Killing Cluedo
Nationale Wiskundedagen 2002

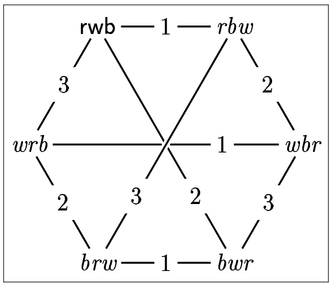
Hans van Ditmarsch *Computer Science, University of Otago, Nieuw Zeeland*

In kennislogica kan beschreven worden wat de informatie is van actoren over een gegeven situatie en over elkaar, in *dynamische* kennislogica kan beschreven worden hoe die informatie verandert ten gevolge van acties. Een voorbeeld is de informatie van spelers in een spelstiuatie, en de veranderingen van informatie als gevolg van zetten in het spel. Een concreet voorbeeld daarvan is het spel Cluedo. Bij dit spel moet je erachter komen welke kaarten omgekeerd op tafel liggen door vragen te stellen over de kaarten van je medespelers. We gaan nu achtereenvolgens in op wat kennislogica is en kan, hoe zetten in spelen in kennislogica kunnen worden beschreven, en tot slot hoe dit voor een analyse van Cluedo gebruikt kan worden.

1 Kennislogica

Cluedo bevat 21 kaarten en is alleen al daarom nogal ingewikkeld te analyseren. Eenvoudiger is de volgende spelsituatie: Er zijn drie spelers; ze heten 1, 2 en 3. Er zijn drie kaarten: rood, wit en blauw (r, w en b); de kleuren van de Nederlandse vlag dus. Iedere speler houdt n kaart vast. Spelers kunnen alleen hun eigen kaarten zien. We nemen aan dat speler 1 rood heeft, speler 2 wit heeft en speler 3 blauw heeft. Ze weten dat iedereen een kaart heeft en welke kaarten er in het spel zijn, maar ze kunnen alleen hun eigen kaart zien. Om te beginnen hebben we een formele taal nodig om dit spel te beschrijven. Dit is de taal van de *kennislogica*. We beperken ons tot de essenti le zaken: r_1 staat voor 'speler 1 heeft rood',$K_1\, r_1$ staat voor speler 1 weet dat hij rood heeft. De operator K_1 heet de modale kennisoperator en staat voor 'weten dat'. Nog een ander voorbeeld: $K_1\, r_1 \wedge \neg K_2\, r_1$ staat voor 'speler 1 weet dat hij rood heeft en speler 2 weet niet dat speler 1 rood heeft'.

Los van een logische taal om dit spel te beschrijven willen we ook modellen om deze taal in te interpreteren. Het model *Hexa* in figuur 1 geeft de kennis van de spelers weer in de beginsituatie van het spel, voor de kaartverdeling *rwb*. Hierbij representeert *rwb* de kaartverdeling waarbij speler 1 rood heeft, speler 2 wit, en speler 3 blauw. In de figuur staat deze kaartverdeling in een ander lettertype om aan te geven dat het de echte kaartverdeling is. Deze situatie kan speler 1 niet onderscheiden van de situatie *rbw* waarbij speler 1 nog steeds rood heeft, maar waarbij speler 2 blauw heeft en speler 3 wit. Dit wordt in het plaatje weergegeven door een verbinding van *rwb* naar *rbw* gelabeld met een 1. Als we ook voor de andere spelers weergeven welke kaartverdelingen ze niet van elkaar kunnen onderscheiden, is de figuur compleet. Zo'n *Kripke-model* (genoemd naar de 'uitvinder' Saul Kripke) bestaat dus uit een verzameling objecten, genaamd *werelden*, voor iedere speler een binaire equivalentierelatie tussen die objecten, genaamd *toegankelijkheidsrelatie*, en bovendien per wereld een *waardering* van atomaire beweringen y_x voor 'speler x houdt kaart y vast'.



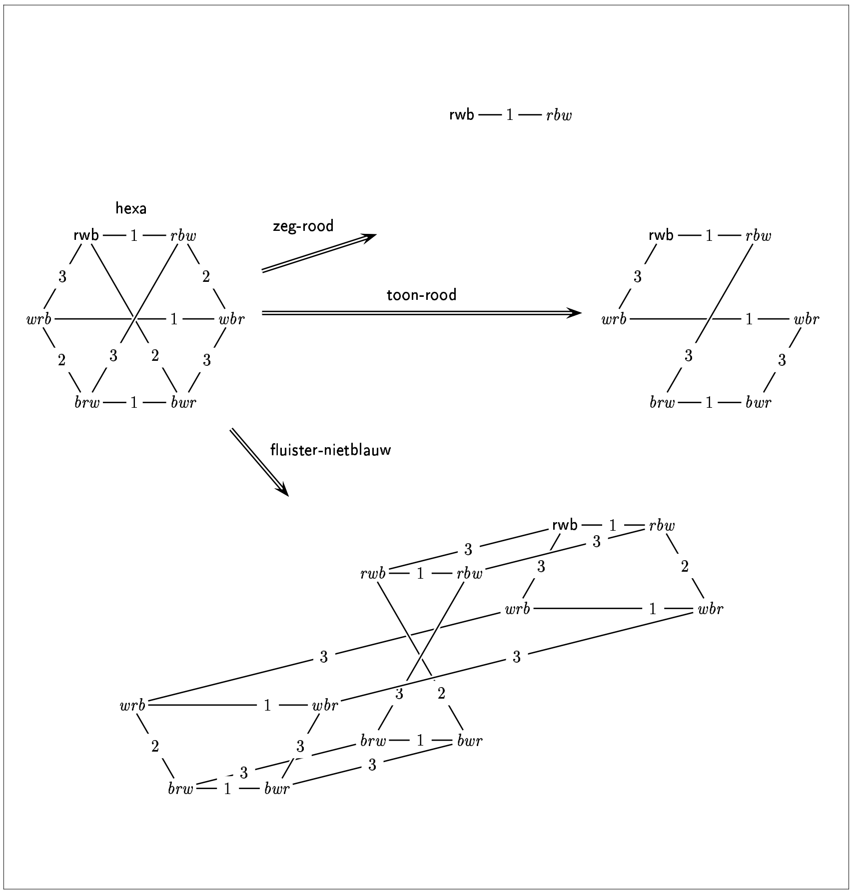
Figuur 1 *Hexa*

We kunnen beweringen in onze logische taal nu interpreteren in dit model *Hexa*. We gaan niet in op de details maar geven alleen wat voorbeelden. In *rwb* weet speler 1 bijvoorbeeld dat speler 2 van zichzelf weet dat hij rood niet heeft, in logica $K_1K_2\neg r_2$. Dit is waar omdat in alle werelden die toegankelijk zijn voor speler 1 (*rwb* en *rbw*) in de werelden die toegankelijk zijn voor speler 2 ((*rwb* en *bwr*) en (*rbw* en *wbr*)) het niet zo is dat speler 2 rood heeft. In het plaatje hierboven werd ook aangegeven wat de echte wereld is. ZoÕn model noemen we een *gepunt* model. De echte wereld is de *punt* van het model.

2 Kennisverandering in kaartspelen

Het gepunte model (*Hexa*, *rwb*) is het model voor de beginsituatie waarin 1 rood heeft, 2 wit en 3 blauw. In deze beginsituatie weet 2 wel dat hij wit vasthoudt, omdat hij deze kaart kan zien, maar weet 2 niet welke kaart 1 heeft. Speler 2 houdt zowel voor mogelijk dat 1 rood heeft als dat 1 blauw heeft. In figuur 1 zagen we dat de toegankelijkheidsrelatie voor speler 2 de zes kaartverdelingen in drie equivalentieklassen verdeelt. De klasse die de werkelijke kaartverdeling *rwb* bevat, bevat eveneens de kaartverdeling *bwr*. In het eerste geval heeft 1 rood, in het tweede geval heeft 1 blauw. Beide zijn dus voor 2 voorstelbaar. Er zijn verschillende manieren waarop speler 2 door een vraag te stellen kan achterhalen welke kaart speler 1 heeft . De vraag moet eerlijk beantwoord worden. Drie van deze manieren zijn:

- Speler 2 vraagt speler 1 zijn kaart. Speler 1 legt als antwoord rood op tafel. Gegeven dat de spelers niet liegen is dit dezelfde actie als die waarin speler 1 *zegt* dat hij rood heeft. Deze actie noemen we (mede daarom) *zeg-rood*.
- Speler 2 vraagt speler 1 zijn kaart. Speler 1 laat alleen speler 2 rood zien, dat wil zeggen: speler 3 ziet niet welke kaart het is, maar ziet wel dat er een kaart getoond wordt. Dit is gemeenschappelijke kennis voor alle spelers. Deze actie noemen we *toon-rood*.
- Speler 2 vraagt speler 1 hem een kaart in het oor te fluisteren die hij niet heeft. Speler 1 fluistert in het oor van speler 2: ÒIk heb blauw nietÓ Deze actie noemen we *fluister-nietblauw*.



Figuur 2 Het resultaat van *zeg-rood*, *toon-rood* en *fluister-nietblauw* in een spelsituatie (*Hexa*, *rwb*) waarin 1 rood, 2 wit en 3 blauw vasthoudt. De toegankelijkheidsrelaties zijn reflexief en transitief.

In figuur 2 zien we het resultaat van deze drie acties in de gegeven speltoestand. Merk op dat in alle drie de gevallen het gewenste resultaat bereikt is dat 2 weet wat de kaart van speler 1 is: vanuit *rwb* zijn er voor speler 2 in geen van de drie modellen alternatieven. De kennis van speler 3 verschilt echter per model: na *zeg-rood* weet behalve 2 ook 3 dat 1 rood heeft. Na *toon-rood* weet 3 niet dat 1 rood heeft maar weet 3 wel dat 2 weet welke kaart 1 heeft. Na *fluister-nietblauw* weet 3 ook niet dat 1 rood heeft maar weet 3 evenmin of 2 weet welke kaart 1 heeft. Merk verder op dat de vraag van speler 3 die tot het antwoord *fluister-nietblauw* leidde niet van risico ontbloot was: in plaats van te antwoorden dat hij blauw niet heeft, had speler 1 ook kunnen antwoorden dat hij wit niet heeft. Dit wist 2 echter al, omdat 2 zelf wit heeft!

Dit soort acties kunnen we weer op verschillende manieren modelleren, en zelfs weer als een soort van Kripke-model. We geven geen details. Net zoals we een logische taal voor beweringen over kennis kunnen interpreteren in modellen, kunnen we ook een logische taal voor dit soort ÔkennisactiesÕ (knowledge actions) defini ren. Dit gebeurt in mijn proefschrift. De actie *toon-rood* waarin speler 1 zijn rode kaart aan speler 2 laat zien, zonder dat 3 ziet welke kaart het is, is in die taal te beschrijven als $L_{123}\ (!L_{12}\ ?r_1 \cup L_{12}\ ?w_1 \cup L_{12}\ ?b_1)$. We lezen dit als volgt:

> 1, 2 en 3 leren dat een van de volgende drie alternatieven is uitgevoerd: (1 en 2 leren dat de test op r_1 slaagt) of (1 en 2 leren dat de test op w_1 slaagt) of (1 en 2 leren dat de test op b_1 slaagt); en het werkelijk uitgevoerde alternatief is (1 en 2 leren dat de test op r_1 slaagt).

De operator L staat voor ÔlerenÕ en de nummers in subscript bij deze operator staan voor de groep die leert; de operator $\cup$ staat voor nondeterministische keuze, het vraagteken dat aan een propositie voorafgaat, geeft aan dat het om een test op die bewering gaat; het uitroepteken voor het eerste alternatief geeft aan dat dit werkelijk uitgevoerd is. Het uitroepteken had dus ook voor een van de andere alternatieven kunnen staan. Aan een kennislogische taal kunnen we *dynamische* operatoren toevoegen voor acties. In deze taal kunnen we als volgt uitdrukken dat speler 2 na *toon-rood* weet wat de kaart van speler 1 is: in (*Hexa*, *rwb*) geldt $[\textit{toon-rood}]K_2 r_1$.

3 Het spelen van een kennisspel

We hebben nu modellen en acties, maar nog geen spel. Een actie zoals *toon-rood* kunnen we zien als een zet in een spel van vragen en antwoorden. Het doel van het spel is om de kaartverdeling als eerste te kennen. Door verschillende soorten vragen toe te staan krijgen we verschillende varianten van het spel. De regels bepalen naar hoeveel kaarten een speler kan vragen. Een speler kan aan een andere speler een specifieke kaart vragen (heb jij rood), of een van twee kaarten (heb jij rood of blauw), of een van drie kaarten (heb jij rood, wit, of blauw). We staan maar twee soorten antwoorden toe op deze vragen: ÒneeÓ of het tonen van een kaart. De vraag naar een van drie kaarten komt dus neer op de vraag Ôtoon me je kaartÕ. Nadat de vraag beantwoord is, mag een speler (maar n keer tijdens het spel) zeggen dat hij weet wat de kaartverdeling is. De eerste speler die dat (terecht) doet, wint het spel. Hiermee hebben we een kennisspel gedefinieerd. Voor 3 spelers en 3 kaarten, en vragen naar n van drie kaarten zijn de strategie n beperkt en is het zelfs onmogelijk voor de eerste speler om te verliezen. We spelen de drie varianten van het spelletje met de kaartverdeling *rwb*, waarbij speler 2 begint:

- De vraag is naar een van drie kaarten. Speler 2 vraagt speler 1 naar zijn kaart. Speler 1 toont (alleen) speler 2 zijn rode kaart. Speler 2 zegt dat hij de kaartverdeling kent. Speler 2 wint het spel. Speler 2 kan het spel niet verliezen!
- De vraag is naar een van twee kaarten. Speler 2 vraagt speler 1 of hij rood of blauw heeft. Speler 1 toont (alleen) speler 2 zijn rode kaart. Speler 2 zegt dat hij de kaartverdeling kent. Speler 2 wint het spel. Speler 2 kan het spel niet verliezen!
- De vraag is naar n kaart. Nu kan speler 2 het spel wel verliezen, door een suboptimale strategie te kiezen. Speler 2 vraagt speler 1 of hij wit heeft. Speler 1 zegt ÔneeÕ, want speler 2 heeft zelf wit. Speler 2 be indigt zijn beurt. (Dat wil zeggen: maakt impliciet publiek dat hij nog niet kan winnen.) Speler 3 is nu aan zet. Speler 3 zegt dat hij de kaartverdeling kent. Speler 3 wint het spel. Speler 2 had het spel ook

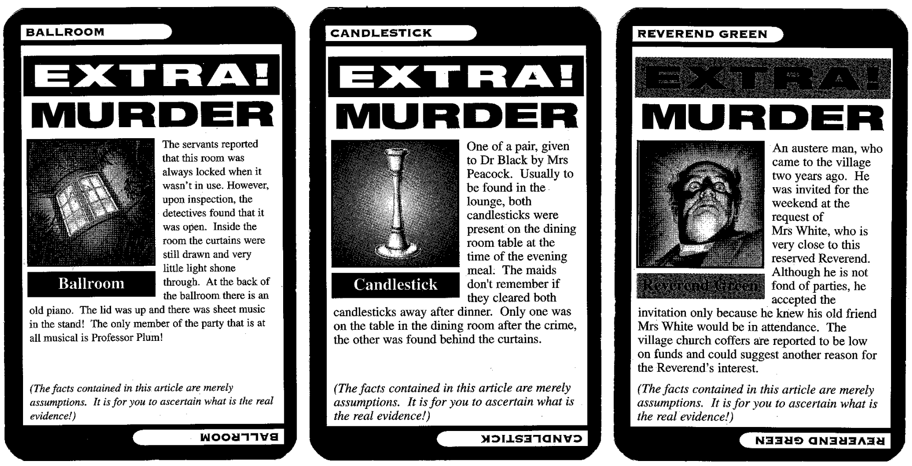
kunnen winnen, namelijk door een andere strategie te kiezen. Zowel de strategie Ôgeb je witÕ als Ôgeb je blauwÕ leidt tot winst.

Merk op dat ook de acties waarin een speler een vraag met ÒneeÓ beantwoordt, of waarin een speler wint, of waarin een speler zijn zet be indigt (dat wil zeggen Ôniet wintÕ), geanalyseerd kunnen worden met de methoden uit de vorige sectie. In alledrie de gevallen gaat het om het publiek maken van een test. In het eerste geval is dit een test op een bewering over kaartbezit, bijvoorbeeld $\neg w_1$ in het geval dat speler 1 ÒneeÓ antwoordt op de vraag van 2 of hij wit bezit. De actie die dit beschrijft is dus $L_{123}\ ?\neg w_1$. In het tweede en in het derde geval is het een test op een bewering over de kennis van spelers. Als in het laatste spelvoorbeeld speler 2 zijn zet be indigt maakt hij publiek dat de test slaagt op de bewering dat hij niet kan winnen, in termen van de actietaal: $L_{123}\ ?\neg\ (K_2\ \delta_{rwb} \vee K_2\ \delta_{rbw} \vee ...)$. Hierin staat bijvoorbeeld δ_{rwb} voor $r_1 \wedge w_2 \wedge b_3 \wedge \neg r_2 \wedge \neg r_3 \wedge \neg w_1 \wedge \neg w_3 \wedge \neg b_1 \wedge \neg b_2$.

4 Cluedo

Als er maar 3 spelers en 3 kaarten zijn, is het een flauw spel. De speler die begint wint altijd. Het wordt interessanter als er meer spelers of meer kaarten zijn. Een echt kennisspel is het spel Cluedo.

In het spel Cluedo is een moord gepleegd. Het doel van het spel is om deze moord op te lossen. Oplossen betekent: ontdekken wie de moord gepleegd heeft, met welk wapen en in welke kamer van een huis met negen kamers. Het wordt gespeeld met een bord waarop dit huis is afgebeeld. Het bord bepaalt wie een vraag mag stellen en over welke kamer. Belangrijk voor onze analyse is, dat er zes spelers zijn, zes verdachten, zes mogelijk moordwapens en negen kamers. Deze worden vertegenwoordigd door speelkaarten. Van ieder type kaart wordt er aan het begin van het spel, na schudden, n apart gelegd, ondersteboven op tafel. Dit zijn de kaarten die staan voor de moordenaar, het moordwapen en de moordkamer. De overige 18 van de in totaal 21 kaarten worden opnieuw geschud en onder de zes spelers verdeeld. De kaarten die je hebt vertellen je dus iets over de moord: die kunnen het niet geweest zijn! Dit is de informatie waarmee je begint. Een zet in het spel bestaat uit een vraag over drie kaarten van de verschillende typen, net alsof hiermee een verdenking geuit wordt, bijvoorbeeld ÔIk denk dat Miss Scarlett het heeft gedaan met het mes in de keukenÕ. De speler aan wie de vraag wordt gesteld kan deze beantwoorden met ofwel ÔneeÕ ofwel met het tonen van een van deze drie gevraagde kaarten. Dit is dus hetzelfde soort actie als *toon-rood*. Zolang het antwoord ÔneeÕ is stel je de vraag aan de volgende speler. Je beurt is voorbij als een speler je een kaart laat zien. Door de antwoorden van de spelers kun je erachter komen welke kaarten op tafel liggen. Tijdens je beurt mag je ook nog een Ôdefinitieve beschuldigingÕ uiten, bijvoorbeeld omdat je weet hoe de moord gepleegd is. Het spel bestaat dus uit een afwisseling van verschillende acties waarvan we de gevolgen precies kunnen modelleren. Er zijn dus precies vier typen acties: een kaart wordt getoond, kaartbezit wordt ontkend, de zet wordt afgerond, een speler wint.



Hiermee is ieder spelverloop van Cluedo uitputtend te beschrijven. Speltheoretici zijn echter voornamelijk ge nteresseerd in winststrategie n. Het is echter nog onduidelijk wat voor kennisspelen en voor Cluedo in het bijzonder goede strategie n zijn. Wat wel duidelijk is, is dat de logische modellering van spelsituaties en zetten onmisbaar is voordat je over strategie n kunt gaan nadenken: alleen op grond van de voorliggende precieze modellering van zetten kunnen we berekenen wat de individuele preferenties van spelers zijn op de verzameling mogelijke vragen op een gegeven spelmoment. Is het bij Cluedo beter om naar drie kaarten te vragen die je niet hebt en nog niet kent, of naar drie kaarten waarvan je er n of twee zelf hebt? En dan hebben we het nog niet eens over optimale respons. Om een voorbeeld te geven: op ieder moment van het spel mag een definitieve beschuldiging door een speler geuit worden, niet alleen als hij het echt weet. Als je vermoedt dat de volgende speler over voldoende informatie beschikt, of gaat beschikken, om het spel te winnen, doe je er natuurlijk beter aan de moordenaar alvast te gokken en dus te riskeren dat je verliest. Er is dus nog genoeg werk te doen.

Referenties

- Johan van Benthem et al. *Logic in Action*. ILLC, Amsterdam 2002. (verschijnt februari)
- Hans van Ditmarsch. *Knowledge Games*, ILLC Dissertation Series DS-2000-06, Amsterdam, 2000. (Online: `www.illc.uva.nl/Publications/`)
- Hans van Ditmarsch. *Killing Cluedo*, Natuur & Techniek 69(11) 32-40, 2001.
- Ronald Fagin, Joseph Y. Halpern, Yoram Moses, Moshe Y. Vardi. *Reasoning about Knowledge*, MIT Press, Cambridge, Massachusetts, 1995.
- Martin J. Osborne en Ariel Rubinstein. *A Course in Game Theory*, MIT Press, Cambridge, Massachusetts, 1994.
- Online *Hexa* updaten: `www.science.uva.nl/projects/opencollege/cognitie/hexagon/`