwichtstrategie van het eenmalige spel speelt. De speler die als eerste afwijkt, heeft daarna geen betere optie meer dan ook maar naar het dreigpunt toe te gaan. Als hij zich echter aan de afspraak had gehouden, had hij meer kunnen verdienen dan in het dreigpunt. De algemene theorie die dit beschrijft wordt ook wel *Folk Theorem* genoemd. Dit omdat niemand zich meer kan herinneren wie het eigenlijk als eerste bedacht had [18].

Vanuit de sociologie weten we dat er nog meer mechanismen zijn die samenwerking bevorderen naast het herhaaldelijk omgaan met dezelfde mensen. Het sociale netwerk kan ook een belangrijke rol spelen in het bevorderen van samenwerking of vertrouwen. Hierin kan reputatie-informatie worden doorgespeeld. We kunnen in de samenleving afspreken wat we doen als we ontdekken dat iemand niet samenwerkt of vertrouwen misbruikt. Zo kunnen we afspreken dat dan niet alleen de directe partner, maar ook alle mensen die ervan horen in hun interacties met de persoon die zich misdroeg, zich gaan gedragen volgens het bovengenoemde dreigpunt. In dat geval wordt de straf voor het niet-coöperatieve gedrag alleen maar zwaarder. Originele analyses van Raub en Weesie [29] laten zien dat hoe beter de informatieverpreiding is in het netwerk hoe meer samenwerking er in een groep mogelijk is. Buskens en Weesie [10] stellen een nog gedetailleerder model voor over vertrouwen en passen daarop basale Markov-theorie toe. Zo laten ze zien dat niet alleen iemand die meer contacten heeft, maar ook iemand die meer contacten heeft die weer meer contacten hebben, makkelijker anderen kan vertrouwen.

Hoewel de speltheoretische resultaten hele mooie algemene resultaten laten zien, zijn de voorspellingen ook niet altijd even robuust. Bijvoorbeeld is het feit dat nooit iemand een foutje maakt cruciaal om samenwerking in stand te houden. Alles valt uit elkaar zo gauw er wel een foutje wordt gemaakt. Nu zijn er allerhande analytische uitbreidingen die hier ook weer rekening mee kunnen houden. Daarbij worden de aannames over hoe individuen zich gedragen in dergelijke situaties echter niet per se plausibeler. We kunnen er nu eenmaal niet van uitgaan dat iedereen zich als een volleerd speltheoreticus gedraagt. Daarom wordt voor deze toepassingen ook vaak gekozen voor modellen waarin de aannames over hoe individuen zich gedragen eenvoudiger zijn. De studie van Axelrod [3] is klassiek in dit opzicht. Deze laat zien dat mensen die de strategie gebruiken van 'oog om oog, tand om tand' ("ik doe tegenover jou vandaag, wat jij gisteren ten opzichte van mij deed") heel succesvol zijn in het spelen van herhaalde prisoner's dilemma's. Deze strategie leidt ook tot veel samenwerking als hij met andere coöperatieve strategieën wordt gecombineerd.

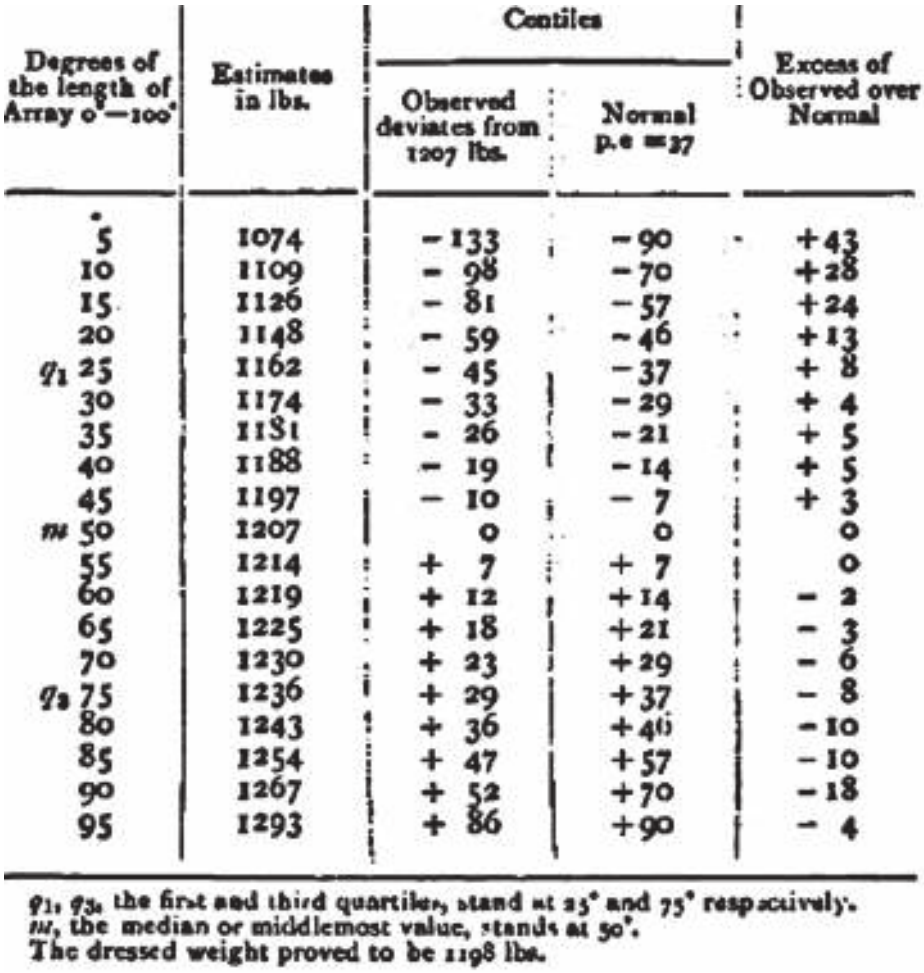
| Herhaald gevangenendilemma | | Speler 2 | |
|---|---|---|---|
| | | Nooit samenwerken | Samenwerken tot teleurstelling |
| Speler 1 | Nooit samenwerken | 2, 2 | 5, 1 |
| | Samenwerken tot teleurstelling | 1, 5 | 6, 6 |

**Tabel 2** Matrixweergave Herhaald gevangenendilemma.

Het type simulatiestudies zoals die van Axelrod, hebben een brede literatuur geïnspireerd over hoe mensen op netwerken dit soort spellen spelen. Iedereen speelt dan het gegeven spel met zijn buren. Een veel gebruikte aanname is dat actoren hun gedrag aanpassen als ze in het eenmalige spel meer kunnen verdienen gegeven het gedrag van al hun buren in de vorige ronde. De tweede auteur heeft aan verschillende van dit soort studies bijgedragen [8, 9]. Deze literatuur heeft wel al veel resultaten opgeleverd voor specifieke spellen op specifieke netwerken, maar meer algemene wiskundige resultaten zijn zeker nog welkom.

## Wijsheid van de massa

Iedereen heeft op een fancy fair wel eens meegedaan aan het gokken hoeveel snoepjes of knikkers er in een bepaalde vaas zitten. Maar zijn we eigenlijk wel goed in het gokken van zo een aantal? Galton [19] liet al in het begin van vorige eeuw zien dat we zelfs als we er als individu niet zo goed in zijn, we er als grote groep mensen wel goed in kunnen zijn. Galton liet bijna $800$ mensen het gewicht van een os inschatten. Er bestond grote variatie in inschattingen en veel mensen zaten er ver vanaf. Echter, de mediane inschatting, zoals Galton 'vox populi' definieerde, week minder dan $1\%$ af van het echte gewicht



| Degrees of the length of Array 0—100 | Estimates in lbs. | Centiles: Observed deviates from 1207 lbs. | Centiles: Normal p.e = 37 | Excess of Observed over Normal |
|---|---|---|---|---|
| 5 | 1074 | −133 | −90 | +43 |
| 10 | 1109 | − 98 | −70 | +28 |
| 15 | 1126 | − 81 | −57 | +24 |
| 20 | 1148 | − 59 | −46 | +13 |
| $q_1$ 25 | 1162 | − 45 | −37 | + 8 |
| 30 | 1174 | − 33 | −29 | + 4 |
| 35 | 1181 | − 26 | −21 | + 5 |
| 40 | 1188 | − 19 | −14 | + 5 |
| 45 | 1197 | − 10 | − 7 | + 3 |
| *m* 50 | 1207 | 0 | 0 | 0 |
| 55 | 1214 | + 7 | + 7 | 0 |
| 60 | 1219 | + 12 | +14 | − 2 |
| 65 | 1225 | + 18 | +21 | − 3 |
| 70 | 1230 | + 23 | +29 | − 6 |
| $q_3$ 75 | 1236 | + 29 | +37 | − 8 |
| 80 | 1243 | + 36 | +46 | −10 |
| 85 | 1254 | + 47 | +57 | −10 |
| 90 | 1267 | + 52 | +70 | −18 |
| 95 | 1293 | + 86 | +90 | − 4 |

$q_1$, $q_3$, the first and third quartiles, stand at 25° and 75° respectively.
*m*, the median or middlemost value, stands at 50°.
The dressed weight proved to be 1198 lbs.

**Figuur 10** Originele tabel uit het artikel van Galton [19].

van de os (zie Figuur 10). Dit fenomeen is bekend geworden als de wijsheid van de massa. Maar gaat het nou echt altijd op? En zo niet, wanneer niet en wanneer wel?

Stel dat de inschattingen van mensen symmetrisch verdeeld zijn rond de correcte uitkomst en mensen dus geen systematische fout maken. Zij over- of onderschatten de ware uitkomst gemiddeld niet. Vanuit statistische theorie moge dan duidelijk zijn dat het gemiddelde of de mediaan van veel onafhankelijke inschattingen dichterbij de ware uitkomst ligt dan veel van de individuele inschattingen. Zo stelde Condorcet in de achttiende eeuw al dat de meerderheidsstem in een grote groep onafhankelijke stemmers bijna altijd correct is, zolang de gemiddelde persoon het vaker juist dan mis heeft [13]. Er zijn echter twee aspecten mogelijk problematisch bij inschattingen die mensen maken in echte situaties. Ten eerste zijn er veel situaties waarin mensen zich systematisch vergissen. En dan zal het gemiddelde van de massa niet dichtbij de echte waarde komen. Het kan niettemin nog steeds zo zijn dat het gemiddelde van de massa beter is dan de meeste individuele keuzes.

Het tweede probleem is afhankelijkheid. Als mensen een keuze maken, kijken ze vaak eerst om zich heen als ze niet meteen kunnen bepalen wat de beste keuze is. Ze laten zich dan bij hun keuze beïnvloeden door de keuzes van anderen. Nu hebben verschillende studies [5, 16, 26] proefpersonen in experimenten iets laten schatten. Voorbeelden zijn het aantal ballen in een vaas of de lengte van de grens van Zwitserland. In een tweede ronde krijgen de proefpersonen in deze studies elkaars aanvankelijke antwoorden te zien. Ze mogen dan nog eens gokken. De studies laten zien dat sociale beïnvloeding nuttig kan zijn om tot goede keuzes te komen. Via sociale beïnvloeding worden op basis van de wijsheid van de massa uit de eerste ronde de keuzes in de tweede ronde in de juiste richting bijgesteld. Zo wordt het proces van tot een goed antwoord komen versneld. Becker e.a. [5] vonden zelfs dat niet alleen de variantie omlaag gaat van ronde 1 naar ronde 2, maar dat ook het gemiddelde dichter bij de juiste waarde komt te liggen. Dit gebeurt omdat zij die het minst accuraat zijn zich in het experiment meer aanpassen aan de meerderheidsopinie dan zij die meteen al warm zijn. Maar in een recent model laat de eerste auteur van dit artikel zien dat er wel een addertje onder het gras zit [31].

Het probleem is dat de opzet in de genoemde experimenten zo is dat iedereen eerst een onafhankelijke schatting maakt. De collectie antwoorden wordt vervolgens gedeeld met iedereen. Hiermee wordt het beïnvloedingsproces een goede start gegeven. De eerste inschattingen vormen een 'wijs anker' op basis waarvan iedereen zijn oorspronkelijke oordeel kan herzien. Echter, als een beïnvloedingsproces juist begint vanaf een punt dat nogal vér van de waarheid verwijderd is, dan kunnen groepskeuzes ook voor langere tijd verder van de waarheid af blijven. In een recente studie vergeleek de eerste auteur twee processen om bij een goed-foutvraag tot het goede antwoord te komen. Een voorbeeld van een goed-foutvraag is welk van twee schilderijen geschilderd is door Vincent van Gogh (zie Figuur 11). In proces 1 laten we in verschillende groepen iedereen onafhankelijk gokken en beschouwen het antwoord dat de meeste mensen in de groep hebben gekozen als het groepsoordeel. In proces 2 laten we de mensen in even grote groepen om de beurt kiezen en vertellen ze steeds hoeveel mensen tot dan toe al voor het ene of het andere schilderij hebben gekozen. Opnieuw beschouwen we het antwoord dat de meeste mensen in de groep geven als het groepsoordeel. In proces 2 is het aannemelijk dat mensen de informatie over de eerdere keuzes van anderen laten meewegen in hun keuze als ze onzeker zijn. Ze zijn dan meer geneigd zich te conformeren aan de eerdere keuzes. Meer mensen zullen dan het goede antwoord kiezen, en deze voorspelling wordt inderdaad bevestigd in het experiment van de eerste auteur. Dit lijkt dus te wijzen in de richting van een positief effect van sociale beïnvloeding op het goede antwoord. Wat blijkt echter, zowel in het model als in het experiment: De eindbeslissing van groepen in proces 2 is minder vaak correct dan in proces 1. Meer groepen komen onder sociale invloed op het verkeerde antwoord uit dan

**Figuur 11** Welk schilderij is van Vincent van Gogh?

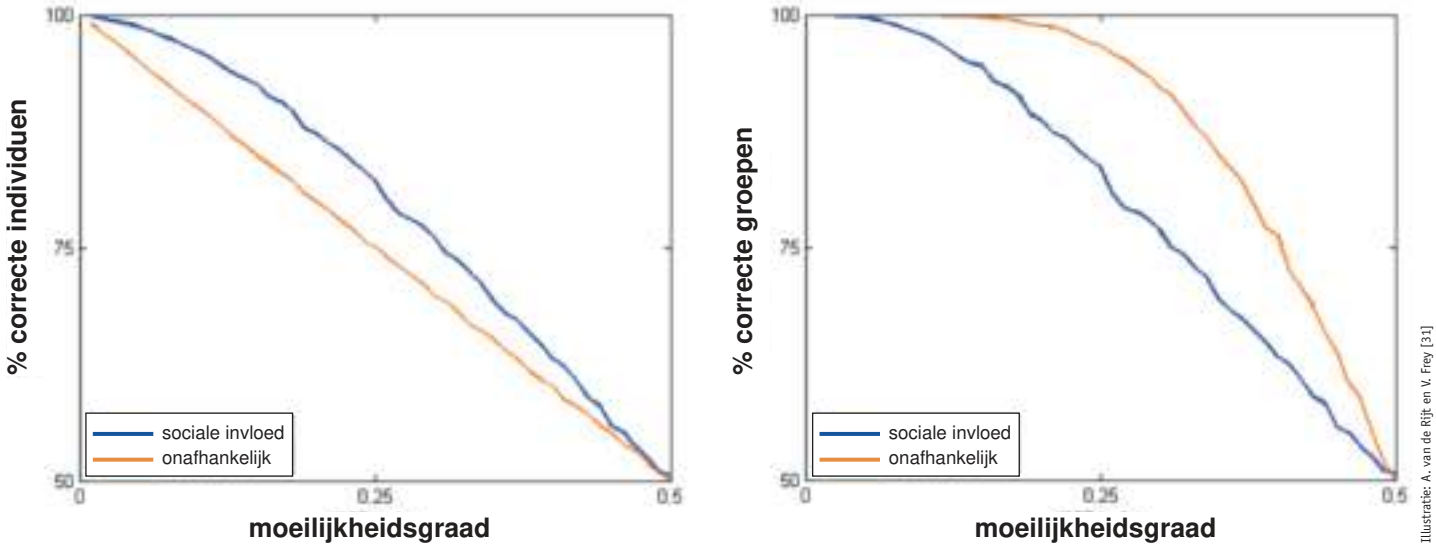
in het geval waarbij de groepsleden onafhankelijke keuze hebben gemaakt. Terwijl in het geval van onafhankelijkheid er in bijna alle groepen wel een meerderheid is die het goede antwoord kiest, worden er in het tweede geval twijfelaars door foute beginkeuzes naar de verkeerde kant getrokken waardoor meer groepen op het verkeerde antwoord komen. Er zijn dus wel individueel meer goede keuzes in het tweede geval, maar die zitten geconcentreerd in minder groepen. Dit modelresultaat is geïllustreerd in Figuur 12. In het model maken individuen $i = 1,...,N$ om de beurt een binaire keuze $C_i \in \{-1,1\}$, waar $1$ goed is en $-1$ verkeerd. $d \in \left[0,\frac{1}{2}\right]$ is de moeilijkheidsgraad van de vraag, en $s_i$ de beïnvloedbaarheid van $i$. Aangenomen wordt dat de kans dat een individu $i$ het juiste antwoord kiest een logistische functie is van de eerdere keuzes van $j < i$ met parameters $d$ en $s_i$:

$$\text{Prob}(C_i = 1) = \left(1 + \frac{d}{1-d} e^{-s_i \frac{\sum_{j<i} C_j}{i-1}}\right)^{-1}.$$

De twee panels laten respectievelijk het percentage correcte individuen (links) en het percentage correcte groepen (rechts) zien voor variabele moeilijkheidsgraden $d$. Aangenomen is dat groepen uit $12$ individuen bestaan en dat $s_i$ normaal verdeeld is met gemiddelde $3$ en standaarddeviatie $1$. Deze parameters komen goed overeen met schattingen uit logistische regressiemodellen van de data uit het laboratoriumexperiment. In het linker panel is te zien dat de mogelijkheid tot sociale beïnvloeding in proces 2 de individuele wijsheid verhoogt, ongeacht de moeilijkheidsgraad. In het rechter panel is daarentegen te zien dat voor alle moeilijkheidsgraden sociale beïnvloeding de wijsheid van groepen verlaagt. Schapen worden slimmer, kuddes worden dommer. Er is reeds levendige literatuur [1, 6] in economics and decision theory die aan dit soort scenario's rekent, met nog veel openstaande problemen voor geïnteresseerde wiskundigen.

## Afsluiting

Dit laatste voorbeeld is een mooi voorbeeld over hoe micro-uitkomsten en macro-uitkomsten uit elkaar kunnen lopen. Op individueel niveau willen we graag gebruikmaken van de informatie van anderen, maar voor de groep zou het beter zijn als iedereen onafhankelijk zou beslissen. Deze uitkomst was moeilijk te doorzien zonder een formeel model te ontwikkelen wat deze uitkomst genereert. Merk op dat we




Illustratie: A. van de Rijt en V. Frey [31]

**Figuur 12** Modelresultaat voor individuen en groepen voor de vraag welk schilderij van Van Gogh is.

daarna wel makkelijker kunnen verwoorden waarom deze uitkomst toch intuïtief is.

Duncan Watts gaf de pakkende titel aan zijn interessante boek over 'common sense': *Everything is Obvious (Once You Know the Answer)* [34]. Dit om aan te geven dat in de sociale wetenschappen veel verbanden aannemelijk te maken zijn, als je eenmaal weet dat ze er zijn. Met behulp van wiskundige formaliseringen kunnen we echter bewijzen hoe verbanden ook binnen de sociale wetenschappen volgen uit aannames die we maken over individueel gedrag. En we kunnen bewijzen hoe condities in de samenlevingen dit gedrag kunnen beïnvloeden. Zo een formalisering bewijst niet dat het verband echt zo tot stand komt. Dat kan alleen door het verband ook empirisch aan te tonen. De formaliseringen helpen wel om logisch correcte afleidingen te maken over waarom een bepaald verband zou kunnen bestaan. En als het empirisch verband inderdaad is aangetoond, helpt het wiskundige model ook om intuïtieve herformuleringen van het mechanisme te geven, die logisch consistent zijn. Die zijn ook door niet-wiskundigen te appreciëren.

De beschreven voorbeelden zijn slechts een kleine steekproef van materiaal uit de mathematische sociologie. Nog veel groter echter is de hoeveelheid interessante sociologische ideeën en inzichten op allerlei toepassingsterreinen die nog niet gemathematiseerd zijn. De auteurs reiken geïnteresseerde wiskundigen de hand.

**Referenties**


1 S. Alpern en B. Chen, Who should cast the casting vote? Using sequential voting to amalgamate information, *Theory and Decision* 83(2) (2017), 259–282.

2 T. Antal, P. L. Krapivsky en S. Redner, Dynamics of social balance on networks, *Physical Review E* 72(3) (2005), 036121.

3 R. Axelrod, *The Evolution of Cooperation*, Basic Books, 1984.

4 X. Ban, J. Gao en A. van de Rijt, Navigation in real-world complex networks through embedding in latent space, *ALENEX10*, SIAM, 2010, pp. 138–148.

5 J. Becker, D. Brackbill en D. Centola, Network dynamics of social influence in the wisdom of crowds, *Proceedings of the National Academy of Sciences* 114(26) (2017), E5070–E5076.

6 S. Bikhchandani, D. Hirshleifer en I. Welch, A theory of fads, fashion, custom, and cultural change as informational cascades, *Journal of Political Economy* 100(5) (1992), 992–1026.

7 P. Bonacich en P. Lu, *Introduction to Mathematical Sociology*, Princeton University Press, 2012.

8 J. Broere, V. Buskens, J. Weesie en H. Stoof, Network effects on coordination in asymmetric games, *Scientific Reports* 7(1) (2017), 17016.

9 V. Buskens en C. Snijders, Effects of network characteristics on reaching the payoff-dominant equilibrium in coordination games: a simulation study, *Dynamic Games and Applications* 6(4) (2016), 477–494.

10 V. Buskens en J. Weesie, Cooperation via social networks, *Analyse & Kritik* 22 (2000), 44–74.

11 D. Cartwright en F. Harary, Structural balance: a generalization of Heider's theory, *Psychological review* 63(5) (1956), 277.

12 J.S. Coleman, *Introduction to Mathematical Sociology*, Free Press of Glencoe, 1964.

13 M.J.A.N. Condorcet, *Essai sur l'Application de l'Analyse à la Probabilité des Décisions Rendues à la Pluralité des Voix*, Imprimerie Royale, Paris, 1785.

14 T. Fararo, *Mathematical Sociology*, Wiley, 1973.

15 T.J. Fararo, Reflections on mathematical sociology, in *Sociological Forum,* Vol. 12, No. 1, Kluwer Academic Publishers-Plenum Publishers, 1997. pp. 73–102.

16 D.P. Foster en H.P. Young, Stochastic evolutionary game dynamics, *Theoretical Population Biology* 38(2) (1990), 219–232.

17 N.E. Friedkin en F. Bullo, How truth wins in opinion dynamics along issue sequences, *Proceedings of the National Academy of Sciences* 114(43) (2017), 11380–11385.

18 J.W. Friedman, A non-cooperative equilibrium for supergames, *Review of Economic Studies* 38 (1971), 1–12.

19 F. Galton, Vox populi (the wisdom of crowds), *Nature* 75 (1907), 450–451.

20 F. Heider, Attitudes and cognitive organization, *Psychol.* 21 (1946), 107–112.

21 R. Kleinberg, Geographic routing using hyperbolic space, *Proceedings of the IEEE INFOCOM 2007 – 26th IEEE International Conference on Computer Communications*, IEEE, 2007, pp. 1902–1909

22 J.S. Kleinfeld, The small world problem, *Society* 39(2) (2002), 61–66.

23 M. Korzec en A. Verbeek, Mathematische sociologie, *Mens en Maatschappij* 53(3) (1978), 323–336.

24 M. Moore, An international application of Heider's balance theory, *European Journal of Social Psychology* 8(3) (1978), 401–405.

25 K. Nakamura, G. Tita en D. Krackhardt, Violence in the 'balance': a structural analysis of how rivals, allies, and third-parties shape inter-gang violence, *Global Crime* (2019).

26 J. Nash, Non-cooperative games, *Annals of Mathematics* 54 (1951), 286–295.

27 J. Lorenz, H. Rauhut, F. Schweitzer en D. Helbing, How social influence can undermine the wisdom of crowd effect, *PNAS* 108(22) (2011), 9020–9025.

28 J.A. Rambaran, J.K. Dijkstra, A. Munniksma en A.H. Cillessen, The development of adolescents' friendships and antipathies: A longitudinal multivariate network test of balance theory, *Social Networks* 43 (2015), 162–176.

29 W. Raub en J. Weesie, Reputation and efficiency in social interactions: An example of network effects, *American Journal of Sociology* 96 (1990), 626–654.

30 A. van de Rijt, The micro-macro link for the theory of structural balance, *The Journal of Mathematical Sociology* 35(1-3) (2011), 94–113.

31 A. van de Rijt en V. Frey, Social influence undermines the wisdom of the crowd in sequential decision-making, Working paper (2019),

32 A.B. Sorensen, Mathematical models in sociology, *Annual Review of Sociology* 4(1) (1978), 345–371.

33 J. Travers en S. Milgram, An experimental study of the small world problem, *Sociometry* 32 (1969), 425–443.

34 D.J. Watts, *Everything Is Obvious: Once You Know the Answer*, Crown Business, 2011.

35 D.J. Watts en S.H. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393(6684) (1998), 440–443.

Casper Albers
*Psychometrie & Statistiek*
*Rijksuniversiteit Groningen*
*c.j.albers@rug.nl*

# Psychologische netwerken – Een introductie

**De afgelopen tien jaar is het gebruik van netwerkmodellen voor psychologische datasets enorm toegenomen. In dit artikel legt Casper Albers de achtergrond van deze modellen uit.**

Meten is weten. In medisch onderzoek wordt veelvuldig gebruik gemaakt van meetinstrumenten om bijvoorbeeld de hartslag te meten of om te tellen hoeveel tumormarkers er in het bloed zitten. Hier is vaak ingewikkelde technologie voor nodig, maar de statistiek is redelijk eenvoudig: je meet wat je wilt meten, met daaroverheen een (doorgaans) normaalverdeelde meetfout.

In de klinische psychologie ligt dit anders. Ziektes als angststoornissen of depressies laten zich niet direct meten: er is geen apparaat dat, na analyse van een druppeltje bloed of urine, iemands depressiescore op een schaal van $0$ tot $100$ uitrekent. Bepalen of iemand klinisch depressief is, gebeurt door te kijken naar de aanwezigheid en ernst van symptomen.

De *Diagnostic and Statistical Manual of Mental Disorders* (DSM-V) geeft de richtlijnen weer waarop psychische aandoeningen gescoord worden. Zo spreekt men (kort door de bocht samengevat) van klinische depressie wanneer minstens vijf van de volgende negen symptomen met grote regelmaat voorkomen: (1) depressieve gedachten, (2) verminderd plezier in de meeste activiteiten, (3) significante verandering in eetlust of gewicht, (4) slapeloosheid of juist overmatige vermoeidheid, (5) rusteloosheid of andere activiteitsproblemen, (6) gebrek aan energie, (7) gevoel van waardeloosheid, (8) concentratieproblemen, en (9) suïcidale gedachten.

**Momentary assessment**

Veel psychologische meetgegevens worden verzameld via vragenlijsten. Door technologische vooruitgang is het mogelijk om dergelijke vragenlijsten met regelmaat af te nemen, bijvoorbeeld met een app op de smartphone. Deze manier van dataverzameling wordt *ecological momentary assessment* [25] genoemd: doordat je 'live' aan iemand kan vragen hoe het op dat moment gaat, is de ecologische validiteit van de meting hoger dan wanneer je achteraf een mening over bijvoorbeeld de afgelopen week vraagt. Niet alleen is de ecologische validiteit hoger: je krijgt ook meer inzicht in iemands gemoedstoestand wanneer je deze persoon over een bepaalde periode monitort. Niemands emotionele gesteldheid is namelijk constant: je zit het ene moment wat beter in je vel dan het andere. Ook volgens de spelregels van de

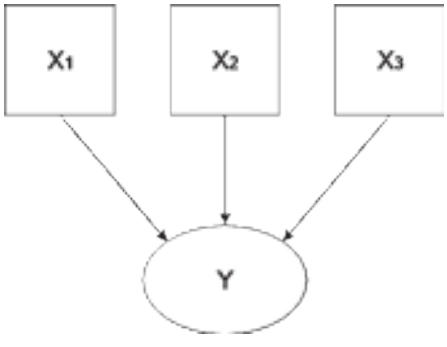
DSM-V moet je minimaal gedurende een periode van twee weken vijf van de bovengenoemde negen symptomen hebben; een enkele slapeloze nacht is nog geen depressie.

Om goed de processen die ten grondslag aan menselijk gedrag liggen te begrijpen, is het dus van vitaal belang om dezelfde variabelen bij dezelfde personen met grote regelmaat te meten. Dit is nodig om inzicht te verkrijgen in de gecompliceerde aspecten van menselijk gedrag. Die complexiteit uit zich door fluctuaties in gedrag over de tijd. Deze fluctuaties hangen af van de context, van inter-individuele verschillen, en van toevallige verstoringen. Het begrijpen van de dynamiek van een psychologisch proces is een essentiële voorwaarde om het proces zelf te begrijpen [1].

**De klassieke aanpak**

Traditioneel worden dergelijke zaken aangepakt met een latente variabelen-model [21]. Een latente variabele is een variabele die je niet zelf kan meten, maar waarvan je de waarde voorspelt aan de hand van observeerbare variabelen. Een voorbeeld is gegeven in Figuur 1. Op basis van de gemeten uitkomsten $x_1,\ldots,x_k$ wordt de latente variabele $y$ samengesteld als gewogen som van de symptomen: $y = \sum a_i x_i$. Hierbij wordt gebruikgemaakt van factoranalyse.

Recentelijk is men afgestapt van deze gedachtenlijn. Pogingen binnen de psychopathologie om psychische aandoeningen terug te brengen tot eendimensionale (biologische) constructen, zijn mislukt [18]. De nieuwe gedachtenlijn is dat we bij psychische stoornissen de symptomen niet moeten zien als *gevolg* van de aandoening, maar als de aandoening zelf: die structurele slapeloosheid, gebrek aan eetlust en terugkerende suïcidale gedachten, dat *is* de depressie. Er is niet een of andere gehypothetiseerde latente variabele 'depressie', de depressie dat is wat je waarneemt.



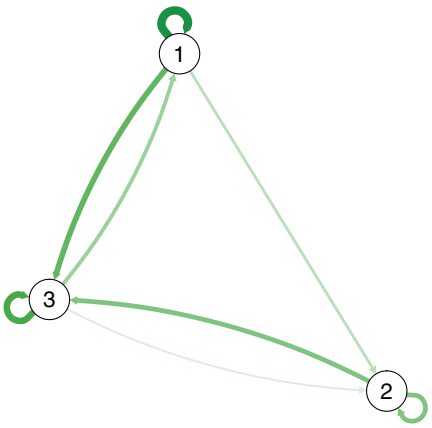

**Figuur 1** Voorbeeld van een latente variabelen-diagram met drie geobserveerde en een latente variabele.

**Psychologische netwerktheorie**

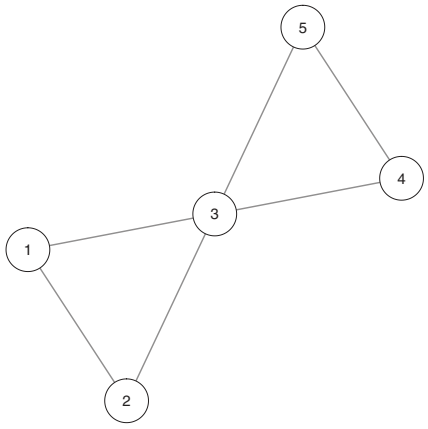
Deze gedachtenlijn, ingebracht door een groep psychometrici van de Universiteit van Amsterdam [5,7], leidt op een natuurlijke manier tot de netwerkmodellen zoals die nu binnen de psychopathologie en klinische psychologie gebruikt worden. In Figuur 2 wordt zo'n netwerk, gebaseerd op drie variabelen, weergegeven. Afhankelijk van wat voor type data verzameld is — meerdere personen op een tijdstip, of een persoon over meerdere tijdstippen — kunnen de zijden van het netwerk een richting hebben of niet. Figuur 2 maakt gebruik van zijden met een richting; hoe dikker de zijde, hoe groter het effect. Een hoge score op variabele $1$ op het ene tijdstip correleert dus flink met een hoge score op diezelfde variabele het volgende tijdstip, maar slechts licht met een hoge score op variabele $2$ op het volgende tijdstip. Zouden niet meerdere tijdspunten maar meerdere personen (eenmalig) gemeten zijn, dan geeft de dikte van de zijden aan hoe sterk variabelen met elkaar correleren tussen personen.

De steeds groter wordende onderzoeksgroepen en datasets hebben ertoe geleid dat klinisch psychologen meerdere aandoeningen simultaan willen onderzoeken. Doorgaans staan aandoeningen niet compleet los van elkaar, maar zijn er overlappende symptomen. In Figuur 3 worden twee aandoeningen beschouwd: de ene aandoening bestaat uit variabelen $1$, $2$ en $3$; de ander uit $3$, $4$ en $5$. Variabele $3$ slaat een brug tussen beide aandoeningen en wordt derhalve een *bridge variable* genoemd.

De brugvariabelen tussen klinische depressie en een angststoornis zijn de symptomen slapeloosheid, vermoeidheid en concentratieproblemen [5]. Dit is relevant bij de behandeling van patiënten met ofwel depressie ofwel angststoornissen: het is een bekend gegeven uit de literatuur dat de behandeling van meerdere stoornissen tegelijk — comorbiditeit geheten — veel ingewikkelder is dan de behandeling van een enkele stoornis. Het geheel is meer dan de som der delen, wat hier dus een nadelig effect heeft. De behandeling van een psychische stoornis gebeurt vaak door het tegengaan van een of meerdere symptomen, en in dit geval




**Figuur 2** Voorbeeld van een netwerkdiagram met drie variabelen.




**Figuur 3** Voorbeeld van een 'bridge variable', variabele 3.

wordt er voorrang verleend aan het tegengaan van slapeloosheid, vermoeidheid en concentratieproblemen, omdat dit de brugvariabelen zijn.

Brugvariabelen zijn bij niet-psychologische netwerken vaak makkelijker te conceptualiseren. Denk bijvoorbeeld aan het netwerk van NS treinstations. Al het verkeer vanuit Groningen, Friesland en Drenthe naar bijvoorbeeld de Randstad moet via station Zwolle: er is geen andere route. Zwolle is daarmee de brugvariabele tussen het westen en het noorden. Wanneer er problemen zijn op station Zwolle, heeft dat direct invloed op al het treinverkeer van het westen naar het noorden en vice versa. Andere stations vervullen die rol veel minder. Als er vanwege spoorproblemen geen treinen langs Zoetermeer kunnen, is dat uiterst vervelend voor degenen met Zoetermeer als beginpunt of eindpunt van de reis. Maar reizigers van bijvoorbeeld Den Haag naar Utrecht kunnen met minieme vertraging omrijden via Leiden. Iets soortgelijks speelt dus bij de symptomen van psychologische aandoeningen.

### Het Gaussische Grafische Model (GGM)

Voor een groot deel is de psychologische netwerktheorie nieuwe wijn in oude zakken. Het onderliggende netwerkmodel is namelijk niet nieuw; het is de methodologische manier van denken over psychische aandoeningen die nieuw is.

Het meestgebruikte netwerkmodel binnen de psychologische toepassingen is het Gaussische Grafische Model (GGM). Dit model werd in de jaren zeventig geïntroduceerd [8] en sindsdien veelvuldig toegepast. Het boek *Graphical Models* van de Deense statisticus Steffen Lauritzen [20] geldt als hét naslagwerk voor GGMs. Binnen de sociale wetenschappen, worden GGMs, en andere netwerkmodellen, al enkele decennia succesvol toegepast bij sociale netwerkanalyse [9, 24], maar dus pas sinds een jaar of tien voor psychologische netwerken.

Het standaard GGM, zonder tijdsdimensie, werkt als volgt. Gegeven is een $n \times p$ datamatrix $X$: $p$ variabelen zijn $n$ keer gemeten. De eerste letter G uit het GGM verwijst naar de aanname van normaliteit van de data:

$$X \sim N(\mu, \Sigma).$$

Neem als voorbeeld een dataset waarbij van $n = 88$ studenten de cijfers op vijf wiskundevakken gemeten is [23]. Doel van een netwerkmodel is om verbanden, correlaties dus, te visualiseren. Tabel 1 laat de Pearson product-momentcorrelaties tussen de vijf vakken zien. Alle vakken hebben een positieve samenhang, met correlaties van $0{,}39$ tot $0{,}71$. In Tabel 1 zijn er $5 \times 4/2 = 10$ unieke correlaties en dit is nog enigszins overzichtelijk. Wanneer je met grotere datasets werkt, stijgt dit aantal snel: bij $p$ variabelen heb je $p(p-1)/2$ correlaties (en, als er een tijdsdimensie aanwezig is zelfs $p^2$ unieke variabelen) en dan is niet meer (makkelijk) te zien waar de sterke verbanden zitten. De meerwaarde van de GGM ligt met name in het visualiseren van de verbanden.

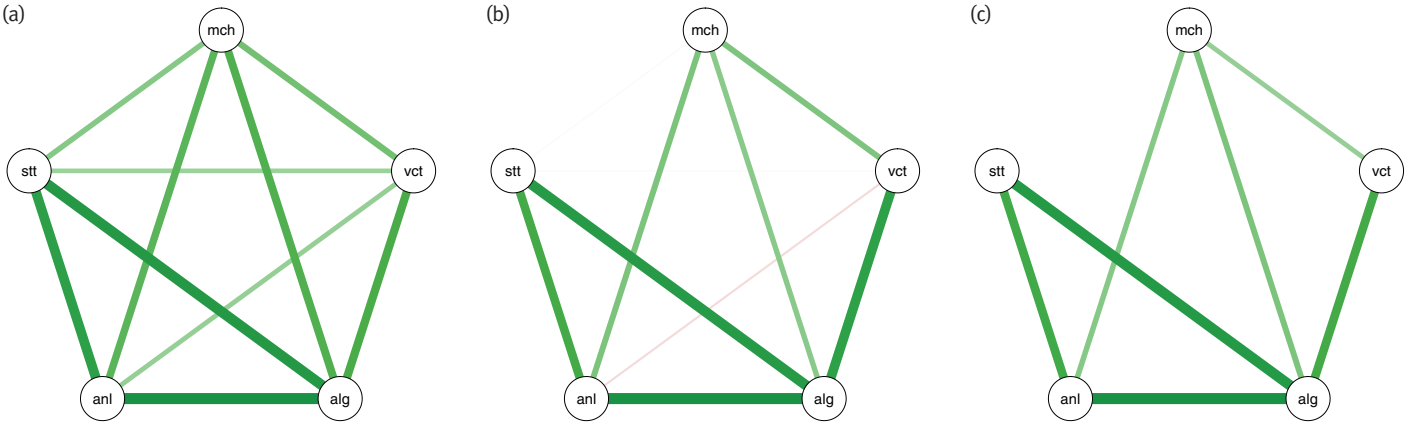
| | mec | vec | alg | ana | sta |
|---|---|---|---|---|---|
| **mechanica** | 1,00 | 0,55 | 0,55 | 0,41 | 0,39 |
| **vectoralgebra** | 0,55 | 1,00 | 0,61 | 0,48 | 0,44 |
| **algebra** | 0,55 | 0,61 | 1,00 | 0,71 | 0,66 |
| **analyse** | 0,41 | 0,48 | 0,71 | 1,00 | 0,61 |
| **statistiek** | 0,39 | 0,44 | 0,66 | 0,61 | 1.00 |

**Tabel 1** Correlaties tussen de cijfers van $n = 88$ studenten op vijf wiskundevakken [23].

Figuur 4(a) laat zo'n visualisatie zien. Elke variabele wordt weergegeven door een node (een cirkel) en de correlatie tussen twee variabelen, een zijde (lijnstuk tussen beide variabelen): een groene zijde voor een positief verband en een rode bij eventuele negatieve verbanden. De dikte van de zijde is proportioneel met de correlatie. De geschatte correlatiematrix, $\hat{\Sigma}$, geeft dus direct de informatie die nodig is voor de visualisatie. In een oogopslag is te zien dat de vakken statistiek, analyse en algebra onderling de sterkste samenhang hebben.

Nadeel van Figuur 4(a) is dat elke variabele wel samenhang vertoont met elke andere. Dit is bij psychologische datasets niet anders. Dat bij de symptomen van klinische depressie slapeloosheid, gebrek aan energie en concentratieproblemen sterk samenhangen is evident. Het is informatiever om de *partiële* correlaties, gegeven door $\hat{\Sigma}^{-1}$ te visualiseren; zie Figuur 4(b). Een groot voordeel van het werken met partiële correlaties bij normaalverdeelde data is dat de grafiek nu conditionele onafhankelijkheid laat zien. Zo is de partiële correlatie tussen statistiek en vectoralgebra (nagenoeg) nul, wat impliceert dat de scores op deze vakken (nagenoeg) onafhankelijk zijn, conditioneel op de andere vakken. Oftewel: als je van een student de cijfers op analyse, algebra en mechanica weet, biedt het geen meerwaarde meer om het statistiekcijfer te kennen wanneer je het vectoralgebra-cijfer wilt voorspellen.

Het is wenselijk om het aantal zichtbare zijdes nog verder terug te brengen: als een (partiële) correlatie te dicht bij nul ligt, is deze toch niet praktisch relevant. Dit terugbrengen kan op twee manieren: via *thresholding* en via de *graphical lasso*-methode. Bij thresholding zet je simpelweg alle correlaties waarvoor geldt dat $|r| < \theta$, met $\theta$ bijvoorbeeld $0{,}2$, gelijk aan nul. Bij de graphical lasso-methode (GLASSO [16]) wordt de partiële correlatiematrix geschat onder $L_1$-penalties. De waarden dicht bij nul worden daardoor richting nul getrokken, en de overige waarden worden geschat conditioneel op die nulwaarden. Met een tuning-parameter kan je instellen hoe




**Figuur 4** Netwerken gebaseerd op de gegevens van Tabel 1: (a) Correlatie-netwerk, (b) Partiële correlatie-netwerk, (c) GLASSO-netwerk.

sterk de $L_1$-penalty moet optreden. Zoals te zien in Figuur 4(c) levert dit een duidelijkere visualisatie op.

Gerelateerd aan het concept van brugvariabelen is centraliteit. Met centraliteitsmethoden wordt gekeken welke variabelen het belangrijkst zijn. De meestgebruikte centraliteitsmaat is die van 'node strength', gedefinieerd als de som van alle absolute waarden van de zijdes van/naar een node. In Figuur 4(c) heeft algebra de hoogste en mechanica de laagste centraliteit. Hoewel gebruikt binnen de psychologie, zijn er kritische stemmen over de interpretatie van centraliteit in een klinisch psychologische context [6].

**Netwerken voor tijdreeksdata**

In de vorige paragraaf ging het over netwerken voor data met per persoon eenmalig een meting. Het is voor psychologisch onderzoek, zoals omschreven onder het kopje 'Momentary assessment', juist interessant om iemand meerdere keren over de tijd te meten. In principe krijg je zo nog steeds een datamatrix $X$ van afmeting $n \times p$, alleen bestaat de $i$-de rij nu niet uit de data van persoon $i$ maar van tijdstip $i$. Via een tijdreeksmodel is dan te beschouwen hoe de metingen op tijdstip $i$ afhangen van die op tijdstip $i-1$. Hiervoor wordt meestal het vector autoregressieve model van de eerste orde (VAR(1)) gebruikt:

$$X_i = BX_{i-1} + \varepsilon_i.$$

Hierin is $B$ een matrix van regressiegewichten waarbij $B_{ab}$ aangeeft hoe de meting van variabele $a$ afhangt van de waarde van variabele $b$ op het vorige moment. Om de link met conditionele onafhankelijkheid te bewaren, wordt van de residuen $\varepsilon$ aangenomen dat ze onderling onafhankelijk zijn en een $N(0, \text{x}^{-1})$-verdeling volgen.

Net als bij het niet-tijdsafhankelijke GGM is het hier gebruikelijk om via $L_1$-penalties het aantal zijden laag te houden. Het meestgebruikte algoritme hiervoor [26] zit ingebouwd in de standaard packages in R voor dit soort analyses [10, 11]. Figuur 2 is een voorbeeld van het soort visualisatie dat dit oplevert.

**'Klassiek' versus psychologisch**

Zoals gezegd zijn psychologische netwerken sterk gebaseerd op andere netwerkmodellen. Ze hebben echter ook een aantal methodologische eigenschappen die ze uniek maken.

1. Psychologische netwerken zijn netwerken van *variabelen*, bij andere modellen zijn de nodes 'dingen'. Bij het eerder genoemde NS-netwerk waren de nodes treinstations: deze zijn statisch. Ook bij sociale netwerken zijn de nodes statische entiteiten, bijvoorbeeld personen.
2. Die modellen kunnen worden uitgebreid door bijvoorbeeld op een bepaald moment een persoon toe te voegen of weg te halen uit het model; maar bij psychologische netwerken zijn de nodes — slapeloosheid, concentratieproblemen, et cetera — per definitie variabel. De nodes in een psychologisch netwerk zijn dus toevalsvariabelen, in andere modellen is dit niet zo.
3. De zijden in een niet-psychologisch netwerk kunnen geobserveerd worden: de treinstations waartussen een trein rijdt, zijn verbonden; de personen die elkaars Facebook-vrienden zijn, zijn verbonden, enzovoort. Bij psychologische netwerken dienen de zijden geschat te worden en bevatten dus statistische onzekerheden. Denk bijvoorbeeld aan de correlatie tussen vakken: bij een andere steekproef zouden we andere correlaties gemeten hebben.
4. Bij psychologische netwerken is het hoofddoel om te laten zien welke nodes tegelijk optreden. Bij niet-psychologische netwerken is het hoofddoel om de aanwezigheid/afwezigheid van zijden te bepalen, de meest centrale node te vinden, enzovoort.

**Schaalvalidatie**

Psychologische netwerken kunnen niet alleen gebruikt worden als vervanger van de latente variabelen-modellen; maar ook als vervanger van de schaalvalidatie-technieken uit de klassieke psychometrie. Psychologische vragenlijsten bestaan doorgaans uit meerdere vragen die hetzelfde concept meten. Via instrumenten als Cronbachs alfa is vervolgens te meten of er niet een vraag is die duidelijk afwijkt van de rest.

Dit kan ook via een netwerkvisualisatie. In [3] laten we zien hoe dit moet. Een uitgebreide vragenlijst met 65 vragen in 22 domeinen is afgenomen onder 694 deelnemers. Een netwerkvisualisatie van alle 65 vragen laat duidelijk zien dat de vragen uit hetzelfde domein dichter bij elkaar liggen dan bij vragen uit andere domeinen: aan een van de voorwaarden voor een valide meetschaal is daarmee voldaan.

**Vergelijken/samenvoegen van netwerken**

Psychologische netwerken worden, zeker bij *single case designs*, vaak vooral gebruikt voor exploratieve doeleinden. Het is echter goed mogelijk om ook confirmatieve conclusies te trekken. Via bootstrapmethoden kan worden gekeken of bepaalde zijden significant afwijken van nul of van andere zijden en welke nodes een significante centraliteitsmaat hebben [12]. Soms wil men ook verschillende netwerken met elkaar vergelijken: bijvoorbeeld de netwerken van een groep personen gediagnosticeerd met klinische depressie en die van een groep zonder die diagnose. Zijn de netwerken daadwerkelijk verschillend of zijn de visuele verschillen het resultaat van toeval? Voor zulke vergelijkingen is de Network Comparison Test [4] geschikt. Deze test is gebaseerd op de Manteltest [22] om een tweetal correlatiematrices te vergelijken.

Men kan ook proberen de netwerken van verschillende personen samen te voegen. Een manier om dit te doen is met behulp van multilevel-modellen, ook bekend als hiërarchische modellen [13]. Bij dergelijke modellen is het model voor individu $k$ gegeven door

$$X_t^{(i)} = B^{(k)} X_{i-1} + \varepsilon_i^{(k)}$$

met

$$B^{(i)} \sim N(B, \Sigma_B) \quad \text{en} \quad \varepsilon_i^{(k)} \sim N(0, \Sigma_\varepsilon).$$

Een andere aanpak is het gebruik van clusteringtechnieken. Hierbij worden personen met gelijksoortige karakteristieken in eenzelfde cluster gezet. Dit kan of in een tweestapsmodel, waarbij eerst de individuele karakteristieken worden bepaald en deze daarna worden geclusterd [19]. Het kan echter ook in een geïntegreerd model waarbij beide stappen simultaan gezet worden [14].

**Verder lezen**

De afgelopen pagina's heb ik geprobeerd een korte introductie te geven van de psychologische netwerktheorie. De geïnteresseerde lezer kan uit een breed scala aan literatuur kiezen om meer te lezen. Het boek van Lauritzen [20] bevat alle (wiskundige) achtergrond voor GGM's. Psychologische netwerken worden uitgelegd in verschillende artikelen [3, 12, 17] en websites [15].

## Referenties


1 Albers, C.J. (2019), Toegepaste statistiek: van ontwikkeling tot communicatie, Oratie, Rijksuniversiteit Groningen, *Nieuw Archief voor Wiskunde* 5/20(2), 94–100.

2 American Psychiatric Association (2013), *Diagnostic and Statistical Manual of Mental Disorders* (vijfde editie).

3 Bhushan, N., Mohnert, F., Sloot, D., Jans, L., Albers, C.J. en Steg, L. (2019), Using a Gaussian Graphical Model to Explore Relationships Between Items and Variables in Environmental Psychology Research, *Frontiers in Psychology* 10, 1050.

4 Borkulo, C.D. van, Waldorp, L.J., Boschloo, L., Kossakowski, J., Tio, P., Schoevers, R. en Borsboom, D. (2016), Comparing network structures on three aspects: A permutation test, Preprint op ResearchGate.

5 Borsboom, D. en Cramer, A.O.J. (2013), Network analysis: an integrative approach to the structure of psychopathology, *Annual Review of Clinical Psychology* 9, 91–121.

6 Bringmann, L.F., Elmer, T., Epskamp, S., Krause, R.W., Schoch, D., Wichers, M., Wigman, J. en Snippe, E. (2018), What do centrality measures measure in psychological networks? Te verschijnen in *Journal of Abnormal Psychology*.

7 Cramer, A.O.J., Waldorp, L., Maas, H. van der en Borsboom, D. (2010), Comorbidity: A network perspective, *Behavioral and Brain Sciences* 33(2–3), 137–150.

8 Dempster, A.P. (1972), Covariance selection, *Biometrics* 28(1), 157–175.

9 Duijn, M.A. van en Vermunt, J.K. (2006), What is special about social network analysis? *Methodology* 2(1), 2–6.

10 Epskamp, S., Cramer, A.O.J., Waldorp, L.J., Schmittmann, V.D. en Borsboom, D. (2012), qgraph: Network visualizations of relationships in psychometric data, *Journal of Statistical Software* 48(4), 1–18.

11 Epskamp, S. (2018), graphicalVAR: Graphical VAR for experience sampling data, https://CRAN.R-project.org/package=graphicalVAR.

12 Epskamp, S., Borsboom, D. en Fried, E. (2018), Estimating psychological networks and their accuracy: a tutorial paper, *Behavioural Research Methods* 50(1), 195–212.

13 Epskamp, S., Waldorp, L.J., Mõttus, R. en Borsboom, D. (2018), The Gaussian graphical model in cross-sectional and time-series data, *Multivariate Behavioral Research* 53(4), 453–480.

14 Ernst, A.F., Timmerman, M.E., Jeronimus, B. en Albers, C.J. (2019), Inter-individual differences in multivariate time series: dynamic adaptive cluster modelling based on finite mixtures of vector-autoregressive processes. Ingestuurd voor publicatie.

15 Fried, E.I. (2019), Psych networks website, https://psych-networks.com/tutorials.

16 Friedman, J., Hastie, T. en Tibshirani, R. (2008), Sparse inverse covariance estimation with the graphical lasso, *Biostatistics* 9(3), 432–441.

17 Jones, P.J., Mair, P. en McNally, R.J. (2018), Visualizing psychological networks: a tutorial in R, *Frontiers in Psychology* 10, 3389.

18 Kendler, K.S. (2005), Toward a philosophical structure for psychiatry, *American Journal of Psychiatry* 162, 433–440.

19 Krone, T., Albers, C.J., Kuppens, P. en Timmerman, M.E. (2017), A multivariate statistical model for emotion dynamics, *Emotion* 18(5), 739–754.

20 Lauritzen, S.L. (1996), *Graphical Models*, Clarendon Press.

21 Loehlin, J.C. (2004), *Latent Variable Models. An Introduction to Factor, Path and Structural Equation Analysis* (vierde editie), Psychology Press.

22 Mantel, N. (1967), The detection of disease clustering and a generalized regression approach, *Cancer Research* 27(2).

23 Mardia, K.V., Kent, J.T. en Bibby, J.M. (1979), *Multivariate Analysis*, Academic Press.

24 Scott, J. (1988), Social network analysis, *Sociology* 22(1), 109–127.

25 Shiffman, S., Stone, A.A. en Hufford, M.R. (2008), Ecological momentary assessment, *Annual Review of Clinical Psychology* 4, 1–32.

26 Wild, B., Eichler, M., Friederich, H.C., Hartmann, M., Zipfel, S. en Herzog, W. (2010), A graphical vector autoregressive modelling approach to the analysis of electronic diary data, *BMC Medical Research Methodology*, 10(1), 28.

Denny Borsboom
*Faculteit der Maatschappij- en Gedragswetenschappen*
*Universiteit van Amsterdam*
*d.borsboom@uva.nl*

Maarten Marsman
*Faculteit der Maatschappij- en Gedragswetenschappen*
*Universiteit van Amsterdam*
*m.marsman@uva.nl*

# Latente variabelen en netwerken: verschillende benaderingen van psychometrische data

**Psychometrische gegevens spelen een steeds belangrijker rol in onze maatschappij. Van IQ-score tot Citotoets en van persoonlijkheidstest tot psychodiagnostiek; iedereen krijgt ergens in zijn of haar leven wel eens met een psychologische test te maken. De statistische analyse van testscores, die op deze manier verzameld worden, is het domein van de psychometrie. De centrale taakstelling van de psychometrie is het ontwikkelen van modellen die hiervoor optimaal geschikt zijn. In dit artikel geven Denny Borsboom en Maarten Marsman een overzicht van de belangrijkste ideeën en wiskundige representaties die geassocieerd zijn met netwerkbenaderingen van de psychometrie, en zij illustreren de hieruit voortkomende technieken met empirische data. Daarnaast bespreken zij openstaande kwesties en te verwachten ontwikkelingen.**

Hoe men de psychometrische analyse van testgegevens aanpakt, en welk model men daarvoor kiest, hangt deels af van de vraag hoe de relatie tussen die gegevens en de eigenschap die men wil bestuderen in elkaar steekt. Daar komen naast statistische ook inhoudelijke en wetenschapsfilosofische vragen bij kijken. Wie denkt dat depressie een meetbare eigenschap is zal de data anders analyseren dan iemand die denkt dat het maar een naam is voor een groepje symptomen. In dit stuk zullen we twee manieren bespreken om over deze kwestie na te denken. In de klassieke opvatting zijn psychometrische gegevens (bijvoorbeeld IQ-scores, symptomen van depressie, observaties van menselijk gedrag in verschillende situaties) metingen van een 'onderliggende' eigenschap (intelligentie, depressie, persoonlijkheid). Gedurende de afgelopen vijftien jaar is daar ook een nieuwe opvatting tegenover geplaatst, namelijk dat de in psychologische tests uitgevraagde eigenschappen beter gezien kunnen worden als complexe systemen waarin allerlei verschillende gevoelens, gedachten, en gedragingen interacteren. Deze gedachte heeft aanleiding gegeven tot het ontstaan van de netwerkbenadering van de psychometrie.

## Latente variabelen

In de psychometrie was lange tijd het belangrijkste statistische model het latente-variabelenmodel. Dit model is gebaseerd op de constatering dat we de eigenschappen waarin we in de psychologie geïnteresseerd zijn (zoals bijvoorbeeld depressie) niet direct kunnen waarnemen. Voor het bepalen van de mate waarin iemand depressief is zijn we aangewezen op variabelen die wel min of meer direct waarneembaar zijn; bijvoorbeeld of iemand slaapgebrek heeft, zich niet kan concentreren, zegt zich neerslachtig te voelen, of zelfs geprobeerd heeft zich te suïcideren. In de psychometrie wordt dit type variabelen gezien als *manifest*, terwijl de theoretische eigenschap die men via zulke variabelen wil bepalen als *latent* [3, 8, 15].

De hoofdstrategie in de psychometrie was gedurende de twintigste eeuw om een wiskundig model aan te leggen dat het functionele verband tussen manifeste en latente variabelen specificeert, en met dit model manieren te vinden om op basis van de waargenomen scores op manifeste variabelen tot een schatting van iemands waarde op de latente variabele te komen. Zo'n schatting kan worden gezien als een (indirecte) meting van depressie, en zodoende wordt het latente-variabelenmodel vaak een *meetmodel* genoemd. (Het idee dat via dit type modellen inderdaad — kwantitatieve — metingen kunnen worden gerealiseerd is overigens niet onomstreden. Zie Michell [28,29] voor een aanval op dit idee, en Borsboom [6] voor een verdediging ervan.) Wiskundig gezien komt dit latente-variabelenmodel neer op een factorisatie van de gemeenschappelijke kansverdeling van de manifeste variabelen. In het geval van een aantal binaire manifeste variabelen en een continue latente variabele is deze factorisatie bijvoorbeeld van de vorm

$$P(\mathbf{x}) = \int_{-\infty}^{\infty} P(\mathbf{x} \mid \theta) f(\theta)\, d\theta,$$

waarin $\mathbf{x}$ een vector van realisaties is van een set binaire random variabelen $\mathbf{X}$, $\theta$ een continue latente variabele weergeeft, $P(\mathbf{x} \mid \theta)$ de kansfunctie geeft die de

kans op de waargenomen data $\mathbf{x}$ aan $\theta$ relateert, en $f(\theta)$ de dichtheid van de latente variabele weergeeft. De functie $f(\theta)$ representeert de verdeling van de gemeten latente variabele in de populatie. De functie $P(\mathbf{x} \mid \theta)$ kent voor iedere waarde van de gemeten eigenschap $\theta$ een specifieke kans toe aan alle mogelijke responspatronen $\mathbf{x}$. In het geval van depressie zou deze functie bijvoorbeeld voor ieder niveau van depressie de kans bepalen dat iemand de symptomen neerslachtigheid, vermoeidheid, en concentratieproblemen al dan niet zou hebben.

De vorm van de kansfunctie $P(\mathbf{x} \mid \theta)$ wordt bepaald door de zogenaamde Item Respons Functie (IRF). De IRF kan in principe alle vormen aannemen [27], maar meestal wordt de voorkeur gegeven aan een logistische functie. Deze functie specificeert bijvoorbeeld dat de kans dat persoon $i$ met waarde $\theta_i$ op de latente variabele een item goed maakt (in het geval van vaardighedentests), het betreffende gedrag laat zien (in het geval van persoonlijkheidstest), of het symptoom in kwestie heeft (in het geval van psychopathologie), de volgende vorm heeft:

$$P(X_{ij} = 1 | \theta_i) = \frac{e^{\alpha_j(\theta_i - \beta_j)}}{1 + e^{\alpha_j(\theta_i - \beta_j)}}.$$

In deze kansfunctie wordt de kans dat de manifeste variabele $X_{ij}$ (de respons van persoon $j$ op item $i$) de waarde '1' aanneemt dus gemodelleerd als een functie van een latente eigenschap van de persoon ($\theta_i$) en twee eigenschappen van het item (een interviewvraag, opgave, of vragenlijst-item): een locatieparameter $\beta_j$ en een discriminatieparameter $\alpha_j$. De locatieparameter geeft aan hoe hoog de waarde op de latente trek moet zijn om een kans van $\frac{1}{2}$ op een waarde 1 te genereren. Deze wordt bij het meten van cognitieve vaardigheden bijvoorbeeld geïnterpreteerd als de moeilijkheid van het item, en bij het meten van psychopathologie als de ernst van het symptoom. De parameter $\alpha_j$ is de richtingscoëfficiënt van de IRF geëvalueerd op het punt $\theta_i = \beta_j$. Deze parameter geeft aan hoe sterk het item discrimineert tussen verschillende waarden van $\theta$. In de meeste testtheoretische toepassingen streeft men naar een goede spreiding van de waardes van $\beta$ en een zo hoog mogelijke waarde van $\alpha$.

Aan de keuze voor de logistische functie liggen principiële en praktische redenen ten grondslag. Een belangrijke principiële reden is dat bepaalde specifieke invullingen van het zo ontstane kansmodel zeer aantrekkelijke meeteigenschappen opleveren. Indien de discriminatieparameters allen gelijk zijn ontstaat het zogenaamde Rasch-model, vernoemd naar de Deense wiskundige Georg Rasch [33]. Onder dit model is de totaalscore van een sufficiënte statistiek voor de latente variabele en daardoor kunnen de itemparameters bepaald worden onafhankelijk van de vorm van de dichtheid $f(\theta)$. Omdat de latente variabele latent is, hebben we zelden sterke a priori redenen om verdelingsaannames te motiveren, en daarom is dit een aantrekkelijke eigenschap.

In het bovenstaande voorbeeld zijn de manifeste variabelen dichotoom, en is de latente variabele continu, maar er bestaan soortgelijke modellen voor allerlei soorten en maten manifeste en latente variabelen [27]. Hoewel het latente-variabelenmodel een psychologische uitvinding is [41] wordt het tegenwoordig ook steeds belangrijker in velden als *machine learning* en bij modellen die gebruik maken van neurale netwerken als in *deep learning* (ook het neurale netwerkmodel vindt overigens zijn oorsprong in de psychologie [21]).

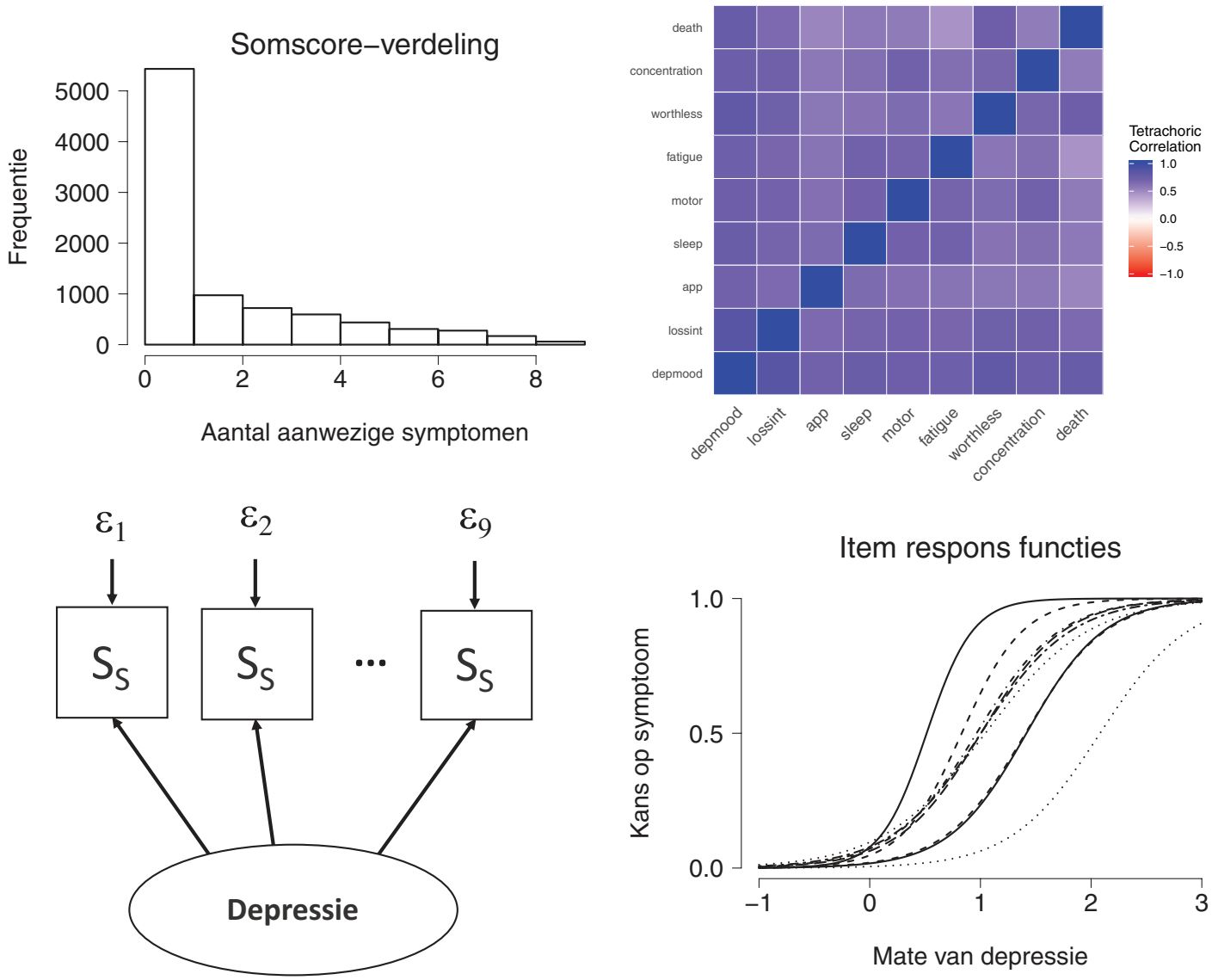
Tegenwoordig is het gemakkelijk het latente-variabelenmodel te schatten op basis van data. Dit levert voor iedere manifeste variabele een discriminatie- en locatieparameter op, en voor iedere persoon een schatting van de positie van die persoon op de latente variabele. Een voorbeeld van een model, geschat met het *ltm*-pakket [35] in het *open source* statistisch softwarepakket R [32] is weergegeven in Figuur 1. De analyse is uitgevoerd op een dataset van 8973 personen in de zogenaamde VATSPSUD-studie [23].

De verdeling van depressiesymptomen is scheef naar rechts, en de items discrimineren dus alleen aan de bovenkant van de populatie. Dit is ook wat men verwacht van een diagnostisch instrument. De symptomen vertonen verder de *positive manifold*: alle correlaties zijn positief. De *positive manifold* is een interessante eigenschap van psychometrische data die zich ook voordoet bij cognitieve vaardigheden (alle vaardigheidstests correleren in de algemene populatie positief), persoonlijkheidstests (vrijwel alle persoonlijkheidseigenschappen correleren positief als men deze zo codeert, dat een hogere score een wenselijker gedragsprofiel indiceert) en psychopathologie in het algemeen (symptomen correleren bijna altijd positief, met uitzondering van enkele symptomen die met manie te maken hebben — deze correleren negatief met de inactiviteit die bij typische depressies hoort). Dit robuuste fenomeen, dat honderden keren gerepliceerd is, vormt een van de belangrijkste ontdekkingen van de psychometrie.

## Netwerken

Het meetmodel dat hierboven is uiteengelegd is van ontzagwekkend belang voor de organisatie van onderzoek in de psychopathologie. Als het model klopt, dan kunnen we namelijk de telling van symptomen gebruiken als *meting* van een theoretisch relevante variabele als depressie, net zoals we andere tellingen kunnen gebruiken als metingen van angststoornissen, fobieën, schizofrenie, et cetera. Inderdaad is het onderzoek langs deze psychometrische lijnen georganiseerd. Als men bijvoorbeeld leest over de neurale grondslag voor depressie, dan gaat het om vergelijkingen van hersendata tussen mensen die hoog of laag op deze telling scoren. Therapeutisch effectonderzoek richt zich op de relatieve afname van de gemiddelde totaalscore als functie van een controle dan wel experimentele conditie. De erfelijkheid van stoornissen (die overigens zeer substantieel is) wordt bepaald door sterkte van de correlaties tussen symptoomtellingen te vergelijken over paren van één- en twee-eiige tweelingen. De zoektocht naar de genetische basis van stoornissen is gebaseerd op het regresseren van symptoomtellingen op variaties in basenparen op het DNA.

Maar de conclusie dat wij de mate van de depressie kunnen meten door de toepassing van een statistisch model is niet gratis. Wij moeten haar 'kopen' met zware aannames over de processen die ten grondslag liggen aan de waargenomen symptomatologie. De belangrijkste daarvan is de aanname dat de symptomen bepaald worden door één en de dezelfde onderliggende grootheid — de latente variabele. Zoals weergegeven in Figuur 1 betekent dit dat aan stoornis als depressie de *gemeenschappelijke oorzaak* van de symptomen is, net zoals bijvoorbeeld een tumor in de longen de gemeenschappelijke oorzaak van symptomen als pijn op de borst, hoesten, en bloed opgeven kan zijn.




**Figuur 1** Het latente-variabelenmodel toegepast op depressie. Linksboven de verdeling van de somscore (het aantal aanwezige symptomen in de data), rechtsboven een heatmap van de tetrachorische correlaties (allen positief), linksonder de structuur van het latente-variabelenmodel (waarin symptomen S1–S9 gerelateerd worden aan de latente variabele 'depressie'), en tot slot, rechtsonder, de Item Respons Functies die de bijbehorende modelschatting oplevert.

Deze structuur verraadt zich door de wijze waarop de kansverdeling gefactoriseerd wordt. De factorisatie heeft de vorm van een product en berust dus op de aanname dat de manifeste variabelen statistisch onafhankelijk zijn, gegeven de latente variabele. Deze aanname staat bekend als *lokale onafhankelijkheid*. Het model zegt in wezen dat het statistisch verband tussen symptomen als vermoeidheid en insomnia in essentie 'spurieus' is. Net zoals de correlatie tussen het aantal ooievaars en het aantal geboortes in Macedonische dorpjes berust op hun gemeenschappelijke afhankelijkheid van het inwonertal (meer babies ← meer stelletjes ← meer inwoners → meer schoorstenen → meer ooievaars), zo berust de correlatie tussen vermoeidheid en insomnia op hun gemeenschappelijke afhankelijkheid van de mate van depressie. Wanneer men voor deze gemeenschappelijke oorzaak controleert, dan vallen de correlaties tussen de manifeste variabelen weg, net zoals de correlatie tussen het aantal ooievaars en het aantal geboortes wegvalt als men voor de populatiegrootte van de dorpjes controleert. Samen met de aanname dat de kansverdeling van iedere manifeste variabele monotoon is in de latente variabele — hoe depressiever, hoe groter de kans op ieder van de symptomen — levert dit een model op waarin de manifeste variabelen tot op zekere hoogte uitwisselbaar zijn [22]

Dit is weliswaar een mooi statistisch model, maar inhoudelijk gezien is dit model niet altijd even voor de hand liggend. Een goed voorbeeld van waar het misgaat is het domein van de psychopathologie. De symptomen van psychologische stoornissen lijken op het eerste gezicht immers geen plausibele kandidaten voor uitwisselbaarheid (bij depressie is het bijvoorbeeld lastig vol te houden dat symptomen als suicidaliteit en gebrek aan eetlust uitwisselbaar zijn). Bovendien zijn die symptomen niet zuiver en alleen gecorreleerd vanwege een gemeenschappelijke afhankelijkheid van latente variabelen. De correlatie tussen insomnia en vermoeidheid berust bijvoorbeeld naar alle waarschijnlijkheid wel op een direct verband tussen deze symptomen: wie niet slaapt, wordt moe. Hetzelfde geldt voor allerlei andere symptomen die

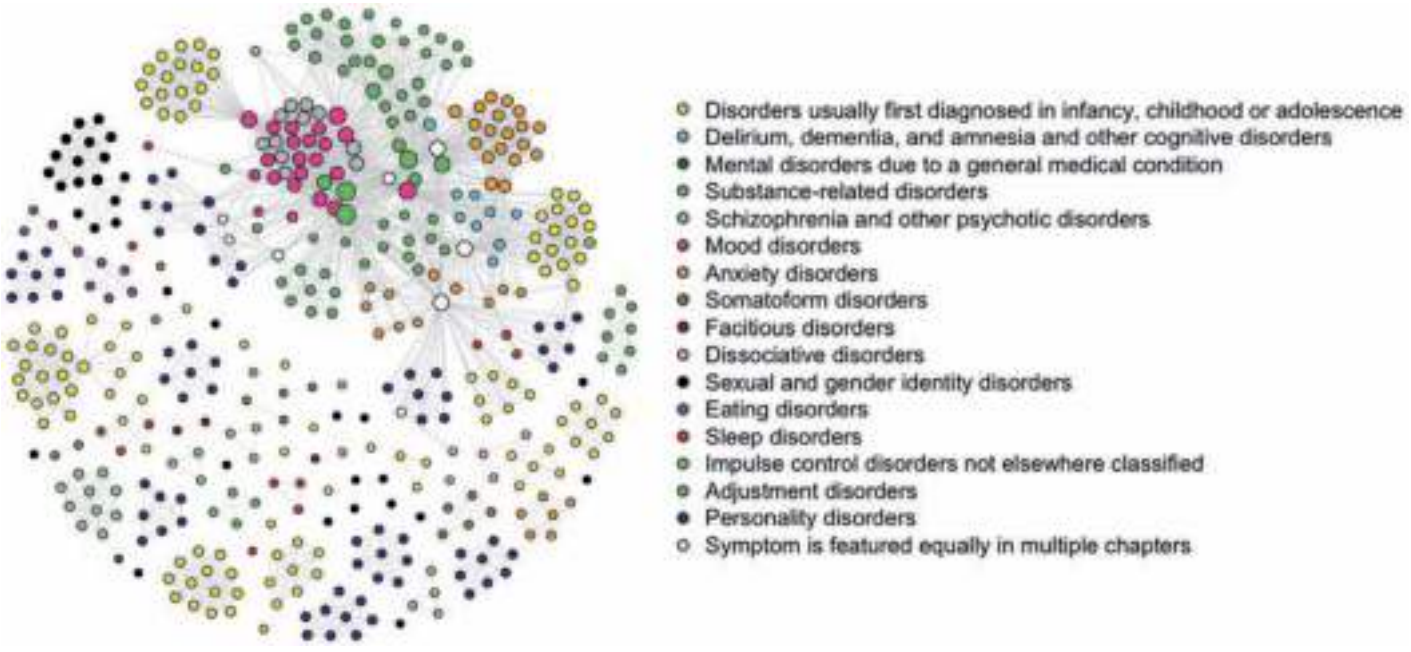
we aantreffen in almanakken als de DSM (het belangrijkste diagnostisch handboek in de psychiatrie [1]). Hallucinaties leiden tot angsten, drugsgebruik tot tolerantie en onthoudingsverschijnselen, gebrekkige energie tot zelfverwijt, de overtuiging dik te zijn tot excessief afvallen, angst tot vermijding van datgene waar men bang voor is, enzovoort. Symptomen van psychische stoornissen zijn dus hoogstwaarschijnlijk betrokken in allerlei directe causale relaties. Vaak worden symptomen daarbij ook nog gekenmerkt door *feedback*-relaties: angst voor spinnen leidt tot vermijding van spinnen, wat weer leidt tot instandhouding van de angst; wie overtuigd is dat de CIA achter hem aan zit ziet overal camera's, wat de betreffende waan in stand houdt; de gokker die geld verliest aan de gokautomaat zal vaak dat geld terug proberen te verdienen met gokken, wat weer tot meer schulden leidt.

Het beeld dat hieruit oprijst is dat van symptomen die samenklonteren in allerlei patronen van al dan niet wederzijdse bekrachtiging. Dat staat op gespannen voet met de causale interpretatie van het latentevariabelenmodel. In het onderzoeksprogramma, dat we in 2007 gestart zijn aan de Universiteit van Amsterdam, besloten we daarom de zaak om te draaien. In plaats van aan te nemen dat alle associaties tussen symptomen van stoornissen ontstaan doordat die symptomen veroorzaakt worden door dezelfde stoornis, gingen wij uit van een situatie waarin symptomen eigenstandige causale entiteiten zijn, die netwerken vormen van interacties.

In een van onze eerste studies [7] besloten we als vertrekpunt de *Diagnostic and Statistical Manual of Mental Disorders* [1] te nemen. Het leek ons plausibel dat symptomen, die in de loop van de eeuw in dezelfde stoornis waren gecategoriseerd, sterkere verbindingen in het netwerk zouden moeten hebben. Om een eerste zicht te krijgen op de structuur van het netwerk van psychopathologie als geheel produceerden we een afbeelding van de DSM, waarin ieder symptoom is weergegeven als een knoop, en twee symptomen verbonden zijn als ze als diagnostische criteria voor dezelfde stoornis in de DSM figureren (dit is dus de projectie van een *bipartite* netwerk). Dit netwerk is weergegeven in Figuur 2. Interessant is dat de graaf een zogenaamde *giant component* laat zien (bijna $50\%$ van de symptomen is verbonden).

Gegeven de netwerkhypothese levert dit patroon eigenlijk een 'gratis' verklaring voor de enorme comorbiditeit tussen mentale stoornissen. Een groot deel van de mensen die aan de diagnostische criteria voldoen voor een stoornis (bijvoorbeeld depressie), voldoen óók aan de criteria voor een andere stoornis (bijvoorbeeld de paniekstoornis), en dit patroon doet zich voor bij vrijwel alle stoornissen in de DSM. Dit wordt vaak gezien als een meetprobleem, maar in de netwerktheorie is dat een verkeerde opvatting. Als stoornissen ontstaan door interacties tussen symptomen, dan houden die interacties natuurlijk niet op bij de grenzen van een diagnostische categorie. Wanneer iemand bijvoorbeeld niet van de drank af kan blijven, en daardoor ontslagen word en ruzie krijgt met zijn familie (symptomen van alcoholmisbruik), dan kan die persoon zich daardoor waardeloos gaan voelen (symptoom van depressie) of zich excessief zorgen gaan maken (symptoom van gegeneraliseerde angststoornis). Zo kan comorbiditeit ontstaan via causale paden die symptomen van verschillende stoornissen verbinden. De gemiddelde kortste padlengte tussen de symptomen van twee stoornissen in het netwerk correleren dan ook sterk negatief ($r = -0{,}7$) met de empirisch bepaalde mate van comorbiditeit tussen de betreffende stoornissen, zelfs als daarbij gemeenschappelijke symptomen niet worden meegenomen.

Met behulp van een netwerkstructuur als die in Figuur 2 kan men simulatiemodellen opstellen om het ontstaan en beloop van stoornissen na te bootsen. In deze modellen, die verwant zijn aan epidemiologische modellen voor virussen, kunnen actieve symptomen hun buren 'aansteken'. Zo kan de symptoomactivatie




**Figuur 2** Het DSM-IV-netwerk op basis van de structuur van het diagnostische systeem, waarin twee symptomen verbonden zijn als ze bij dezelfde stoornis zijn ingedeeld (boven) en voor een deel van de symptomen op basis van empirische data, waarin twee symptomen verbonden zijn als ze conditioneel geassocieerd zijn, gegeven de andere symptomen.

zich verspreiden over het netwerk. Zulke simulaties laten zien dat dit proces geloofwaardige eigenschappen heeft: het leidt ertoe dat symptoomactiviteit zich zal concentreren rond hecht verbonden groepjes symptomen ('stoornissen') die evenwel niet goed afgescheiden zijn, waardoor symptoomactiviteit van de ene naar de andere stoornis over kan springen ('comorbiditeit') via symptomen die met beide groepjes symptomen verbonden zijn (deze noemen we 'brugsymptomen' [14]). Bovendien kan zulke symptoomactiviteit er bij hecht verbonden netwerken toe leiden dat het netwerk blijft plakken in een patroon van zichzelf in stand houdende symptomen, omdat er *hysterese* optreedt: het netwerk heeft niet-lineaire eigenschappen, in de zin dat een kleine perturbatie kan leiden tot een systematische symptoomactivatie, die dan niet uit zichzelf meer kan verdwijnen. Het netwerk heeft dus een *tipping point*, zoals dat ook bekend is bij allerlei andere complexe dynamische systemen [38, 39]

Onder de versimpelde aanname dat symptomen zich inderdaad uitsluitend organiseren langs de lijnen van paarsgewijze interacties in een netwerk, kan men de structuur en parameters van zo'n netwerk ook schatten. Technisch gezien is een systeem van binaire variabelen die paarsgewijs geassocieerd zijn statistisch equivalent met het bekende Ising-model uit de natuurkunde [19, 26]. De kans dat we het systeem in de configuratie $\mathbf{x}$ aantreffen wordt in het Ising-model gekarakteriseerd door de formule

$$p(\mathbf{x}) = \frac{\exp(\mathbf{x}^T\boldsymbol{\mu} + \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x})}{\sum_{\mathbf{x}}\exp(\mathbf{x}^T\boldsymbol{\mu} + \mathbf{x}^T\boldsymbol{\Sigma}\mathbf{x})},$$

waarbij de som in de noemer over alle mogelijke configuraties van de vector $\mathbf{x}$ loopt. De hoofdeffecten $\boldsymbol{\mu}$ bepalen de kans dat knopen in het netwerk geactiveerd zijn, buiten de invloed van de andere knopen in het netwerk om; hogere waarden leiden tot een hogere kans tot activatie. De invloed van de andere knopen worden uitgedrukt in de matrix met paarsgewijze interacties, $\boldsymbol{\Sigma}$, waarbij een hogere waarde van een interactie-parameter $\sigma_{ij}$ leidt tot een hogere kans dat een paar variabelen $X_i$ en $X_j$ beide geactiveerd zijn. Deze interacties worden in de psychometrie dus gebruikt om, bijvoorbeeld, de wederkerige interactie tussen symptomen te kenmerken.

De structuur van het Ising-model in termen van hoofdeffecten en paarsgewijze interacties lijkt in veel opzichten op dat van een andere bekende multivariate verdeling, de multivariate normale verdeling. Maar waar het doorgaans eenvoudig rekenen is met de normale verdeling, is het Ising-model notoir lastig. De som in de noemer, bijvoorbeeld, kan slechts in uitzonderlijke gevallen worden versimpeld, wat er in de praktijk op neer komt dat alle $2^p$ configuraties geëvalueerd dienen te worden, en met slechts $p = 20$ variabelen zijn dit meer dan een miljoen termen. Dit is extra kostbaar voor de iteratieve procedures die gebruikt worden om de waarden van modelparameters te bepalen uit geobserveerde data. Om deze reden zijn er allerlei trucs bedacht om de parameters te schatten zonder het model in zijn volledigheid te hoeven evalueren.

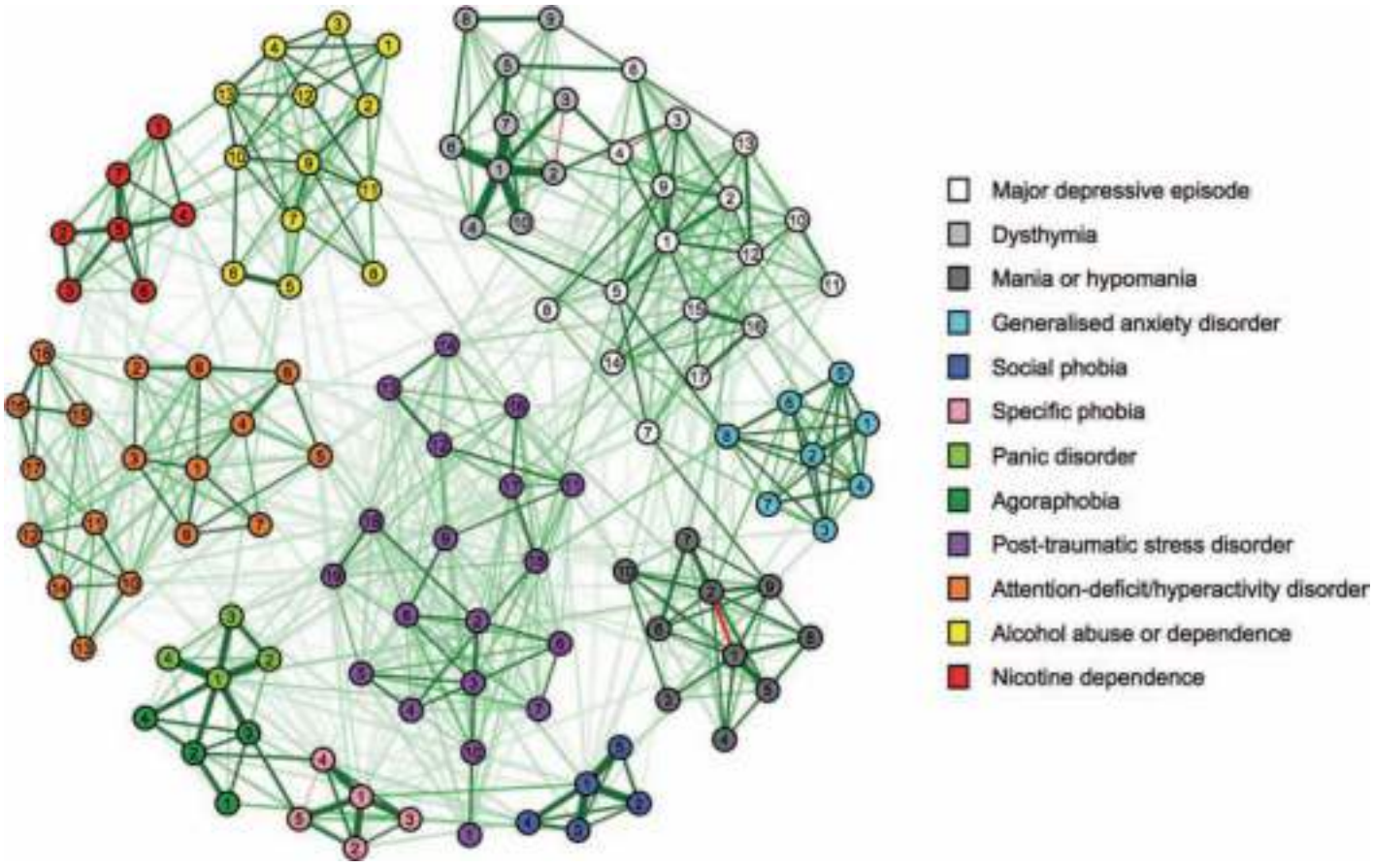
Een belangrijke schattingsmethode gebruikt de zogeheten pseudo-likelihood. Hierbij optimaliseert men niet de parameters in de gezamenlijk verdeling, $P(\mathbf{X} = \mathbf{x})$, maar optimaliseert men de conditionele verdeling van één variabele, zeg $X_i$, gegeven de rest van de variabelen in het netwerk. De kans dat we variabele $i$ in de toestand $x_i = 1$ observeren gegeven de overige variabelen $\mathbf{x}^{(i)}$ wordt gegeven door de formule

$$P\big(X_i = 1 \,|\, \mathbf{x}^{(i)}\big) = \frac{\exp\big(\mu_i + 2\sum_{j\neq i} x_j\sigma_{ij}\big)}{1+\exp\big(\mu_i + 2\sum_{j\neq i} x_j\sigma_{ij}\big)},$$

ofwel de regressie van $X_i$ op $\mathbf{x}^{(i)}$. Hier zien we ook direct hoe de interacties gemodelleerd worden in het Ising-model; als $\sigma_{ij} = 0$ dan zijn de variabelen $i$ en $j$ onafhankelijk van elkaar gegeven de rest van het netwerk.

De afgelopen jaren zijn schattingsmethoden waarmee deze netwerken geschat kunnen worden [34] door onze groep in Amsterdam verder ontwikkeld en geïmplementeerd in het vrij beschikbare softwareprogramma R. Een voorbeeld van een netwerk dat daarmee is geschat op basis van empirische gegevens is weergegeven in Figuur 3. In dit netwerk wordt ieder symptoom weergegeven als een knoop, en zijn twee knopen verbonden als het model met die verbinding beter op de data fit, dan het model zonder de verbinding [4]. Dit betekent in de praktijk dat variabelen verbonden raken als het verband tussen die variabelen niet 'wegverklaard' kan worden door voor de andere variabelen in het netwerk te controleren. De schattingsmethode benadert zo het Markov Random Field [24], waarin twee niet direct verbonden variabelen onafhankelijk zijn, gegeven hun gemeenschappelijke buren. Deze structuur van conditionele afhankelijkheden is nauw gerelateerd aan de moderne causaliteitsopvattingen van bijvoorbeeld Pearl [31]. Het is overigens belangrijk om op te merken dat een verbinding tussen variabelen in een netwerk niet betekent dat aangetoond is dat die variabelen elkaar beïnvloeden; een verbinding kan ook ontstaan door gemeenschappelijke afhankelijkheid van een latente variabele of door conditioneren op een gemeenschappelijk effect [16]. Netwerkmethodologie zoals in Figuur 3 weergegeven moet men dus in eerste instantie zien als een hypothese-genererende techniek waarmee inzicht in de structuur van de data kan worden gegeven, en niet zozeer als een test van de netwerkhypothese zelf.

Toen de eerste netwerkrepresentaties van psychometrische variabelen beschikbaar werden via het programma *qgraph* van Sacha Epskamp [18] en in het verlengde daarvan schattingsmethoden op basis van geregulariseerde regressie geïmplementeerd werden door Claudia van Borkulo [4] was dat eigenlijk meteen een enorm succes. Ook onafhankelijk van de onderliggende theorie geeft een netwerk als in Figuur 3 een inzichtelijke visualisatie van complexe multivariate structuren in datasets, die er ook nog aantrekkelijk uitziet. Toen wij deze methodologie introduceerden leidde dat dan ook tot een explosie van toepassingen in allerlei velden van de psychologie en daarbuiten. In het kielzog van die populariteit hebben we de netwerkmethodologie in allerlei richtingen verder ontwikkeld. De huidige programmatuur is in staat netwerken te schatten voor binaire [4], continue [17], en gemengde variabelen [20], zowel op basis van data die bij een groot aantal mensen eenmalig verzameld zijn (zogenaamde cross-sectionele data) als op basis van herhaalde metingen [13, 16]. Bovendien hebben we systematische methodologie ontwikkeld om de stabiliteit en robuustheid van de geschatte netwerken te bepalen [17]. Op dit moment zijn de meeste standaard datasets hierdoor eenvoudig om te zetten in een netwerkrepresentatie.




**Figuur 3** Een empirisch geschat Ising-model voor de DSM-IV-symptomatologie geschat op basis van een grote steekproef van mensen bij wie de aan- dan wel afwezigheid van de symptomen vastgesteld is [9]. Het model is geschat op basis van geregulariseerde regressiemodellen [4].

## Conclusie

In dit artikel hebben we de belangrijkste uitgangspunten en eigenschappen uiteengezet van twee benaderingen in de psychometrie: de klassieke benadering, waarin manifeste variabelen worden gezien als metingen van een onderliggend theoretisch construct, en de netwerkbenadering, waarin manifeste variabelen worden gezien als onderdeel van een gecompliceerd systeem waarin gevoelens, gedachten, en gedragingen elkaar beïnvloeden. Het latente-variabelenmodel is de uitkomst van een eeuw aan psychometrisch onderzoek, en is statistisch gezien grotendeels af, terwijl de jeugdiger netwerkbenadering nog maar net op stoom is en stormachtige ontwikkelingen doormaakt.

Het is de vraag waar die ontwikkelingen ons zullen brengen. Nu het mogelijk is de topologie van psychometrische netwerken te bepalen op basis van data, ontstaat de vraag wat die topologie ons eigenlijk oplevert. Zijn meer centrale variabelen in een netwerk bijvoorbeeld op een of andere manier belangrijker? Is de netwerkstructuur bij mensen met psychopathologie anders dan die van de algemene populatie? Kunnen netwerkstructuren helpen bij de keuze waar in het systeem moet worden geïntervenieerd? Er lijken inderdaad aanwijzingen te zijn dat de topologie van netwerken ertoe doet. De positie van knopen in psychopathologienetwerken blijkt geassocieerd met hun voorspellend vermogen: hoe centraler de knoop, hoe beter deze de toekomstige ontwikkeling van de stoornis voorspelt [10, 37]. Daarnaast zijn groepsverschillen in netwerkstructuren in verband gebracht met relevante externe variabelen. Zo lijken mensen die blijven hangen in een depressie een dichter verbonden structuur te hebben dan mensen die herstellen [5, 40] en zijn de negatieve emoties van mensen met verhoogde kans op depressie sterker met elkaar verbonden [12, 30]. Sommige experimentele interventies lijken op sommige variabelen in het netwerk sterker aan te grijpen dan op andere, wat zich kan geven op de nog grotendeels onbekende werkingsmechanismen van psychotherapie [2]. Karakteristieke patronen die in verband zijn gebracht met fase-overgangen in netwerken en hysterese, zoals het vertragen van het systeem voorafgaand aan een fase-overgang [39] worden ook gezien in psychopathologie-onderzoek [42]. En de eerste therapeutische interventietechnieken op basis van netwerkanalyse worden inmiddels onderzocht [25].

De ontwikkelingen gaan in feite zo snel, dat er inmiddels ook allerlei waarschuwingen in de literatuur opduiken die kanttekeningen plaatsen bij de validiteit van conclusies die op basis van netwerkanalyse worden getrokken. Zo zijn centraliteitsmaten heel populair, terwijl niet altijd even duidelijk is wat deze meten [11, 16]. Daarnaast zijn sommige netwerkschattingen gevoelig voor steekproeffluctuaties [17], en het effect daarvan is niet altijd even gemakkelijk te bepalen. Zulke vragen zijn natuurlijk zeer belangrijk en zullen de komende jaren ongetwijfeld veel aandacht krijgen.

## Referenties


1 American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed., APA, 2013.

2 T.F. Blanken, T. van der Zweerde, A. van Straten, E.J.W. van Someren, D. Borsboom en J. Lancee, Introducing network intervention analysis to investigate sequential, symptom-specific treatment effects: A demonstration in co-occurring insomnia and depression, *Psychotherapy and Psychosomatics* 88 (2019), 52–54.

3 K.A. Bollen, Latent variables in psychology and the social sciences, *Annual Review of Psychology* 53 (2002), 605–634.

4 C.D. van Borkulo, D. Borsboom, S. Epskamp, T.F. Blanken, L. Boschloo, R.A. Schoevers en L.J. Waldorp, A new method for constructing networks from binary data, *Scientific Reports* 4 (2014), 5918.

5 C.D. van Borkulo, L. Boschloo, D. Borsboom, B.W.J.H. Penninx, L.J. Waldorp en R.A. Schoevers, Association of symptom network structure with the course of longitudinal depression, *JAMA Psychiatry* 72 (2015), 1219–1226.

6 D. Borsboom, *Measuring the Mind: Conceptual Issues in Contemporary Psychometrics*, Cambridge University Press, 2005.

7 D. Borsboom, A.O.J. Cramer, V.D. Schmittmann, S. Epskamp en L.J. Waldorp, The small world of psychopathology, *PLoS ONE* 11 (2011), e27407.

8 D. Borsboom, G.J. Mellenbergh en J. van Heerden, The theoretical status of latent variables, *Psychological Review* 110 (2003), 203–219.

9 L. Boschloo, C.D. van Borkulo, M. Rhemtulla, K.M. Keyes, D. Borsboom en R.A. Schoevers, The network structure of symptoms of the Diagnostic and Statistical Manual of Mental Disorders, *PLoS One* 10 (2015), e0137621.

10 L. Boschloo, C.D. van Borkulo, D. Borsboom en R.A. Schoevers, Letter to the editor: A prospective study on how symptoms in a network predict the onset of depression, *Psychotherapy and Psychosomatics* 85 (2016), 183–184.

11 L.F. Bringmann, T. Elmer, S. Epskamp, R.W. Krause, D. Schoch, M. Wichers, J.T.W. Wigman en E. Snippe, What do centrality measures measure in psychological networks? *Journal of Abnormal Psychology* (2019), in press.

12 L.F. Bringmann, M.L. Pe, N. Vissers, E. Ceulemans, D. Borsboom, W. Vanpaemel, F. Tuerlinckx en P. Kuppens, Assessing temporal emotion dynamics using networks, *Assessment* 23 (2016), 425–435.

13 L.F. Bringmann, N. Vissers, M. Wichers, N. Geschwind, P. Kuppens, F. Peeters, D. Borsboom en F. Tuerlinckx, A network approach to psychopathology: New insights into clinical longitudinal data, *PLoS ONE* 8 (2013), e60188.

14 A.O.J. Cramer, L.J. Waldorp, H.L.J. van der Maas en D. Borsboom, Comorbidity: A network perspective, *Behavioral and Brain Sciences* 33 (2010), 137–150.

15 L.J. Cronbach en P.E. Meehl, Construct validity in psychological tests, *Psychological Bulletin* 52 (1955), 281–302.

16 S. Epskamp, *Network Psychometrics*, proefschrift Universiteit van Amsterdam, 2017, http://sachaepskamp.com/dissertation/EpskampDissertation.pdf

17 S. Epskamp, D. Borsboom en E.I. Fried, Estimating psychological networks and their accuracy: A tutorial paper, *Behavior Research Methods* 50(1) (2018), 195–212.

18 S. Epskamp, A.O.J. Cramer, L.J. Waldorp, V.D. Schmittmann en D. Borsboom, Qgraph: Network visualizations of relationships in psychometric data, *Journal of Statistical Software* 48 (2012), 1–18.

19 S. Epskamp, G.K.J. Maris, L.J. Waldorp en D. Borsboom, Network psychometrics, in: P. Irwing, D. Hughes en T. Booth (Eds.), *Handbook of Psychometrics*, Wiley, 2018.

20 J.M.B. Haslbeck en L.J. Waldorp, mgm: Estimating time-varying mixed graphical models in high-dimensional data, *Journal of Statistical Software* (2019), in press.

21 D.O. Hebb, *The Organization of Behavior*, Wiley, 1949.

22 B.W. Junker en J.L. Ellis, A characterization of monotone unidimensional latent variable models, *The Annals of Statistics* 25 (1997), 1327–1343.

23 K.S. Kendler en C.A. Prescott, *Genes, Environment, and Psychopathology: Understanding the Causes of Psychiatric and Substance Use Disorders*, Guilford Press, 2006.

24 R. Kindermann en J.L. Snell, *Markov Random Fields and their Applications*, American Mathematics Society, 1980.

25 R. Kroeze, D.C. van der Veen, M.N. Servaas, J.A. Bastiaansen, R.C. Oude Voshaar, D. Borsboom, H.G. Ruhe, R.A. Schoevers H. Riese, Personalized feedback on symptom dynamics of psychopathology: A proof-of-principle study, *Journal for Person-Oriented Research* 3 (2017), 1–10.

26 M. Marsman, D. Borsboom, J. Kruis, S. Epskamp, R. van Bork, L.J. Waldorp, H.L.J. van der Maas en G.K.J. Maris, An introduction to network psychometrics: Relating Ising network models to item response theory models, *Multivariate Behavioral Research* 53 (2018), 15–35.

27 G.J. Mellenbergh, Generalized linear item response theory, *Psychological Bulletin* 115 (1994), 300–307.

28 J. Michell, Quantitative science and the definition of measurement in psychology, *British Journal of Psychology* 88 (1997), 355–383.

29 J. Michell, *Measurement in Psychology: A Critical History of a Methodological Concept*, Cambridge University Press, 1999.

30 M.L. Pe, K. Kircanski, R.J. Thompson, L.F. Bringmann, F. Tuerlinckx, M. Mestdagh, J. Mata, S.M. Jaeggi, M. Buschkuehl, J. Jonides, P. Kuppens en I. H. Gotlib, Emotion-network density in major depressive disorder, *Clinical Psychological Science* 3 (2015), 292–300.

31 J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed., Cambridge University Press, 2009.

32 R Core Team R: A language and environment for statistical computing, R Foundation for Statistical Computing, 2014, www.R-project.org.

33 G. Rasch, *Probabilistic Models for Some Intelligence and Attainment Tests*, Paedagogiske Institut Copenhagen, 1960.

34 P. Ravikumar, M.J. Wainwright en J.D. Lafferty, High-dimensional Ising model selection using l1-regularized logistic regression, *The Annals of Statistics* 38 (2010), 1287–1319.

35 D. Rizopoulos, ltm: An R package for latent variable modeling and item response theory analyses, *Journal of Statistical Software* 17 (2006), 1–25.

36 D.J. Robinaugh, N.J. Leblanc, H.A. Vuletich en R.J. McNally, Network analysis of persistent complex bereavement disorder in conjugally bereaved adults, *Journal of Abnormal Psychology* 123 (2014), 510–522.

37 T.L. Rodebaugh, N.A. Tonge, M.L. Piccirillo, E. Fried, A. Horenstein, A.S. Morrison, P. Goldin, J.J. Gross, M.H. Lim, K.C. Fernandez, C. Blanco, F.R. Schneier, R. Bogdan, R.J. Thompson en R.G. Heimberg, Does centrality in a cross-sectional network suggest intervention targets for social anxiety disorder? *Journal of Consulting and Clinical Psychology* 86(10) (2019), 831–844.

38 M. Scheffer, J.E. Bolhuis, D. Borsboom, T.G. Buchman, S.M.W. Gijzel, D. Goulson, J.E. Kammenga, B. Kemp, I.A. van de Leemput, S. Levin, C.M. Martin, R.J.F. Melis, E.H. van Nes, L.M. Romero en M.G.M. Olde Rikkert, Quantifying resilience of humans and other animals, *Proceedings of the National Academy of Sciences* 115 (2018), 11883–11890.

39 M. Scheffer, S.R. Carpenter, T.M. Lenton, J. Bascompte, W. Brock, V. Dakos, J. van de Koppel, I.A. van de Leemput, S.A. Levin, E.H. van Nes, M. Pascual en J. Vandermeer, Anticipating critical transitions, *Science* 338 (2012), 344–348.

40 L. Schweren, C.D. van Borkulo, E. Fried en I.M. Goodyer, Assessment of symptom network density as aprognostic marker of treatment response inadolescent depression, *JAMA Psychiatry* 75 (2018), 98–100.

41 C. Spearman, General intelligence, objectively determined and measured, *American Journal of Psychology* 15 (1904), 201–293.

42 M. Wichers, P. Groot, C.J.P. Simons, F. Peeters, Psychosystems Group, ESM Group en ESW Group, Critical slowing down as a personalized early warning signal for depression, *Psychotherapy and Psychosomatics* 85 (2016), 114–116.

Herbert Hamers
*Department of Econometrics and Operations Research and TIAS School for Business and Society*
*Tilburg University*
*h.j.m.hamers@tilburguniversity.edu*

Bart Husslage
*Afdeling Data Science*
*CZ, Tilburg*

Roy Lindelauf
*Section Intelligence and Security*
*Nederlandse Defensie Academie, Breda*
*rha.lindelauf.01@mindef.nl*

# Coöperatieve speltheorie en terroristische netwerken

**Een terroristisch netwerk is een voorbeeld van een verborgen netwerk: een netwerk waarvan weinig bekend is over de architectuur. Geheimhouding en organisatorische veerkracht zijn belangrijke kenmerken. In dit artikel beschrijven Herbert Hamers, Bart Husslage en Roy Lindelauf hoe coöperatieve speltheorie gebruikt kan worden voor de analyse van terroristische netwerken.**

Het is een jaar voor de tragische gebeurtenissen op 9/11, ergens in een trainingskamp bij Kaboel. Een roodharige instructeur in spijkerbroek geeft les aan rekruten van de Afghaanse jihad. Hij noemt zich Abu Musab al-Suri (de Syriër). Op een whiteboard tekent hij een boom, een graaf waarin elk tweetal personen is verbonden door een uniek pad. "Zo moet je organisatie er niet uitzien", zegt hij in een instructievideo die later is gevonden door Amerikaanse soldaten na de invasie van Afghanistan. Mustafa Setmariam Nasar, zoals hij werkelijk heet, is een van de architecten van de globale jihad en wordt verdacht van banden met Al Qaida, de Taliban en andere terroristische organisaties. Hij is de auteur van het bekende *Oproep tot Internationale Islamitisch Verzet*, een boekwerk van 1600 bladzijden waarvan de Engelse vertaling is te vinden op de internetbibliotheek archive.org. Geheimhouding en organisatorische veerkracht zijn volgens al-Suri de belangrijkste kenmerken van een verborgen organisatie. In de video legt hij uit hoe een organogram van een terroristische organisatie eruit moet zien. In plaats van een boomstructuur, die eenvoudig kan worden geïnfiltreerd en ontmanteld, pleit hij voor een netwerk van cellen. De organisatie moet decentraal zijn in plaats van hiërarchisch.

Een foto van Abu Musab al-Suri in 2004 uit een opsporingsbevel van de Amerikaanse overheid. Hij werd gearresteerd in 2005 in Pakistan en overgedragen aan de Amerikaanse autoriteiten. Wat er daarna met hem is gebeurd is niet bekend. Hij zit waarschijnlijk in een Syrische gevangenis.

Na 9/11 vielen de Verenigde Staten Afghanistan binnen en verwijderden de Taliban in een poging om Al Qaida te ontmantelen. Tot dan toe was de wereld gewend aan terroristische organisaties met een verticale commando structuur, zoals de IRA, de PLO, Hamas, Hezbollah en de Rode Brigades. Al-Suri, een werktuigbouwkundig ingenieur uit Aleppo, had een duidelijk andere visie. Analisten van de veiligheidsdiensten kregen te maken met een nieuwe organisatiestructuur, waarin leiders en volgers maar moeilijk van elkaar waren te onderscheiden. Het werd moeilijker om te bepalen welke individuen in de gaten moesten worden gehouden, waardoor destabilisatie of ontmanteling onmogelijk was. Na 9/11 werd Al Qaida het archetype van de moderne terroristische organisatie, waarvan nauwelijks betrouwbare data viel te verzamelen.

Ondertussen waren wetenschappers bezig om de structuur van verschillende sociale en biologische netwerken te ontrafelen [8, 16, 19]. Men realiseerde zich al

snel dat deze inzichten belangrijk zouden kunnen zijn voor de analyse van terroristische netwerken [22]. Eerder onderzoek van terroristische netwerken richtte zich op destabilisatie [5,7], de karakterisatie van de organisatie [6,13,14] en methoden om sleutelfiguren te onderscheiden [1,4,18]. De netwerktopologie in deze onderzoeken was meestal een vast gekozen model. Soms werd historische data gebruikt, maar die is niet betrouwbaar en anekdotisch en bovendien zijn de datasets van de veiligheidsdiensten niet vrij toegankelijk.

De architectuur van decentrale netwerken is goed bestudeerd [20,21], maar dit betreft alleen openbare netwerken. Van verborgen netwerken is weinig bekend. De gebruikelijke sociale netwerken zijn te karakteriseren als *small-world networks*, met een hoge clustering en een kleine gemiddelde afstand. Het is maar de vraag of dit ook geldt voor verborgen netwerken en wat dit betekent voor de analyse ervan. Vanwege het gebrek aan goede data werd het noodzakelijk om theoretische modellen van verborgen netwerken te ontwikkelen. Wij benaderen dit probleem vanuit de *coöperatieve speltheorie*.

**Nash-onderhandelingsoplossing**

Geheime diensten beschikken ook over heimelijke netwerken die werken voor de overheid. De bekendste voorbeelden zijn de CIA, de KGB en de Mossad. Soms wordt een heimelijk netwerk ontdekt, zoals in de operatie Susannah, een Israëlische geheime missie in Egypte gedurende het begin van de jaren vijftig. Al snel na aanvang van de operatie werd het netwerk opgerold, omdat de Egyptische geheime dienst achter de namen kwam van de contacten van één verdacht persoon. Het probleem van de operatie was dat iedereen elkaar kende. Het netwerk was een volledige graaf. Dit bevordert de operationele slagkracht, maar het is niet goed voor de geheimhouding. De Israëlische minister van defensie, Pinhas Lavon, werd gedwongen tot opstappen als uiteindelijke verantwoordelijke van deze mislukte operatie. Dit staat nu bekend als de Lavon-affaire.

Het is duidelijk dat al-Suri een afweging maakte tussen geheimhouding en slagkracht van verborgen organisaties. Een dergelijke afweging wordt ook gemaakt in de *onderhandelingstheorie*, een onderwerp binnen de speltheorie. Als twee spelers een taart moeten verdelen, dan

Pinhas Lavon werd verantwoordelijk gehouden voor de operatie Susannah, waarin geheim agenten bomaanslagen zouden plegen in Egypte. De schuld zou worden gegeven aan de Moslimbroederschap.

ligt de helft voor beiden het meest voor de hand. Maar hoe is de verdeling bij verschillende preferenties? Een van de spelers is misschien op dieet en mag hoogstens een kwart taart. Of hoe verdeel je de taart als sommige delen slagroom bevat maar niet gelijk verdeeld over de hele taart? We zeggen dan dat de te verdelen taart asymmetrisch is. Uiteraard is ook een mogelijke uitkomst dat er geen verdeling wordt gevonden zodat beide spelers helemaal geen taart krijgen. In een onderhandelingsprobleem wordt een optimale verdeling $(\bar{u},\bar{v})$ in een verzameling $F$ van alle mogelijke verdelingen $(u,v)$ gezocht waarbij er een dreigpunt bestaat als de spelers niet tot een verdeling komen. Een optimale verdeling moet voldoen aan een aantal redelijke aannames. De volgende axioma's zijn opgesteld door John Nash, Nobelprijswinnaar in 1994 voor zijn werk op het gebied van de speltheorie, waarvan de onderhandelingstheorie een onderdeel is:

- *Pareto-optimaal*. Er is geen $(u,v) \in F$ met $u \geq \bar{u}$ en $v \geq \bar{v}$ anders dan $(\bar{u},\bar{v})$. Ofwel, als een oplossing Pareto-optimaal is dan kan een speler zichzelf alleen verbeteren ten koste van de andere speler.
- *Symmetrie*. Als $F$ invariant is onder de spiegeling $(u,v) \mapsto (v,u)$ dan $\bar{u} = \bar{v}$. Ofwel, als de verzameling van alle mogelijke verdelingen symmetrisch is dan verdelen de spelers gelijk.
- *Onafhankelijkheid van onbelangrijke alternatieven*. Als $G \subset F$ en als $(\bar{u},\bar{v}) \in G$, dan is dit ook de optimale verdeling voor $F$. Ofwel, als de optimale oplossing in een deelverzameling zit van alle mogelijke verdelingen, dan zijn alle mogelijke verdelingen buiten deze deelverzameling irrelevant. Immers, de oplossing wordt al gevonden uit alle mogelijke verdelingen in deze deelverzameling.
- *Schaalinvariant*. Als $\tau(x,y) = (ax,bx)$ voor positieve $a$ en $b$, dan is de optimale verdeling in $\tau(F)$ gelijk aan $\tau(\bar{u},\bar{v})$. Ofwel, het verhogen van alle verdelingen met dezelfde factoren verhoogt tevens het optimum met deze factoren.

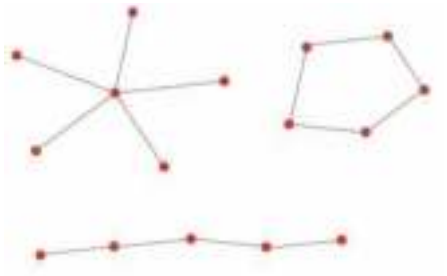
Nash introduceerde de Nash-onderhandelingsoplossing [15] als de oplossing die het product $uv$ maximaliseert. Bovendien toonde hij aan dat de Nash-onderhandelingsoplossing de enige oplossing is die aan de vier genoemde axioma's voldoet. In de volgende paragraaf zullen we de geheimhouding en de slagkracht van een netwerk definiëren. Hierbij gebruiken we de kennis van de Nash-onderhandelingsoplossing waarbij een optimaal netwerk het product van deze twee kenmerken maximaliseert.

**Geheimhouding en informatie in een graaf**

Een netwerk of graaf $G$ bestaat zoals welbekend uit een verzameling knopen $N$ en zijden $E$ tussen de knopen. In ons geval staat $N$ voor de leden of voor afzonderlijke cellen in een terroristisch netwerk. De zijden $E$ staan voor de directe informatiekanalen binnen het netwerk. Binnen het netwerk wordt informatie uitgewisseld en we nemen aan dat de tijd die het duurt om informatie te sturen van $i$ naar $j$ proportioneel is met de afstand van $l_{ij}$ (het minimaal aantal zijden tussen $i$ en $j$). Hoe groter de afstand, hoe waarschijnlijker de onderschepping van de informatie. We stellen $n$ gelijk aan het aantal knopen en $T(G)$ gelijk aan de som over alle afstanden $\sum_{i,j \in N} l_{ij}$. We definiëren de informatie-afstand in de graaf als

$$I(G) = \frac{n(n-1)}{T(G)}.$$

Met andere woorden, $I(G)$ is de reciproke waarde van de gemiddelde afstand. Merk op dat $0 < I(G) \leq 1$ en dat $I(G) = 1$ dan en slechts dan als $G$ een volledige graaf is.



**Figuur 1** Stergraaf, cyclische graaf en lineaire graaf.

**Voorbeeld.** Beschouw de stergraaf $S_n$, de cyclische graaf $C_n$ en de lineaire graaf $L_n$ zoals in Figuur 1. Als $n$ oneven, dan zijn de informatie-afstanden van deze grafen gelijk aan

$$I(S_n) = \frac{n}{2(n-1)},$$
$$I(C_n) = \frac{4}{n+1},$$
$$I(L_n) = \frac{3}{n+1}.$$

Nu definiëren we de mate van geheimhouding. Hierbij spelen twee factoren een rol. Ten eerste de kans op ontmaskering van een individueel lid. Ten tweede, de kans dat bij ontmaskering ook andere leden van het netwerk in beeld komen. Neem bijvoorbeeld het geval van Nawaf al-Hazmi, een van de kapers van 9/11. In het rapport van de Nationale Commissie Terreuraanslagen [9] staat: "U.S. intelligence would analyze communications associated with Mihdhar whom they identified during this travel, and Hazmi, whom they could have identified but did not." Khalid al-Mihdhar was een boezemvriend van Nawaf, die wel was afgereisd naar de Verenigde Staten, maar op het laatste moment afzag van de 9/11-aanslag. Hij was onder surveillance, maar toch ontsnapte Nawaf aan de aandacht. We hebben hier te maken met een kans op detectie $\alpha_i$ van knoop $i$. In geval van detectie wordt ook een fractie $1-u_i$ van het netwerk zichtbaar. De kans op detectie is klein en we nemen aan dat er slechts één knoop kan worden gedetecteerd. We definiëren de geheimhouding $S(G)$ van de graaf als de verwachte fractie die onzichtbaar blijft, gegeven dat één van de knopen wordt gedetecteerd,

$$S(G) = \sum_{i \in N} \alpha_i u_i.$$

Volgens het Nash-onderhandelingsprincipe zijn $I(G)$ en $S(G)$ optimaal als het product maximaal is. Daarom definiëren we

$$\mu(G) = S(G)I(G) \quad (1)$$

als de prestatiemaatstaf van de graaf. Een optimaal verborgen netwerk maximaliseert $\mu(G)$.

**Enkele optimale verborgen netwerken**

We beschouwen enkele voorbeelden. Stel dat de ontmaskering van individu $i$ ook leidt tot het openbaar maken van al zijn $d_i$ buren. De fractie ongedetecteerde knopen is dan

$$1 - \frac{d_i + 1}{n}.$$

Als de kans op ontmaskering voor elke knoop even groot is, dan is de geheimhouding gelijk aan

$$S_1(G) = \sum_{i \in N} \frac{1}{n}\left(1 - \frac{d_i+1}{n}\right) = \frac{n^2 - n - 2m}{n^2},$$

waarbij $m$ het aantal zijden is. De prestatiemaat is in dit geval

$$\mu_1(G) = S_1(G)I(G) = \frac{n^2-n}{n^2} \cdot \frac{n^2-n-2m}{T(G)}.$$

Het optimale netwerk voor deze maat is de stergraaf.

Als de ontmaskering van een knoop leidt tot het openbaar maken van een buur met kans $p$, dan krijgen we te maken met de binomiale verdeling op $d_i$. De prestatiemaat is nu

$$\mu_2(G) = S_2(G)I(G) = \frac{n^2-n}{n^2} \cdot \frac{n^2-n-2pm}{T(G)}.$$

Voor lage waarden van $p$ is de volledige graaf optimaal. De geheimhouding is dan uitstekend en de informatiestroom is maximaal voor de volledige graaf. De hieronder vermelde resultaten kunnen worden gevonden in [11]:

**Stelling.** *Als $p \leq \frac{1}{2}$ dan is de volledige graaf optimaal onder $\mu_2$. Als $p \geq \frac{1}{2}$ dan is de stergraaf optimaal onder $\mu_2$.*

We vinden hier dus ofwel een volledige graaf ofwel een boom. De ene kennen we uit de Lavon-affaire en de andere wordt afgekeurd door al-Suri. De beschouwde prestatiematen zijn relevant voor de initiele fase van een operatie. Dan zijn leden van het netwerk passief met een minieme kans op detectie. Gedurende de operatie

**Figuur 2** Optimale netwerken tot en met $n = 7$.

worden sommige leden meer zichtbaar. We modelleren deze 'informatie-centraliteit' via de evenwichtsverdeling van een standaard random walk op een graaf. De prestatiemaatstaf wordt dan

$$\mu_3(G) = \frac{2m(n-2)+n(n-1)-\sum_{i \in N} d_i^2}{n(2m+n)} \cdot \frac{n(n-1)}{T(g)}.$$

De optimale netwerken zijn nu minder eenvoudig te beschrijven. Figuur 2 toont de optimale netwerken tot en met $n = 7$ personen.

**Centraliteit in een netwerk**

Naast de structuur van verborgen netwerken is het ook van belang te bepalen welke knopen cruciaal zijn. Welke individuen moeten in de gaten worden gehouden? Het is onmogelijk om iedereen te surveilleren dus moet er een selectie worden gemaakt. Hoe meet men het belang van een knoop in een netwerk? Een van de eerste onderzoe-

Kurt Lewin (1890–1947) was een van de grondleggers van de sociale psychologie. Hij introduceerde het begrip groepsdynamica. Zijn student Alex Bavelas probeerde dit in de praktijk meetbaar te maken.

Lloyd Shapley (1923–2016) was een Amerikaans wiskundige. Hij kreeg de Nobelprijs voor de economie in 2012 vanwege zijn werk aan coöperatieve speltheorie. Deze theorie raakt aan combinatorisch optimaliseren. Een bekend voorbeeld is de Edmonds–Shapley-stelling over het basispolyeder van een submodulaire functie. In coöperatieve speltheorie representeert dit polyeder alle 'redelijke' verdelingen van $v(N)$. Het heet de *core* van het spel.

kers die zich dit afvroeg was Alex Bavelas, een socioloog van MIT die grafentheorie gebruikte om groepen te analyseren [2, 3]. Hij definieerde de centraliteit van een knoop als de reciproke waarde van de gemiddelde afstand tot alle andere knopen. Sindsdien zijn er veel van dit soort *centraliteitsmaten* gedefinieerd, een bekend voorbeeld is de PageRank. De meeste centraliteitsmaten zijn niet bruikbaar voor de analyse van verborgen netwerken, omdat ze te veel kennis van het netwerk eisen. Hier presenteren we een benadering van centraliteit via de *coöperatieve speltheorie*.

Een coöperatief spel kan worden gedefinieerd als een paar $(N,v)$, waarbij $N$ de verzameling spelers is en $v$ een functie die een waarde $v(S)$ toekent aan elke coalitie $S \subset N$. Vaak worden extra eisen opgelegd aan $v$. Het is gebruikelijk dat $v(\emptyset) = 0$ en dat $v$ monotoon is, dat wil zeggen, $v(A) \leq v(B)$ als $A \subset B$. Daarmee is de waarde van de grote coalitie $N$ de maximaal haalbare opbrengst. De kernvraag is dan hoe $v(N)$ moet worden verdeeld over alle spelers. Met andere woorden, wat is de individuele bijdrage van elke speler aan $v(N)$. Lloyd Shapley stelde voor om deze waarde als volgt uit te rekenen [17]. Laat $\pi$ een random permutatie van $N$ zijn. Met andere woorden, zet de spelers random op een rij. Speler $i$ staat op plaats $\pi(i)$ en speler $j$ staat voor speler $i$ als $\pi(j) < \pi(i)$. Definieer nu $F(i)$ als de verzameling spelers die voor $i$ staan. De marginale bijdrage van $i$ in deze permutatie $\pi$ is $m_i^{\pi}(v) = v(F(i) \cup \{i\}) - v(F(i))$. De *Shapley-waarde* van $i$ is dan de verwachte marginale bijdrage van speler $i$. Ofwel, de gemiddelde marginale bijdrage van $i$ wanneer we alle mogelijke permutaties beschouwen. In formule, voor alle $i \in N$,

$$\phi_i(N,v) = \frac{1}{n!} \sum_{\pi \in \Pi(N)} m_i^{\pi}(v),$$

waarbij $\Pi(N)$ de verzameling van alle permutaties van $N$ is.

Om een waarde $v$ aan iedere coalitie $S$ toe te kennen, kiezen we een functie $v$ die het onderliggende netwerk, de personen in het netwerk en de aanwezige informatie over personen en relaties meeneemt. In [12] wordt een gewogen graaf met gewichten $w_i$ op de knopen en $k_{ij}$ op de zijden beschouwd. Om deze gewichten mee te kunnen nemen in de analyse is gekozen voor het definiëren van een zogenaamde *weighted connectivity game* op de gewogen graaf.

Als de coalitie $S \subset N$ samenhangend is, dus alle personen in de coalitie kunnen direct of indirect met elkaar communiceren, dan definiëren we

$$f(S) = \left( \sum_{i \in S} w_i \right) \cdot \max_{ij \in E_S} k_{ij},$$

anders is $f(S) = 0$. Definieer nu $v(S) = \max_{T \subset S} f(S)$. Ofwel, kies een samenhangende coalitie binnen $S$ met de hoogste waarde $f(T)$, Als centraliteitsmaat voor persoon $i$ kunnen we dan de Shapley-waarde van deze persoon berekenen.

In dit model moeten de gewichten op een of andere manier worden bepaald uit gegevens, zoals individuele vaardigheden, financiën, frequentie van communicatie en dergelijke. Dit is het werk voor analisten van de veiligheidsdienst.

**Shapley-waarde toegepast op Al Qaida**

De berekening van de Shapley-waarde kan soms lastig zijn afhankelijk van het spel. De reden is dat het aantal permutaties exponentieel toeneemt. Bijvoorbeeld, voor een spel met $50$ spelers ($n = 50$) heeft een redelijk snelle computer al jaren nodig om de Shapley waarde te berekenen. Bij $100$ spelers ($n = 100$) is er geen enkele computer op aarde die de Shapley waarde kan vinden binnen duizenden jaren.

Echter, soms heeft een spel een mooie structuur. Dit is gelukkig ook hier het geval. Echter de wiskunde voor deze methode gaat iets te ver om in dit artikel te vertellen. Er wordt bij deze methode gebruikgemaakt van de *treewidth* van een netwerk. Als deze niet al te groot is, en dat is het geval voor dit Al Qaida netwerk, dan kunnen we de Shapley waarde snel en exact uitrekenen [23].

Het terroristisch netwerk van de aanslag van 9/11 bestond uit 69 spelers. In Tabel 1 staan de $15$ leden met de hoogste waarde.

De top 10 bevat bekende namen zoals Mohamed Atta (piloot van vlucht $11$, WTC

| Ranking | Naam | Geschatte Shapley-waarde |
|---|---|---|
| 1 | Mohamed Atta | 0,114099473 |
| 2 | Essid Sami Ben Khemais | 0,111249806 |
| 3 | Hani Hanjour | 0,110702087 |
| 4 | Djamal Beghal | 0,107273424 |
| 5 | Khalid Almihdhar | 0,107153568 |
| 6 | Mahmoun Darkazanli | 0,106635302 |
| 7 | Zacarias Moussaoui | 0,101188122 |
| 8 | Nawaf Alhazmi | 0,099594346 |
| 9 | Ramzi Bin al-Shibh | 0,098434678 |
| 10 | Raed Hijazi | 0,094777059 |
| 11 | Hamza Alghamdi | 0,008901234 |
| 12 | Fayez Ahmed | 0,008790340 |
| 13 | Marwan Al-Shehhi | 0,004533725 |
| 14 | Satam Suqami | 0,003693424 |
| 15 | Saeed Alghamdi | 0,003666945 |

**Tabel 1** De vijftien leden van het 9/11-netwerk met de hoogste Shapley-waarde.

De vier kopstukken achter de 9/11-aanslag, volgens de coöperatieve speltheorie.

North), Hani Hanjour (piloot van vlucht 77, Pentagon), Khalid Almihdhar and Nawaf Alhazmi (beiden kapers van vlucht 77). Zacarias Moussaoui, die werd gearresteerd voor de WTC aanval plaats vond en die bekend staat als de twintigste kaper, staat ook in de top 10. Naast de actieve leden, de piloten en de kapers, staan ook andere leden van het netwerk als Essid Sami Ben Khemais en Djamal Beghal in de top 10. De eerste was hoofd operaties voor Al-Qaida in Italië en de tweede was betrokken bij aanslagen in Frankrijk. Beiden zijn geïdentificeerd door de Amerikaanse overheid als organisatoren van aanvallen op diverse Amerikaanse ambassades.

## Slot

In dit artikel zijn we kort ingegaan op de bijdrage die wiskundige modellen kunnen leveren aan de kwantitatieve analyse van terroristische netwerken. Het grote probleem met dit soort heimelijke netwerken, net zoals heimelijke criminele en andersoortige netwerken, is dat ze in de schaduw opereren en zodanig hun structuur, communicatiepatronen en *modus operandi* hebben ingericht. Elementen uit de coöperatieve speltheorie zijn de revue gepasseerd en het nut ervan in deze context is geïllustreerd. Zowel in het geval dat er geen of onvoldoende complete informatie over dergelijke netwerken beschikbaar is en voor de analyse van deze incomplete informatie kunnen technieken als Nash-onderhandeling en speltheoretische centraliteit ingezet worden. Deze technieken ondersteunen de inlichtingenanalist bij het complexe besluitvormingsproces over heimelijke organisaties: wat is de meest waarschijnlijke organisatievorm van een dergelijk netwerk? Welke personen moeten (meer) onder surveillance komen? Hoe kunnen we beter zicht krijgen op heimelijke netwerken en hoe destabiliseren wij ze? De heterogene en incomplete data die dergelijke heimelijke netwerken genereren kunnen niet alleen handmatig en kwalitatief geanalyseerd en geduid worden. Daarvoor is het probleem te diffuus. Modellen uit de speltheorie en netwerkanalyse leveren belangrijke bouwstenen voor de data-analyse ten behoeve en bevordering van het inlichtingenproces.

## Referenties


1 C. Ballester, A. Calvó-Armengol en Y. Zenou, Who's who in networks. Wanted: the key player, *Econometrica* 74(5) (2006), 1403–1417.

2 A. Bavelas, A mathematical model for group structures, *Applied Anthropology* 7 (1948), 16–30.

3 A. Bavelas, Communication patterns in task-oriented groups, *The Journal of the Acoustical Society of America* 22 (1950), 725–730.

4 S.P. Borgatti en M.G. Everett, A graph-theoretic framework for classifying centrality measures, *Social Networks* 28 (2006), 466–484.

5 K.M. Carley, J. Reminga, N. Kamneva, Destabilizing terrorist networks, in: *NAACSOS Conference Proceedings*, Pittsburgh, 2003.

6 W. Enders en X. Su, Rational terrorists and optimal network structure, *Journal of Conflict Resolution* 51(1) (2007), 33–57.

7 J.D. Farley, Breaking Al Qaeda cells: a mathematical analysis of counterterrorism operations, *Studies in Conflict and Terrorism* 26 (2003), 399–411.

8 B. Jasny en B. Ray, Life and the art of networks, *Science* 301 (2003), 1863.

9 T.H. Kean, L.H. Hamilton, R. Ben-Veniste, B. Kerrey, F.F. Fielding, J.F. Lehman, J.S. Gorelick, T.J. Roemer, S. Gorton en J.R. Thompson, *The 9/11 Commission Report, Final Report of the National Commission on Terrorist Attacks upon the United States*, W.W. Norton and Company, 2002.

10 V.E. Krebs, Uncloaking terrorist networks, *First Monday* 7 (2002), 1–10.

11 R.H.A. Lindelauf, P. Borm en H. Hamers, The influence of secrecy on the communication structure of covert networks, *Social Networks* 31 (2009), 126–137.

12 R. Lindelauf, H.J.M. Hamers en B.G.M. Husslage, Game theoretic centrality analysis of terrorist networks: the cases of Jemaah Islamiyah and Al Qaeda, *European Journal of Operational Research* 229(1) (2013), 230–238.

13 G. McCormick en G. Owen, Security and coordination in a clandestine organization, *Mathematical and Computer Modelling* 31(6–7) (2000), 175–192.

14 C. Morselli, C. Giguère en K. Petit, The efficiency/security trade-off in criminal networks, *Social Networks* 29(1) (2007), 143–153.

15 J. Nash, Two person cooperative games, *Econometrica* 21 (1953), 128–140.

16 M.E.J. Newman, Analysis of weighted networks, *Physical Review E* 70 (2004), 56–131.

17 L. Shapley, A value for $n$-person games, *Annals of Mathematics Studies* 28 (1953), 307–317.

18 M. Sparrow, The application of network analysis to criminal intelligence: an assessment of the prospects, *Social Networks* 13 (1991), 251–274.

19 S. Strogatz, Exploring complex networks, *Nature* 410 (1991), 268–276.

20 D. Watts en S. Strogatz, Collective dynamics of 'small-world' networks, *Nature* 393 (1998), 440–442.

21 D. Watts, The 'new' science of networks, *Annual Review of Sociology* 30 (2004), 243–270.

22 G.L. Zacharias, *Behavioral Modeling and Simulation: From Individuals to Societies*, National Academy of Sciences, 2008.

23 T. van der Zanden, H. Bodlaender en H. Hamers, 'Efficiently computing the Shapley value of connectivity games in low-tree-width graphs', working paper, 2019.

S. Naranan
*Tata Institute of Fundamental Research, Mumbai*
*snaranan@gmail.com*

T.V. Suresh
*TAO Consulting Systems, Chennai*
*sureshtao@gmail.com*

Swarna Srinivasan
*Tata Consultancy Services, Chennai*
*swarna.srinivasan@gmail.com*

# Kolam designs based on Fibonacci series

**In this article S. Naranan, T.V. Suresh and Swarna Srinivasan demonstrate the connection between the beauty of Indian kolams with the elegance of Fibonacci numbers. Kolams are decorative designs in South Indian folk art. They are drawn as lines and curves around a grid of dots. An essential feature is the rotational symmetry of the design — four-fold for squares and two-fold for rectangles. Any set of four integers ($a\ b\ c\ d$) with the property $c = a + b$ and $d = c + b$ can be the basis for kolams based on the Fibonacci series. Symmetry properties of these kolams arise naturally from the recursive property of the Fibonacci series. Hence, the Fibonacci recurrence is an elegant mathematical tool for design of kolams. A figure shows how this recurrence can be used for construction of square and rectangular kolams of different sizes. As examples, kolams with Fibonacci quartets ($1\ 2\ 3\ 5$), ($2\ 3\ 5\ 8$) and ($5\ 8\ 13\ 21$) are shown. A desirable feature of kolams is that they are 'single-loop'. Since each Fibonacci kolam has at least five modules, a single loop can only be achieved with specific splices between modules. Using probability theory and matrices, we examine the relationship between the numbers of splices and loops. Results show that regardless of the initial number of splices, the end result is one, three or five loops. Fibonacci kolams can be drawn on the computer using eight basic shapes derived from 8 equations. Such computer generated kolams are attractively usable as a 14 x 14 board game.**

Every morning before sunrise, women in the south of India clean their doorsteps and draw elaborate figures with rice powder in front of their homes. Through the day, these drawings get walked on, washed out in the rain, or blown away by the wind. A new figure is drawn on the next day. These drawings are thought to bring prosperity to the homes and they are a sign of welcome. They are known as *kolams* and the mathematics behind this folk art is fascinating.

Special kolams, large with bewildering complexity are drawn on festival days. On such occasions, kolam competitions are held in temples. The folk art is handed down through generations of women from historic times, dating perhaps thousand years or more. Today they are nurtured by housemaids and housewives both in rural and urban areas. A kolam is usually drawn around dots, which are trickled on the floor in a regular grid, as a template. The designs are intricate and creative, governed by some broad rules. One common rule is four-fold symmetry: viewed from all the four sides North, East, South and West the kolam appears the same. A Kolam may have multiple loops (the large kolam in Figure 1 has three) but a single loop is considered special and is harder to achieve, see wikipedia.org/wiki/Kolam and www.ikolam.com.

The simplest geometrical shape with four-fold symmetry is the square. Rectangles have only two-fold rotational symmetry. They serve as modules in building larger squares. More complex shapes can be broken down into squares and rectangles. Therefore we consider only square and rectangular kolams and we will see how Fibonacci recurrence can be used to generate them.

Photo: Wikimedia Commons, Arulraja

**Figure 1** Various Kolams in front of a house in Tamil Nadu (India). The three larger Kolams are of the type considered in this paper. The two small Kolams on the red tiles are of a different style.

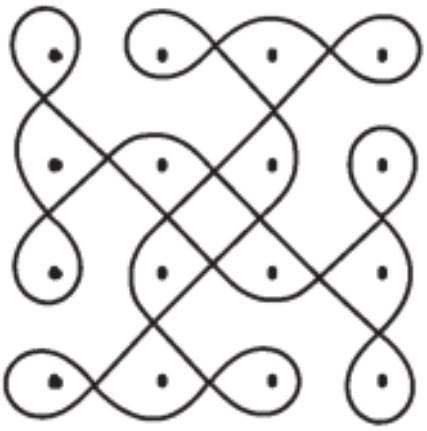


**Figure 2** A square kolam with four-fold symmetry.

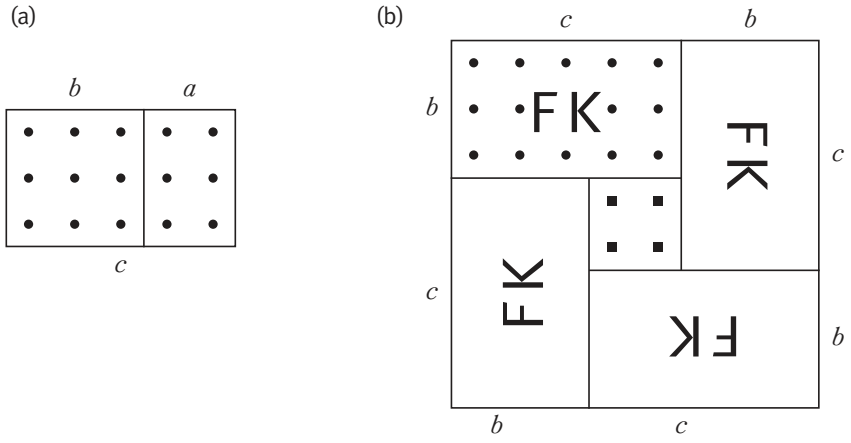



**Figure 3** The quartet $(2\,3\,5\,8)$ generates a template for a kolam.

### Fibonacci recurrence

In Fibonacci recurrence, a series of integers is generated in which every integer is a sum of two preceding integers. Given two seed numbers all the subsequent integers are determined. The simplest such series starts with seeds $0$, $1$ and is familiar to all of us:

$$0\ 1\ 1\ 2\ 3\ 5\ 8\ 13\ 21\ 34\ 55\ 89\ 144\ldots \quad (1)$$

These Fibonacci numbers occur in many branches of science. In geometry they are related to pentagons, decagons and the 3-dimensional Platonic solids. Rectangles with sides as consecutive Fibonacci integers are known as *golden rectangles*. The ratio of the sides approaches a limiting value $\phi = (1+\sqrt{5})/2 = 1.61803\ldots$ called the *golden ratio*. This ratio is ubiquitous in nature: in the branching of trees, arrangement of leaves, seeds and flower petals, spiral patterns of florets in sun flowers, spiral shapes of sea-shells et cetera. The golden ratio is also believed to figure prominently in Western art: in architecture (pyramids, Parthenon) paintings (Leonardo da Vinci) sculpture, poetry (Virgil) and music [4]. Many claims are controversial.

The Fibonacci series appeared in the book *Liber Abaci* (1202) by the Italian mathematician Leonardo of Pisa, also known as Fibonacci. It is claimed that the numbers appear in the analysis of prosody of Sanskrit poetry by Acharya Hemachandra in 1150 A.D., 52 years before *Liber Abaci* [8]. For more on Fibonacci numbers see Martin Gardner [2].

For kolam designs we are interested in a set of four consecutive integers in a generalized Fibonacci series such as $(3\,5\,8\,13)$, $(3\,4\,7\,11)$. In a quartet $Q(a\,b\,c\,d)$ we have

$$\begin{aligned} c &= a+b, \\ d &= b+c = a+2b, \end{aligned} \quad (2)$$

and we leave it to the reader to verify that the $a\,b\,c\,d$ are related as follows:

$$bc = b^2 + ab, \quad (3a)$$
$$d^2 = a^2 + 4bc. \quad (3b)$$

Expressed in standard notation equations (3a) and (3b) are

$$F_{n-1}F_n = F_{n-1}^2 + F_{n-1}F_{n-2} \quad (n>1), \quad (4a)$$
$$F_n^2 = F_{n-3}^2 + 4F_{n-2}F_{n-1} \quad (n>3), \quad (4b)$$

for given $F_0$ and $F_1$.

Equations (3a) and (3b) are the basis of Fibonacci kolams. They have geometrical counterparts. In Figure 3(b) a square $d^2$ has a smaller concentric square $a^2$ and the space between the two is filled up with four rectangles $(b\times c)$ placed in a cyclical pattern. This construction is the essence of Fibonacci kolams. It ensures four-fold symmetry. The rectangles are identical and need have no symmetry property, but the central square $a^2$ should have four-fold symmetry, as in Figure 3(b). It contains golden rectangles with the sides in ratio $\phi = (1+\sqrt{5})/2$ just as in the case of the Fibonacci series (1). Using equations (3a) and (3b) one can build a hierarchy of square and rectangular kolams leading up to any desired size.

### Rangoli

The traditional South Indian kolam is based on a grid of points and is known as PuLLi (dot) kolam or NeLi (curve) kolam in Tamil Nadu. There are also kolams that are free geometric shapes that usually has bright colours. Such drawings are called *rangoli* and they are popular in North India. Special kolams, large with bewildering complexity are drawn on festival days. On such occasions, kolam competitions are held in temples.

A rangoli at Chennai (flickr.com, B. Balaji)

### The Zeckendorf array

Edouard Zeckendorf was an army doctor from Liège with broad scientific and artistic interests. In 1947 he was sent to Pakistan as a member of the United Nations peace keeping force. He stayed there for several years. During that time he wrote a paper on the Fibonacci numbers. The Zeckendorf array is named after him. Each next row in the array starts by selecting the first and third number that have not been used yet, and proceeds by Fibonacci recurrence. To construct a square Fibonacci kolam, one can choose a quartet from this array.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 5 | 8 | 13 | 21 | … |
| 4 | 7 | 11 | 18 | 29 | 37 | 66 | … |
| 6 | 10 | 16 | 26 | 42 | 68 | 110 | … |
| 9 | 15 | 24 | 39 | 63 | 102 | 165 | … |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋱ |

### Square Fibonacci kolams

Using equations (3a) and (3b) a sequence of square kolams can be built in a hierarchical scheme as shown in the 'skeleton diagrams' in Figure 4. The larger square consists of a smaller square and four cyclic rectangles. Assuming the five constituents are all single loop, our task is to merge them together at splicing points which are symmetrically placed to preserve four-fold symmetry and create a single loop. In general, when all the splices are completed the final outcome will be multiple loops. This is achieved by a careful choice of splicing points.

*Some examples of Fibonacci kolams*

The construction is based on $Q(1\,2\,3\,5)$ with $5^2 = 1^2 + 4(2\times 3)$. In Figure 5(a) the square $5^2$ contains four $2\times 3$ rectangles around a central dot. The rectangles are merged around the central dot in a four-way splice ($\diamond$) which results in a single loop kolam. In Figure 5(b) there are eight additional splicing points marked in blue and green. This gives us another single loop Kolam. To illustrate the effect of the choice of splicing points on the number of loops, all the blue splicing points are omitted in Figure 5(c). Here we get a kolam with five loops.

The number of possible $5^2$ Fibonacci kolams depends on the number of distinct $2\times 3$ rectangles and the splicing choices. Below we shall see that there are $30$ distinct $2\times 3$ rectangles each giving a different $5^2$ kolam.

A $3\times 5$ Fibonacci kolam can be obtained by splicing together a $3^2$ and a $2\times 3$ rectangle ($3\times 5 = 3^2 + 2\times 3$). In Figure 6(a), the $3^2$ has three loops but if it is merged with a $2\times 3$ rectangle at three points the result is a single loop $3\times 5$ with two-fold symmetry as in Figure 6(b).

The $8^2$ Fibonacci kolam is based on $Q(2\,3\,5\,8)$: $8^2 = 2^2 + 4(3\times 5)$. A $3\times 5$ Kolam is now available from Figure 6. To get $2^2$ can $Q(0\,1\,1\,2)$ be used? Figure 7(a) is the first attempt, but it has an empty unit cell at the centre. The generator is $2^2 = 0^2 + 4(1\times 1)$. In Figure 7(b) the four $1\times 1$ 'rectangles' are the four circled dots and the $0^2$ can be imagined as a small empty square at the centre with area approaching $0$, i.e. the square collapses into a point with dimension $0$. The clover-leaf pattern with four leaves touching at the centre is the obvious answer (Figure 7(c)).

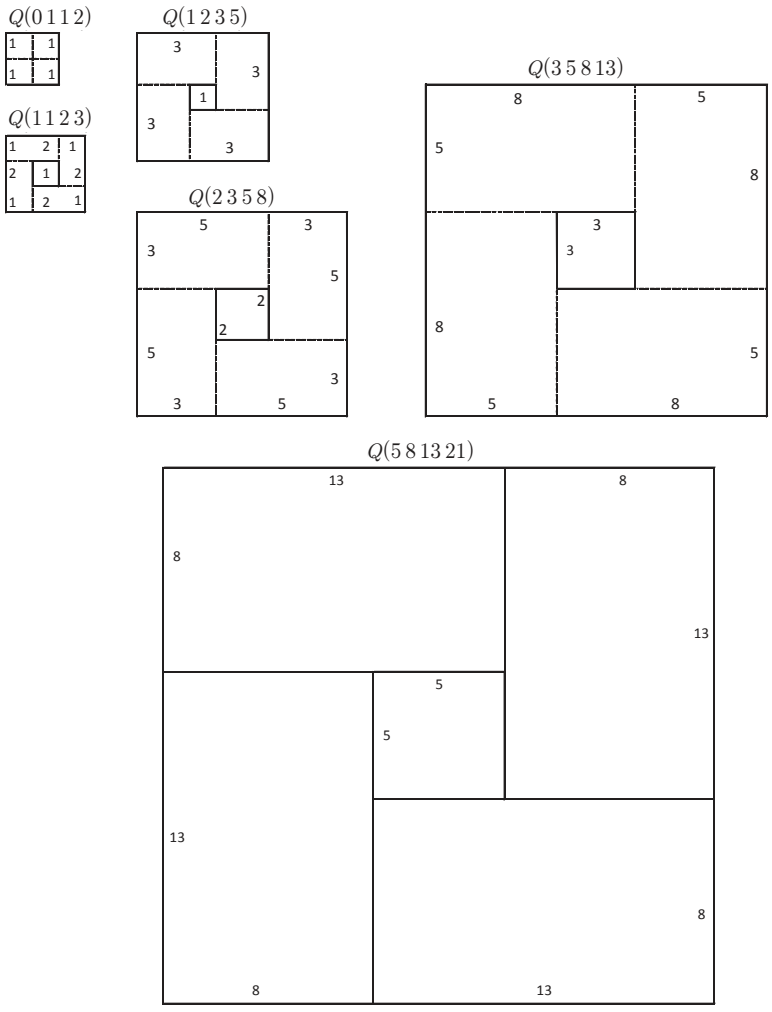



**Figure 4** Skeletons of squares for Fibonacci kolams

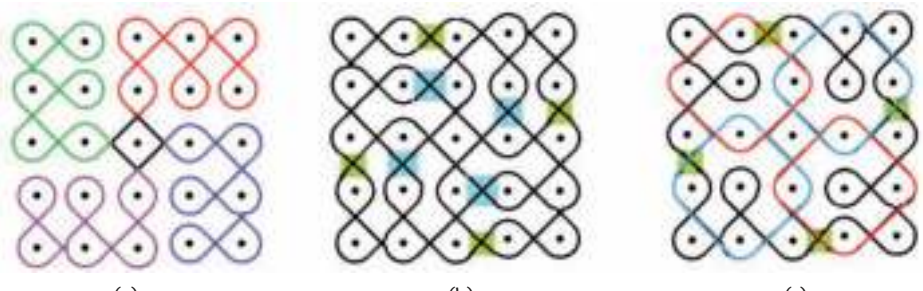



**Figure 5** Three $5\times 5$ kolams with a four-fold symmetry. (b) and (c) are created from (a) by splicing and unsplicing.

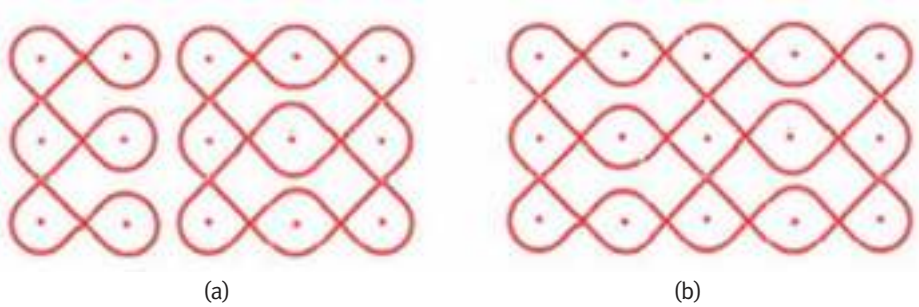



**Figure 6** Creating a $3\times 5$ kolam with a single loop from two separate kolams.

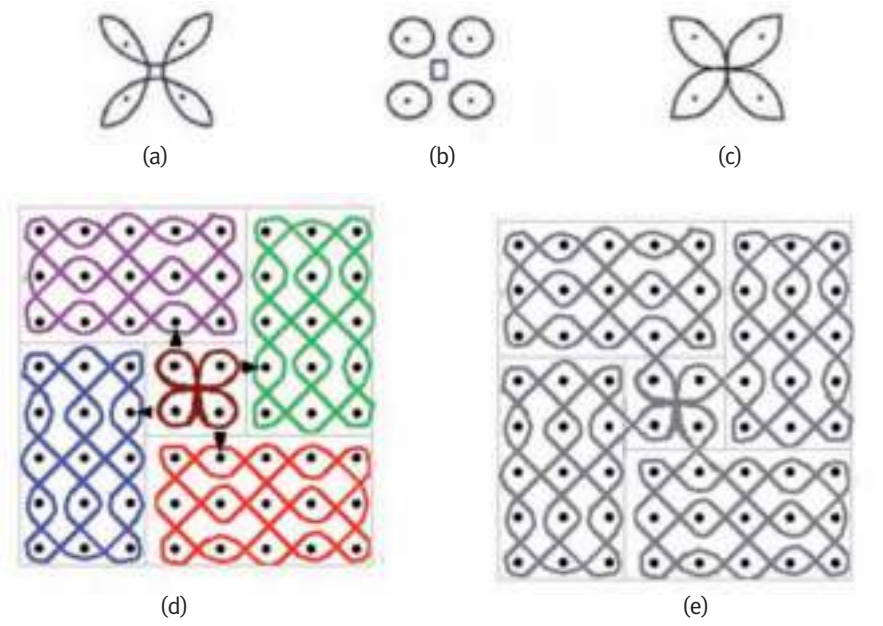



**Figure 7** A Fibonacci $8 \times 8$ kolam.

This pattern occurs rarely in traditional kolams. Clover-leaf patterns generally occur in square Fibonacci kolams whose sides are even and enhance the overall aesthetics of the kolams.

*General construction of Fibonacci kolams*

Every Fibonacci kolam — square or rectangle — is composed of a square and rectangles. This is seen in the recursive equations (4a) and (4b). They can be used to generate square and rectangular kolams of any order. The procedure is graphically presented in Figure 8. There are three blocks (I, II, III) each with two columns. The middle block (II) contains the modules used for building the 'square Fibonacci kolams' in the right block (III) and the 'rectangular Fibonacci kolams' in the left block (I). Modules are shown in italics in block II. For instance, to find the composition of a $13^2$ Fibonacci kolam (in III) follow the arrows

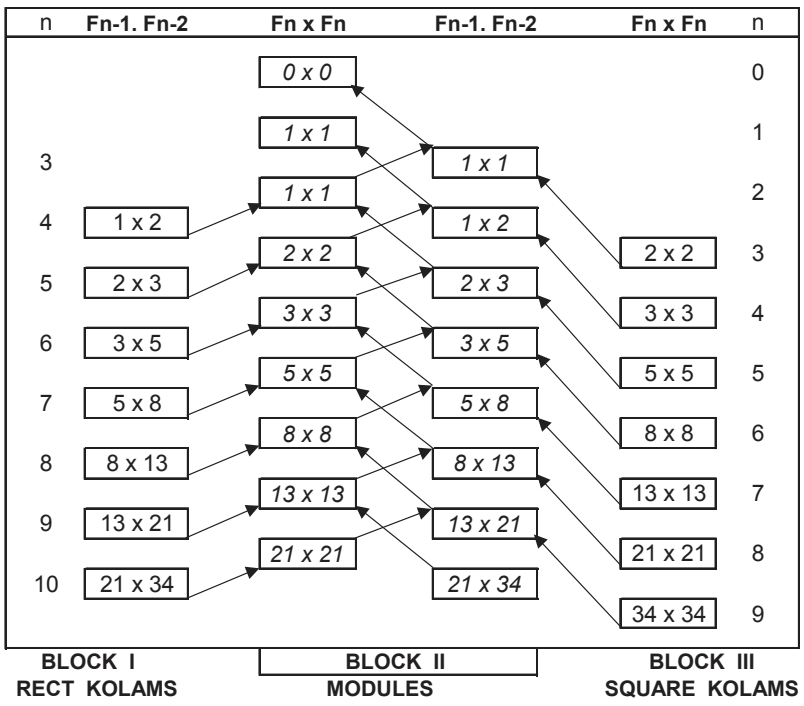



**Figure 8** Recursive procedure for construction of square and rectangular Fibonacci kolams.

leading to module II ($5 \times 8$ and $3^2$). Similarly the $8 \times 13$ rectangle in block I traces its composition to the module $8^2$ and $5 \times 8$ in block II. The origins of all square and rectangular kolams can be traced to $F_1^2 = 1^2$ or $F_2^2 = 1^2$.

Every third Fibonacci number $F_{3m}$ ($m = 1,2,3,\ldots$) is even since in $Q(a\,b\,c\,d)$ $d - a = b + c - a = 2b$ is an even number. Therefore $d$ and $a$ are both either even or odd. Since $F_3 = 2$ is even, so are $F_6$, $F_9$, $F_{12}$,… (8, 34, 144,…). Since $F_1 = F_2 = 1$, all the other Fibonacci numbers are odd.

Figure 9 shows how to build a Fibonacci kolam recursively. The $21^2$ Fibonacci kolam is constructed from an arrangement of four $8 \times 13$ kolams around a $5^2$ kolam. They are spliced together at six sets of points (24 in all). Each $8 \times 13$ kolam on its turn is a union of an $8 \times 8$ and a $5 \times 8$ kolam.

The Fibonacci kolams described so far are all based on the Fibonacci series (1). This is a particular example of a generalized Fibonacci series $G_n$ ($n = 0,1,2,3,\ldots$) where the first two starting numbers $G_0$ and $G_1$ are $\alpha$ and $\beta$. The Fibonacci series corresponds to $\alpha = 0$ and $\beta = 1$. If $\alpha = 2$ and $\beta = 1$, we get the series

$$2\ 1\ 3\ 4\ 7\ 11\ 18\ 29\ 47\ 76\ldots,$$

known as *Lucas series*. As long as the Fibonacci recursion

$$G_n = G_{n-1} + G_{n-2} \quad (n > 1)$$

is used, the quartet relation $Q(a\,b\,c\,d)$ with equation (3a) and equation (3b) applies. In other words in equations (4a) and (4b) $F_n$ can be replaced by $G_n$. This permits generalized Fibonacci kolams of any desired order, i.e. $n \times n$ or $m \times n$ for all $m$, $n$. In particular in the Lucas series 4 and all the succeeding numbers are different from Fibonacci numbers. For example a $4 \times 4$ generalized Fibonacci kolam can be based on $Q(2\,1\,3\,4)$: $4^2 = 2^2 + 4(1 \times 3)$. This is illustrated in Figures 10(a) and 10(b). A variant of $4^2$ based on $Q(0\,2\,2\,4)$: $4^2 = 0^2 + 4(2 \times 2)$ is shown in Figures 10(c)–(f).

How many distinct kolams exist of a given size? This is a problem in combinatorics since kolams are built from smaller sub-units. Starting with the smallest square $2^2$ single-loop kolams, the only symmetric version is the 'cloverleaf' pattern in Figure 11(a). The asymmetric versions in Figure 11(b) are used to build $2 \times 3$ rectangles (as $2^2$ and $1 \times 2$). After enumerating all possible $2 \times 3$ kolams it

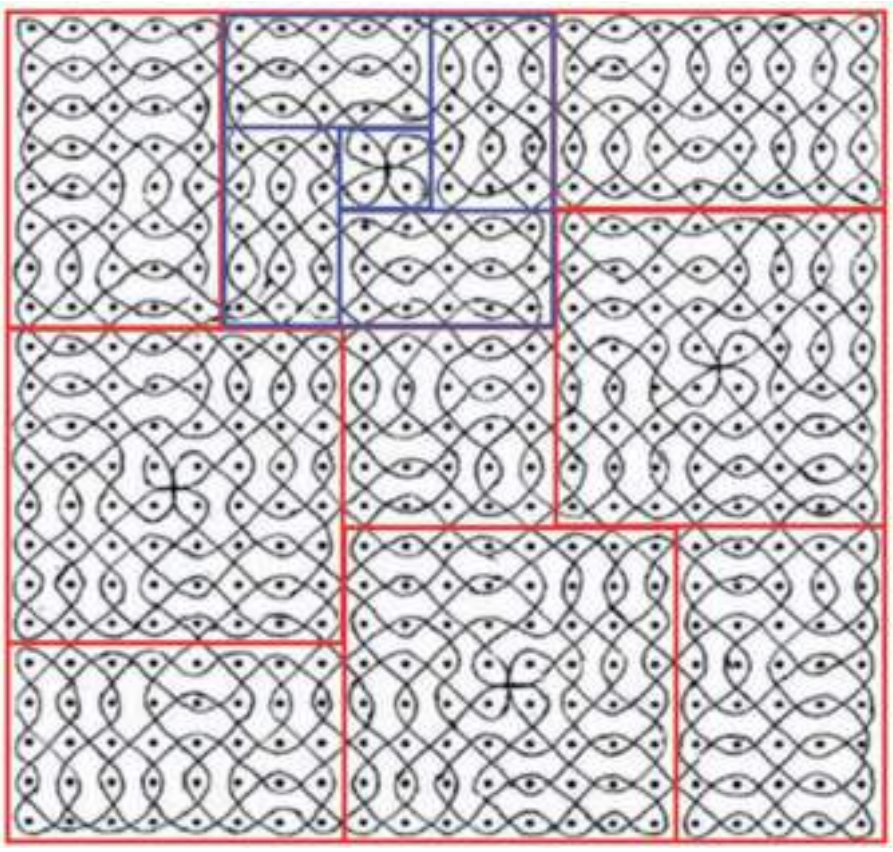


**Figure 9** A $21 \times 21$ Fibonacci kolam drawn on a skeleton diagram.

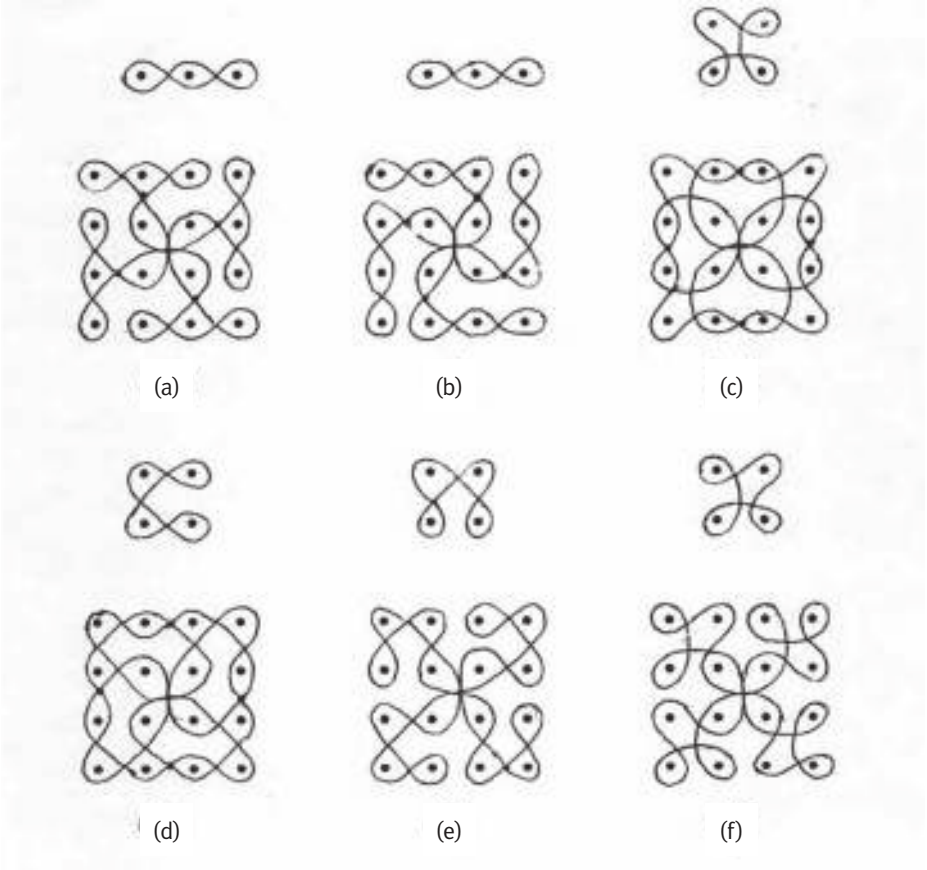



**Figure 10** Building $4 \times 4$ kolams from $3 \times 1$ or from $2 \times 2$ kolams.

was found that the six kolams shown in Figure 11(c) are suitable as basic modules. They are labelled E, R, G, H, U, S, so that the alphabet patterns reflect the shape and symmetry property of the kolams. Enumeration of larger kolams is a complex problem.

*Symmetry properties*

The dihedral group $D_4$ is the symmetry group of the square. It contains the rotations by $0°$, $90°$, $180°$, $270°$ which we label by $I$, $R(90)$, $R(180)$, $R(-90)$. The reflection operators, or mirrors, are $M(X)$, $M(Y)$, $M(45)$ and $M(-45)$ for reflections about $X$-axis, $Y$-axis, diagonal $Y = X$ and the anti-diagonal ($Y = -X$). These eight operators

$$I \;\; R(90) \;\; R(180) \;\; R(-90)$$
$$M(X) \;\; M(Y) \;\; M(45) \;\; M(-45)$$

form the dihedral group. How do these operators alter the shape of ERGHUS kolams? In Figure 12, the basic kolam shapes ($K$) are in the top box. Successive rows show the effect of different operators on $K$. Some interesting features are as follows:

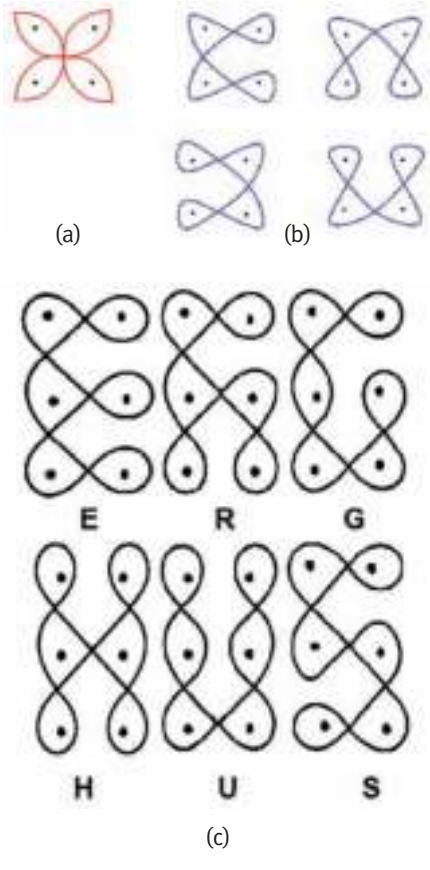



**Figure 11** The $2 \times 2$ and $2 \times 3$ kolams – building blocks of the Fibonacci kolams.

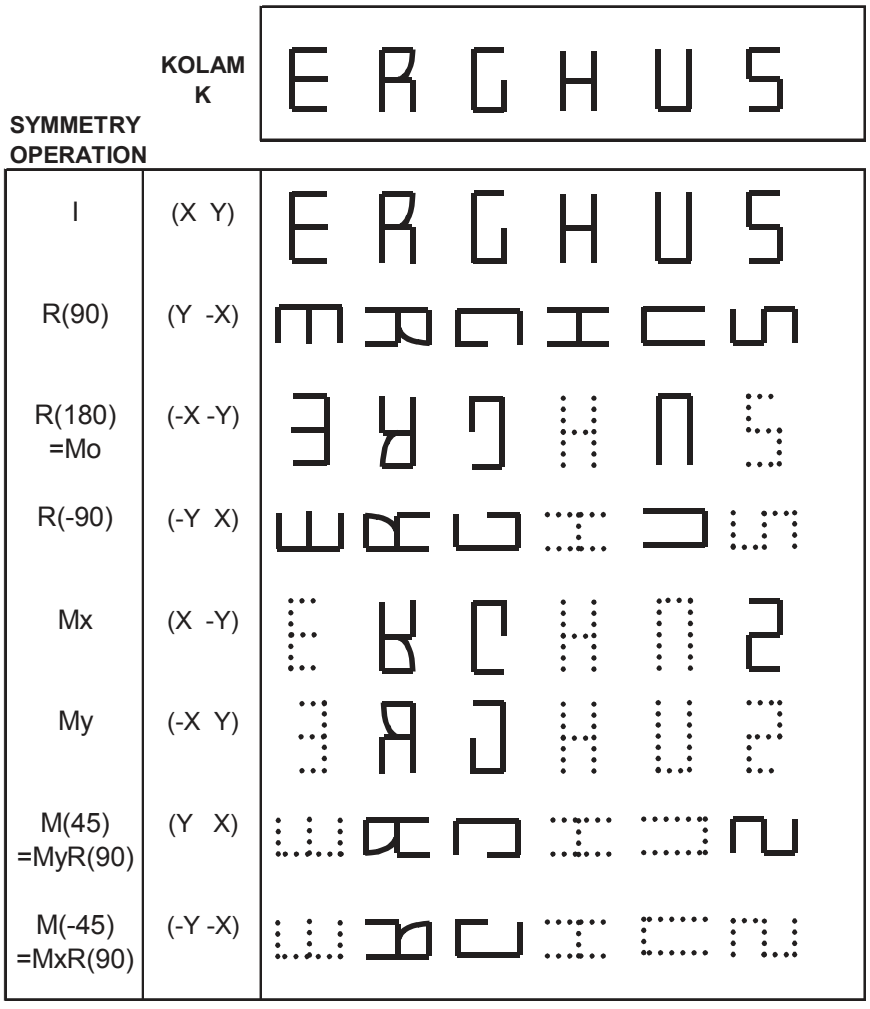
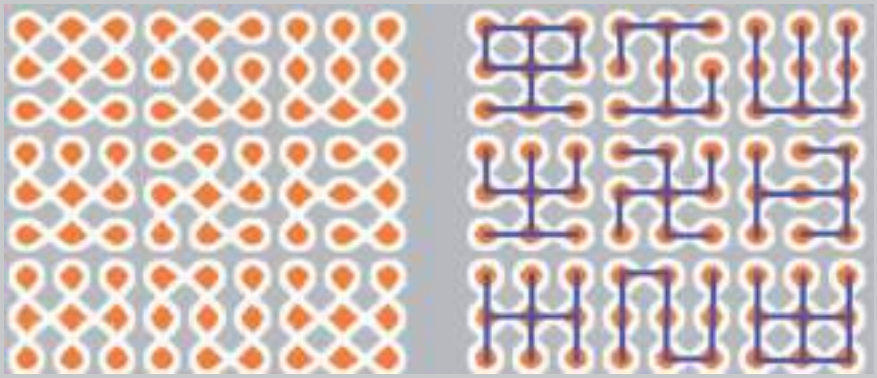



**Figure 12** Transformations of the ERGHUS kolams.

**Nine goddesses**

The Hindu festival of Navaratri, which means *nine nights* in Sanskrit, celebrates the nine incarnations of Shakti. There is a different kolam for each incarnation, all single loops around nine dots. Each kolam has a symmetry. Seven kolams have a single symmetry, a mirror or a point reflection, but two have more. The central kolam has a four-fold symmetry. It is a swastika, which means *good fortune* in Sanskrit. It is not hard to find other single loops around nine dots. However, it is hard to find another single loop which has a symmetry.



Nine kolams for nine incarnations on the left. Graphs display the symmetries of the kolams on the right.

1. Of the total $48$ $(6\times 8)$ patterns only $30$ are distinct. The remaining $18$ are repetitions shown as dotted patterns.
2. Only for R and G all the eight operators result in distinct patterns (columns 2 and 3). For E, U, S only four are distinct. For H only two give distinct patterns [$I$ and $R(90)$].
3. Symmetric kolams often have a special meaning, as described in the text box on the nine goddesses of Navaratri.

**Other Fibonacci kolams**

In generating square kolams invariably rectangular kolams appear as constituents. These rectangles are golden rectangles (ratio of sides $=\phi=1.618..$), but they lack the two-fold rotational symmetry (appearing the same viewed from north or south). 'Rectangles' with two-fold symmetry can be built with two Fibonacci quartets. In analogy with equation (2),

$$Q_1(a_1\, b_1\, c_1\, d_1), \qquad Q_2(a_2\, b_2\, c_2\, d_2),$$
$$c_1=a_1+b_1, \quad d_1=c_1+b_1=a_1+2b_1, \tag{5a}$$
$$c_2=a_2+b_2, \quad d_2=c_2+b_2=a_2+2b_2, \tag{5b}$$

An identity relating all the numbers, in analogy with equation (3b) is

$$d_1 d_2 = a_1 a_2 + 2b_1 c_2 + 2c_1 b_2. \tag{6}$$

The rectangle of sides $d_1 d_2$ is composed of a concentric rectangle $a_1 a_2$ and two pairs of rectangles $b_1 c_2$ and $c_2 b_1$. In each pair, one is rotated with respect to the other to ensure two-fold rotational symmetry. In addition, if rectangle $a_1 a_2$ has two-fold symmetry, the overall kolam $d_1 d_2$ has two-fold symmetry. The construction is illustrated in a $6\times 7$ kolam in Figure 13 using

$$Q_1(4\,1\,5\,6), \qquad Q_2(1\,3\,4\,7).$$

This $6\times 7$ kolam is named *Kaprekar kolam* [5,6] as all the four digits of the Kaprekar constant $6174$ appear in the quartets $Q_1(4\,1\,5\,6)$ and $Q_2(1\,3\,4\,7)$.

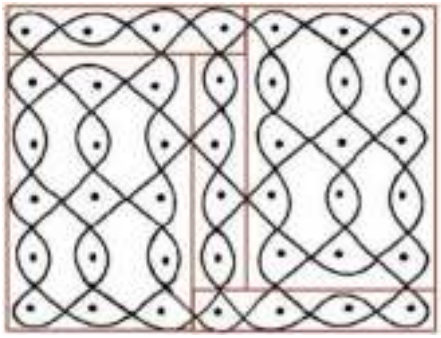


**Figure 13** Kaprekar kolam drawn on its skeleton diagram.

**Kaprekar's constant**
Take any four digits, not all equal. Arrange them in descending and ascending order. Subtract to get four new digits. Repeat. D.R. Kaprekar, an Indian school teacher who published a great number of recreational math papers, discovered that this process always ends up at $6174$. This is now called Kaprekar's constant.

$$8765 - 5678 = 3087$$
$$8730 - 0378 = 8352$$
$$8532 - 2358 = 6174$$
$$7641 - 1467 = 6174$$

Rectangular kolams with two-fold symmetry can be drawn with any desired $d_1$ and $d_2$. A rectangle of sides $7 \times 22$ has the ratio $22/7 = 3.1428\ldots$ which is a very good approximation to $\pi$. A rectangle with sides $7 \times 19$ has the ratio $19/7 = 2.7412\ldots$ which is a close approximation to $e$, the natural base for logarithms. Both $\pi$ and $e$ along with the golden ratio $\phi$, are among the most important mathematical constants.

The $\pi$-kolam is coded by the quartets

$$Q_1(1\,3\,4\,7),$$
$$Q_2(6\,8\,14\,22).$$

At the centre is the linear string $1 \times 6$ and the enveloping rectangles are $3 \times 14$ and $4 \times 8$ (in pairs). They serve as modules for the $7 \times 22$ kolam in Figure 14.

In Figure 15 is drawn a $9 \times 13$ kolam based on the quartets

$$Q_1(5\,2\,7\,9),$$
$$Q_2(7\,3\,10\,13).$$

The constituent rectangles $5 \times 7$ (at the centre) and the surrounding rectangles $2 \times 10$ and $7 \times 3$ are spliced together at $2 \times 10$ points to form a $9 \times 13$ single-loop kolam.

If we remove the central $5 \times 7$ rectangle from this kolam, then we get the 'window-frame' kolam in Figure 16. A careful inspection reveals that it has two loops. It can be shown that a window-frame kolam can never be single loop whatever be the choices of quartets and splices. This is due to the fact that there is no central rectangle. Since the starting configuration has four loops, an even number, parity conservation dictates the minimum number of loops as two. We consider the number of loops in the next section.

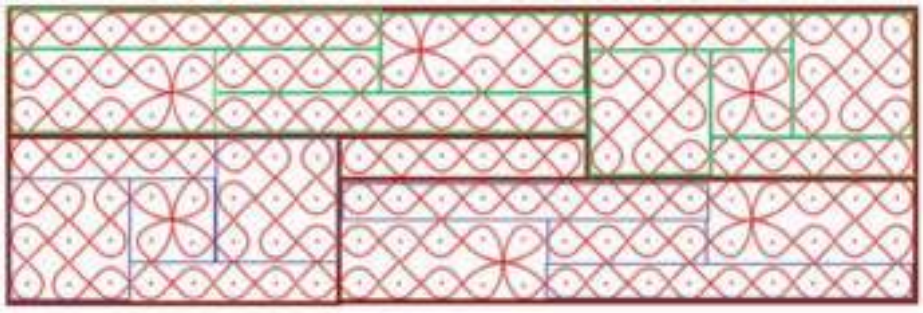


**Figure 14** The $\pi$-kolam drawn on its skeleton diagram.

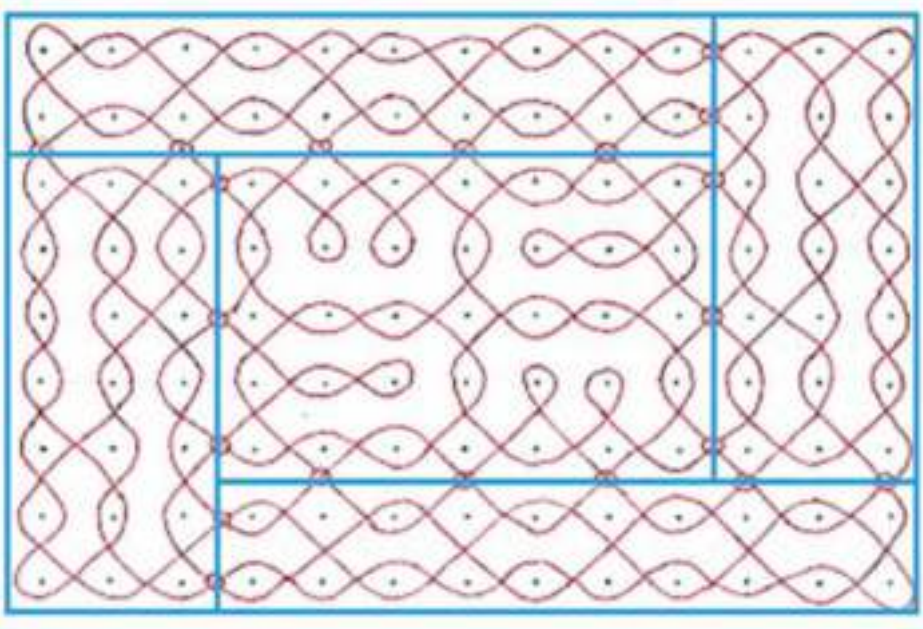


**Figure 15** A single loop $9 \times 13$ kolam.

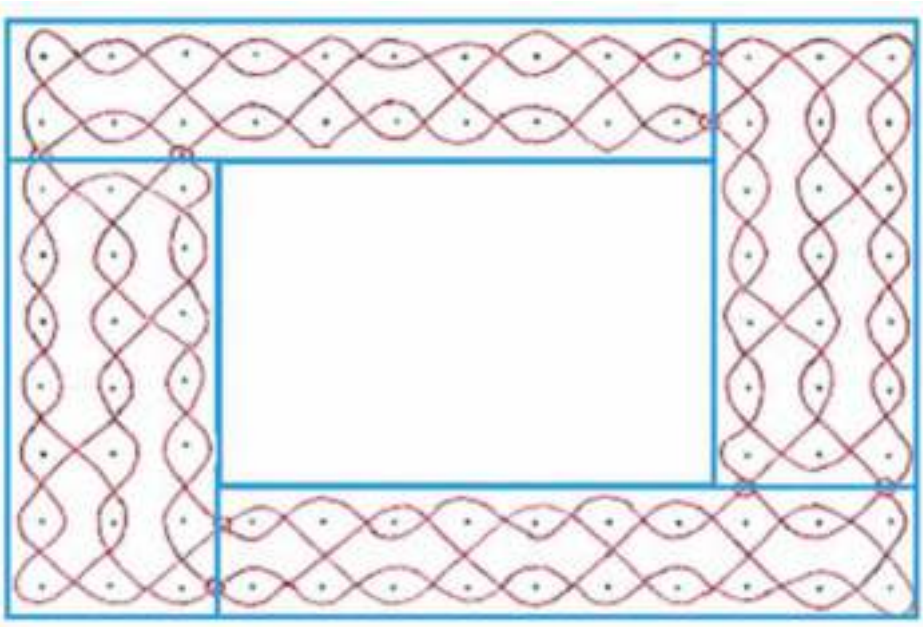


**Figure 16** A window frame kolam with two loops.

**Evolution of loops**

A Fibonacci kolam is assembled by merging or splicing five smaller modules at a set of points. In general, the final kolam has multiple loops. A single loop is achieved by a suitable choice of splices. Another requirement prompted by aesthetic reasons is that 'four-sided islands' are not allowed (see section on 'Square Fibonacci kolams'). To elaborate: islands are empty bounded regions without a dot. A four-sided island appears like a diamond without a dot (◊). Four-sided islands arise mainly when splices are made at adjacent points. So generally splices are made at alternate points. They occur along the edges of the constituent rectangles. In practice, a good strategy to obtain maximal splicing consistent with a single loop is the following. Splice all allowed splicing points at one 'go'. If the number of loops is one, the task is done. If not, unsplice one or more sets of four points to make the kolam single loop; the choice of the sets will require some 'trial and error' experimentation.

*Splicing and unsplicing rules*

From a detailed analysis of loop evolutions the following empirical rules emerge.

1. A splice between two loops gives one loop.
2. A splice within a single loop splits it into two loops.

As a converse to above, the unsplicing rules are the following.

1. Unsplicing at the intersection of two loops gives one loop.
2. Unsplicing within a single loop will split it into two loops.

If there are $l\,(>1)$ loops, an operation (splice or unsplice) results in $l-1$ or $l+1$ loops, a binary option. A set of four splices can be modelled as a binary tree with four nodes.

*Loop evolution as a binary tree*

In Figure 17 a 'top-down' tree has its root at the top which stands for $1$ (number of loops). After the first splice there are two loops. The second splice alters the number of loops to one or three at the next level depending on whether the splice is in a single loop or between two loops. We assume this is random and assign equal probabilities $(\frac{1}{2})$ to the two branches. If the third splice occurs in a single loop the result is two loops with probability one. If it occurs in one of the three loops the outcome is two or four loops with probabilities $\frac{2}{3}$ and $\frac{1}{3}$, respectively. Among $l$ loops if a random splicing point is chosen between two loops, the loops are likely to be the same loop with probability $\frac{1}{l}$ and different with probability $\frac{l-1}{l}$. Hence the branches $3\rightarrow 2$ and $3\rightarrow 4$ are assigned probabilities $\frac{2}{3}$ and $\frac{1}{3}$. After the fourth and last splice the end points are $1$, $3$ or $5$. Their relative probabilities are calculated by multiplying the probabilities assigned to each evolutionary path. For example $l=1$ is the end point of two pathways 1-2-1-2-1 or 1-2-3-2-1. Their respective probabilities are $1\times\frac{1}{2}\times 1\times\frac{1}{2}=\frac{1}{4}$ and $1\times\frac{1}{2}\times\frac{2}{3}\times\frac{1}{2}=\frac{1}{6}$. Together they add up to $\frac{5}{12}\,(\approx 0.417)$. A similar calculation for $l=3$ as the end point shows three pathways 1-2-1-2-3, 1-2-3-2-3 and 1-2-3-4-3 with probabilities $\frac{1}{4}$, $\frac{1}{6}$, $\frac{1}{8}$ adding up to $\frac{13}{24}\,(\approx 0.542)$. Finally the probability of $l=5$ is $1\times\frac{1}{2}\times\frac{1}{3}\times\frac{1}{4}=\frac{1}{24}\,(\approx 0.041)$. Summarising, starting from a single loop, after a set of four splices the numbers of loops is $1$, $3$ or $5$ with probabilities $0.417$, $0.542$ and $0.041$, respectively.

Notice that a Fibonacci kolam is built from five modules and therefore we need to splice five loops. The evolution of three loops is shown in Figure 18. We leave the evolution of five loops as an exercise to the reader. As for $i=1$, here too the branches are labelled with probabilities. Starting with $i=3$, the probability of final $j=7$ is $\frac{1}{3}\times\frac{1}{4}\times\frac{1}{5}\times\frac{1}{6}=\frac{1}{360}\;(\approx 0.0028)$ less than $0.3\%$. How do the loops evolve further when another four-splice is added? Ultimately we have to deal with a large number of four-splices. This can be handled elegantly using matrices of loop probabilities.

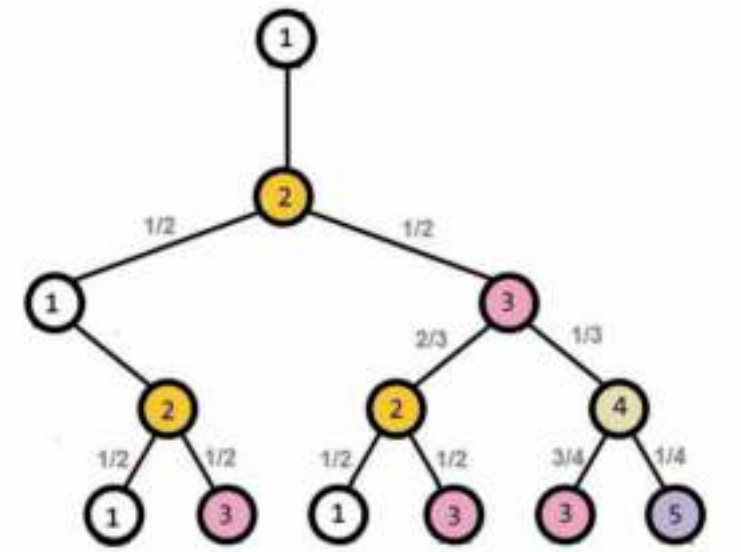



**Figure 17** The binary tree associated to splicing a single loop kolam four times, including the transition probabilities. Four-fold symmetry requires four splices.

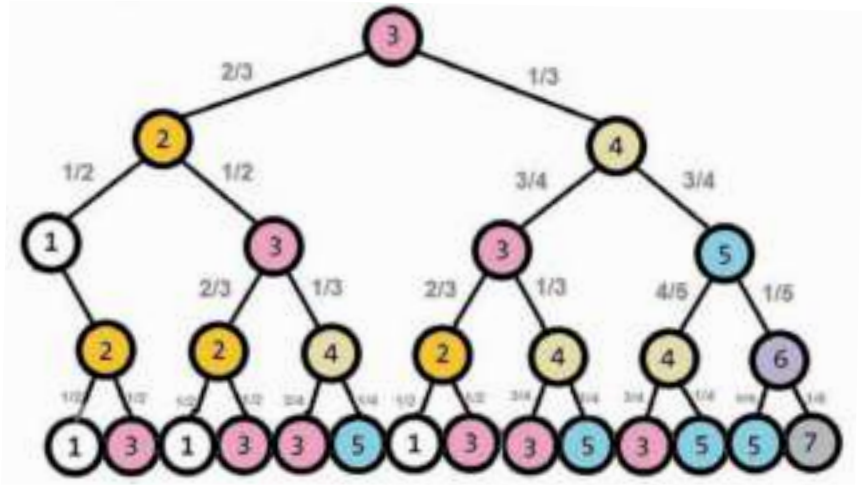
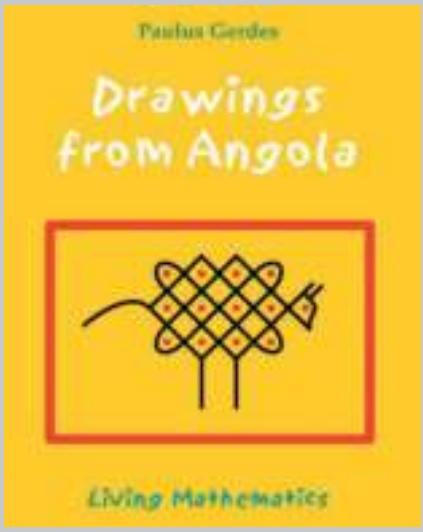



**Figure 18** The binary tree for splicing three loops illustrates the bias towards a small number of loops.

**Curvy loops around the globe**

Drawing loops around a dotted grid is a common practice in several places around the globe. In South-West Africa, the Chokwe and Luchazi people draw curves called *lusona* in the sand. These curves can depict fables or riddles or animals. Similar curves can be found on the Vanuata Islands (New Hebrides) in the Pacific Ocean. Paulus Gerdes, a Dutch mathematician who lived and worked in Mozambique, wrote several books and papers about them. Scientific works such as [3], but also children's books. A kolam or a lusona is drawn by simple hand movements, turning left and right. It is natural to study the movements from an algorithmic point of view. This was done by Gabrielle Allouche, Jean-Paul Allouche and Jeffrey Shallit. They constructed a kolam from the Thue–Morse sequence in [1].




The cover of a children's book by Paulus Gerdes

*Loop probability matrix.*

Let $p(ij)$ be the probability of initial $i$ loops leading to $j$ loops for $i,j = 1,3,5$ as in Figure 17. The probabilities define a $3\times 3$ matrix $S$ of nine elements

$$\begin{matrix} p(11) & p(13) & p(15) \\ p(31) & p(33) & p(35) \\ p(51) & p(53) & p(55) \end{matrix}$$

We have already determined $p(11) = 0.417$, $p(13) = 0.542$, $p(15) = 0.041$. The full matrix is

| $i\backslash j$ | 1 | 3 | 5* |
|---|---|---|---|
| 1 | 0.417 | 0.542 | 0.041 |
| 3 | 0.361 | 0.557 | 0.082 |
| 5 | 0.200 | 0.570 | 0.230 |

*These probabilities include small contributions from $j = 7,9$.

How do the loop probabilities change after two sets of splices? Let the new probabilities be $q(ij)$. Then

$$q(ij) = p(i1)p(1j) + p(i3)p(3j) + p(i5)p(5j) \quad \text{for } i = 1,3,5. \quad (7)$$

Equation (7) is exactly the same as the rule for multiplication of matrix $S$ by itself. Matrix $|q(ij)|$ is simply $S * S\,(= S2)$, where $*$ denotes matrix multiplication. To find the probability after three sets of four-splices multiply $S2$ by $S$ to get $S3$, etcetera. These matrices are presented in Figure 19. The matrix elements quickly converge to nearly constant values.

$$p(11) = 0.37, \quad p(13) = 0.55, \quad p(15) = 0.08.$$

$p(ij)$ is the limiting value of probability after a large number of four-splices. Notice all the rows are nearly the same. The probability of loops after a large number of four-splices is independent of the starting number. It is the stationary distribution of this Markov chain.

The most probable number of loops at the end-point is three with a probability of $0.55$ and the probability of the desired single loop is $0.37$. This leaves a small probability of $0.08$ for the probability of five or more loops. The relative probabilities for $j = 1$ and $3$ are roughly $2{:}3$. These results are consistent with an empirical observation that three loops is the most likely end result even for large kolams with a large number of splices.

One can estimate the maximum number of splicing points available in constructing a Fibonacci kolam. Splices occur along edges of the rectangles shown in Figure 3. Since splices are chosen at alternate points along the edge, the maximum number of splices along an edge is $\approx c/2$ for square kolams. Since four-fold symmetry forces splicing at four symmetrically placed points, the maximum, number of splices is $\approx 2c$. For rectangular kolams the maximum number of splices $\approx c_1 + c_2$. In large kolams of size $d^2 (d = 20$–$30)$, $c \approx 0.6d$ and the maximum number of splices is $1.2d$ or $24$–$36$. The number of four-splices is $6$–$9$.

It is a remarkable fact that even with $\approx 30$ splices the final number of loops $j$ is most likely $1$, $3$ or $5$ with relative probabilities $0.37$, $0.55$ and $0.08$. This is because

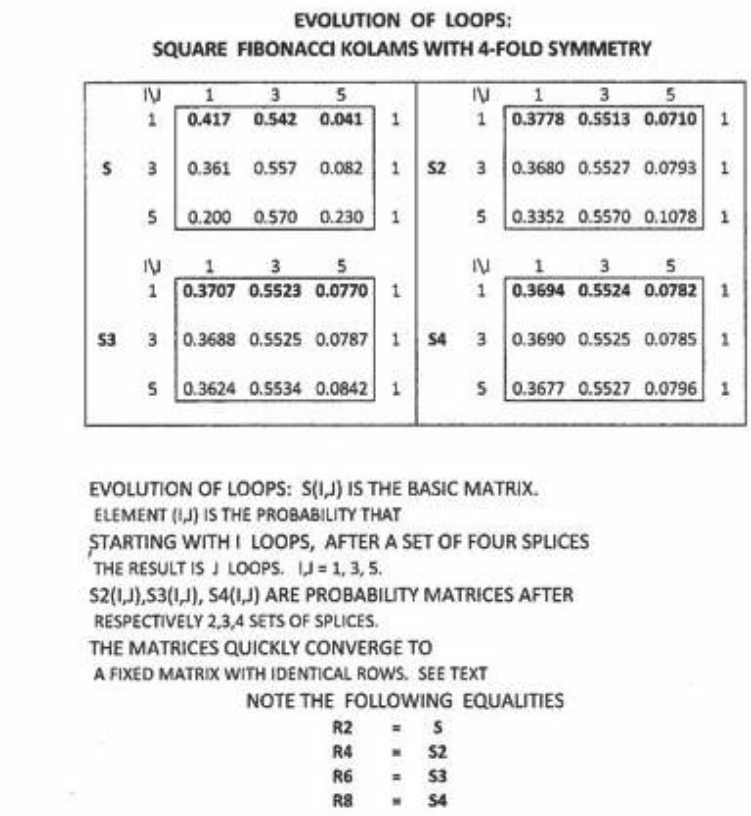



**Figure 19** Convergence of the transition matrices towards the stationary distribution.

the pathways of loop evolution tend to alternately rise and fall keeping the final $j$ to low values as illustrated by the binary tree diagrams Figures 17 and 18. What goes up must come down!

The basic principles underlying the Fibonacci kolams are well understood and the rules for splicing/unsplicing the parts into a whole are simple. More information can be found in [7].

**Concluding remarks and summary**

Kolams are artistic geometrical drawings with curves and loops around dots in a square grid. Fibonacci Recurrence is used to create a new family of kolams with ground rules — single loop, symmetry (four-fold for square, and two-fold for rectangle) and avoidance of four-sided islands (previous section) — dictated by aesthetics. Using generalized Fibonacci series, kolams of any desired size are possible.

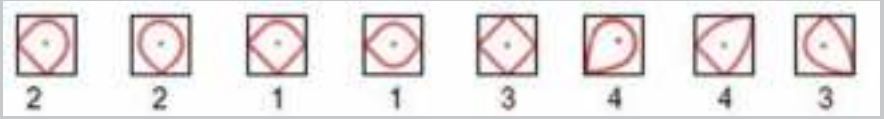
Although squares and rectangles are described in separate sections, in practice they are inseparable. Square kolams as well as rectangular kolams are composed of smaller squares and rectangles as implied by equations (3a) and (3b). The rectangles in section 'Square Fibonacci kolams' are not two-fold symmetric since they are based on only one quarter $Q(a\,b\,c\,d)$. Rectangles in section 'Other Fibonacci kolams' are two-fold symmetric since they are based on two quartets. Hierarchical structures involving squares and rectangles are possible in building kolams of any arbitrary size. The basic modules $2^2$, $3^2$, $2\times3$ are based on quartets $Q(0\,1\,1\,2)$ and $Q(1\,1\,2\,3)$. The $2^2$, the clover-leaf pattern is unique among kolam patterns. Enumeration of the number of these basic modules invokes symmetry operators of group theory.

Merging five modules of a Fibonacci kolam at splicing points along their edges to produce a single loop, is governed by simple rules. As the number of splices increases the number of loops goes up and down resulting in a small number of loops $1$, $3$ or $5$ at the end. This can be shown as a result of the fact that as each splice is added the number of loops is increased or decreased by one. This leads to modelling the evolution of loops as a binary tree and expressing the loop probabilities as elements of a $3\times3$ matrix (previous section). The relative probabilities of the number of loops $1$, $3$, $5$ converge to limiting values of $0.37$, $0.55$ and $0.08$, respectively, as the number of splices becomes large. After each set of four-splices the number of loops changes by $0$, $2$ or $4$, so that the parity of initial and final number of loops is conserved. Since the starting number of loops is odd $(5)$ the final number of loops is also odd $(1,\ 3$ or $5)$. When the starting number is even $(4)$, as in the case of window-frame *kolam*, the final number of loops is $2$ since the parity has to be even.

It is remarkable that all Fibonacci kolams can be assembled with only eight basic shapes (see box above). Rotations and reflections give $31$ distinct shapes. This enables designing a board game for assembling kolam designs on a board with a square grid. It can be played by one or more players. The game has been adapted for play on a computer. Feedback from more than 800 users has been positive. The game is also physically realised with a $14\times14$ square grid made of thick cardboard and $300$ white tiles with kolam shapes etched on them.

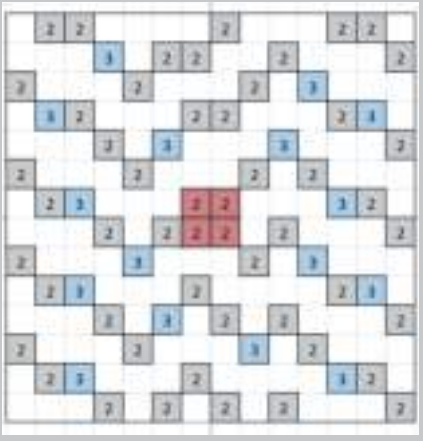
**The Kolam Game**

Kolams are built from single cells with a single dot encircled by a loop that touches one or more sides. There are eight basic cell patterns. All tiles have a mirror symmetry, except for the last one.




The 8 cells as tiles for a board game.

The *Kolam Game* is played with ten tiles, the eight basic tiles and the mirror of the unsymmetric tile and a blank. Each tile counts for a number of points, as shown below the tiles, with no points for the blank. Players draw tiles from the bank and place them on the board. Tiles have to be contiguous. The game is played on a board, shown on the right. The initial tile has to be placed on one of the central red squares. As in the game of *Scrabble*, there are special squares which multiply the value of the tile. By now, this game has been tested by more than 800 users. For details on the rules of the game, see [7].



The board of the game (14×14).

**Acknowledgment**

This paper is mainly based on a series of papers on Fibonacci Kolams on the website of the author, see vindhiya.com/snaranan/fk [7].

**References**

1 G. Allouche, J.-P. Allouche en J. Shallit, Kolam indiens, dessins sur le sable aux îles Vanuatu, courbe de Sierpiński et morphismes de monoïde, *Ann. Inst. Fourier (Grenoble)* 56(7) (2006), 2115–2130.

2 Martin Gardner, *The Mathematical Circus*, The Mathematical Association of America, 1992.

3 P. Gerdes, Interweaving geometry and art: examples from Africa. *Aesthetics of Interdisciplinarity: Art and Mathematics*, Birkhäuser/Springer, 2017, pp. 181–195.

4 Mario Livio, *Golden Ratio: Story of Phi the World's Most Outstanding Number*, Broadway Books 2003.

5 S. Naranan, A variant of Kaprekar Iteration and Kaprekar Constant, https://vindhiya.com/snaranan/mki/mkirevised3.pdf.

6 S. Naranan, Kaprekar Constant 6174, https://vindhiya.com/Naranan/kc.

7 S. Naranan, Kolam designs based on Fibonacci numbers, Parts I–V and Overview, https://vindhiya.com/snaranan/fk.

8 Paramananda Singh, Acharya Hemachandra and the so called Fibonacci Numbers, *Math. Education* 20 (1986), 28–30.

Daan Mulder
*mulderdaan@live.nl*

## Het keerpunt van Marjolein Kool

# "Kom maar op, Kool, waar blijft jouw bijdrage?"

**Wie wiskunde heeft geleerd met de methode *Getal & Ruimte*, moet ze haast wel zijn tegengekomen: de vrolijke wiskundegedichten van Marjolein Kool (1958). Een gedicht over bewijzen uit het ongerijmde (dat inderdaad niet rijmt), bijvoorbeeld, of een gedicht over een kubus die liever als piramide was geboren. Kool was (en is nog steeds) wiskundedocent aan de pabo van de Hogeschool Utrecht en een succesvol schrijfster van light verse, maar pas na een toevallige ontmoeting met Drs. P werd ze op het idee gebracht om wiskundegedichten te gaan schrijven. Ter gelegenheid van de 100ste geboortedag van Drs. P verscheen een heruitgave van hun gezamenlijke dichtbundel '*Wis- en natuurlyriek*', met 'verse verzen' van 'K.', zoals Kool, ook nog gepromoveerd Neerlandicus, zichzelf bescheiden genoeg noemt.**

*Hoe ben je met het schrijven van gedichten begonnen?*
"Dat deed ik al vanaf de lagere school. Ik heb nog een oud schriftje met gedichten die ik in de zesde klas, groep acht heet dat nu, heb geschreven. Dat waren toen ook al grappige versjes. Jaren later mocht ik in de introductieweek van de Universiteit Utrecht optreden, weer met komische gedichten. Toen was er een journalist, Nico Scheepmaker, die mij hoorde. Hij schreef daarover in zijn column Trijfel, en daar reageerden uitgevers op. Ik heb nog met hem samen een selectie uit mijn hele pak versjes gemaakt. Het wrange was dat hij is overleden aan een hartstilstand toen de proefdrukken werden klaargemaakt, in 1990. Dat was heel heftig. Hij had het immers allemaal voor mij geregeld, maar mocht het eindresultaat niet meemaken. Die publicatie ging wel gewoon door en het boek verkocht goed, dus de uitgever zei: 'Maak nog maar een boek als je wilt.' Dat is wel een gouden start geweest.

Op een poëzieavond waar ik optrad heb ik op een gegeven moment Drs. P ontmoet. Dan spreek je elkaar achter de schermen, dat vond ik al heel bijzonder. Hij vroeg wat ik deed en toen hij hoorde dat ik wiskundeleraar was, dacht hij dat ik vast ook wel wiskundegedichten had geschreven. Maar dat had ik tot dan toe nog nooit gedaan! Hij zei: 'Jij schrijft de wiskundegedichten, ik doe natuur- en scheikunde en dan maken we een boek.' Ik dacht: 'Dat meen je niet, mijn grote idool wil een boek met mij schrijven! Misschien heeft hij wel teveel gedronken, morgen is hij het weer vergeten.' Maar na een week of drie lag er al een stapel gedichten van hem op de mat. 'Kom maar op, Kool, waar blijft jouw bijdrage?' Dat betekende dat ik linksom of rechtsom die wiskundegedichten moest gaan schrijven. Zonder Drs. P was ik dat denk ik nooit gaan doen."

Marjolein Kool

*Terwijl je toch ook al wiskundeleraar was.*
"Ja, en ik gaf ook nog Nederlands. En naast het lesgeven ben ik ook Middelnederlandse letterkunde gaan studeren aan de Universiteit Utrecht. Als doctoraalscriptie deed iedereen *Karel ende Elegast* of *Van den vos Reynaerde* of *Mariken van Nimwegen*. Ik wist niet zo goed wat daar nog over te onderzoeken zou zijn. Toen een van die hoogleraren, professor Gerritsen, hoorde dat ik wiskundeleraar was, zei hij: 'Oh wiskunde! Er is een rekenboekje gevonden uit 1532, het ligt in Gent en ik zou het wel leuk vinden als een van mijn studenten daar op zou afstuderen, maar niemand wil het.' Nou, ik wilde wel! Dus ik ben naar Gent gegaan om dat te bekijken. Ik zat er een hele dag en toen had ik één pagina van het manuscript ontcijferd. Het was moeilijk, maar zo leuk dat ik dacht: 'Nou, kom maar op.' Dus dat is mijn doctoraalonderzoek geworden.."

*Wat was er aan die ene pagina zo leuk?*
"Ik weet nog dat het ging over een vrouwtje met een mandje eieren. Toen ze die eieren twee aan twee in de mand had gelegd hield ze er één over, drie aan drie hield ze er ook één over, enzovoort, tot zeven, geloof ik. En daaruit kun je afleiden hoeveel eieren ze dan bij zich kan hebben gehad. In dit soort boekjes gaat het eigenlijk heel veel over praktisch rekenwerk uit de realiteit: Een koopman verkoopt zo veel baal katoen voor die prijs. Hij wil daar peper voor die prijs voor terug hebben. Hoeveel peper krijgt hij voor zijn katoen? Maar er zitten ook vraagstukken bij die juist heel

onrealistisch zijn. Later heb ik ontdekt dat dat vaak de klassiekers zijn, die stonden al in heel erg oude opgaveverzamelingen, en die zijn dan ook weer overgenomen in die praktische rekenboekjes uit de zestiende eeuw.

Ik heb vooral de woordenschat met betrekking tot rekenen uit de zestiende eeuw bestudeerd. Die praktische rekenboekjes waren bijvoorbeeld voor kooplieden, klokkengieters, timmerlieden en boekhouders, die meestal geen Latijn kenden. Dus werden die boekjes vertaald naar het Nederlands, maar er waren nog geen Nederlandse woorden voor optellen, vermenigvuldigen, enzovoort. In het begin werden de Latijnse termen overgenomen. Later gingen vertalers pogingen doen om ze te vertalen. Voor optellen gebruikten ze bijvoorbeeld 'vergaderen', 'samendoen', 'bijeennemen' en op een keer ook wel 'optellen'. Maar het is niet zo dat toen 'optellen' eenmaal werd gebruikt, dat dat vanaf dat moment de standaardterm was. Van Simon Stevin wordt altijd gezegd dat hij de Nederlandse wetenschappelijke of wiskundige woordenschat heeft bedacht. Voor een deel klopt dat, maar voor een deel heeft hij vaktermen gekozen die al beschikbaar waren. En omdat hij aanzien had, is de rekenwoordenschat die hij heeft gekozen voor het grootste deel nu nog steeds in gebruik."

*Kom je uit een wiskundegezin?*
"Nou, niet uitgesproken. Mijn ouders waren wel slim maar ze moesten na de lagere school aan het werk. Ik kom uit een gezin van zes, en de meiden waren de oudsten. Voor ons was studeren in die tijd ook nog niet helemaal vanzelfsprekend. Mijn oudste zus werd eerst secretaresse — die heeft later nog wel gestudeerd —, mijn andere zus werd basisschoolleerkracht, en ik gaf les op de middelbare school. Je mocht wel studeren, maar de universiteit was misschien niet voor ons soort mensen. En toen kwamen de jongens, en die gingen ineens wel! Dus toen ik de lerarenopleiding af had, ben ik naast mijn baan ook aan de universiteit gaan studeren. Mijn ouders hadden hun twijfels: 'Zou je dat nou wel doen? Straks lukt het niet en wij kunnen je niet helpen, hoor.' Dat neem ik hen niet kwalijk, ik snap dat wel. Ze kenden de academische wereld niet. En ik evenmin. Van mijn broers is er een accountant, een andere heeft Wageningen gedaan, en weer een andere Delft. Die studies liggen inderdaad allemaal wel in de exacte hoek."

**Nul**

*Nul is het aantal golven in een vijver vol met ijs.*
*Nul is het aantal kippen met een geldig rijbewijs.*
*Nul is het aantal M&M's dat Rembrandt heeft gegeten.*
*Nul is het aantal smartphones dat Piet Hein heeft stukgesmeten.*
*Nul is het aantal keepers in een badmintonpartijtje.*
*Nul is het aantal x'en in het woord 'fazanteneitje'.*
*Nul is het aantal cirkels met wel zeven stompe hoeken.*
*Nul is het aantal pinguïns dat de Noordpool wil bezoeken.*
*Nul is het aantal duo's dat muziek maakt met z'n drieën.*
*Nul is het aantal taartjes met slechts zeven calorieën.*
*Nul is het aantal aardbeien in pruimenconfiture.*
*Nul is het aantal wilde haren in Beatrix' coiffure.*
*Nul is zo ontzettend veel, dus neem nul niet te licht.*
*Nul is het aantal regels dat nog rest van dit gedicht.*

Gedicht van Marjolein Kool uit *Wis- en natuurlyriek*, haar gezamenlijke bundel met Drs. P

*Waar ben je nu mee bezig?*
"Ik geef op de pabo rekenen-wiskunde, vooral de didactiek daarvan. Maar we werken ook aan de wiskundige kennis en vaardigheden van de studenten. Sommige studenten zijn er al best goed in. Maar er zijn er ook die ontzettend onzeker zijn: 'Ik wil leuk met kinderen omgaan, maar dat rekenen is niet echt mijn ding.' Ze leren het liefst regeltjes uit hun hoofd. Als ze een probleem niet herkennen weten ze niet wat ze moeten doen en gaat de deur dicht. Ze moeten een heel andere attitude aanleren: 'Als ik het niet weet, dan ga ik eens op onderzoek, puzzelen, en dan kom ik er misschien toch wel uit.' Dat valt niet mee als je achttien of twintig bent. Je hebt een leven lang gehoord of jezelf wijsgemaakt: 'Ik kan dat niet.' Krijg die deur dan maar weer eens op een kiertje.

Omdat ik hogeschoolhoofddocent ben mag ik ook onderzoek doen. Het onderzoek waar ik nu mee bezig ben richt zich op het ontwikkelen van probleemoplossend vermogen. Dus dat je je kennis op een nieuw vraagstuk kan toepassen. Ga eerst goed kijken voor je gaat rekenen: Wat is de vraag? Wat is het probleem? Welke vraag moet je straks kunnen beantwoorden? Probeer iets te tekenen is ook wel een tip die vaak werkt. Ik stimuleer studenten ook om na het oplossen van een probleem met elkaar erover in gesprek te gaan. Sommige studenten ontwikkelen zo een zelfverzekerde en onderzoekende houding. Andere studenten vinden het vreselijk. Die zijn al blij dat ze een probleem opgelost hebben, en willen geen andere aanpak horen, omdat ze bang zijn dat ze daarvan in de war raken. In mijn onderzoek probeer ik ze te verleiden dat los te laten, bijvoorbeeld met reflectiefilmpjes, tips, adviezen, stappenplannen."

*Komen er nog meer gedichten?*
"Vroeger ging het wel, om dwars door alles heen ook nog gedichten te schrijven. Nu is het een beetje het laatste wat ik op mijn lijstje heb staan. Ik verzamel ideeën, maar om die uit te werken moet je echt gaan zitten en schaven en slijpen en kapotscheuren. Mensen zeggen: 'Het ziet er makkelijk uit, je schudt het uit je mouw.' Maar juist als een vers lekker soepel huppelt, heb ik er waarschijnlijk flink op zitten zweten. Bij *Wis- en natuurlyriek* had ik een heel erg goede redactie-assistent. Zij zei soms: 'Volgens mij moet het accent hier op een andere plek liggen. Het is niet Bloemendáál, maar Blóémendaal. En als de rest van het vers dan al helemaal vast staat, dan moet het hele parket eruit om één plankje recht te krijgen. Ik dicht minder dan ik zou willen, maar hé, over zes jaar ben ik gepensioneerd. Dan gaan we los! "

Goede suggesties voor een Nederlandse wiskundige met een keerpunt in zijn of haar carrière zijn welkom via keerpunt@nieuwarchief.nl.

Wim Caspers
*Lyceum Ypenburg, Den Haag, en*
*Faculteit EWI en Lerarenopleiding, TU Delft*
*w.t.m.caspers@tudelft.nl*

**Onderwijs** Bespreking examens vwo wiskunde B en C 2019

# Als het niet lukt, is het niet erg

**Wiskunde C is het wiskundevak op het vwo dat gekozen kan worden door leerlingen in het profiel Cultuur en Maatschappij (kort door de bocht samengevat: het minst technische/exacte profiel). Het kent eigen domeinen als Logisch Redeneren en Vorm en Ruimte, maar heeft ook onderwerpen uit de domeinen Verbanden, Veranderingen en Statistiek en Kansrekening met wiskunde A gemeenschappelijk. Wiskunde B is de wiskunde die hoort bij het profiel Natuur en Techniek (het meer technische/exacte profiel). In zekere zin bevinden de twee vakken zich dus aan de uitersten van het spectrum aan wiskundevakken op het vwo (natuurlijk is er ook nog wiskunde D als profielkeuzevak binnen Natuur en Techniek, maar dat vak kent geen centraal examen). Wim Caspers belicht de centrale examens van beide vakken en duidt hierbij hun verschillen en overeenkomsten.**

Het vwo in Nederland heeft afgelopen zomer weer een nieuwe lichting studenten afgeleverd. De tweede lichting van leerlingen die examen deden volgens de vernieuwde wiskundeprogramma's. Meteen na het centraal examen verscheen op internet een plaatje waarmee iemand een verschil tussen wiskunde B en wiskunde C onder de aandacht wilde brengen [1] — zie Figuur 1. Te zien zijn de enigszins abstract ogende opgave 12 uit het wiskunde B-examen en het wat meer toepassingsgerichte vraagstuk 6 waarin de wiskunde C-kandidaten uitgenodigd werden een vlakverdeling in te kleuren. De maker van het plaatje voegde bij de wiskunde C-opdracht zelf de teksten "Zet hem op!" en "En vergeet niet: als het niet lukt is het niet erg, meedoen is het belangrijkste!" toe en ook de vrolijke krijtjes en enthousiaste kinderen. Misschien ter geruststelling: de meme doet geen recht aan het examen wiskunde C, alleen al omdat 20 tot 25 procent van het examen overeenkomt met het wiskunde A-examen.

### Vwo wiskunde C

De kleurplaat was onderdeel van de eerste opgave van het examen, getiteld 'Mondriaan'. Een kunstenaar wil, volgens het verhaal in deze opgave, in navolging van Mondriaan een schilderij bestaand uit vlakken kleuren met de kleuren rood, blauw en wit. (Dat is handig omdat in het basispakket hulpmiddelen dat kandidaten meenemen naar het examen een blauw en een rood kleurpotlood zitten.) Het aantal mogelijke kleuringen $M$ wordt vooraf gegeven door de formule

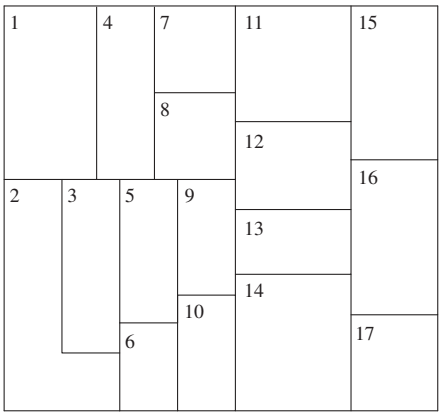
$$M = 3^V$$

waarbij $V$ het aantal vakken is. Dat de formule niet afgeleid hoeft te worden is waarschijnlijk ingegeven door het schrappen van het domein 'Tellen' uit de centraal-examenstof vorig jaar en dit jaar. De kunstenaar wil graag minimaal vijf miljoen mogelijkheden hebben; waarom wordt overigens niet duidelijk. De eerste vraag is hoeveel vlakken daarvoor nodig zijn. Kandidaten mogen de vergelijking

$$3^V = 5 \cdot 10^6$$

grafisch met een rekenmachine oplossen, wat sommige kandidaten er niet van weer-


**Figuur 1** Vergelijking examens wiskunde B en wiskunde C [1].




**Figuur 2** Vlakvulling uit vraag 3 tot en met 6.

houdt om toch netjes een logaritme met grondtal $3$ aan te roepen. De tweede vraag was om te onderzoeken of voor een verdubbeling van het aantal kleuringen een verdubbeling van het aantal vlakken nodig is. Een getallenvoorbeeld met bijvoorbeeld $V=3$ respectievelijk $V=6$ (of eigenlijk $V=5$ of mogelijk zelfs een nog kleinere waarde) is voldoende om duidelijk te maken dat een verdubbeling van $V$ niet nodig is.

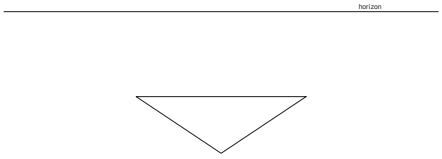
Dan wordt de werkelijke reden voor het opnemen van deze opgave in het examen duidelijk: het domein Logisch Redeneren wordt getoetst. De kunstenaar kiest een vlakverdeling (Figuur 2) en besluit, anders dan Mondriaan, dat aan elkaar grenzende vlakken niet eenzelfde kleur mogen krijgen.

Er wordt een notatie geïntroduceerd. $W_5$ betekent bijvoorbeeld dat vlak nummer $5$ wit wordt gekleurd. De kunstenaar zegt: "Vlak nummer $1$ is rood, dus vlak nummer $4$ is blauw of wit." De kandidaten wordt gevraagd die mededeling te "vertalen in logische symbolen", gebruikmakend van de afgesproken notatie. In de vraag erna is besloten dat vlak $4$ wit wordt en dienen vier redeneerstappen die daaruit volgen (zie Figuur 3) te worden gegeven in gewone zinnen.

En dan volgt de opdracht om gebruikmakend van de afgesproken notatie en logische symbolen een redenering te geven, bestaande uit een aantal redeneerstappen,

- $(R_1 \wedge W_4) \Rightarrow B_3$
- $B_3 \Rightarrow (\neg B_5 \wedge \neg B_2)$
- $(R_1 \wedge \neg B_2) \Rightarrow W_2$
- $(W_2 \wedge B_3) \Rightarrow R_6$

**Figuur 3** Redeneerstappen uit vraag 4.




**Figuur 4** Bijlage bij opgave 22.

waaruit blijkt dat kleuren van vlak nummer $5$ problemen op gaat leveren. De hele opgave graaft dus iets dieper dan men in eerste instantie zou denken.

Om het bovengenoemde probleem op te lossen wordt een nieuwe kleur geïntroduceerd: vlak $5$ krijgt de kleur geel. De gewraakte vraag 6 waar de kleurplaat verder ingekleurd moet worden met blauw, wit en rood vormt het sluitstuk, ook tot verbazing van sommige kandidaten. Velen zullen het als een zegen ervaren hebben dat met deze redelijk overzichtelijke openingsvragen het domein Logisch Redeneren werd afgedekt. Landelijk werd $84\%$ van de punten voor de eerste zes vragen gescoord. Veel slechter werd er gescoord op de vragen uit het domein Vorm en Ruimte. Vooral het tekenen van een regelmatige zeshoek in een gegeven gelijkzijdige driehoek in tweepuntsperspectief (Figuur 4) leverde problemen op ($30\%$ van de punten landelijk behaald).

**Vwo wiskunde B**

In de internetmeme werd de kleuropgave uit het wiskunde C-examen vergeleken met vraag 12 uit het wiskunde B-examen. Die vraag was onderdeel van de opgave 'Gebroken goniometrische functie' (Figuur 5); inderdaad een van de lastigere opgaven uit het examen. Vraag 10 bijvoorbeeld doet een behoorlijk beroep op algebraïsche vaardigheden, door te vragen naar het oplossen van een gebroken vergelijking met goniometrie en kwaadaardige wortels erin:

$$\frac{\cos(x)}{-\sin^2(x)} = \sqrt{2}\,.$$

De grafische rekenmachine mag niet aangewend worden, want de vergelijking moet exact opgelost worden. Dan volgt een onderzoek naar het bestaan van perforaties van de grafiek van de functie $f_p$ met

$$f_p(x) = \frac{\cos(x)}{p-\sin^2(x)}.$$

Ondankbaar werk, want voor geen enkele waarde van $p$ blijken ze er te zijn. De grafiek van de voor de hand liggende kandidaat $f_1$ blijkt een verticale asymptoot op de bedoelde plek van de perforatie te herbergen. Veel kandidaten zullen daar niet op bedacht geweest zijn. Vraag 12 valt dan nog wel mee. Van de punten $P$, $Q$ en $R$ op de grafiek van $f_p$ wordt de $x$-coördinaat gegeven en vervolgens wordt gevraagd exact de waarden van $p$ te berekenen waarvoor $PQ \perp QR$. Wanneer deze vraag bij leerlingen de associatie "product richtingscoëfficiënten gelijk aan $-1$" of "inproduct gelijk aan $0$" oproept, is de oplossing niet ver weg.

Het examen bleek achteraf de nogal hoge N-term 1,8 toegekend te krijgen; het cijfer dat de kandidaten krijgen is dan 0,8 hoger dan de standaardnormering.

Die hoge N-term kan ook mede tot stand zijn gekomen vanwege de betrekkelijk lastige vragen in de opgave 'Een logaritmische functie en haar afgeleide'. Om te beginnen worden de functies $f$ en $g$ gegeven door

$$f(x) = x\ln(x) - x + 1$$

en

$$g(x) = f'(x)$$

om vervolgens te vragen naar de $x$-coördinaten van snijpunten van de grafieken van $f$ en $g$ (niet een erg interessante vraag). Het leidt tot de op te lossen vergelijking

$$x\ln(x) - x + 1 = \ln(x)$$

waar toch enige handigheid voor vereist is. Waarom er niet gewoon gevraagd wordt om op te lossen

$$f(x) = f'(x)$$

en in plaats daarvan $g$ op zo'n vreemde manier geïntroduceerd wordt, blijkt uit de volgende vraag waar gevraagd wordt een waarde voor $p$ te vinden zodanig dat

$$\int_{p}^{2p} g(x)\,dx = 0.$$

Dat leidt dan weer tot de vergelijking

$$2p\ln(2p) - p\ln(p) - p = 0,$$

ook geen alledaagse verschijning. Vermeldenswaardig is dat $F(x) = x\ln(x) - x$ niet meer als primitieve van $f(x) = \ln(x)$ tot het examenprogramma behoort. Is deze opgave misschien een aangepaste versie van een ouder ontwerp?

Een vraag die een uitdaging opleverde voor de examinatoren vanwege het grote aantal oplossingsmethoden is vraag 14. Gegeven lijn $k$ door de punten $A(0,10)$ en $B(40,0)$ en lijn $m$ als baan van een punt $P$

## Gebroken goniometrische functie

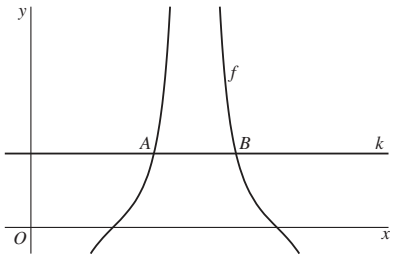
De functie $f$ is gegeven door:

$$f(x)=\frac{\cos(x)}{-\sin^2(x)}$$

Lijn $k$ is de lijn met vergelijking $y=\sqrt{2}$.

Lijn $k$ en de grafiek van $f$ hebben oneindig veel snijpunten. De punten $A$ en $B$ zijn de twee snijpunten met de kleinste positieve $x$-coördinaten. Deze zijn in figuur 1 aangegeven.

**figuur 1**




6p **10** Bereken exact de $x$-coördinaten van $A$ en $B$.

Voor elke waarde van $p$ is de functie $f_p$ gegeven door:

$$f_p(x)=\frac{\cos(x)}{p-\sin^2(x)}$$

6p **11** Onderzoek of er waarden van $p$ zijn waarvoor de grafiek van $f_p$ perforaties heeft.

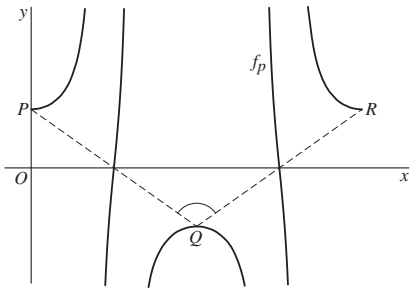
In de rest van de opgave beperken we ons tot waarden van $p$ waarvoor geldt: $p\neq 0$

De punten op de grafiek van $f_p$ met $x$-coördinaten $0$, $\pi$ en $2\pi$ noemen we respectievelijk $P$, $Q$ en $R$.

In figuur 2 is voor een waarde van $p$ de grafiek van $f_p$ weergegeven.

Ook zijn de lijnstukken $PQ$ en $QR$ weergegeven.

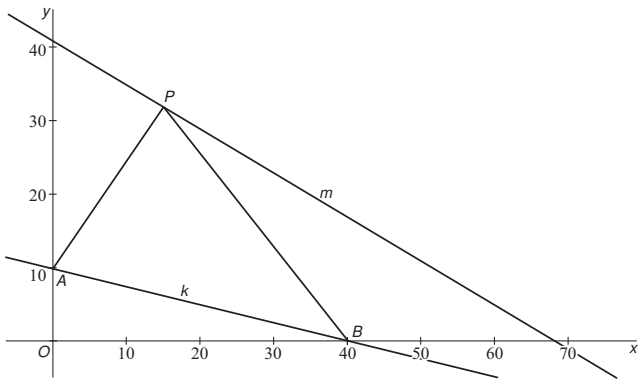
**figuur 2**



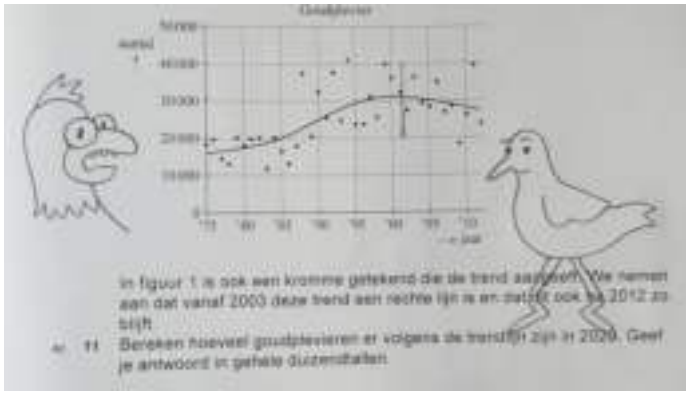

Er zijn waarden van $p$ waarvoor $PQ$ en $QR$ loodrecht op elkaar staan.

4p **12** Bereken exact deze waarden van $p$.

**Figuur 5** Opgave 'Gebroken goniometrische functie'.




**Figuur 6** Figuur bij opgave 14.




**Figuur 7** Illustratie bij opgave 11 wiskunde C.

met bewegingsvergelijkingen

$$\begin{cases}x=18+5t,\\ y=30-3t,\end{cases}$$

wordt gevraagd naar een positie van $P$ zodat driehoek $ABP$ (zie Figuur 6) gelijkbenig en rechthoekig is.

Het is een opgave die volop ruimte biedt om een eigen manier van oplossen te vinden. Van die ruimte is goed gebruikgemaakt door de kandidaten.

### Conclusie

Uit de hier beschreven opgaven blijkt in ieder geval dat wiskunde C en B andere doelgroepen bedienen en op een andere manier voorbereiden op vervolgopleidingen. Het moge duidelijk zijn dat wiskunde B-leerlingen wiskundig handiger zijn dan C-leerlingen. De manier waarop wiskunde gedaan wordt verschilt overigens niet per se in enthousiasme, creativiteit of inventiviteit. Wiskunde C-leerlingen deden wel allemaal examen in een kunstvak en zullen bij de eerste opgave in hun wiskundeexamen meewarig het hoofd geschud hebben. Over Mondriaan wordt in de opgave namelijk gemeld dat zijn latere schilderijen bestaan uit zwarte lijnen en rode, gele, blauwe en witte vlakken, maar in die werken spelen zwarte en grijze vlakken ook bijna altijd een rol. En laten we ook een anonieme C-leerling prijzen die om eens rustig na te denken het examen voorzag van illustraties (Figuur 7). Kom daar maar eens om, bij wiskunde B.

Alle opgaven van de examens wiskunde voor havo en vwo, eerste en tweede tijdvak zijn te vinden via de website Examenblad.nl.

### Referentie

1 Gebruiker @Schoolstress2019 op Instagram, https://www.instagram.com/p/BxsnsLLibPb, 20 mei 2019.

Wieb Bosma
*IMAPP*
*Radboud Universiteit Nijmegen*
*bosma@math.ru.nl*

**Stoffelijke resten** tastbare herinneringen aan wiskundigen

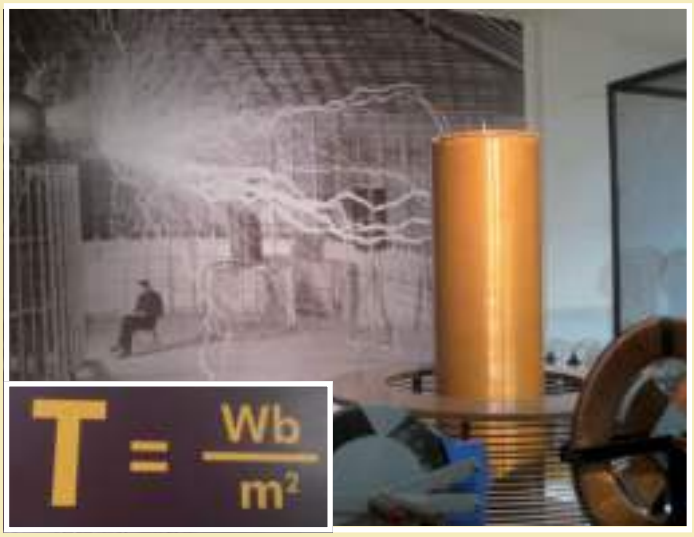
# Zie zo Zagreb

Zagreb is de hoofdstad van Kroatië. In omvang vergelijkbaar met Amsterdam, in drukte niet: je kunt er een kanon afschieten (en dat gebeurt ook dagelijks om 12 uur precies)! Het valt niet mee er iets wiskundig interessants te vinden, en we rekken dat begrip dan maar aanzienlijk op. Om Nikola Tesla (1856–1943), bijvoorbeeld, kunnen we in de stad vrijwel niet heen; zijn band met de wiskunde is niet alleen flinterdun, ook die met Kroatië is niet onbetwist: we bevinden ons in de Balkan... hij werd weliswaar in Smiljan, in het huidige Kroatië geboren, maar dat was destijds Oostenrijk. In dat land kreeg hij ook zijn opleiding, en werkzaam was de uitvinder en ontwikkelaar vooral in de Verenigde Staten. Maar Kroatië was een deel van de twintigste eeuw onderdeel van Joegoslavië en dus claimt ook hoofdstad Belgrado (nu in Servië) zijn nalatenschap.


In Zagreb vinden we een groot Tesla-monument aan de Tesla-straat, ...

... een plaquette op het voormalige stadhuis, in de oude bovenstad ...


... en zijn afbeelding in het Technisch Museum, dat naar hem vernoemd is.




Een permanente tentoonstelling in het Technisch Museum toont een deel van de opstellingen in het lab van Tesla, en legt zijn experimenten met wisselstroom en (draadloze) communicatie uit. De formule geeft de Tesla als eenheid van magnetische flux-dichtheid. Verder staat het museum vol met vliegtuigen, auto's, naaimachines, en ook een kleine afdeling rekenapparatuur. Maar de schitterende Calcorex hierboven vonden we gewoon buiten op de zondagse rommelmarkt.

Elders in het museum staat ook de klok die op vele straathoeken te vinden is: de afgeknotte kubus bestaat uit 6 regelmatige achthoeken en 8 driehoeken. In de hoop de regelmatige achthoek ook architectonisch aan te treffen, zochten we een befaamd gebouw, het Oktogon ...

... Helaas, de achthoekige vide van dit Art Nouveau-monument staat juist wegens renovatie in de steigers.

Er schijnt ook een museum met (Optische) Illusies te zijn, maar wij vonden deze 'trompe l'oeuil' in de open lucht al mooi genoeg.

Zoals overal staat ook deze stad vol met standbeelden van 'bekende' schrijvers, dichters, staatsmannen, politici. De enige Kroatische geleerde (en, vooral, wiskundige) met een eigen beeld is Rugjer Boscovic (1711–1787), wel omschreven als 'de Kroatische Leibniz'. Zijn buste staat pal naast het gebouw van de Akademie van Kunsten en Wetenschappen. Hij was geboren in Dubrovnik, kreeg een opleiding als Jezuïet in Rome, en werd naast theoloog een moderne wetenschapper van zijn tijd, die zich bezighield met astronomie (planeetovergangen), aardwetenschap en driehoeksmeting (het opmeten van een stuk van een meridiaan), en atoomtheorie (waar hij inging tegen Newton en atomen niet zag als elastische balletjes maar als puntmassa's waarop krachten werken en die elkaar op microscopische schaal afstoten terwijl macroscopisch de zwaartekracht overheerst). Als hij al herinnerd wordt als wiskundige gaat het over zijn methode (voorloper van de kleinste kwadraten) te schatten door absolute waarden van afwijkingen te minimaliseren. Laplace (lees het nieuwe boek van Andriesse!) bekritiseerde zijn methode om een komeetbaan op grond van drie waarnemingen te bepalen. Daar staat tegenover dat filosofen erop wijzen dat Boscovic al veel eerder dan Laplace een versie van de later zo genoemde 'demon van Laplace' formuleerde: als je alle posities van en krachten op alle deeltjes in het universum op een tijdstip kent, kun je die in de toekomst ook bepalen — determinisme dus.

Het gebouw achter de Akademie is de bibliotheek van de Akademie. Onder de daklijst staan namen van buitenlandse scheikundigen, natuurkundigen, (Lavoisier, Thomson, Dalton, Berzelius, Gay-Lussac, Avogadro, Bojle (!), Cavendish) en zowaar ook een wiskundige: Pierre Alphonse Laurent (1813–1854), die van de reeksen.

In hetzelfde park staat ook een prachtig meteorologisch monumentje uit 1884 met achter glas aan de vier zijden een thermometer, barometer, klok, en een thermo/baro/hydrograaf.





Verder is er een heel bijzonder kunstwerk dat een model op schaal (1 op 680 miljoen) vormt van ons zonnestelsel: een zon (van bijna 2 meter doorsnee) in het centrum (oorspronkelijk uit 1971) en 9 planeten (inclusief Pluto) als kleine metalen bolletjes (toegevoegd in 2004) daar op afstanden variërend van enkele tientallen meters tot meer dan 7 kilometers vandaan — wij zijn niet verder dan tot Mars gekomen.

Een laatste redmiddel voor het vinden van wiskundig-historisch interessante monumenten is een bezoek aan het kerkhof. Het trof dat de grote begraafplaats Mirogoi van Zagreb wordt aangeprezen als een toeristische attractie: enorm groot, verschillende religies door elkaar. Nu bleek vooral de buitenmuur met toegangspoort en koepels bezienswaardig, inclusief de vele monumenten aan de binnenzijde daarvan. Speurwerk vooraf wees uit dat Oton Kučera (1857–1931) en Svetozar Kurepa (1929–2010) hier begraven moesten liggen. De eerste was een vooraanstaand astronoom en popularisator van exacte wetenschap, de tweede de wiskundige gespecialiseerd in functionaalanalyse (wiens, beroemdere, oom Duro in Belgrado overleed). Inderdaad wist een uiterste behulpzame Modrić-lookalike begraafplaats-beambte ons te leiden naar de juiste baaien om deze graven te kunnen lokaliseren, in deze onafzienbare zee van zerken.


K. P. Hart
*Faculteit EWI*
*TU Delft*
*k.p.hart@tudelft.nl*

**Research**

# Machine learning and the Continuum Hypothesis

**In January 2019 the journal *Nature* reported on an exciting development in Machine Learning: the very first issue of the journal *Nature Machine Intelligence* contains a paper that describes a learning problem whose solvability is neither provable nor refutable on the basis of the standard ZFC axioms of Set Theory. In this note K. P. Hart describes what the fuss is all about and indicates that maybe the problem is not so undecidable after all.**

In the paper [1], in *Nature Machine Intelligence*, its authors exhibit an abstract machine-learning situation where the learnability is actually neither provable nor refutable on the basis of the axioms of ZFC. This was deemed so exciting that the mother journal *Nature* devoted *two* commentaries to this: [9] and [3].

The first of these, [9], is rather matter-of-fact in its description of the problem but the second, [3], manages, in just a few lines, to mix up Gödel's Incompleteness Theorems and the undecidability of the Continuum Hypothesis. It misstates the former — "Gödel discovered logical paradoxes" — and misinterprets the latter: "a paradox known as the Continuum Hypothesis".

No, Gödel did not discover paradoxes; he proved a (highly) technical result about formal proofs. That result shows that under certain circumstances a first-order theory will be incomplete, that is, there is a formula $\varphi$ such that there is no formal proof of $\varphi$ nor of its negation. The formula $\varphi$ constructed by Gödel asserts, indirectly, "there is not formal proof for $\varphi$" and as such looks a bit like "this sentence is false", which can be construed as a version of the Liar's Paradox. There is a difference however: the formula $\varphi$ does not refer directly to itself and this prevents it from being a paradox.

And, no, the Continuum Hypothesis is not a paradox either. It is 'simply' a statement about subsets of the real line that exhibits a concrete incompleteness of ZFC Set Theory. That theory is subject to Gödel's incompleteness theorem, hence it comes with its own version of the formula $\varphi$. Both $\varphi$ and the Continuum Hypothesis show that ZFC is incomplete, the difference between these formulas is that the Continuum Hypothesis is interesting and $\varphi$ is not. This is not meant in a pejorative way; as Gödel's construction applies to potentially very many different theories one would not expect $\varphi$ to say something very specific in the theory that it is constructed for.

The set theory in [1] is related to Cantor's original formulation of the Continuum Hypothesis [2]: if one declares two sets to be equivalent if there is a bijection between them then the infinite subsets of $\mathbb{R}$ are divided into *two* equivalence classes, those of the sets equivalent to $\mathbb{N}$ and those of the sets equivalent to $\mathbb{R}$.

The learning problem from [1] is equivalent to a weaker version: the number of equivalence classes is finite. For the rest of this note we shall refer to that statement as the Weak Continuum Hypothesis. This statement is, like the Continuum Hypothesis neither provable not refutable on the basis of the axioms of ZFC.

For later use we abbreviate "there is a bijection between $X$ and $Y$" as $X \equiv Y$. Thus, the Continuum Hypothesis states that if $X$ is an infinite subset of $\mathbb{R}$ then either $X \equiv \mathbb{N}$ or $X \equiv \mathbb{R}$.

In the next section I will summarize the description from [1] of the so-called *EMX learning problem*. The section that follows contains the translation from [1] of the learning problem into a purely combinatorial problem about functions between powers of the unit interval and an explanation of why that translation is equivalent to the Weak Continuum Hypothesis. In the section thereafter we shall see that the combinatorial part is related to a result of Kuratowski from 1951 [6], that characterizes when, given $k \in \mathbb{N}$, a set has (at most) $k+1$ equivalence classes of infinite sets under the equivalence relation $\equiv$ discussed above. In the last section I will show why I think that the problem is not undecidable at all: there is no algorithm that solves this particular learning problem.

## The learning problem

This is a summary of the parts of [1] that lead to the undecidability result.

The authors start with the following real-life situation as an instance of their general learning problem. A website has a collection of advertisements that it can show to its visitors; each advertisement, $A$, comes with a set, $F_A$, of visitors for whom it is of interest: say if $A$ advertis-

es running shoes then $F_A$ contains avid runners (or people who just like snazzy shoes). Choosing the optimal advertisement to display amounts to choosing a finite set from a population while maximizing the probability that the visitor is actually in that set. The problem is that the probability distribution is unknown.

Rather than dwell on this particular example the authors make an abstraction: Given a set $X$ and a family $\mathcal{F}$ of subsets of $X$ find a member of $\mathcal{F}$ whose measure with respect to an unknown probability distribution is close to maximal. This should be done based on a finite sample generated i.i.d. from the unknown distribution.

The undecidability manifests itself when we let $X$ be the unit interval $\mathbb{I}$ and $\mathcal{F}$ the family $\operatorname{fin}\mathbb{I}$ of finite subsets of $\mathbb{I}$.

*Learning functions*

In the general situation the abstract problem described above is made more explicit and quantitative as follows.

For the unknown probability distribution $P$ on $X$ find $F\in\mathcal{F}$ such that $E_P(F)$ is quite close to $\operatorname{Opt}(P)$, which is defined to be $\sup_{Y\in\mathcal{F}}E_P(Y)$.

To quantify this further a *learning function* for $\mathcal{F}$ is defined to be a function

$$G\colon \bigcup_{k\in\mathbb{N}} X^k \to \mathcal{F}$$

with certain desirable properties.

Say, if $S\in X^k$ represents a sample of visitors then $G(S)$ would be a set of visitors from which the next visitor is very likely to come. As $G(S)$ belongs to $\mathcal{F}$, there is an advertisement $A$ such that $G(S)=F_A$ and this $A$ will be displayed on the website.

The desirable properties, e.g., the 'very likely' in the example above, are captured in the following definition of an $(\epsilon,\delta)$-*EMX learner* for $\mathcal{F}$. This is a function $G$ as above such that for some $d\in\mathbb{N}$, depending on $\epsilon$ and $\delta$, the following inequality holds

$$\Pr_{S\sim P^d}\left[E_P(G(S))\leqslant \operatorname{Opt}(P)-\varepsilon\right]\leqslant\delta$$

for all distributions $P$ with finite support.

The letters EMX abbreviate 'estimating the maximum'.

*Combinatorics*

It seems a nigh on hopeless task to say anything sensible when there are so many possible probability distributions to consider. However, as we shall see, the existence of EMX learning functions is equivalent to the existence of maps between finite powers of $X$ that mention no probabilities at all but instead are required to satisfy a few simple inclusion relations. These will prove to be much more amenable to set-theoretic investigations.

**A combinatorial translation**

The authors of [1] do not waste a lot of time and formulate, without much ado, the combinatorial statement equivalent to the existence of an EMX learner.

This statement involves what the authors call monotone compression schemes. Their formulation needs the following piece of notation: For a set $X$ and a natural number $n$ we use $[X]^n$ to denote the family of $n$-element subsets of $X$.

**Definition 1.** Let $m$ and $d$ be two natural numbers with $m>d$. An $m\to d$ *monotone compression scheme* for a family $\mathcal{F}$ of finite subsets of a set $X$ is a function $\eta\colon[X]^d\to\mathcal{F}$ such that whenever $A$ is an $m$-element subset of $X$ it has a $d$-element subset $B$ such that $A\subseteq\eta(B)$.

This is slightly different from the formulation of Definition 2 in [1], which leaves open the possibility that $|A|<m$ and that $|B|<d$, as it uses indexed sets. It is clear from the results and their proofs that our definition captures the essence of the notion.

The idea here is that someone, Alice say, thinks of an $m$-element set $A$ and provides their friend Bob with a $d$-element subset $B$ of $A$. The function $\eta$ helps Bob to recover some information about $A$, namely that it is a subset of the member $\eta(B)$ of the family $\mathcal{F}$.

There is a second unnamed function implicit in Definition 2: the choice of the subset $B$ of $A$; this function we shall call $\sigma$.

So a scheme consists of a pair of functions: $\sigma\colon[X]^m\to[X]^d$ and $\eta\colon[X]^d\to\mathcal{F}$; these should satisfy $A\subseteq(\eta\circ\sigma)(A)$ for all $A$. In fact, as we shall see in the next section, the function $\sigma$ is more convenient to work with.

The translation is now as follows.

**Lemma 1** [1, Lemma 1.1]. *For an upward-directed family $\mathcal{F}$ of finite sets the existence of a $(\frac{1}{3},\frac{1}{3})$-EMX learning function is equivalent to the existence of a natural number $m$ and an $(m+1)\to m$ monotone compression scheme for $\mathcal{F}$.*

The proof of necessity takes the natural number $d$ in the learning function and produces a monotone $(m+1)\to m$ compression scheme with $m=\left\lceil\frac{3}{2}d\right\rceil$.

*Undecidability*

At this point the authors turn to the aforementioned special case of the unit interval $\mathbb{I}$ and its family $\operatorname{fin}\mathbb{I}$ of finite subsets and prove the following.

**Theorem 1.** *There is a monotone $(m+1)\to m$ compression scheme for $\operatorname{fin}\mathbb{I}$ for some $m\in\mathbb{N}$ if and only if the Weak Continuum Hypothesis holds.*

As the Weak Continuum Hypothesis is both consistent with and independent of the axioms of ZFC the same holds for the existence of a compression scheme and for the existence of a $(\frac{1}{3},\frac{1}{3})$-EMX learning function. Theorem 1 is an immediate consequence of the set of equivalences in the following theorem.

**Theorem 2** [1, Theorem 1]**.** *Let $k\in\mathbb{N}$ and let $X$ be a set. There is a $(k+2)\to(k+1)$ monotone compression scheme for the finite subsets of $X$ if and only the infinite subsets of $X$ are divided into $k+1$ (or fewer) equivalence classes by the relation $\equiv$.*

Indeed, the Weak Continuum Hypothesis holds iff the infinite subsets of $|\mathbb{I}|$ are divided into $k+1$ equivalence classes for some $k\in\mathbb{N}$.

In the next section we take a closer look at monotone compression schemes and point out a connection with an old result of Kuratowski's.

**Compression schemes and decompositions**

In the general case considered above the function $\eta$ is important because of its codomain $\mathcal{F}$: Bob is required to choose a member of that family. It turns out that in the case considered by the authors of [1], namely the family of *all* finite subsets of a set $X$, it is the function $\sigma$ that is more interesting. This is borne out by the following proposition.

**Proposition 1.** *Let $m$ and $d$ be natural numbers and let $X$ be a set. There is an $m\to d$ monotone compression scheme for the finite subsets of $X$ if and only if there is a finite-to-one function $\sigma\colon[X]^m\to[X]^d$ such that $\sigma(x)\subseteq x$ for all $x$.*

*Proof.* If the pair $\langle \eta, \sigma \rangle$ determines an $m \to d$ monotone compression scheme then $\sigma$ is finite-to-one. For let $y \in [X]^d$ then $\sigma(x) = y$ implies $x \subseteq \eta(y)$, hence there are at most $\binom{M}{m}$ such $x$, where $M = |\eta(y)|$.

Conversely, if $\sigma$ is as in the statement of the proposition then we can let $\eta(y) = \bigcup \{x : \sigma(x) = y\}$. □

*Kuratowski's decompositions*

To do justice to Kuratowski's results and because the proofs require it we will use standard set theoretic notions and notations. We shall need the first countably many infinite cardinal numbers $\aleph_k$ ($k \in \mathbb{N}$) and the ordinal numbers $\omega_k$. What we also need to know is that $\omega_k$ is the 'standard' well-ordered set of cardinality $\aleph_k$.

Above I formulated Kuratowski's 1951 result in terms of the equivalence relation "there is a bijection" but as the title of [6] indicates the original formulation involved the cardinal numbers $\aleph_k$. To be precise the papers characterizes when a set has cardinality at most $\aleph_k$ in terms of its $(k+2)$-nd power. The very definition of the cardinal numbers $\aleph_k$ makes it clear that a set has cardinality at most $\aleph_k$ if and only if there are at most $k+1$ equivalence classes under the equivalence relation $\equiv$. From now on we let $|X|$ denote the cardinality of the set $X$, so that $|X| \leqslant \aleph_k$ abbreviates that the cardinality of $X$ is at most $\aleph_k$.

It should come therefore as no big surprise that Kuratowski's results and Theorem 2 are related.

We start by quoting the following theorem from [6], it provides one direction in the aforementioned characterization.

**Theorem 3.** *The power $\omega_k^{k+2}$ can be written as the union of $k+2$ sets, $\{A_i : i < k+2\}$, such that for every $i < k+2$ and every point $\langle x_j : j < k+2 \rangle$ in $\omega_k^{k+2}$ the set of points $y$ in $A_i$ that satisfy $y_j = x_j$ for $j \neq i$ is finite.*

In Kuratowski's words "$A_i$ is finite in the direction of the $i$th axis".

*Sketch of the proof.* The case $k = 0$ is easy: $\omega_0$ is the first infinite ordinal, therefore $A_0 = \{\langle m, n \rangle : m \leq n\}$ and $A_1 = \{\langle m, n \rangle : m > n\}$ are as required.

The rest of the proof proceeds by induction on $k$. We give the step from $k = 0$ to $k = 1$ in some detail and leave the other steps to the reader.

To decompose $\omega_1^3$ into three sets $A_0$, $A_1$ and $A_2$ we apply the Axiom of Choice to choose (simultaneously) for each infinite ordinal $\alpha$ in $\omega_1$ a decomposition $\{X(\alpha,0), X(\alpha,1)\}$ of $(\alpha+1)^2$, say by choosing well-orders of type $\omega$ and then using the decomposition obtained for $k = 0$.

- One puts $\langle \alpha, \beta, \gamma \rangle$ into $A_0$ if $\beta$ is the largest coordinate and $\langle \alpha, \gamma \rangle \in X(\beta, 0)$ or if $\gamma$ is the largest coordinate and $\langle \alpha, \beta \rangle \in X(\gamma, 0)$.
- One puts $\langle \alpha, \beta, \gamma \rangle$ into $A_1$ if $\alpha$ is the largest coordinate and $\langle \beta, \gamma \rangle \in X(\alpha, 0)$ or if $\gamma$ is the largest coordinate and $\langle \alpha, \beta \rangle \in X(\gamma, 1)$.
- One puts $\langle \alpha, \beta, \gamma \rangle$ into $A_2$ if $\alpha$ is the largest coordinate and $\langle \beta, \gamma \rangle \in X(\alpha, 1)$ or if $\beta$ is the largest coordinate and $\langle \alpha, \gamma \rangle \in X(\gamma, 0)$.

To see that $A_0$ is finite in the direction of the 0th coordinate take $\langle \beta, \gamma \rangle \in \omega_1^2$, then $\langle \alpha, \beta, \gamma \rangle \in A_0$ implies $\beta$ is largest and $\langle \alpha, \gamma \rangle \in X(\beta, 0)$, or $\gamma$ is largest and $\langle \alpha, \beta \rangle \in X(\gamma, 0)$; in either case $\alpha$ belongs to a finite set.

A similar argument works for $A_1$ and $A_2$ of course.

The inductive steps for larger $k$ are modeled on this step. □

We now show how Theorem 3 can be used to prove sufficiency in Theorem 2.

Here and in later sections it will be convenient to identify $[X]^m$, the family of $m$-element subsets of $X$, with a subset of the product $X^m$. In the cases of interest the set $X$ has a (natural) linear order $\prec$; we use this to let a set correspond with its monotone enumeration:

$$[X]^m = \{x \in X^m : (i < j < m) \to (x_i \prec x_j)\}$$

*Constructing a compression scheme from a decomposition.* From a decomposition as in Theorem 3 we construct a finite-to-one function $\sigma : [\omega_k]^{k+2} \to [\omega_k]^{k+1}$ such that $\sigma(x) \subseteq x$ for all $x$. We assume, without loss of generality, that the sets $A_i$ are disjoint.

Let $x \in [\omega_k]^{k+2}$ (so $i < j < k+2$ implies $x_i < x_j$). Take (the unique) $i$ such that $x \in A_i$ and let $\sigma(x)$ be the point in $\omega_k^{k+1}$ that is $x$ but without its coordinate $x_i$. In terms of sets we would have set $\sigma(x) = x \setminus \{x_i\}$.

This function is finite-to-one: if $y \in [\omega_k]^{k+1}$ then for each $i < k+2$ there are only finitely many $x$ in $A_i$ with $y = \sigma(x)$. □

As mentioned above Kuratowski's result works both ways: if $X^{k+2}$ admits a decomposition as above for $\omega_k^{k+2}$ then $|X| \leq \aleph_k$. This suggests that the necessity in Theorem 2 is related to the converse of Theorem 3. This is indeed the case: one can construct a Kuratowski-type decomposition from a compression scheme, but because of our definition of the schemes we only get a decomposition of the subset $[\omega_k]^{k+2}$ of the whole power. This can be turned into one for the whole power but the process is a bit messy so we leave it be.

The proof of necessity from [1] closes the circle of implications that proves the following.

**Theorem 4.** *For a set $X$ and a natural number $k$ the following are equivalent:*

1. $|X| \leq \aleph_k$;
2. *$X^{k+2}$ admits a Kuratowski-type decomposition into $k+2$ sets*;
3. *there is a $(k+2) \to (k+1)$ monotone compression scheme for the finite subsets of $X$.*

For completeness sake I sketch the proof of that last implication. Both it and Kuratowski's necessity proof use a form of the following lemma. Its proof uses some elementary cardinal arithmetic for infinite cardinals numbers.

**Lemma 2.** *Let $k$, $l$, and $m$ be natural numbers with $m > l$. Assume $\sigma : [\omega_{k+1}]^{m+1} \to [\omega_{k+1}]^{l+1}$ determines an $(m+1) \to (l+1)$ monotone compression scheme. Then there is an $m \to l$ monotone compression scheme for the finite subsets of $\omega_k$.*

*Proof.* We start by determining an ordinal $\delta$ as follows. Let $\delta_0 = \omega_k$. Given $\delta_n$ use the fact that $\sigma$ is finite-to-one to find an ordinal $\delta_{n+1} > \delta_n$ such that every $x \in [\omega_{k+1}]^{m+1}$ that satisfies $\sigma(x) \in [\delta_n]^{l+1}$ is in $[\delta_{n+1}]^{m+1}$.

In the end let $\delta = \sup_n \delta_n$. Then $\delta$ satisfies: every $x \in [\omega_{k+1}]^{m+1}$ that satisfies $\sigma(x) \in [\delta]^{l+1}$ is in $[\delta]^{m+1}$.

We define an $m \to l$ monotone compression scheme for $\delta$. If $x \in [\delta]^m$ then $y = x \cup \{\delta\}$ is in $[\omega_{k+1}]^{m+1}$ and so $\sigma(y) \subseteq y$. By the choice of $\delta$ it is not possible that $\sigma(y) \subseteq x$ hence $\delta \in \sigma(y)$ and so setting $\varsigma(x) = \sigma(y) \setminus \{\delta\}$ defines a map $\varsigma : [\delta]^m \to [\delta]^l$. This map is finite-to-one and satisfies $\varsigma(x) \subseteq x$ for all $x$. □

To finish the proof of necessity we argue by induction and contradiction. If $|X| = \aleph_{k+1}$ and there is a finite-to-one $\sigma : [X]^{k+2} \to [X]^{k+1}$ with $\sigma(x) \subseteq x$ for all $x$ then there is a subset $Y$ of $X$ with $|Y| = \aleph_k$ and a finite-to-one $\varsigma : [Y]^{k+1} \to [Y]^k$ with $\varsigma(x) \subseteq x$ for all $x$. This would contradict the obvious inductive assumption. We leave it as an exercise to the reader to ponder what absurdity would arise in the case $k = 0$ and provide the basis for the induction.

**Algorithmic considerations**

In this section we address a point already raised by the authors in [1]: the functions that are used in the previous sections are quite arbitrary and not related to any recognizable algorithm. Indeed, the constructions of the compression schemes for uncountable sets blatantly applied the Axiom of Choice: once by assuming that the underlying sets were well-ordered and again when in every step of the induction a choice of well-orders of type $\omega_k$ needed to be made.

One may therefore wonder what happens if we impose some structure on the maps in question. One possible way of separating out 'algorithmic' functions is by requiring them to have nice descriptive properties. If 'nice' is taken to mean 'Borel measurable' then the desired functions do not exist.

*Continuity and Borel measurability*

Here we show, for arbitrary $m \in \mathbb{N}$, that there does not exist an $(m+1) \to m$ monotone compression scheme for the finite subsets of $\mathbb{I}$ where the function $\sigma$ is Borel measurable. Remember that we identify $[\mathbb{I}]^k$ with the open subset of the $k$-cube $\mathbb{I}^k$ consisting of its strictly increasing elements. As such it inherits a metric and a Borel structure from that cube. We consider continuity and Borel measurability with respect to these structures. Let $m$ be a natural number and let $\sigma : [\mathbb{I}]^{m+1} \to [\mathbb{I}]^m$ be a function such that $\sigma(x) \subseteq x$ for all $x$.

*If $\sigma$ is continuous then $\sigma$ is not finite-to-one* To see this let $x \in [\mathbb{I}]^{m+1}$ and assume for notational convenience that $\sigma(x) = \langle x_i : i < m \rangle$, i.e., that the coordinate $x_m$ is left out of $x$ when forming $\sigma(x)$.

Let $\varepsilon = \frac{1}{3}\min\{x_{i+1} - x_i : i < m\}$ and let $\delta > 0$ be such that $\delta \leq \varepsilon$ and for all $y \in [\mathbb{I}]^{m+1}$ with $\|y - x\| < \delta$ we have $\|\sigma(y) - \sigma(x)\| < \varepsilon$.

Now if $y \in [\mathbb{I}]^{m+1}$ and $\|y - x\| < \delta$ then $|y_i - x_i| < \varepsilon$ for all $i \leq m$. Also, when $i < j$ we have $x_j - x_i > 3\varepsilon$. It follows that $y_m - x_i > \varepsilon$ for all $i < m$. This implies that $\sigma(y) = \langle y_i : i < m \rangle$ for all $y$ with $|y - x| < \delta$.

This shows that for every $i$ the set $O_i = \{x \in [\mathbb{I}]^{m+1} : \sigma(x) = x \setminus \{x_i\}\}$ is open. Because $[\mathbb{I}]^{m+1}$ is connected there is one $i$ such that $O_i = [\mathbb{I}]^{m+1}$. This shows that $\sigma$ cannot be finite-to-one. □

The above proof can be used/adapted to show that if $\sigma$ is Borel measurable it is not finite-to-one either.

*If $\sigma$ is Borel measurable then $\sigma$ is not finite-to-one*. There is a dense $G_\delta$-set $G$ in $[\mathbb{I}]^{m+1}$ such that the restriction of $\sigma$ to $G$ is continuous, see [7, Section 31 II].

Let $x \in G$. As in the previous proof we assume $\sigma(x) = \langle x_i : i < m \rangle$ and we obtain a $\delta > 0$ such that $\sigma(y) = \langle y_i : i < m \rangle$ for all $y \in G$ that satisfy $\|y - x\| < \delta$.

By the Kuratowski–Ulam theorem [8] we can find a point $y$ in $G$ with $\|y - x\| < \delta$ such that the set of points $t$ in the interval $(x_m - \delta, x_m + \delta)$ for which $y_t = \sigma(y) * \langle t \rangle$ belongs to $G$ is co-meager. But for every such point we have $\sigma(y_t) = \sigma(y)$ and this shows that $\sigma$ is not finite-to-one. □

*EMX learning is impossible*

As we saw above a learning function is a function $G$ from the union $\bigcup_{k \in \mathbb{N}} \mathbb{I}^k$ to the family of finite subsets of $\mathbb{I}$. We can call such a function continuous or Borel measurable if its restriction to each individual power is.

In the construction of an $(m+1) \to m$ compression scheme from a learning function the authors use its restriction to just one of these powers $\mathbb{I}^d$, where $d \leqslant m$. The definition of $\eta(S)$ involves taking the union of $G(T)$ for all $d$-element subsets $T$ of $S$, hence a union of $\binom{m}{d}$ many sets.

The definition of $\sigma$ involves choosing one $m$-element subset with a certain property from of a given $m+1$-element set.

The latter choice can be made explicit using a Borel linear order on the family of all finite subsets of $\mathbb{I}$, or just on $[\mathbb{I}]^m$.

An analysis of this procedure shows that if $G$ is Borel measurable then so are $\sigma$ and $\eta$. The results of this section then imply that a Borel measurable learning function does not exist. In this author's opinion that means that the title of [1] should be emended to "EMX learning is impossible".

*On the other hand...*

One may argue that the choice of the unit interval in [1] is a bit of a red herring. None of the arguments in the paper use the structure of $\mathbb{I}$ in any significant way.

In the step from the problem of the advertisements to the more abstract problem there is no real need to go to the unit interval. One may equally well use the set of rational numbers to code or rank the elements of the learning set.

In that case there is, as we have seen, a $2 \to 1$ monotone compression scheme for the finite subsets of $\mathbb{N}$: simply let $\sigma(x) = \max x$; the corresponding function $\eta$ is defined by $\eta(n) = \{i : i \leqslant n\}$.

It is an easy matter to transfer this scheme to the family of finite subsets of the rational numbers. Whether this scheme gives rise to a useful EMX learning function remains to be seen.

**References**


1 Shai Ben-David, Pavel Hrubeš, Shay Moran, Amir Shpilka and Amir Yehudayoff, Learnability can be undecidable, *Nature Machine Intelligence* 1 (2019), 44–48.

2 Georg Cantor, Ein Beitrag zur Mannigfaltigkeitslehre, *Crelles Journal für Mathematik* 84 (1878) 242–258.

3 Davide Castelvecchi, Machine learning leads mathematicians to unsolvable problem, *Nature* 565 (2019), 277.

4 Ryszard Engelking, *General Topology,* Sigma Series in Pure Mathematics 6, Heldermann Verlag, 1989, 2nd ed..

5 Kenneth Kunen, *Set Theory. An Introduction to Independence Proofs,* Studies in Logic and the Foundations of Mathematics 102, North-Holland, 1980.

6 Casimir Kuratowski, Sur une caractérisation des alephs, *Fundamenta Mathematicae* 38 (1951), 14–17.

7 K. Kuratowski, *Topology. Vol. I*, Academic Press and Państwowe Wydawnictwo Naukowe, 1966.

8 C. Kuratowski and St. Ulam, Quelques propriétés topologiques du produit combinatoire, *Fundamenta Mathematicae* 19 (1932), 247–251.

9 Lev Reyzin, Unprovability comes to machine learning, *Nature* 565 (2019), 166–167.

Azmy S. Ackleh
*Department of Mathematics*
*University of Louisiana at Lafayettee*
*ackleh@louisiana.edu*

Rinaldo M. Colombo
*INdAM Unit*
*University of Brescia*
*rinaldo@ing.unibs.it*

Paola Goatin
*Team ACUMES*
*INRIA*
*paola.goatin@inria.fr*

Sander Hille
*Mathematical Institute*
*Leiden University*
*shille@math.leidenuniv.nl*

Adrian Muntean
*Department of Mathematics & Computer Science*
*Karlstad University*
*adrian.muntean@kau.se*

**Event** Workshop Lorentz Center, 3–7 December 2018

# Mathematical modeling with measures

**In December 2018 a workshop entitled 'Mathematical Modeling with Measures: where Applications, Probability and Determinism Meet' was held at the Lorentz Center in Leiden. This workshop brought together mathematicians from countries in Eastern and Western Europe and North and South America with different expertise including, modeling, analysis and probability to discuss state of the art research results for a common mathematical language that is used in these fields, which is modeling with measures, and to initiate new collaborations in an attempt to solve some open problems that were posed by the workshop participants. Azmy Ackleh, Rinaldo Colombo, Paola Goatin, Sander Hille en Adrian Muntean report about this event.**

Having in mind the use of measures as key modeling tool, workshop participants considered a wide variety of applications: from structured population models, to selection-mutation models, to vehicular/traffic traffic flows, to balance equations, to probability, ...

### A selection model

Let us begin with a simple model studied two decades ago by Ackleh et al. in [1]. Therein, the authors consider the integro-differential equation on the state space of continuous functions:

$$\frac{d}{dt}x(t,q) = x(t,q)\,[q_1 - q_2 X(t)]. \qquad (1)$$

Here, $x(t,q)$ is the density of individuals having trait $q = (q_1, q_2) \in Q = [a_1, b_1] \times [a_2, b_2]$ a rectangle in the interior of $\mathbb{R}^2_+$, where $q_1$ describes the growth rate and $q_2$ describes the mortality rate. This model implies that individuals with trait $q$ produce individuals with the same trait (i.e., pure selection) and that individuals compete for resources since the mortality term is dependent on the level of the total population $X(t) = \int_Q x(t,q)\,dq$. By letting $q^*$ be the unique trait value attaining $\max_{q\in Q} q_1/q_2 = b_1/a_2$ and using the Lyapunov function $L(t) = x(t,q)^{1/q_1} / x(t,q^*)^{1/q_1^*}$ one can show that $\frac{d}{dt}L(t) = (q_2^*/q_1^* - q_2/q_1)X(t)L(t)$. Since, $(q_2^*/q_1^* - q_2/q_1) < 0$, establishing boundedness of $X(t)$, one can deduce that for any $\varepsilon$ radius ball in $Q$ centered at $q^*$ denoted by $B_\varepsilon$, $\int_{Q/B_\varepsilon} x(t,q)\,dq \to 0$, as $t \to \infty$. Using this result one can show that $x(t,q) \to \frac{b_1}{a_2}\delta_{q^*}$ in the weak* topology. Biologically this implies that the fittest trait is $q^*$ and the long term dynamics of this model is that of competitive exclusion where the fittest trait survives and all other traits go to extinction. However, from a mathematical point of view it is clear that this Dirac limit $\frac{b_1}{a_2}\delta_{q^*}$ is not an element of the state space considered in [1] which is $C(Q)$ but is an element of the more natural space for formulating this model which is the space of finite signed measure $M(Q)$. Indeed, the authors in [2] reformulated a more general model which includes (1) as a special case on the space of finite signed measures.

### Shepherd dogs and fairy tales

A recurrent intriguing question is: how can a leader drive a multitude towards a given goal?

With respect to the multitude, the leader can be *attractive* or *repulsive*, for instance. It is easy to refer to these two sample cases respectively as to the *pied piper* that attracts mice (e.g. *De rattenvanger van Hamelen*, see [9]) or to a shepherd dog herding sheep. On this basis, a variety of new control problems can be formalized. Can we characterize the *best* strategy for the leader? Clearly, here *best* may mean fastest, or cheapest, or simplest, or ...

Further questions arise when we start thinking at a team of leaders cooperating

towards the same goal. How much are $2$ shepherd dogs more efficient than only one? Does there exist an *optimal* number of pied pipers to gather mice in a given region?

The next step is even more intriguing. Indeed, we may pass from a control problem, where one goal has to be achieved, to a *game*, where different goals are sought by different (teams of) competing leaders. Does there exist a *winning* strategy? Can we characterize Nash [12] equilibria? (These are strategies that each controller adopts because other choices would be less convenient, see also [6].)

When we get to the formalization of these problems, we find an exciting wealth of tools that mathematics offers to describe these situations. During the meeting at the Lorentz center, focus was mostly on descriptions based on conservation laws. Call $\rho = \rho(t,x)$ the time ($t$) and space ($x$) dependent density of individuals, think at $\rho$ as at the number of mice/sheep per square meter. Let $P_1(t),\dots,P_n(t)$ be the time dependent positions of the pipers/shepherd dogs and describe the pipers-mice or dogs-sheep interactions through the speed $v = v(t,x,\rho,P_1,\dots,P_n)$. We thus describe the whole dynamics through the continuity equation

$$\partial_t \rho + \operatorname{div}(\rho v(t,x,\rho,P_1,\dots,P_n)) = 0$$

while the controls or strategies are the speeds $u_1,\dots,u_n$ of the leaders, so that

$$P_i(t) = P_i^o + \int_0^t u_i(\tau)\,d\tau \quad \text{with} \| u_i \| \le U,$$

$U$ being the maximal leaders' speed and $P_i^o$ the initial position.

Basic well posedness and stability results for the resulting problem were obtained in [7]. The goals of the leaders can be easily formalized through suitable integrals of $\rho$.

At the time of this writing, the existence and the characterization of optimal controls is an open problem, as also any information on Nash equilibria. We expect that these questions have to be answered prior to suggesting optimal escape strategies to mice and sheep...

**Nash and Braess close roads**

The Braess Paradox, see [4] is a famous example, showing that the dynamics on networks can significantly differ from that

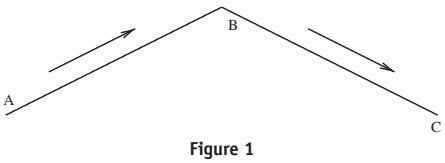


**Figure 1**

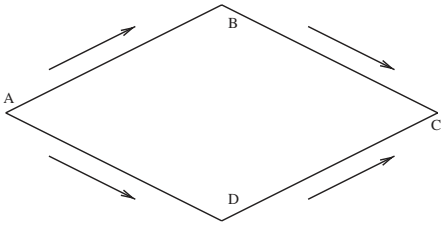


**Figure 2**

along a single road, displaying surprising properties.

Cities A and C are connected by two roads that pass through city B (here and in what follows all roads are one way). The travel time along AB depends on traffic, say it equals the number of vehicles traveling along that road divided by $100$. On the contrary, the road from B to C is so wide that the travel time along this segment is $45$, regardless of traffic. See Figure 1. Clearly, if $4000$ commuters need to go from A to C every day, their travel time is $\frac{4000}{100} + 45 = 85$.

New roads are build, so that A is connected to C also through D. To ease computations, we assume that the road AD has the same travel time as that between B and C, while the road DC is identical to AB. See Figure 2. How will our $4000$ commuters distribute between the two routes? Clearly, the evident symmetry between ABC and ADC suggests that half will go through B and half through D. The resulting travel times are

$$\begin{array}{rcl} \text{AB} & + \text{BC} = \text{ABC} \\ \frac{2000}{100} & + 45 = & 65 \end{array} \quad \text{and} \quad \begin{array}{rcl} \text{AD} & + & \text{DC} & = \text{ADC} \\ 45 & + & \frac{2000}{100} & = & 65. \end{array}$$

Note that this configuration is an example of a Nash equilibrium. Indeed, look at each driver as to a *player*, whose objective is to minimize his/her travel time. Imagine that one of the commuters passes from using

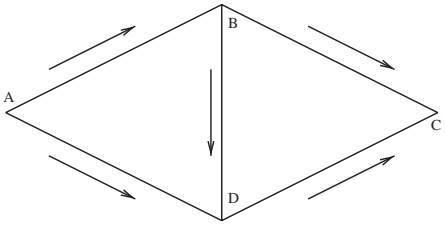


**Figure 3**

route ABC to route ADC. Then, the new travel times become

$$\begin{array}{rcl} \text{AB} & + \text{BC} = \text{ABC} \\ \frac{1999}{100} & + 45 < & 65 \end{array} \quad \text{and} \quad \begin{array}{rcl} \text{AD} & + & \text{DC} & = \text{ADC} \\ 45 & + & \frac{2001}{100} & > & 65, \end{array}$$

showing that this change of mind is not convenient for our commuter.

Due to the intensive use of this network, a new road is now built connecting B to D. This brand new road allows for infinite speed, so that commuters reach D from B in time $0$. See Figure 3. Now, to go from A to C three choices are available: ABC, ABDC and ADC. How will commuters distribute among these three different routes? In other words, is there a new Nash equilibrium? And which is it?

A moments' thought reveals that if $x$ commuters travel along ABC, $y$ along ABDC and $z$ along ADC, then the three travel times are

$$\begin{aligned} \text{ABC} &= \text{AB} + \text{BC} &&= \frac{x+y}{100} + 45 \\ \text{ABDC} &= \text{AB} + \text{BD} + \text{DC} &&= \frac{x+y}{100} + 0 + \frac{y+z}{100} \\ \text{ADC} &= \text{AD} + \text{DC} &&= 45 + \frac{y+z}{100}. \end{aligned}$$

The equality of the travel times yields $x+y = y+z = 4500$, which is inconsistent with the total number of commuters being $x+y+z = 4000$. Thus, no equilibrium exists when all routes are used. (Here, an *equilibrium* configuration is a distribution of drivers yielding the same travel time along all used routes, see for instance [6].)

A reasonable guess is now that the Nash equilibrium prior to the construction of BD still is a Nash equilibrium in presence of BD. But it is not the case because it is convenient for a driver to pass from, say, route ABC to route ABDC. Indeed, the travel time corresponding to $x = 2000$, $y = 0$ and $z = 2000$ is $65$ along both routes ABC and ADC. On the other side, setting $x = 1999$, $y = 1$ and $z = 2000$ results in the travel times

$$\text{ABC} = 65,\ \text{ABDC} = 40.01,\ \text{ADC} = 65.01.$$

Thus, we imagine that as soon as BD is opened to commuters, they find convenient to use it. As more and more commuters drive along the new route ABDC, the corresponding travel time increases, but remains lower than those along ABC and ADC. Thus, the new Nash equilibrium corresponds to all commuters driving along ABDC, that is $x = 0$, $y = 4000$ and $z = 0$.

Here comes the paradox: the new travel time is $\frac{4000}{100}+0+\frac{4000}{100}=80$! It is *higher* than prior to the construction of $\mathrm{BD}$!

In other words, adding a road to a network may make the network less efficient, even if the new road is, by itself, extremely efficient (our $\mathrm{BD}$ segment is traveled at infinite speed!).

This remark is clearly counter-intuitive, paradoxical, but it is *real*. Here, we refer to the famous closing of 22nd street in New York that took place on 22 April 1990, see [11], and to [3]. The reader is invited to personally search for other real examples.

The literature on Braess paradox [4] is vast and currently comprises a variety of phenomena not necessarily related to traffic on road networks. During the meeting at the Lorentz center, some discussions centered about the possibility that PDE based macroscopic models are able to capture this paradox. A first result in this connection is [6], but several questions remain unanswered. For instance, can the dynamics of PDE describe the *insurgence* of a Braess-like regime in a network? It goes without saying that this descriptive ability is preliminary to tackling optimal management problems.

A somewhat *inverse* Braess paradox is reported to happen in flocks of sheep passing through narrow gates [8], see also https://www.youtube.com/watch?v=mkqhYhdgvGg. Similarly, Hughes [10] suggested that an obstacle, suitably placed in front of an exit, may increase the through flow of pedestrians, thus reducing evacuation time. Indeed, even if the presence of obstacles may be seen as leading to a worse condition, it may reduce the inter-pedestrian pressure near the exit and prevents it from blocking. This effect opens interesting perspectives for safety and efficient evacuation of buildings and other confined environments. It is thus of great importance to develop mathematical models capable to describe these phenomena. In this perspective, various studies [13] showed that models based on the classical mass conservation equation

$$\partial_t \rho + \mathrm{div}(\rho\, V(\rho)\, \vec{\mu}) = 0, \qquad (2)$$

where $\vec{\mu}$ is the vector field of the trajectories followed by pedestrians (possibly dependent on the density distribution through an eikonal equation), are not able to capture this behavior. One needs either to add a momentum balance equation describing acceleration, or to account for inter-pedestrian interactions through non-local terms. In the latter case, the velocity vector field is given in the form

$$\vec{\mu}(t,x) = \vec{v}(x) - \varepsilon \frac{\nabla(\rho * \eta)}{\sqrt{1+\| \rho * \eta \|^2}}, \qquad (3)$$

where $\vec{v}$ is the vector field of the preferred path, for example the shortest to destination, tempered by the latter non-local term, which pushes pedestrians towards low density regions through the action of a suitable positive mollifier $\eta$ (here $\rho * \eta$ is the usual convolution product), accounting for the local density distribution.

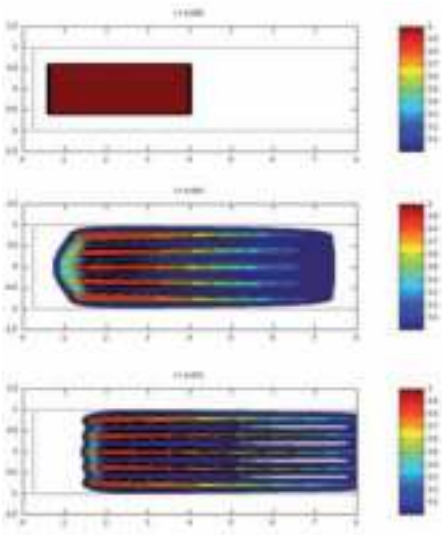


**Figure 4**

Model (2), (3) is able to display not only Braess paradox but also lane formation [5]. Indeed, Figure 4 shows the movement of a crowd. Initially, individuals are uniformly distributed in the rectangle above. At time $t=0$, they move to the right, the ones in front being faster. Without any ad hoc prescription, individuals form five lanes of higher density (see [5] for details about the numerical integration).

**Acknowledgement**
The authors acknowledge the financial support from the Lorentz Center in Leiden for the organized workshop, which was co-financed also by INRIA (France) and Stiftelsen GS Magnusons fond (MG2018-0019; Sweden).

**References**

1 A.S. Ackleh, D.F. Marshall, H.E. Heatherly and B.G. Fitzpatrick, Survival of the fittest in a generalized logistic model, *Mathematical Models and Methods in Applied Sciences* 9 (1999), 1379–1391.

2 A.S. Ackleh, B.G. Fitzpatrick and H.R. Thieme, Rate distributions and survival of the fittest: A formulation on the space of measures, *Discrete and Continuous Dynamical Systems Series B* 5 (2005), 917–928.

3 L. Baker, Removing roads and traffic lights speeds urban travel, *Scientific American*, 20–21 February 2009.

4 D. Braess, Über ein Paradoxon aus der Verkehrsplanung, *Unternehmensforschung* 12 (1968), 258–268.

5 R.M. Colombo, M. Garavello and M. Lécureux-Mercier, A class of non-local models for pedestrian traffic, *Mathematical Models and Methods in Applied Sciences* 22(4) (2012), 1793–6314.

6 R.M. Colombo, H. Holden, On the Braess paradox with nonlinear dynamics and control theory, *J. Optimization Theory and Applications* 168 (2016), 216–230.

7 R.M. Colombo, M. Lécureux-Mercier, An analytical framework to describe the interactions between individuals and a continuum, *J. Nonlinear Sci.* 22(1) (2012), 39–61.

8 A. Garcimartín, J.M. Pastor, L.M. Ferrer, J.J. Ramos, C. Martín-Gómez and I. Zuriguel, Flow and clogging of a sheep herd passing through a bottleneck, *Phys. Rev. E* 91(2) (2015), 022808–022815.

9 J. Grimm and W. Grimm, *Deutsche Sagen*. Nicolaische Verlagsbichhandlung, second edition, 1865.

10 R. Hughes, The flow of human crowds, *Annual Review of Fluid Mechanics* 35 (2003), 169–182.

11 G. Kolata. What if they closed 42d Street and nobody noticed? *New York Times*, 25 December 1990.

12 J. Nash, Non-cooperative games, *Annals of Mathematics* 54 (1951), 286–295.

13 M. Twarogowska, P. Goatin and R. Duvigneau, Macroscopic modeling and simulations of room evacuation, *Appl. Math. Model.* 38(24) (2014), 5781–5795.

Klaas van Harn
*Afdeling Wiskunde, FEW*
*Vrije Universiteit Amsterdam*
*k.van.harn@vu.nl*

**In Memoriam** Piet Holewijn (1935–2019)

# Inspirerend en enthousiasmerend in de waarschijnlijkheidsrekening

**Op 4 februari 2019 overleed Piet Holewijn, emeritus hoogleraar waarschijnlijkheidsrekening aan de Vrije Universiteit Amsterdam. Hij werd 83 jaar. Klaas van Harn, met wie Holewijn tijdens zijn laatste levensfase nog een boek heeft afgerond, kijkt terug op het werk en leven van zijn collega en vriend.**

**Proloog**

Op 4 februari 2019 overleed Petrus Johannes (Piet) Holewijn, na jaren van afnemende gezondheid. Piet was emeritus hoogleraar van de Faculteit der Exacte Wetenschappen van de Vrije Universiteit Amsterdam. Hij hield zich met name bezig met de fundamentele waarschijnlijkheidsrekening en richtte zich gaandeweg steeds meer op het schrijven van notities, syllabi en boeken die van nut waren bij de begeleiding van studenten en promovendi. In het navolgende levensbericht beschrijf ik, na een korte schets van Piets wetenschappelijke carrière, vooral deze laatste activiteiten van hem en de wijze waarop hij hier ook in het contact met zijn omgeving invulling aan gaf. Als achtergrond moge dienen dat ik veertig jaar met Piet heb samengewerkt, twintig jaar voor en twintig jaar na zijn emeritaat, vooral op het gebied van onderwijs. Bij deze beschrijving heb ik met dankbaarheid kunnen gebruikmaken van bijdragen vanuit Piets verdere 'wiskundige' vriendenkring: zijn Delftse studiegenoot Frans Schurer en latere collega Boele Braaksma, zijn promovendi Frans Boshuizen, José Gouweleeuw en Pieter Allaart, en zijn laatste groepje studenten, bestaande uit Hendrik Blauwendraat, Alex van den Brandhof, Laura Spierdijk en Martijn de Vries.

**Wetenschappelijke carrière**

Piet werd geboren in Utrecht op 19 mei 1935. Na het behalen van zijn middelbare schooldiploma studeerde hij aan de Technische Hogeschool Delft (THD, later TUD). Toentertijd moest je, alvorens aldaar een wiskundestudie te kunnen beginnen, eerst een tweejarige propedeuse doen in een technische richting; Piet deed werk-

Foto: Alex van den Brandhof

Piet Holewijn in zijn tuin in Nederhorst den Berg, op 27 juli 2018

tuigbouwkunde en studeerde vervolgens, in 1962, af als wiskundig ingenieur bij prof. dr. L. Kuipers. Hij werd vervolgens benoemd tot wetenschappelijk ambtenaar bij de Afdeling der Algemene Wetenschappen — waartoe de Onderafdeling der Wiskunde behoorde — van de THD. Bij genoemde hoogleraar promoveerde hij, in 1965, op een onderwerp uit de functietheorie; zie [7]. Het proefschrift bevatte ook alternatieve bewijzen op basis van kanstheoretische resultaten, zoals de continuïteitsstelling voor karakteristieke functies. Tot 1971 bleef Piet aan de THD verbonden. In zijn Delftse periode had hij, behalve met zijn latere collega Boele Braaksma, vooral contact met zijn studiegenoten Frans Schurer, Pleun van der Steen en Han Lemei; de twee laatstgenoemden zijn inmiddels overleden. Een dankbaar, tot overdrijving uitnodigend, onderwerp van hun gesprekken was onder andere de kwaliteit, of het vermeende gebrek eraan, van de hoogleraren zuivere wiskunde toentertijd aan de THD. Intussen was Piet ook getrouwd, met Ineke Nelemans, en kregen zij twee zoons, Bart en Peter. Het hele gezin ging in 1966/1967 een jaar naar Ann Arbor in de Verenigde Staten voor een studieverblijf.

In 1971 kreeg Piet een aanstelling bij de Afdeling Wiskunde van de Vrije Universiteit (VU) in Amsterdam, eerst als lector en later als hoogleraar in de waarschijnlijkheidsrekening. Vanaf het begin beijverde hij zich voor het verwerven van een zelfstandige positie van zijn vakgebied binnen de wiskunde. De studenten waardeerden zijn inspanningen zeer; uit eigen gelederen hebben vijf promovendi onder Piets begeleiding een proefschrift geschreven. Piet was een onafhankelijke geest met een eigen visie op het universitair bedrijf en de positie van zijn vakgebied daarbinnen en met een groot relativerend gevoel voor humor. In september 1999 nam Piet afscheid van de VU met een rede getiteld ‘Wat ik nog opmerken wil’.

**Promovendi en hun onderzoek**

Piet heeft niet veel artikelen gepubliceerd in (internationale) wiskundetijdschriften. Wel schreef hij vele ‘oriënterende studies’: gedetailleerde omschrijvingen van interessante problemen met specifieke onderzoeksvragen en een eerste aanzet tot oplossingen. En dat alles in zijn karakteristieke, maar nette en duidelijke handschrift, ook toen het nieuwe tekstverwerkingssysteem LaTeX allang zijn intrede had gedaan. Zo effende hij het pad voor zijn vijf promovendi, voor wie hij misschien meer coach was dan mede-onderzoeker. Hij legde contacten in binnen- en buitenland en drong erop aan om resultaten te delen met andere onderzoekers op hetzelfde of gerelateerd terrein. Via deze contacten zorgde Piet ervoor dat hun resultaten in vruchtbare bodem vielen, met als gevolg uitnodigingen op buitenlandse universiteiten en voor congressen om lezingen te geven. Maar ook op de eigen afdeling bood Piet hun volop gelegenheid om voor medewerkers en gevorderde studenten onderzoek ‘heet van de naald’ te presenteren.

Een blik op de titels van de proefschriften van de vijf promovendi, Henry Berbee, Don van der Vecht, Frans Boshuizen, José Gouweleeuw en Pieter Allaart, (zie respectievelijk [2], [9], [3], [4] en [1]) geeft een duidelijk beeld van 25 jaar promotieonderzoek onder leiding van Piet. Centrale thema’s waren stochastische wandelingen, vernieuwingstheorie, inbedding in de Brownse beweging, optimale stopproblemen, bereik van vectormaten en optimaal partitioneren. Bij dit promotieonderzoek waren steeds vooraanstaande buitenlandse onderzoekers, als copromotor of anderszins, betrokken; aanvankelijk met name Isaco Meilijson (Tel Aviv), later vooral David Gilat (Tel Aviv) en Ted Hill (Atlanta). In de zomerperioden zorgde Piet er vaak voor dat deze en andere internationale coryfeeën naar Nederland kwamen en organiseerde hij een feestje bij hem thuis in Nederhorst den Berg; dat waren inspirerende tijden. Overigens had Piet ook goede contacten met Nederlandse kansrekenaars, zoals Arie Hordijk, Mike Keane, Theo Runnenburg, Fred Steutel, Guus Balkema, Wim Vervaat en Frank den Hollander. Piets onderzoeksactiviteiten werden op passende wijze afgesloten met een symposium bij zijn afscheid in 1999, waar niet alleen zijn promovendi Boshuizen, Gouweleeuw en Allaart spraken, maar ook hun begeleiders Meilijson, Gilat en Hill.

**Studenten en hun onderwijs**

Dat Piet gemakkelijk promovendi uit eigen gelederen kon rekruteren, had alles te maken met het feit dat hij een voortreffelijk en inspirerend docent was (en dat hij veel zorg besteedde aan zijn onderwijsmateriaal; zie daarvoor de volgende paragraaf). Ook vele elders gepromoveerden in de waarschijnlijkheidsrekening, zoals Rob van der Mei, Harry van Zanten, Michel Mandjes, Sandjai Bhulai, Nam Kyoo Boots en Bert Zwart, hadden een degelijke basis en motivatie gevonden in Piets onderwijs. Bij de colleges voor gevorderde studenten, met kleinere groepen, was er ruimte voor de werkgroepsvorm en werden de studenten bij het presenteren van de stof en het uitwerken van de opgaven betrokken. Tijdens pauzes knoopte Piet graag een gesprekje aan: waar kom je vandaan, waarmee houd je je bezig buiten de studie — Piet was oprecht geïnteresseerd. Zijn colleges waren doorspekt met zelfspot en humor, vaak ook nog gerelateerd aan het onderwerp. Soms had hij in een bewijs behoefte aan een ‘woest stuk van de uitkomstenruimte’ of, ter afronding van het bewijs, aan een stukje bordruimte van de maat nul, en kon hij concluderen dat het bewijs ‘ontroerend van eenvoud’ was. Het laatste was nogal eens het geval, omdat Piet altijd op zoek was naar het meest elegante bewijs; wiskunde werd zo bij hem een kunstvorm.

Met een goed gevoel voor theater bediende Piet in de grote collegezalen het knoppenpaneel vóór in de zaal. Met hulp van studenten lukte het hem meestal wel de overheadprojector aan de praat te krijgen. Dat ‘machien’, zoals hij ieder technisch apparaat noemde, was voor hem in het lagerejaars onderwijs een belangrijk object. De collegestof werd hoofdzakelijk gepresenteerd op een flink aantal handgeschreven sheets, die in hoog tempo na elkaar op de projector werden gelegd, steeds uitvoerig voorzien van commentaar. Niet iedere student wist daar even goed raad mee. Moest je nu lezen, luisteren of vooral de tekst op de sheets overnemen? Een driftig schrijvende student verzuchtte daarom eens tijdens een college dat Piet zo ontzettend veel praatte. “Ik praat misschien een hoop, maar ik zeg weinig”, was het geruststellend bedoelde antwoord. Indien nodig wist Piet als geen ander met een kwinkslag de kou uit de lucht te halen. Overigens had hij met eerstejaars nog wel eens moeite; hij kon zich verbázen over hun gedrag. Op zo’n moment citeerde hij eens, vanuit zijn gereformeerde achtergrond, de twee laatste regels van Psalm 118, vers 11 (oude berijming): “Het is een wonder in onz’ ogen; wij zien het, maar doorgronden het niet.” Piet had wel meer de neiging om te overdrijven, maar deed dat altijd op een humorvolle wijze.

Piet Holewijn spreekt José Gouweleeuw toe na afloop van de verdediging van haar proefschrift, op 1 september 1994.

Piet nam de voorbereiding van zijn colleges heel serieus. Hij richtte zich niet slechts op het feitelijk overbrengen van kennis, maar wilde zich telkens inleven in het onderwerp. Je kon hem ijsberend op de gangen tegenkomen, waarbij hij handgebaren maakte alsof hij dingen aan het uitleggen was. Geen wonder dat hij wel gezien werd als een tovenaar die de magische uitstraling van zijn vak op zijn studenten overbracht.

**Auteur van leerboeken en syllabi**

Zoals gezegd heeft Piet zich vanaf het begin van zijn aanstelling aan de VU, in 1971, sterk gemaakt voor de kansrekening als zelfstandige discipline binnen de wiskunde. Toen ik eind 1978 na mijn promotie aan de THE (later TU/e) medewerker van Piet werd, was hij volop bezig met het moderniseren van het kansrekening-onderwijs. Bij mijn komst had hij al een syllabus over Markov-ketens geschreven als ook een syllabus over fundamentele (maattheoretische) waarschijnlijkheidsrekening, en richtte hij zich op de elementaire kansrekening in het eerste jaar, die toen nog samen met de statistiek, onder de naam 'stochastiek', werd gegeven. Vanaf het begin betrok hij mij daarbij en gaf mij volop ruimte mijn Eindhovense ervaringen in te brengen. Zo ontstond een prettige en zeer vruchtbare samenwerking. Daarbij had Piet de gave om een overleg, ook als we het nog niet eens waren, altijd positief af te sluiten, al was het maar met de vaak gehanteerde uitspraak: "We zullen naar bevind van zaken handelen." Piet en ik hadden overigens dezelfde onderwijsvisie; beiden werkten we zoveel mogelijk met 'bord en krijt', want, zoals Piet ooit zei: "Wiskunde bedrijven is een ambachtelijk werk."

De opkomst van LaTeX in het begin van de jaren negentig was voor ons aanleiding om van de Markov-syllabus een boekje te maken, dat in 1991 verscheen in de bekende Epsilon-serie met vooral Nederlandstalige uitgaven en waarvan een tweede druk werd vervaardigd in 2003; zie [5]. Op verzoek van vooral de studenten BWI, Bedrijfswiskunde en informatica, werd vervolgens ook een LaTeX-syllabus geschreven voor de kansrekening in het eerste jaar. Jarenlang is daar met veel plezier mee gewerkt, ook bij de afdeling Econometrie, waar we een goed contact hadden met de meer toegepaste kansrekenaars Henk Tijms en Rein Nobel. In de loop van de jaren negentig trokken Piet en ik het onderwijs in de maat- en integratietheorie naar ons toe. Ook over dit onderwerp schreven we een Epsilon-boek, in 2005, met een tweede druk in 2008; zie [8]. Met dit boek was een mooie basis gelegd voor de meer fundamentele kansrekening, die aan de orde kwam bij het vak Grondslagen van de waarschijnlijkheidsrekening. Na enige aarzeling besloot Piet ook mee te werken aan het vastleggen van onze jarenlange ervaringen met dit vak in een boek, dat onlangs — helaas net na het overlijden van Piet — verschenen is bij VU University Press; zie [6]. De drie hiervoor genoemde boeken, alle met een uitgebreide collectie opgaven, zijn in wiskundig Nederland goed ontvangen. Recensenten spraken bij het Markov-boek van een cursus die didactisch goed in elkaar zit en duidelijk in de praktijk beproefd is, bij het Maat en integraal-boek van bewondering voor het monnikenwerk, van ouderwets goed Nederlands en van uitvoerigheid die niet ontaardt in wijdlopigheid, en bij het Waarschijnlijkheidsrekening-boek van een glasheldere en precieze schrijfstijl. De opmerking over het Nederlands zou een dubbele bodem gehad kunnen hebben; Piets taalgebruik was af en toe lichtelijk archaïsch, maar ik had niet de neiging om in die gevallen alternatieven voor te stellen omdat ik dat soort taal wel bij het vakgebied vond passen.

Piet heeft ook jarenlang geschreven en geschaafd aan twee syllabi die gebruikt werden bij keuzevakken in de masterfase: Voortgezette waarschijnlijkheidsrekening, met de theorie van voorwaardelijke verwachtingen en martingalen, en Brownse beweging, het proces van Wiener. Pas in de laatste paar jaar voor zijn emeritaat was Piet er dermate tevreden over dat hij er getypte (geen LaTeX-) versies van wilde laten maken. Zoals gezegd waren de studenten heel positief over dit onderwijs. Als een misschien niet geheel representatieve illustratie daarvan citeer ik een briefje van de in de proloog genoemde 'bende van vier' dat ik bij het opruimen van Piets archief vond en dat kennelijk bij de overhandiging van een fles wijn was bijgevoegd: "Hartelijk dank voor de enthousiaste begeleiding van de cursus Voortgezette waarschijnlijkheidsrekening. We hebben er allemaal buitengewoon veel plezier aan beleefd. Het samen wiskunde beoefenen was een ervaring om nooit te vergeten. We hopen u nog regelmatig te zien."

**Actief tot het einde**

De vier studenten van Piets laatste college — 'luitjes', zoals ze ook op college door hem werden aangesproken — hebben inderdaad contact gehouden. Regelmatig kwamen ze, gezamenlijk of afzonderlijk, in Nederhorst den Berg op bezoek. Wanneer Piet zich goed voelde — en dat was helaas niet altijd het geval —, zat hij al gauw op

zijn praatstoel en waren de anekdotes niet van de lucht. Een geliefd doelwit waren de eigenaardigheden van bekende wiskundigen die hij tijdens zijn werkzame leven had ontmoet, zoals Benoît Mandelbrot met zijn grote schoenmaat en de legendarische Paul Erdős, die goed kon tafeltennissen en tijdens een van de jaarlijkse bijeenkomsten voor stochastici in Lunteren Piet complimenteerde met diens sterke backhand. Maar al mocht Piet graag praten, uiteindelijk was hij bescheiden in de mate waarin hij over zichzelf sprak. En voortdurend was er de belangstelling voor zijn voormalige pupillen. Dat betrof ook zijn laatste drie promovendi, Frans, José en Pieter. Het contact met Pieter, die na zijn promotie naar Texas vertrok, was incidenteel, maar in het voorjaar van 2018 was er nog een mooie ontmoeting toen Pieter zijn sabbatical doorbracht in Utrecht, de plaats waar Piet geboren en getogen was. Hij leefde ook erg mee met Frans en José — zij trouwden met elkaar — en hun twee kinderen; toen deze nog klein waren, werd er nog wel eens met 'opa Piet' in de bossen gewandeld en pannenkoeken gegeten. Ook de vier al eerder genoemde vrienden uit de Delftse periode werden gastvrij in Nederhorst den Berg ontvangen; oude herinneringen werden opgehaald onder het genot van een stevig glas wijn. En over zijn ergernissen in een revalidatiekliniek waar hij noodgedwongen enige tijd had doorgebracht, kon Piet zo geestig vertellen dat je die ellende bijna vergat.

Een ieder heeft in het leven te kampen met tragiek en tegenslagen. Ook Piet zijn die niet bespaard gebleven; te denken valt aan de scheiding met Ineke en het in december 1996 zo plotselinge overlijden van zijn nieuwe partner, Corry van Rossum, misschien wel de grootste klap die hij te verduren heeft gehad. Corry, jarenlang secretaris van de Afdeling Wiskunde van de VU, was Piets grote steun en toeverlaat. Met name de eerste jaren na zijn emeritaat waren moeilijk voor hem. Mede door het wiskundig actief zijn, ook buiten de kansrekening — hij hield zich een tijdje bezig met Galois-theorie — , kreeg hij weer een doel in het leven. Zijn gezondheid liet echter steeds meer te wensen over. Maar telkens krabbelde hij weer op en gaf hij blijken van grote veerkracht.

Onder deze wisselende omstandigheden zijn we jarenlang, met name de vijf jaar na mijn pensionering in 2013, met het schrijven van het boek [6] over maattheoretische waarschijnlijkheidsrekening bezig geweest. Talloze malen ben ik vanuit mijn woonplaats Hilversum of op de terugweg vanaf de VU bij Piet langs gefietst voor besprekingen en om hem te voorzien van nieuwe stukken tekst, die ik, zoals Piet dat verwoordde, uit het 'machien' getoverd had; aan mailverkeer deed hij allang niet meer. Piet wilde graag een completerend hoofdstuk over stochastische wandelingen toevoegen, maar het kostte ons nogal wat tijd om zijn aantekeningen hiervan om te zetten in een begrijpelijke tekst en die te voorzien van bijpassende opgaven. Ook werd Piets energie gaandeweg steeds minder, dit ten gevolge van problemen met zijn hart. Overigens had Piet het zelden over 'mijn hart'; vaak sprak hij over 'het hart', alsof dat een entiteit was buiten hemzelf. Zo omschreef hij eens op zijn geheel eigen wijze het feit dat het die dag goed met hem ging, als volgt: "Het hart gedraagt zich vandaag buitengewoon coöperatief."

Piet heeft het boek helaas niet in handen mogen houden. Wel heeft hij tot het einde toe eraan bijgedragen; hij overleed op de avond van de dag dat het manuscript in finale vorm naar de uitgever ging. Maar eigenlijk was de tekst al vastgesteld tijdens ons laatste overleg in Nederhorst den Berg op 13 december 2018, een datum die Piet herinnerde aan Corry's begrafenis. Misschien daardoor liet hij mij toen, eigenlijk voor het eerst, iets blijken van zijn levensvisie: hij was bezig met het lezen van het laatste Bijbelboek, Openbaring. Ik kreeg de indruk dat hij zich identificeerde met de apostel Johannes, niet vanwege zijn naam, maar door het feit dat Johannes op het eiland Patmos in eenzaamheid woorden van Jezus ontving. Op mijn opmerking dat die woorden heel perspectiefrijk zijn, reageerde Piet dat ze ook wel heel moeilijk te begrijpen zijn. Dat was typisch Piet; hij maakte het zich nooit gemakkelijk. Voor ik vertrok, op die dertiende december, vertelde hij nog dat hij erg uitzag naar een verblijf in Drenthe, met zijn zoons Bart en Peter, in de week voor kerst. Toen ik vervolgens de oprit bij zijn huis affietste, stond hij zoals altijd in de kamer klaar om als groet zijn hand op te steken, en dacht ik, niet wetende maar wel vrezende dat dit de laatste keer zou zijn: vaarwel, Piet; misschien tot ziens; bye bye.

**Epiloog**

Uit het bovenstaande moge duidelijk zijn dat Petrus Johannes Holewijn gedurende een groot deel van zijn leven in de wereld van de waarschijnlijkheidsrekening een inspirerende en enthousiasmerende rol heeft gespeeld. Hij was correct, welbespraakt en sympathiek. Hij had echte belangstelling voor de ander en was solidair met mensen die het moeilijk hadden in het leven. Daarbij was hij een meester in het beeldend en humorvol beschrijven van gebeurtenissen en situaties, ook wanneer die om kritiek vroegen; hij kon 'literair mopperen' zoals iemand eens opmerkte. Zijn heengaan is een groot verlies voor Bart en Peter en zijn verdere familie, maar ook vele anderen zullen hem node missen. Wij — degenen die hebben bijgedragen aan dit levensbericht, dit eerbetoon aan Piet — verloren een dierbare vriend.

**Referenties**


1 P.C. Allaart, *Ranges of Vector Measures and Optimal-Partitioning Inequalities,* Dissertation, Vrije Universiteit Amsterdam, 1998.

2 H.C.P. Berbee, *Random Walks with Stationary Increments and Renewal Theory,* Mathematical Centre Tracts 112, Mathematisch Centrum, Amsterdam, 1979.

3 F.A. Boshuizen, *Prophet and Minimax Problems in Optimal Stopping Theory,* Dissertation, Vrije Universiteit Amsterdam, 1991.

4 J.M. Gouweleeuw, *On Ranges of Vector Measures and Optimal Stopping,* Dissertation, Vrije Universiteit Amsterdam, 1994.

5 K. van Harn en P.J. Holewijn, *Markov-ketens in discrete tijd,* Epsilon-Uitgaven 21, Utrecht, 1991, 2de druk 2003.

6 K. van Harn en P.J. Holewijn, *Waarschijnlijkheidsrekening – maattheoretische uitgangspunten en fundamentele eigenschappen,* VU University Press, Amsterdam, 2019.

7 P.J. Holewijn, *Contributions to the Theory of Asymptotic Distribution Modulo 1,* Dissertation, Technische Hogeschool Delft, 1965.

8 P.J. Holewijn en K. van Harn, *Maat- en integratietheorie – met basiselementen van de waarschijnlijkheidsrekening,* Epsilon-Uitgaven 58, Utrecht, 2005, 2de druk 2008.

9 D.P. van der Vecht, *Inequalities for Stopped Brownian Motion,* CWI Tracts 21, CWI, Amsterdam, 1985.

Johan van Benthem
*ILLC*
*Universiteit van Amsterdam*
*j.vanbenthem@uva.nl*

Dick de Jongh
*ILLC*
*Universiteit van Amsterdam*
*d.h.j.dejongh@uva.nl*

**In Memoriam** Anne Sjerp Troelstra (1939–2019)

# Creating order in a vast and diverse area

**On 7 March 2019 Anne Sjerp Troelstra passed away at the age of 79. Troelstra was emeritus professor at the Institute for Logic, Language and Computation of the University of Amsterdam. He dedicated much of his career to intuitionistic mathematics and wrote several influential books in this area. His former colleagues Johan van Benthem and Dick de Jongh look back on his life and work.**

Anne Troelstra was born on 10 August 1939 in Maartensdijk. After obtaining his gymnasium beta diploma at the Lorentz Lyceum in Eindhoven, in 1957, he enrolled as a student of mathematics at the University of Amsterdam — where eventually his interests converged on intuitionism with Arend Heyting as his teacher. Students he was close with include Olga Bakker and E.W. Beth's students Dick de Jongh and Hans Kamp. With Dick de Jongh, he wrote a pioneering paper on intuitionistic propositional logic, published only in 1966, that contained the first definition of the central notion of a $p$-morphism between relational frames, widely used today in intuitionistic and modal logics, as well as the simplest form of the duality between Heyting algebras and relational frames. After obtaining his master's degree in 1964, Anne at once became an assistant professor, according to a custom of the time. It took him just two years from there on to finish his dissertation, supervised by Heyting. Besides intuitionism, a main interest of Heyting was geometry and perhaps not accidentally, Anne's PhD thesis was a study of intuitionistic topology. This subject made him aware of the role of continuity in intuitionistic mathematics, a concept that was to play an important part in his research in the years to come, in many different forms.

**Stanford**

Anne then obtained a scholarship to Stanford to visit Georg Kreisel, a leading scholar in intuitionism and proof theory, and spent the academic year 1966–1967 there, together with Olga whom he had married the year before. In this period, Anne sharpened and modified Kreisel's ideas on choice sequences, the basic notion underlying the real number continuum in an intuitionistic perspective. Working together they created formalizations of analysis resulting in a large article where the typically intuitionistic concept of a lawless sequence of numbers that successfully evades description by any fixed law, reached its final form. Though the latter notion is very characteristic for intuitionistic mathematics as based on free creation of objects, lawless sequences are one of the rare concepts in intuitionism not already introduced by Brouwer himself.

Anne Troelstra

**Constructivism**

In August 1968, Anne played a central role in the famous Buffalo Conference on Intuitionism and Proof Theory, a meeting of all important logicians of the day with an interest in constructivism. He gave a series of ten lectures there, which turned into his first book, published in 1969, that contained the core of his seminal ideas on intuitionistic formal systems and their metamathematical investigation. Back home in 1968, he became a lector (associate professor) in 1968, and a full professor in 1970. Further early recognition was to follow. In 1976, he became a member of the Dutch Royal Academy of Sciences. As a full pro-

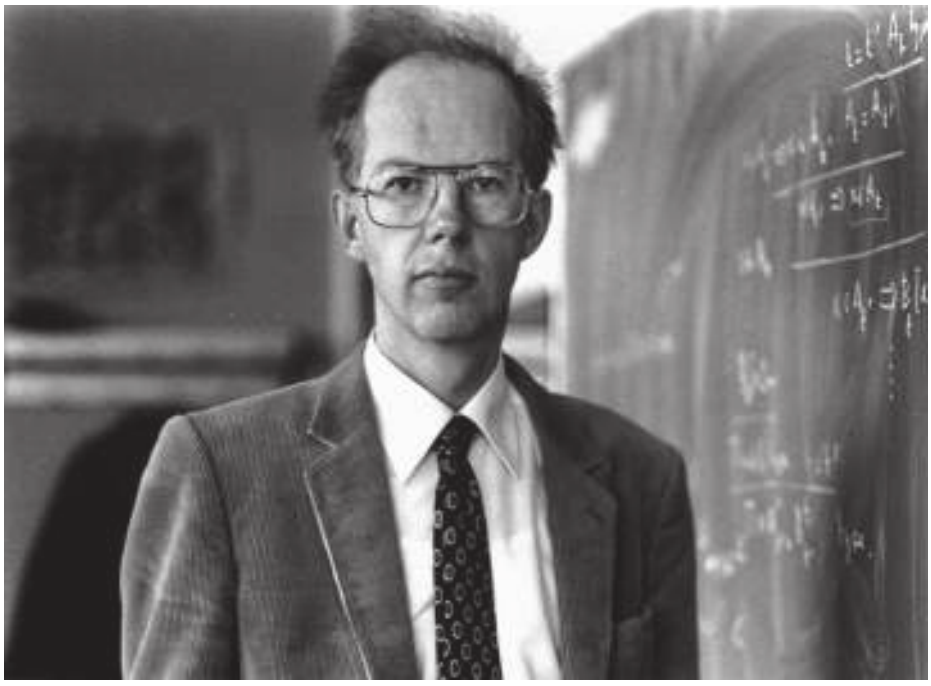


Photo: School en Universiteit

Anne Troelstra in 1985

fessor, Anne was the successor of Arend Heyting, himself a successor to L. E. J. Brouwer. Each was in his time the principal constructivist in the world.

**Realizability**

The metamathematics of intuitionistic systems was a chaotic jumble of results when Anne entered it. Here he showed his greatest strength: creating order in a vast and diverse area. In 1973, the rigorous order was there in his book *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Especially striking are the clarification of the properties of different models for constructive mathematics, several types of realizability, and various functional interpretations. The last chapters on special topics were written by Jeff Zucker, Craig Smorynski and Bill Howard, but the lion's share had been written by Anne, editor and architect of the whole. The book, known in the community as 'Springer 344', still functions as a landmark for researchers in the subject.

The metamathematical style of investigation in this work represents a significant broadening of earlier research. In tandem with developing branches of intuitionistic mathematics such as constructive arithmetic, analysis, topology or algebra, one now also studies the resulting formal theories as mathematical objects, using the tools of logic such as proof theory, recursion theory, and model theory. Perhaps the most prominent among the many results obtained by Anne in this style concern 'realizability', a mathematical analysis of constructive truth introduced by Stephen Kleene in the 1940s. Its purpose was to explain what is 'constructive' about constructive mathematics in terms of recursive functions on natural numbers, then and now a standard paradigm for effective computability. Long viewed as an outside look at intuitionism from the standpoint of classical logic, Anne took realizability to the heartland of intuitionism. Anne's studies of the topic had already resulted in a long article in the proceedings of the Second Scandinavian Logic Symposium of 1971, and they were now expanded in Springer 344.

Troelstra's celebrated theorem characterizing realizability states that the formulas that are provably realizable in Heyting Arithmetic (intuitionistic natural number theory, arguably the most basic formal system of intuitionistic mathematics) are exactly the ones provable from Heyting's Arithmetic plus the so-called Extended Church Thesis. This result displays some typical features of the metamathematical style of investigation. First, the equivalence is formulated in a proof-theoretic setting. Secondly, informal claims such as Church's Thesis saying that every intuitively computable function can be computed by a recursive function, or equivalently a Turing machine, are formulated within number theory. Church's Thesis says in this format that, for any true intuitionistic quantifier combination $\forall x\, \exists y\, B(x,y)$ (the typical sort of statement where constructive mathematics demands a procedure to produce the $y$'s from the $x$'s), there exists a recursive function $f$ such that $\forall x\, B(x,f(x))$. From an informal thesis, Church's thesis has now become a precise axiom that can be true or false in intuitionistic mathematics. In the Extended Church Thesis the universal quantifier is relativized to antecedents $A$ that have to be special 'almost negative formulas', untouched by realizability: If $\forall x\,(A(x) \to \exists y\, B(x,y))$, then $\forall x\,(A(x) \to B(x,f(x)))$ for some recursive function $f$. This is a typical form of careful metamathematical attention to syntactic fine-structure of mathematical assertions that has become standard in the subsequent literature.

Troelstra's result was the first of its kind, and it is widely seen as making the force and meaning of realizability clear within the context of intuitionism. The influence of this work shows in a broad literature analyzing other important notions of realizability with similarly structured characterizations. This study of realizability later developed broader mathematical connections with category theory and topos theory. An interest in this topic remained with Anne for life. In 1998, a chapter on realizability by his hand came out in the *Handbook of Proof Theory*. Characteristically, Anne's text had been finished a few years before, faithfully meeting his deadline, but delays by other authors kept him updating, somewhat grumblingly, with all new results in the area. What he published, on this and in fact on any topic, had to be the complete state of the art.

**Choice sequences**

A next landmark was Anne's book *Choice Sequences* from 1977, exploring the heartland of intuitionistic analysis in mathematical depth and with conceptual finesse. A principal result from this book is Troelstra's Elimination Theorem, which says, essentially, that any statement quantifying over lawless sequences in formalized intuitionistic theories is provably equivalent to a statement that is only about lawlike sequences and objects. This is reminiscent of well-known elimination results in mathematics concerning, say, the reduction from numbers to sets, though the proof

in Troelstra's case is more complex. Later, he went on to prove similar theorems for other types of choice sequences. The Elimination Theorem clarifies the position of lawless and choice sequences in important formal systems, but it does not make the concept of a lawless sequence lose its value: lawless and other choice sequences remain crucial to guarantee a really intuitionistic analysis. Troelstra's result rather means that these concepts can be based on a theory that does not contradict standard mathematics. How the concept of choice sequence remains crucial is shown very clearly by Anne in an extensive article in 1983 in the *Journal of Philosophical Logic*. He shows how lawless sequences and other choice sequences arise by informal, yet rigorous concept analysis (often called 'informal rigor'), the main source of knowledge in intuitionistic mathematics in general and about choice sequences in particular.

**History of intuitionism and proof theory**

Other topics pursued in depth by Anne from the 1970s onward are the history of intuitionism and its philosophical basis. Over time, all this work, consolidated in his monographs and articles, fed into the magisterial *Constructivism in Mathematics*, in two volumes co-authored with Dirk van Dalen, published in 1988, which remains the standard text on constructivism right until today.

Moving beyond intuitionism proper over the years, Anne broadened his scope to proof theory in general and wrote two more books that again created new order in developing fields. In 1992, his textbook *Lectures on Linear Logic* proposed improved formats for the then still only partially understood new paradigm of linear logic merging proof theory and computation. Anne also contributed the majority of the chapters in a book with Helmut Schwichtenberg called *Basic Proof Theory* (1996) that is still a standard resource.

**Students**

An important part of Anne's life were his PhD students, of whom he supervised seventeen, and with many of whom he maintained a close relationship. His first PhD student Daniel Leivant finished a thesis in 1975 on the metamathematics of intuitionistic arithmetic, and later made his career in computer science. Initially a scarce commodity, in the 1980s, the number of PhD students increased, and Anne's students wrote on a broad variety of topics, such as intuitionistic metamathematics, combinatory algebra, category theory, Martin-Löf type theory, bounded arithmetic, linear logic, and provability logic. Many of these topics reflected the introduction by Anne, often in a close collaboration with Dirk van Dalen in Utrecht, of new topics on the Dutch scene. These students then carried the torch further by themselves. For instance, Ieke Moerdijk became an international leader at the interface of topos theory, logic and category theory generally, while Jaap van Oosten became a worldwide authority on realizability. In the Netherlands alone, four of Anne's students have become full professors, in mathematics, computer science, AI and philosophy. But Anne was a dedicated teacher at all academic levels, whose precision, clarity and scholarship influenced generations of students in Amsterdam.

**Natural history**

Anne retired in 2000, but not to rest. All his life, he had a deep interest in natural history and a wide knowledge of the plants of the Netherlands and abroad. His annual linocuts of plants discovered on his travels in Europe were famous. To those on a walk with Anne, listening to him turned what looked like an ordinary city street to the untrained eye into a rich landscape of flora, history, and natural wonders. This very year 2019, an article by him will appear on new species of blackberries, his special interest. Anne also made a further name for himself as an author on natural history travel narratives, chronicling the exotic characters and adventures of the past denied to the average academic of today. *Tijgers op de Ararat* from 2003 is a typical sample of this mix of information and human interest. Characteristically, he managed to create order even in this new field. His major *Bibliography of Natural History Travel Narratives* was published in 2016.

**A very special person**

Anne will be missed in the first place because he will no longer be there to tell us what he thinks about a question that you may have about constructivism. You always knew that you would get a completely honest answer from somebody who knew all the issues and had already thought much further than you. But it is just as much the personal qualities that will be missed. Anne was a very special, and to some, occasionally intimidating person: penetrating, honest, critical, ironic, sharp at times, but always open to arguments and unfailingly supportive of his students and colleagues. He will be deeply missed by all.

Anne Troelstra in 1988 with Boris Kushner from Moscow (later Pittsburgh)

Anne Troelstra in 1998 with Justus Diller from Münster

# In de verdediging

| In defence

*Pas gepromoveerden brengen hun werk onder de aandacht. Heeft u tips voor deze rubriek of bent u zelf pas gepromoveerd? Laat het weten aan onze redacteur.*

*Redacteur: Nicolaos Starreveld*
*FNWI, Universiteit van Amsterdam*
*Postbus 94214*
*1090 GE Amsterdam*
*verdediging@nieuwarchief.nl*


**Tweet Your #MO and Save a BRO – Micro-Mobilization Dynamics and Outcomes of Online Social Movement Campaigns**
*Anna Priante*

In March 2019 Anna Priante from the University of Twente successfully defended her PhD thesis with title *Tweet Your #MO and Save a BRO — Micro-Mobilization Dynamics and Outcomes of Online Social Movement Campaigns*. Anna carried out her research under the supervision of prof. dr. A. Need, dr. ir. D. Hiemstra, dr. M. L. Ehrenhard from the University of Twente and dr. ir. T.A. van den Broek from the Free University of Amsterdam.

**Online social movement campaigns**

Anna's dissertation starts from a social problem (cancer) that calls for social change and analyzes what actions people and organizations take (social movements, campaigns) to solve the problem and how technologies (social media) can be used as a possible way to find solutions to the problem.

Phrased in slightly more technical terms, Anna investigated whether micro-mobilization dynamics do explain the effectiveness of online social movement campaigns in achieving social change. Social movement organizations widely use social media to organize collective action for social change such as health awareness campaigns. Social change can be achieved by promoting online conversations of impact and by inspiring people to move from their armchair to the street. In order to understand how the transition from online action to meaningful (offline) action occurs Anna investigated various issues one of which is how individuals interact and communicate with each other before, during and after the campaign.

When people use social media during campaigns they create relations with others via communication processes, which produce communication networks. Such communication networks represent the flow of information during the time period before a campaign has started, during the campaign and afterwards. Social change is then related to a significant amount of individuals interacting, and probably to a high flow of information in the network during some time period.

Such communication networks can be mathematically represented using graphs, nodes represent individuals and edges between two nodes correspond to the exchange of information between individuals. For example, when people use Twitter in campaigns to promote health or in protests to achieve political change, they create relations with others by sending messages called tweets. But what kind of structures are observed in communication networks related to online movement campaigns? Generally the number of nodes is very large, which is due to the fact that online communication has grown immensely during the last decades. Owing to their large size,

online networks are generally very sparse and fragmented because of high levels of local clustering, where people engage in communication processes in and between small groups.

When mobilization is the result of a planned effort, as in the case with campaigns organized by social movement organizations, these networks tend to be highly centralized and have a 'star-shape', where the campaign organizer usually occupies the center of the network. In other cases, high centralization is due to the presence of a small group of highly connected and committed (individual or organizational) actors.

During her research Anna realized that communication networks related to online campaigns have one more additional and essential feature, they evolve over time, both structurally and socially. She studied the 2014 US Movember health campaign on Twitter organized by the Movember Foundation, a social movement organization that raises awareness of prostate and testicular cancer. In her dissertation Anna adopted a multidisciplinary, mixed-method approach combining theories and methods from sociology, social psychology, communication science, and computational social science.

In order to study the evolution of the Movember communication network the chronological data was decomposed into four subsequent periods:

T1: the pre-campaign, or launching, phase from 15 to 31 October;
T2: the campaign's first two weeks, from 1 to 15 November;
T3: the campaign's second two weeks, from 16 to 30 November;
T4: and the post-campaign phase, from 1 to 15 December.

In Figure 1 a representation of the communication network during these four periods is presented. The campaign's communication network has a characteristic three-layer structure comprising:

1. a network core surrounded by;
2. a constellation of smaller groups; and
3. a periphery of isolated nodes.

Inside this structure, movement members mostly converse with each other instead of remaining isolated. Although the three-layer structure is maintained over time, it turns out that its robustness fades. As the campaign unfolds, the communication network goes through a latency phase during which people decrease their active participation and move to the periphery of the communication or even leave the campaign network.

Analyzing the Twitter data set it appeared that the network was at its largest during the first two campaign weeks (T2), after which the number of nodes decreased by half (T3). Not surprisingly, the post-campaign phase (T4) is the smallest one and shows a decrease of participation after the formal end of the campaign (30 November). The Movember communication network was invariably very sparse over time: Network density was constantly very low, as is typical of large, online networks. Another observation is that a small group of central users is present with a much higher degree than all the other nodes. Among them, the official Twitter account of the Movember Foundation has the highest (in)degree value over time, as it is the most frequent target of movement members' mentions, replies, and retweets. This pattern not only shows the typical power-law distribution of large networks but also suggests the presence of a centralized network structure that is maintained over time.

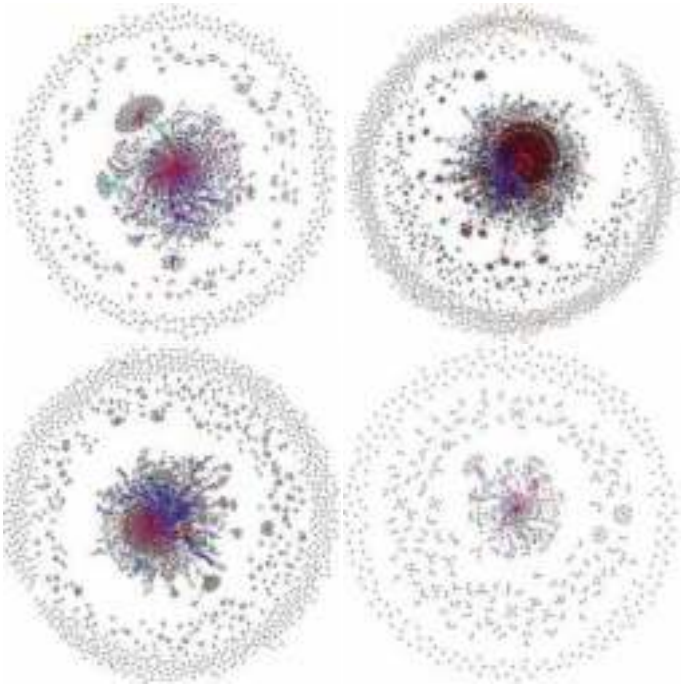


**Figure 1** Black disks are nodes, whereas edges are colored according to the type of communication strategy: red for mentions, blue for retweets, and green for replies. Self-loops derived from regular tweets are identified by a thin, orange line around the node. The network visualizations were realized using Fruchterman Reingold, Force Atlas 2, and Yifan Hu Proportional algorithms in Gephi. Taken from Anna's dissertation.

**The more personal aspect**

Behind all dissertations there always is a person, with flesh and bones, who has endured the long path of a PhD trajectory and has produced the work at hand.

*Anna, how did you experience this path of doing a PhD?* "I had a project-based PhD and only three years to finish it which made it challenging to pursue parallel projects during my PhD. I was mostly, full time into my research. Nonetheless, I was able to do some guest lectures at the University of Twente and in my alma matter university. It was nice to be back to the place where my university education began. I was also part of the board and one of the founders of the PhD Initiative Group (now BMS PhDs for PhD), a voluntary supportive and networking group for PhD students of the Faculty of Bahavioral, Management and Social Sciences, University of Twente. One of our core goals was to facilitate multidisciplinary research and this was an important aspect for me as it helped to forge my research vision based on multidisciplinary collaborations."

*And any stories from conferences you would like to share?* "I really liked the meeting at the Movember Foundation in Cilver City, Los Angeles. I could present my work to them twice during my PhD: they were very nice to me, always showing enthusiasm on my research, and I had good advice from them too!"

*Did you enjoy it working in the Netherlands?* "Coming from abroad, I have always felt welcomed by the people of both departments I have been working in. I am very grateful for having a unique, diverse and incredible supervision team, from which I learned a lot both personally and academically, while also having some fun together outside the office walls. I am very glad I completed the goal of my PhD and I am very grateful I was surrounded by many great people."

# Problemen

| Problem Section

This Problem Section is open to everyone; everybody is encouraged to send in solutions and propose problems. Group contributions are welcome. We will select the most elegant solutions for publication. For this, solutions should be received before **15 October 2019**. The solutions of the problems in this issue will appear in the next issue. Problem C, part b is a star exercise; we do not know a solution to this problem.

**Problem A** (proposed by Hendrik Lenstra)

Let $\tau : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be the map such that $\tau(n)$ is the number of positive divisors of $n$ for any $n \in \mathbb{Z}_{>0}$. Show that there are uncountably many maps $f : \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ such that $f \circ f = \tau$.
Note: This is a follow-up to Problem A of the March 2018 edition.

**Problem B** (proposed by Daan van Gent)

Let $X$ be a set and $* : X^2 \to X$ a binary operator satisfying the following properties:
1. $(\forall x \in X)\ x * x = x$;
2. $(\forall x,y,z \in X)\ (x * y) * z = (y * z) * x$.

Show that there exits an injective map $f : X \to 2^X$ that for all $x,y \in X$ satisfies $f(x * y) = f(x) \cap f(y)$.

**Problem C** (proposed by Onno Berrevoets)

a. Does there exist an infinite set $X \subset \mathbb{Z}_{>0}$ such that for all pairwise distinct $a,b,c \in X$ and all $n \in \mathbb{Z}_{>0}$ we have $\gcd(a^n + b^n, c) = 1$?

b* Does there exist an infinite set $X \subset \mathbb{Z}_{>0}$ such that for all pairwise distinct $a,b,c,d \in X$ and all $n \in \mathbb{Z}_{>0}$ we have $\gcd(a^n + b^n + c^n, d) = 1$?

**Edition 2019-2** We received solutions from Pieter de Groen, Marcel Roggeband, Rik Biel and Alex Heinis.

**Problem 2019-2/A** (proposed by Arthur Bik)

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in \mathrm{SO}(3)$$

be a matrix not equal to the identity matrix. Prove: if the vector

$$\begin{pmatrix} (a_{23} + a_{32})^{-1} \\ (a_{13} + a_{31})^{-1} \\ (a_{12} + a_{21})^{-1} \end{pmatrix}$$

exists, then $A$ is a rotation using this vector as axis.

**Solution** We received solutions from Pieter de Groen and Marcel Roggeband. The solution below is based on the solution by Marcel.
We write

$$v = \begin{pmatrix} (a_{23} + a_{32})^{-1} \\ (a_{13} + a_{31})^{-1} \\ (a_{12} + a_{21})^{-1} \end{pmatrix}.$$

*Redactie: Onno Berrevoets, Rob Eggermont en Daan van Gent*
*problems@nieuwarchief.nl*
*www.nieuwarchief.nl/problems*

Oplossingen | Solutions

Since $A \in \mathrm{SO}(3)$, we know that $A$ is a rotation matrix. Since $v$ exists, we also know $A$ is not the identity matrix. In particular, this means that $A$ has a 1-dimensional eigenspace associated with eigenvalue 1 spanned by the rotation axis. It therefore suffices to show $Av = v$. Since $A$ is orthogonal, the sum of squares of all entries in a row of $A$ equals 1, and likewise the sum of squares of all entries in a column of $A$ equals 1. Comparing the sum of squares of the first row and the first column of $A$ gives us

$$a_{11}^2 + a_{12}^2 + a_{13}^2 = 1 = a_{11}^2 + a_{21}^2 + a_{31}^2,$$

and rearranging terms gives

$$a_{21}^2 - a_{12}^2 = a_{13}^2 - a_{31}^2.$$

Dividing by $a_{13} + a_{31}$ and by $a_{21} + a_{12}$ gives us

$$\frac{a_{21} - a_{12}}{a_{13} + a_{31}} = \frac{a_{13} - a_{31}}{a_{21} + a_{12}}.$$

We can re-arrange this to

$$\frac{a_{21}}{a_{13} + a_{31}} + \frac{a_{31}}{a_{21} + a_{12}} = \frac{a_{12}}{a_{13} + a_{31}} + \frac{a_{13}}{a_{21} + a_{12}}.$$

It follows that the first entry of $Av$ and $A^T v$ are equal. Similar computations for the remaining rows and columns of $A$ give us $Av = A^T v$. Since $A \in \mathrm{SO}(3)$, we have $A^T = A^{-1}$, so we find $A^2 v = v$. Likewise, we have $A^2(Av) = A(A^2 v) = Av$, so both $v$ and $Av$ are eigenvectors of $A^2$ with eigenvalue 1. If $A^2$ is not the identity matrix, it follows that $v$ is the rotation axis of $A^2$. In this case, it is also the rotation axis of $A$, so we find $Av = v$. This leaves the case $A^2 = I$ (but $A \neq I$).

In the case $A^2 = I$, we find that $A$ is a rotation around some vector of $\pi$ radians. In particular, for any vector $u$ in the plane of rotation, we find $(A + I)u = 0$. Moreover, $A + I$ fixes the rotation axis of $A$, so $A + I$ is a matrix of rank one with the rotation axis of $A$ as its image. In particular, all columns of

$$\begin{pmatrix} a_{11}+1 & a_{12} & a_{13} \\ a_{21} & a_{22}+1 & a_{23} \\ a_{31} & a_{32} & a_{33}+1 \end{pmatrix}$$

are multiples of each other. Note that we also have $A^T = A$, so since $v$ exists, we find that no off-diagonal entries are zero. Finally, we observe that in the first column of $A + I$, the ratio between the second and third entry equals the ratio between second and third entry of $v$ (using $a_{12} = a_{21}$ and $a_{13} = a_{31}$), and in the second column of $A + I$, the ratio between the first and third entry equals the ratio between the first and third entry of $v$. Since there are no zeroes involved, we find that $v$ lies in the image of $A + I$, and therefore $A$ must be a rotation around $v$.

An alternative solution by Pieter de Groen uses the fact that $A$ can be described by means of Euler–Rodrigues parameters.

---

**Problem 2019-2/B** (proposed by Onno Berrevoets)

1. Let $k \in \mathbb{Z}_{>0}$ and let $\mathcal{X} \subset 2^{\mathbb{Z}}$ be a subset such that for all distinct $A, B \in \mathcal{X}$ we have $\#(A \cap B) \leq k$. Prove that $\mathcal{X}$ is countable.
2. Does there exist an uncountable set $\mathcal{X} \subset 2^{\mathbb{Z}}$ such that for all distinct $A, B \in \mathcal{X}$ we have $\#(A \cap B) < \infty$?

**Solution** We received solutions from Rik Biel and Alex Heinis. The solution below is based on the solution by Alex.

For the first part of the problem, we replace $\mathbb{Z}$ by $\mathbb{N}$ without loss of generality. Since $\mathbb{N}$ only has countably many finite subsets, it suffices to show that $\mathcal{X}$ only contains countably many infinite sets. Without loss of generality, we assume that $\mathcal{X}$ contains no finite sets. Let $S$ be the collection of subsets of $\mathbb{N}$ of cardinality $k + 1$. Now define $f : \mathcal{X} \to S$ by mapping $A \in \mathcal{X}$ to the set consisting of its smallest $k + 1$ elements. By the assumption that $A, B \in \mathcal{X}$ share at most $k$ elements, the map $f$ must be injective. Since $S$ is countable, $\mathcal{X}$ must also be countable.

Oplossingen

| Solutions

For the second part, the answer is yes. We replace $\mathbb{Z}$ by $\mathbb{N}^2$ without loss of generality. For $a>0$, let $p_n=\lfloor na\rfloor$ for $n\in\mathbb{Z}_{>0}$. This defines a sequence $\frac{p_n}{n}$ that converges to $a$ as $n\to\infty$. Note that for all $n$, we have $0\le a-\frac{p_n}{n}<\frac{1}{n}$. Define $S_a:=\{(p_1,1),(p_2,2),\ldots\}\in 2^{\mathbb{N}^2}$. It is quickly verified that if $b\neq a$, the sets $S_a$ and $S_b$ share only finitely many elements, since if $|b-a|\ge\frac{1}{N}$, we find $\lfloor na\rfloor\neq\lfloor nb\rfloor$ for all $n\ge N$. This implies that for all $a,b>0$ with $a\neq b$, the intersection $S_a\cap S_b$ is finite. In particular, the set $\mathcal{X}=\{S_a: a\in\mathbb{R}_{>0}\}$ is an uncountable set satisfying the desired properties.

---

**Problem 2019-2/C** (proposed by Onno Berrevoets)

Let $A:\mathbb{R}^2\to\mathbb{R}$ be a continuous function such that for every $x,y,z\in\mathbb{R}$ we have
1. $A(x,y)=A(y,x)$,
2. $x\le y\Rightarrow A(x,y)\in[x,y]$,
3. $A(A(x,y),z)=A(x,A(y,z))$,
4. $A$ is not the $\max$ and not the $\min$ function.

Prove that there exists an $a\in\mathbb{R}$ such that for all $x\in\mathbb{R}$ we have $A(x,a)=a$.

**Solution** For simplicity we write $x\oplus y:=A(x,y)$ for all $x,y\in\mathbb{R}$. The binary operator $\oplus$ is then symmetric and associative by properties 1 and 3. We consider three subsets $X_<, X_0, X_>$ of $\mathbb{R}$ defined by

$$\begin{aligned}X_<&:=\{x\in\mathbb{R}\mid\exists y\in\mathbb{R}\mid x\oplus y<x\},\\ X_0&:=\{x\in\mathbb{R}\mid\forall y\in\mathbb{R}\mid x\oplus y=x\},\\ X_>&:=\{x\in\mathbb{R}\mid\exists y\in\mathbb{R}\mid x\oplus y>x\}.\end{aligned}$$

Then it is clear that $X_<\cup X_0\cup X_>=\mathbb{R}$. Moreover, $X_<$ and $X_>$ are open subsets of $\mathbb{R}$ by continuity of $A$. Since $A$ is not the $\min$ function, it follows that there exist $x,y\in\mathbb{R}$ such that $x\le y$ and $x\oplus y>x$. Hence, $X_>\neq\varnothing$. It follows similarly from $A\neq\max$ that $X_<\neq\varnothing$. We will show that $X_>\cap X_<=\varnothing$, which yields the desired result $X_0\neq\varnothing$ because of the connectedness of $\mathbb{R}$.
Suppose that $X_<\cap X_>\neq\varnothing$. We will derive a contradiction. Let $x\in X_<\cap X_>$. Let $y,z\in\mathbb{R}$ be such that $x\oplus y<x$ and $x\oplus z>x$. Without loss of generality we have $x\oplus y\oplus z\ge x$. We also have

$$x\oplus y\oplus y=x\oplus(y\oplus y)=x\oplus y<x.$$

The map $\zeta\mapsto x\oplus y\oplus\zeta$ is continuous since $A$ is continuous, and by the intermediate value theorem we find that $x\oplus y\oplus w=x$ for some $w\in\mathbb{R}$. But now we arrive at a contradiction:

$$x>x\oplus y=(x\oplus y\oplus w)\oplus y=x\oplus(y\oplus y)\oplus w=x\oplus y\oplus w=x.$$

Therefore, $X_<\cap X_>=\varnothing$ and we conclude that $X_0\neq\varnothing$.