Bewijs met volledige inductie

M. Swaen
Hogeschool van Amserdam

Zodra je je in de wiskunde buiten het tuintje van het schoolprogramma begeeft, stuit je al gauw op bewijzen met volledige inductie. Misschien is vanuit die gedachte het onderwerp 'volledige inductie' in het N&T-profiel VWO opgedoken in de voorstellen van de vakontwikkelgroep. Het verraderlijke van volledige inductie in het wiskundeonderwijs is dat het achterliggende principe zo eenvoudig onder woorden gebracht kan worden:

Je wil iets bewijzen voor alle natuurlijke getallen, ga dan als volgt te werk:

1. *Bewijs eerst dat het voor 1 (of 0) geldt (dit heet de basisstap)*
2. *Bewijs dan dat ALS het voor een willekeurig getal n geldt, het ook voor diens opvolger n + 1 moet gelden (dit heet de inductiestap)*

en dat de motivatie ook niet veel omhaal vergt:

dan heb je het voor alle getallen bewezen, want het geldt voor 1, maar dan omdat het voor 1 geldt, ook voor diens opvolger 2, en voor diens opvolger 3, enzovoort.

Verder zijn er genoeg sommetjes te bedenken om het principe te oefenen, zoals bijvoorbeeld de zogenaamde formule van Gauss:

Stelling: $1 + 2 + 3 + \ldots + n = \dfrac{n \cdot (n+1)}{2}$

Bewijs:

1. basisstap:

$$1 = \frac{1 \cdot (1+1)}{2} \text{ klopt}$$

2. inductiestap:
Stel de bewering is waar voor n, dus stel

$$1 + 2 + 3 + \ldots + n = \frac{n \cdot (n+1)}{2}$$

We bekijken de bewering voor $n + 1$:

$$1 + 2 + \ldots + n + (n+1) = (1 + 2 + \ldots + n) + (n+1),$$

omdat $1 + 2 + 3 + \ldots + n = \dfrac{n \cdot (n+1)}{2}$ geldt dan

$$1 + 2 + \ldots + n + (n+1) = \frac{n \cdot (n+1)}{2} + (n+1) =$$

$$= \frac{n \cdot (n+1)}{2} + \frac{(n+1)2}{2} =$$

$$= (n+2)\frac{(n+1)}{2} \text{ klopt.}$$

Dit is ook ongeveer de manier waarop ik tijdens mijn studie heb kennis gemaakt met volledige inductie, en waarschijnlijk met mij de meeste wiskundeleraren. De indruk die ik vooral aan deze kennismaking overhield, was dat volledige inductie een recept is waarmee je duistere betrekkingen met machten, binomiaalcoëfficiënten en sommatietekens te lijf gaat.
Een dergelijke kennismaking behandelt eigenlijk niet meer dan het jargon van een inductiebewijs, en laat dat jargon oefenen aan de hand van opgaven die vaak ook zonder een expliciet beroep op volledige inductie op te lossen zijn, danwel zo gekunsteld zijn dat je niet beseft wat je aan het bewijzen bent. Op universiteit of hogeschool volgen er, als het goed is, op die inleiding vanzelf echte stellingen die alleen met inductie bewezen kunnen worden en krijgt de lege huls in de loop van de studie alsnog een inhoud. Vanwege de beperkte tijd kan deze weg binnen het voortgezet onderwijs niet gevolgd worden.
Willen we (sommige) VWO leerlingen iets laten zien van volledige inductie, dan zou dat niet het lege jargon moeten zijn, maar het betekenisvolle gebruik ervan. Dat kan alleen als je een wiskundige context weet op te bouwen waarbinnen vragen gesteld kunnen worden die inductieachtige redeneringen uitlokken.
Ik wil graag drie voorbeelden geven van dergelijke contexten, zoals wij die op de tweedegraads lerarenopleiding hanteren. Met deze voorbeelden wil ik tevens laten zien dat inductie heel verschillende aspecten heeft. Een vierde voorbeeld laat misbruik van inductie zien.

Hyperkubus

Inductie duikt vaak op als we kwantitatieve vragen stellen over objecten die volgens een recursief proces gevormd worden. Bijvoorbeeld: 'hoeveel stippen zitten er in de driehoekige figuraties van Pythagoras?', 'hoe groot is de zoveelste generatie van de konijntjes van Fibonacci?' of 'hoeveel zetten kost verplaatsing van een toren van Hanoi van zekere hoogte?'(Wells, 1987). Kenmerkend is dat het antwoord voor het volgende object kan gegeven worden in termen van het vorige, en zo krijgen we een recursieve betrekking die we vervolgens proberen om te werken naar een expliciete uitdrukking.

De hyperkubus is een vierdimensionale versie van de kubus. De kubus en de hyperkubus kun je zien als twee leden van een rij van figuren die als volgt ontstaat.
De ééndimensionale kubus K_1 bestaat uit één ribbe, die twee hoekpunten verbindt:

Als K_n de n-de kubus is, dan ontstaat K_{n+1} door naast K_n een kopie van K_n te leggen en de overeenkomstige hoekpunten met elkaar te verbinden:

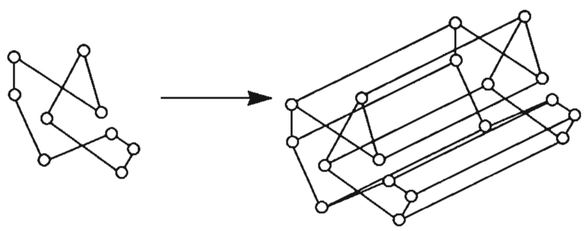


fig. 1

In deze definitie van de n-dimensionale kubus is alleen aandacht besteed aan de hoekpunten en ribben. De eerste leden van de rij zijn:

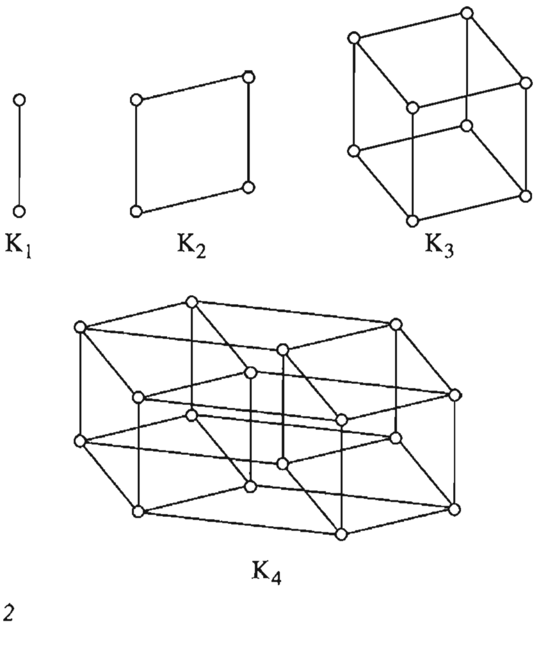


fig. 2

De gewone kubus komen we tegen als K_3 en de hyperkubus als K_4.

Hoeveel ribben en hoekpunten hebben de n-dimensionale kubussen? Noem h_n en r_n respectievelijk de aantallen hoekpunten en ribben van een n-dimensionale kubus. Wij maken een tabelletje van de eerste vier termen van deze rijen:

n	h_n	r_n
1	2	1
2	4	4
3	8	12
4	16	32

K_1 heeft twee hoekpunten, en het aantal hoekpunten verdubbelt iedere keer als we een dimensie hoger gaan.Dus $h_n = 2^n$.
K_1 heeft één ribbe, dus $r_1=1$.
Als we K_{n+1} maken uit K_n dan wordt eerst K_n gekopiëerd, op dat moment verdubbelt het aantal ribben. Daarna worden overeenkomstige hoekpunten door ribben verbonden. Dit levert dus per hoekpunt van K_n een nieuwe ribbe op. Oftewel er komen nog eens h_n dus 2^n ribben bij:

$$r_1=1 \qquad r_{n+1} = 2 \cdot r_n + 2^n.$$

Hiervoor zoeken we een expliciete uitdrukking. Omdat er sprake is van verdubbeling in iedere stap ligt het voor de hand de rij te vergelijken met de rij van machten van 2. Je ziet dan:

$$r_1 = 2^0, \quad r_2 = 2 \cdot 2^1, \quad r_3 = 3 \cdot 2^2 \text{ enzovoort,}$$

zo kom je tot de expliciete formule

$$r_n = n \cdot 2^{n-1}.$$

Reed-Muller codes

Als je wilt bewijzen dat bepaalde objecten bepaalde eigenschappen bezitten, dan is het vaak handig te kijken hoe je de betreffende objecten maakt. Als het goed is, zit de eigenschap bij het maken als het ware ingebakken. Dit aspect laat zich mooi illustreren bij de zogenaamde Reed-Muller codes.

Reed-Muller codes worden gebruikt bij het overzenden van berichten. De berichten worden eerst in nulletjes en eentjes vertaald, en die worden opgedikt tot rijtjes die we codewoorden noemen. Reed-Muller codes zijn hier zeer geschikt voor, zoals we zullen zien.
De codes zijn hier in plaats van met 0 en 1 gegeven in patronen van witte (0) en zwarte (1) hokjes.
We beginnen met één wit hokje: □ . Dit noemen we de eerste serie RM_1. Deze serie heeft maar één woord, na-

melijk 0.
Als je het blok RM_n hebt gemaakt, ontstaat de volgende serie RM_{n+1} door onder en rechts van het blok hetzelfde blok te herhalen en in de overgebleven rechteronderhoek zet je dan het fotonegatief van het blok RM_n (zie fig. 3).

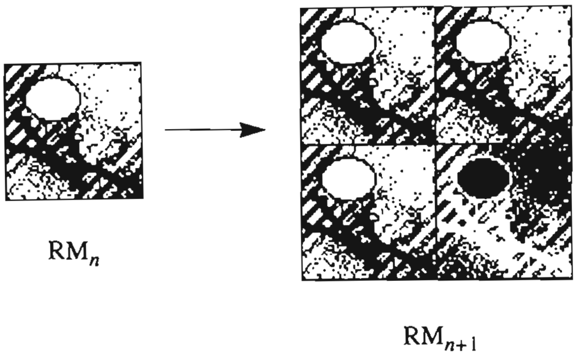


fig. 3

Hieronder (fig. 4) zijn de eerste vijf series Reed-Mullercodes afgebeeld. In het derde blok zie je de vier codewoorden van de derde serie: 0000, 0101, 0011 en 0110.
De Reed-Mullercodes hebben de prettige eigenschap dat ze onderling altijd op de helft van het aantal plaatsen verschillen. Zo verschillen de woorden 0000, 0101, 0011 en 0110 onderling steeds op twee van de vier plaatsen. Mochten er bij het overseinen fouten gemaakt worden, dan leidt dit bij langere woorden toch niet snel tot verwarring, omdat je wel heel veel fouten (een kwart of meer) moet maken voordat het woord op een ander woord gaat lijken.
Om te bewijzen dat de woorden steeds voor de helft verschillend zijn, zullen we het ontstaansproces van de series volgen. Bij de eerste series is snel te zien dat hier de woorden steeds voor de helft verschillen.
Laten we nu onderzoeken hoe deze eigenschap bij het vormen van een volgend blok behouden blijft. In figuur 5 is schematisch een blok RM_{n+1} weergegeven, opgebouwd uit drie kopieën van RM_n en rechtsonder het fotonegatief RM_n^{-1}.

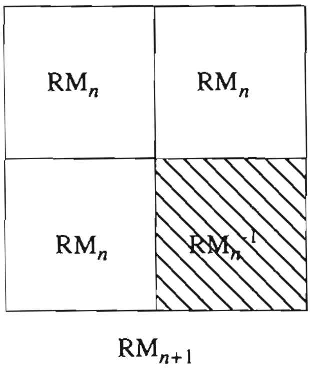


fig. 5

Een woord in het $n+1$-ste blok bestaat uit twee helften, zeg A en B, die ieder zelf regels zijn van de n-de blokken. Kies nu willekeurig twee woorden AB en CD uit het grote blok. Er zijn dan drie mogelijke gevallen:

1. beide woorden komen uit de bovenste helft van het blok
2. beide woorden komen uit de onderste helft
3. het ene woord komt uit de bovenste helft, het andere komt uit de onderste helft.

Geval 1

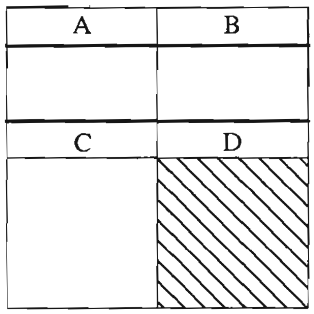


fig. 6

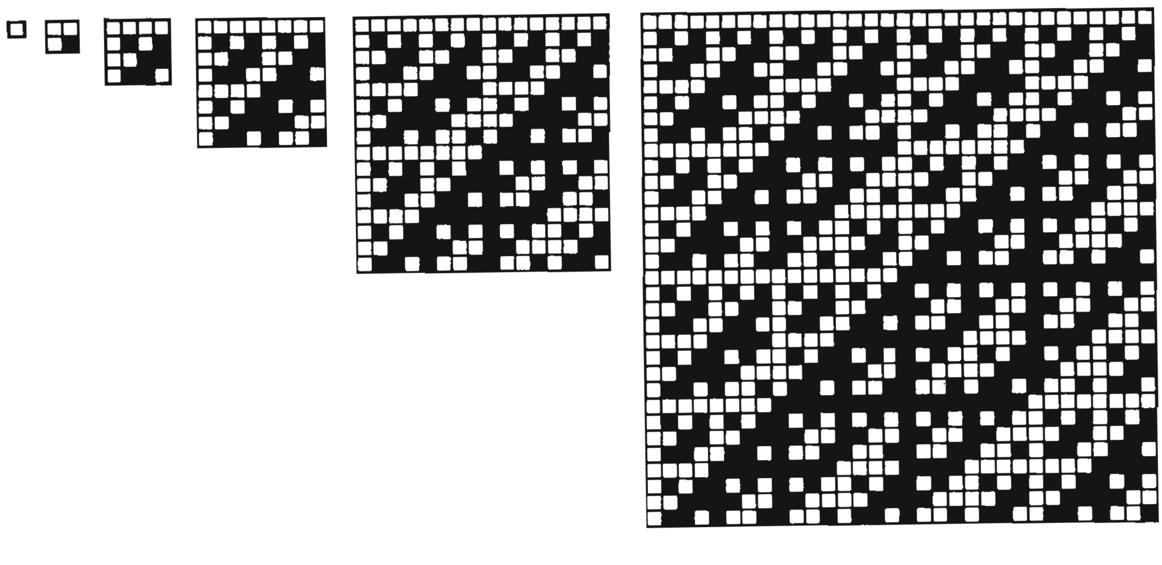


fig. 4

Omdat AB en CD in de bovenste helft van het blok zitten, zijn A en B hetzelfde, en evenzo C en D. Nu zitten A en C in RM_n dus zij verschillen voor de helft. Om dezelfde reden verschillen B en D ook voor de helft, dus zijn AB en CD onderling voor de helft verschillend.

Geval 2

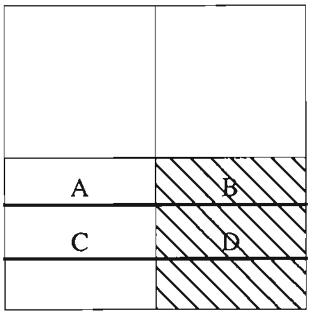


fig. 7

Zitten AB en CD in de onderste helft, dan is B het negatief van A en D is het negatief van C. A en C verschillen voor de helft, dus hun negatieven B en D ook. Dus verschillen AB en CD ook voor de helft.

Geval 3
Er zijn nu twee mogelijkheden:
a. A is hetzelfde als C
b. A is niet hetzelfde als C.

Geval 3a

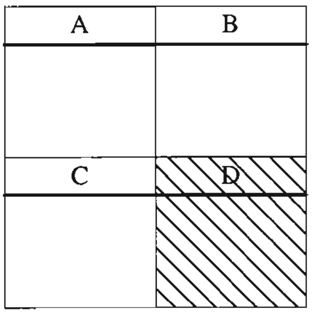


fig. 8

Als A en C hetzelfde zijn, dan is A blijkbaar de kopie van C, dus zit C op dezelfde hoogte als A maar dan in de onderste helft van het grote blok.
Dan is D het fotonegatief van B, en verschilt dus op alle plaatsen van B. Dan verschillen AB en CD dus op de eerste helft helemaal niet en op de tweede helft helemaal wel, dus zijn ze precies voor de helft verschillend.

Geval 3b

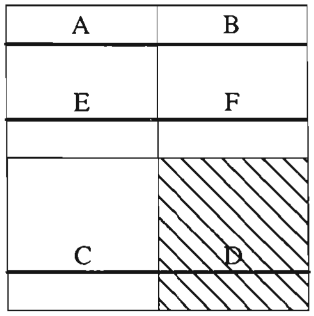


fig. 9

Als A en C niet hetzelfde zijn, zoek dan in de bovenste helft van het blok de regel die ook met C begint, zeg dit is EF. Nu zijn A en E regels in RM_n dus zij verschillen voor de helft. Hun kopieën B en F verschillen dan natuurlijk ook voor de helft. Omdat C een kopie van E is, verschillen dan ook A en C voor de helft. Omdat D het negatief van F is, verschilt D van F op alle plaatsen waar B en F juist hetzelfde zijn en andersom. Nu zijn B en F voor de helft verschillend, dus ook voor de helft gelijk. Daarom verschillen B en D voor de helft. Zo zien wij in dat AB en CD op de helft van het aantal plaatsen verschillen.

Tweekleurbaarheid

Bij de voorgaande inductiebewijzen was er sprake van opklimmen:
vanuit 0 omhoogwerkend volgens $n \rightarrow n + 1$.
In veel inductiebewijzen is de blikrichting echter andersom: om een geval G op te lossen blijkt het voldoende een eenvoudiger geval G' aan te pakken. Herhaalde toepassing van de vereenvoudiging leidt dan tot een basisgeval G_0 waarvan de oplossing duidelijk is. Een karakteristiek voorbeeld is het navolgende bewijs van de tweekleurbaarheid van 'kringels'.

Een 'kringel' is een gesloten kromme in het platte vlak die alleen viersprong-kruisingen bevat, oftewel zoiets als de tekening hieronder.

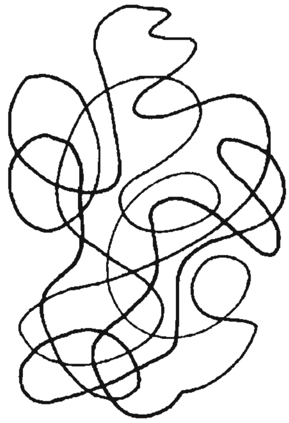
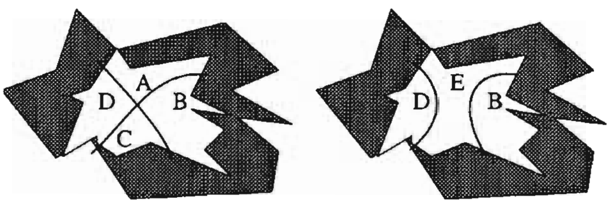


fig. 10

De landkaart die door de kringel gevormd wordt is 'tweekleurbaar', dat wil zeggen dat we ieder gebied zwart danwel wit kunnen maken op zo'n manier dat buurgebieden nooit dezelfde kleur krijgen. De kleuring is het blokpatroon dat bekend is van de tekenles danwel van krabbeltjes op agenda's, proefwerkblaadjes en telefoonboeken.

Waarom is de kringel tweekleurbaar? Om dat aan te tonen zullen we de tekening stap voor stap vereenvoudigen. Eerst zullen we alle kruispunten verwijderen, en wat dan nog overblijft zullen we daarna opruimen. Bij iedere vereenvoudiging zal blijken dat de tweekleurbaarheid niet verandert. Als de vereenvoudigde tekening tweekleurbaar is, dan is de oorspronkelijke het ook.

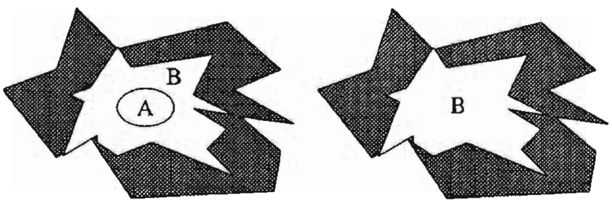
1. Opheffen kruisingen
Kies een kruispunt. Bij dit kruispunt komen vier gebiedsdelen samen, die we A, B, C en D zullen noemen (fig. 11a). Nu versmelten we A met C tot één gebied E, waarbij de kruising wordt opgeheven (fig. 11b). Stel nu dat de nieuwe tekening (rechts) tweekleurbaar is, dan kunnen we die kleuring ook gebruiken voor de oude situatie, waarbij A en C allebei de kleur van E overnemen.



a *b*
fig. 11 a en b

Dit kan alleen maar mis gaan waar er links extra buurrelaties zouden zijn. De gebieden A en C hebben echter alleen maar buren die E ook al had, en ten opzichte van die buren is de kleur van E goed, dus die van A en C ook. Deze procedure herhalen we tot er geen kruisingen meer over zijn. De tekening bestaat dan nog uit een aantal kringen (enkelvoudig gesloten krommen) die elkaar niet snijden.

2. Kringen verwijderen
Neem een kring waar geen kringen meer binnen zitten. Noem het gebied binnen de kring A en het gebied daar direct omheen B (fig. 12a). Verwijder nu de kring (fig. 12b). Stel nu dat de nieuwe tekening (rechts) tweekleurbaar is. Dan kunnen we die kleuring voor de oude tekening overnemen, waarbij je A gewoon de tegenovergestelde kleur van B geeft.
Het enige verschil in buurrelaties is dat A erbij komt met B als enige buur. Maar we hebben A al de goede kleur gegeven ten opzichte van B.



a *b*
fig. 12 a en b

Uiteindelijk, na opheffing van alle kruisingen en verwijdering van alle kringen, is er niets meer overgebleven behalve een leeg vel. En we weten dat als dit lege vel tweekleurbaar is, het voorgaand stadium dat ook is en dus uiteindelijk de oorspronkelijke kringel ook.
Is het lege vel tweekleurbaar? Ja, er is maar één gebied en dit is wit.
Hieronder (fig. 13) is de redenering schematisch weergegeven. Voor de kringel linksboven is de stapsgewijze

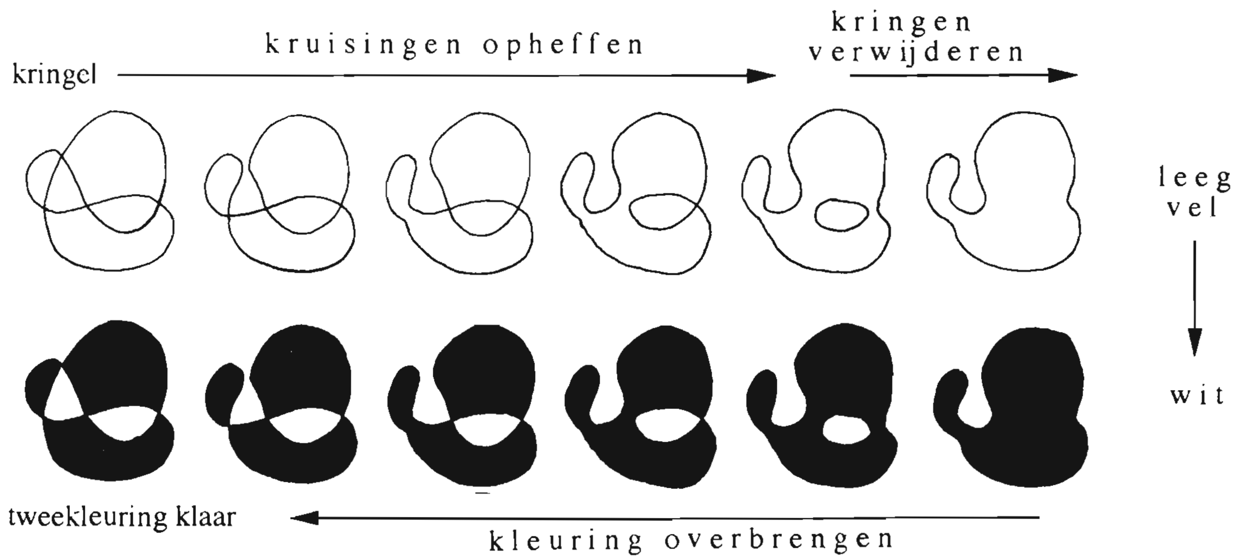


fig. 13

vereenvoudiging getekend, daaronder zien we hoe de kleuring van het lege vel de kleuring van de oorspronkelijke kringel genereert.

Gescheiden onderwijs

Met volledige inductie zijn ook drogredeneringen te voeren. Dergelijke onbewijzen verscherpen het begrip van waar volledige inductie op berust. De Grieken hadden al de sorites-paradox:
Gegeven een hoop zand. Verwijderen we één korrel, dan houden we nog steeds een hoop zand over. Herhalen wij dit (volledige inductie) dan is één korrel zand ook een hoop zand.

Hieronder volgt een 'bewijs' voor de stelling dat er in een klas altijd alleen maar jongens of alleen maar meisjes zitten. Gemengde klassen kunnen dus niet voorkomen.

Bewijs:
(met inductie naar n, het aantal leerlingen in de klas)
Stel in de klas zit maar één leerling, dan is die leerling een meisje óf een jongen, dus die klas is niet gemengd. De bewering klopt voor $n = 1$.
Wij nemen nu aan dat de bewering geldt voor klassen met maar hoogstens n leerlingen. Stel een klas heeft $n + 1$ leerlingen, verdeel de klas dan in twee groepen, maar zorg dat één leerling in beide groepen zit. Die leerling noemen we X. Voor het gemak nemen we aan dat X een meisje is.

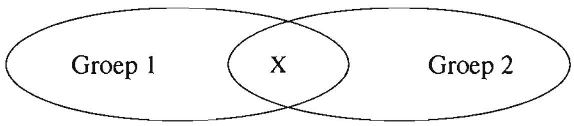


fig. 13

Vorm nu met groep 1 een nieuwe klas. Omdat er leerlingen weg zijn, is dit een klas met maar hoogstens n leerlingen, dus is dit een groep van alleen maar meisjes óf alleen maar jongens (inductie aanname). Omdat meisje X erbij zit, bestaat groep 1 dus uit alleen meisjes.

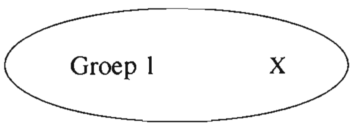


fig. 14

Maak vervolgens een klas van de leerlingen van de tweede groep. Het meisje X zit ook weer in deze klas, deze klas heeft weer maar hoogstens n leerlingen, dus ook groep 2 bestaat uit alleen maar meisjes.

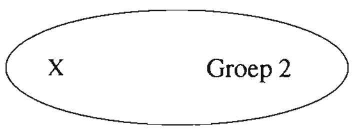


fig. 15

Kortom de hele klas bestond alleen maar uit meisjes. Zou X een jongen zijn geweest, dan zouden er alleen maar jongens in de klas hebben gezeten. QED

Literatuur

Dewdney, A.K. (1989). *The Turing Omnibus: 61 excursions in computer science.* Rockville, MD, Computer Science Press. ISBN 0 7167 8154 9

Wells, David (1987). *Woordenboek van eigenaardige en merkwaardige getallen.* Vertaling Lex Bijlsma. Bert Bakker. ISBN 90 351 0527 3