Speltheorie en Logica

Hans P. van Ditmarsch en Barteld P. Kooi

1 Inleiding

Logica en besluitvormingstheorie komen elkaar tegen als de toestand van de wereld onduidelijk is. Hiervan zullen we in dit artikel twee voorbeelden geven. Het eerste is een voorbeeld van de invloed die besluitvormingstheorie op logica heeft. Hier zijn veel voorbeelden van. Zo zijn er bijvoorbeeld de Ehrenfeucht-Fra ss spelen, waarmee je kunt beoordelen of twee structuren isomorf zijn. Omdat we graag onzekerheid behandelen nemen we de speltheoretische semantiek van Hintikka en Sandu als voorbeeld. Deze komt aan bod in paragraaf 2. We kunnen waarheid opvatten als de optimale uitkomst van een spel tussen twee spelers. De invloed van besluitvormingstheorie op logica komt tot uitdrukking bij een uitbreiding van de predikatenlogica: de zogenaamde ÔIndependence-FriendlyÕlogica, waarbij variabelen ingewikkelder vormen van afhankelijkheid kunnen hebben dan in de standaard predikatenlogica. Deze logica heeft een rijkere taal dan de klassieke logica, en kan worden begrepen door juist de toevoeging van onzekerheid aan de speltheoretische semantiek.

Het tweede voorbeeld laat de invloed van logica op besluitvormingstheorie heeft zien. Het gebruik van logica in speltheorie heeft geresulteerd in de analyse van speltheoretische begrippen zoals speltoestanden, zetten in spelen, spelbomen en equilibria. In ons voorbeeld gaat het om onzekerheid waarbij duidelijkheid moet worden verkregen. Onzekerheid is te beschrijven met behulp van kennislogica, die in paragraaf 3.1 aan bod komt. Om duidelijkheid te verkrijgen dienen we acties te ondernemen. Als er meer acties mogelijk zijn, moeten we daartussen kiezen. Hier speelt de besliskunde een rol. Als je geen rekening hoeft te houden met anderen, kies je gewoon de beste actie (volgens een of ander criterium). Als er echter meerdere actoren acties kunnen ondernemen, is het van belang om niet zomaar een actie te kiezen die voor jou ÔindividueelÕhet beste is, maar een actie die een optimale respons is op wat de andere actoren van jouw gedrag waarnemen. We betreden nu het gebied van de speltheorie. Met speltheorie proberen we te berekenen welke beslissingen optimaal zijn onder bepaalde procedurele omstandigheden. In dynamische kennislogica kan nauwkeurig beschreven worden hoe de informatie van spelers in een spel verandert tengevolge van acties. Een goed voorbeeld van een toepassing hiervan is de analyse van 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. De kennisveranderingen die het gevolg zijn van het laten zien van een kaart aan (alleen) een andere speler, zijn nogal ingewikkeld. In paragraaf 3.2 komt dit voorbeeld uitgebreid aan bod wanneer we nieuwe technieken beschrijven om kennisveranderingen met logica te modelleren. Daarbij zien we dat voor de speltheoretische analyse van dergelijke ingewikkelde vormen van onzekerheid logica onontbeerlijk is.

2 Speltheoretische Semantiek

In het computerprogramma *TarskiÕs World* is predikaatlogische semantiek te zien als een spel. Het programma is een belangrijk onderdeel van het logica-onderwijs dat bij het boek *The Language of First-order Logic* van Jon Barwise en John Etchemendy hoort. Studenten leren met het programma predikatenlogica. Er is een blokkenwereld waarin verschillende objecten kunnen staan met verschillende groottes. Je kunt het programma laten bepalen of een predikaatlogische zin die iets over een blokkenwereld zegt waar of onwaar is. Hoe complexer de zinnen zijn, hoe lastiger het is om te zien waarom een zin waar is of niet. Om hierbij te helpen biedt *TarskiÕs World* de mogelijkheid een spel te spelen. Ofwel je bent de speler die claimt dat de zin waar is of je bent de speler die claimt dat de zin onwaar is. Als je claim klopt kun je het spel winnen. Een dergelijke blik op semantiek is afkomstig van Jaakko Hintikka en Gabriel Sandu. Volgens hen biedt deze blik vele voordelen boven de standaard ÔTarskiaanseÕblik op waarheid. Taal wordt niet opgevat als een regelgestuurd proces, maar als doelgericht proces, waarbij strategie n bepalend zijn. Hintikka beargumenteert in *The Principles of Mathematics Revisited* zelfs dat speltheoretische semantiek een nieuw fundament kan vormen voor de wiskunde. In het hoofdstuk ÔGame-theoretical semanticsÕvan het *Handbook of Logic and Language* leggen Hintikka en Sandu voor een breed publiek uit wat speltheoretische semantiek behelst. In het bijzonder wordt de speltheoretische semantiek van de predikatenlogica behandeld. Dit is het spel dat voor *TarskiÕs World* ge mplementeerd is.

Je speelt het spel tegen de computer om een bepaalde zin bij een bepaalde blokkenwereld. Het is een tweepersoons zero-sum spel. Dat wil zeggen dat er precies twee spelers zijn en dat de belangen van deze spelers tegengesteld zijn. Een voordelige uitkomst voor de ene speler is dus een nadelige uitkomst voor de andere speler. E n speler claimt dat de zin waar is (Helo se) en de andere speler claimt dat de zin onwaar is (Abelard). De verdere regels van het spel liggen vanuit dit perspectief voor de hand. Als de hoofdoperator van de zin een conjunctie is (een Ôen-ÕZin), dan is Abelard aan de beurt en hij kiest een van de conjuncten. Immers, als een conjunctie onwaar is wil dat zeggen dat tenminste een van de conjuncten onwaar is. Abelard moet dus aan kunnen geven waar dat aan ligt. Als Abelard gekozen heeft gaat het spel verder, maar nu met het conjunct dat hij gekozen heeft. Met een disjunctie (een Ôof-ÕZin) is

Helo se aan de beurt. Zij kiest een van de disjuncten en het spel gaat verder met die zin, omdat een disjunctie waar is als tenminste een van de disjuncten waar is. Zinnen met een kwantor als hoofdoperator gaan op dezelfde manier. Bij een universele kwantor (Ôvoor alleÕ) kiest Abelard een object in de blokkenwereld. In het programma doe je dit door er met de muis op te klikken. Dan gaat het spel verder met de zin zonder de universele kwantor, waarbij alle voorkomens van de variabele die door de kwantor gebonden werden vervangen worden door de naam van het gekozen object. Bij een existenti le kwantor gebeurt hetzelfde, maar nu kiest Helo se een object.

Een zin met een negatie (een Ôhet is niet zo datÕ zin) is een apart geval. Hierbij gaat het spel verder om de zin zonder de negatie, maar Abelard en Helo se zijn van rol gewisseld. Dat wil zeggen dat Abelard nu aan de beurt is waar eerst Helo se aan de beurt zou zijn, en andersom. Bovendien wint Abelard nu in het geval dat Helo se gewonnen zou hebben, en andersom. Ook dit ligt voor de hand, omdat claimen dat de negatie van een zin waar is hetzelfde is als te zeggen dat de zin zelf onwaar is. In dit spel komt het claimen dat de negatie van een zin onwaar is dus neer op het claimen dat de zin zelf waar is, hoewel een intu tionist hier misschien bezwaar tegen heeft. Dat brengt ons bij het einde van het spel. Helo se wint als het spel gaat om een atomaire zin (bijvoorbeeld Ôobject *a* is een grote kubusÕ) en deze zin waar is in de blokkenwereld; Abelard wint als deze zin onwaar is. In de laatste stap is deze semantiek dus weer precies gelijk aan TarskiÕs semantiek.

Op deze spelen kunnen we allerlei speltheoretische analyse-technieken loslaten. In het artikel ÔHintikka Self AppliedÕ van Johan van Benthem wordt dit op een zeer heldere wijze gedaan. Ook worden in dit artikel enkele kritische kanttekeningen gemaakt. Allereerst kun je je afvragen wat het wil zeggen dat er een winststrategie is voor een van beide spelers. Welnu, als Helo se een winststrategie heeft voor een zin en een blokkenwereld, wil dat zeggen dat die zin waar is in die blokkenwereld volgens de semantiek van Tarski. Als Abelard een winststrategie heeft wil dat zeggen dat de zin onwaar is.

Een andere eigenschap die deze spelen hebben is dat ze zero-sum zijn. Dat wil zeggen dat de belangen van de spelers tegengesteld zijn. Als Abelard wint verliest Helo se en andersom. Het is dus onmogelijk dat beide spelers tegelijk winnen. Dat betekent dat niet beide spelers een winststrategie kunnen hebben. Oftewel een zin kan niet tegelijk waar en onwaar zijn: het principe van noncontradictie.

Het zijn bovendien spelen van volmaakte informatie. Dat wil zeggen dat iedere speler weet welke zetten er gedaan zijn, zodat beide spelers weten om welke zin gespeeld wordt en met welke blokkenwereld. Ze weten ook precies wat de belangen van de tegenpartij zijn. Bovendien zijn de spelen eindig. Een zin heeft een eindige lengte en na iedere beurt is de zin waarmee verder gespeeld wordt korter. Nu zegt een bekende stelling van Zermelo dat in ieder zero-sum spel met volmaakte informatie precies n van de spelers een winststrategie heeft. Oftewel de zin is waar of onwaar: de wet van de uitgesloten derde.

Zo zie je dat bekende begrippen en stellingen uit de speltheorie inzicht kunnen verschaffen in de logica. Er zijn echter ook uitbreidingen van de predikatenlogica bedacht op basis van deze speltheoretische semantiek. Een van de aspecten die Hintikka en Sandu arbitrair achten is dat het bereik van kwantoren lineair geordend is. Dat wil zeggen dat het bereik van kwantoren ofwel disjunct is, zoals in $\forall x\, \mathit{Kubus}(x) \wedge \forall x\, \mathit{Groot}(x)$, of dat het bereik van een kwantor geheel valt binnen het bereik van een andere, zoals in $\forall x\, \exists y\, (\mathit{Kubus}(x) \wedge \mathit{Groterdan}(x,y))$. Volgens Hintikka en Sandu is er geen enkele reden om deze beperking op te leggen.

> ÒOn a closer look, alas, the requirement of (linear) ordering is seen to be unmotivated by any deeper theoretical reasons and hence dispensible. Every single explanation offered in introductory logic texts or anywhere else of the meaning of the notions of scope and quantifier is applicable irrespective of the restriction. The restriction, we are tempted to say, is simply an outright mistake on FregeÕs and RussellÕs part.Ó

Het voorstel van Hintikka en Sandu is om een nieuw soort kwantoren aan de taal toe te voegen met behulp van een ÔslashÕ waarmee onafhankelijkheid van andere kwantoren kan worden aangegeven. Vanwege de mogelijkheid om onafhankelijkheid uit te drukken wordt deze logica *Independence Friendly* genoemd. $\exists x/\forall y$ wordt bijvoorbeeld gelezen als Ôer is een *x* onafhankelijk van de kwantor $\forall y$ zodat ÉÕ. Een voorbeeld van een zin met zoÕn kwantor is:

$$\forall x \exists y/\forall x\, x=y$$

Voor iedere *x* is er een *y* onafhankelijk van *x* zodat *x* gelijk is aan *y*. Intu tief is deze zin onwaar in alle modellen met tenminste twee objecten. De vraag is echter hoe je deze intu tie om kunt zetten in een formele semantiek. Hintikka en Sandu beargumenteren dat dit eigenlijk volstrekt onmogelijk is met behulp van Tarskiaanse semantiek. Speltheoretische semantiek biedt hier wel een mogelijkheid voor. Als je aan de spelen denkt die hierboven ge ntroduceerd werden is het eenvoudig om een goede interpretatie te vinden. Wat de zin hierboven betreft, betekent het dat Abelard eerst een object voor *x* kiest; vervolgens kiest Helo se een object voor *y* zonder dat ze weet welke keuze Abelard voor *x* heeft gedaan. Tenslotte wordt in het model gekeken of Abelard en Helo se een zelfde keuze gemaakt hebben. Als dat zo is wint Helo se, ander wint Abelard. Wat dit in speltheoretische termen betekent is dat het nu geen spelen van volmaakte informatie meer zijn. Helo se weet immers niet welke zet Abelard gedaan heeft. Hoewel het nog steeds een zero-sum spel is (we hebben dus nog steeds het principe van non-contradictie), zijn we tertium non datur

kwijt, omdat het mogelijk is dat noch Abelard, noch Helo se een winststrategie heeft. Dit is in het spel dat hierboven beschreven werd het geval. Abelard weet niet welk object Helo se gaat kiezen en Helo se weet niet welk object Abelard gekozen heeft. Zo zijn dus niet alle klassiek logische redeneerpatronen geldig in Independence Friendly logica. De gewone eerste orde logica is er echter wel een deel van. Hintikka en Sandu laten bovendien zien dat Independence Friendly logica een werkelijk grotere uitdrukkingskracht heeft dan de gewone predikatenlogica.

Met een speltheoretische blik naar logica kijken is dus meer dan naar hetzelfde met een andere bril op kijken. Speltheorie biedt, zoals in het geval hierboven, soms meer dan standaardtechnieken uit de logica. Dit is een voorbeeld van de invloed van speltheorie op logica. Nu gaan we kijken naar de invloed van logica op speltheorie.

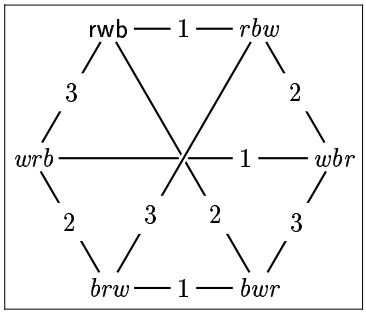
3 Kennis en Verandering

3.1 Kennislogica

In veel situaties ben je onzeker over wat het geval is. Dit is typisch voor besluitvormingstheorische problemen: doordat je niet precies weet wat het geval is weet je niet wat de gevolgen zijn van je handelingen. Een van de vragen die centraal staan in besluitvormingstheorie is hoe je met deze onzekerheid omgaat. Maar wat betekent het eigenlijk dat je onzeker bent? Kennislogica geeft een antwoord op deze vraag.

Kennislogica is zowel ontwikkeld in de speltheorie als in de filosofie. We vinden het dan ook terug in standaard inleidingen in de speltheorie zoals in *Fun and Games* van Ken Binmore en *A Course in Game Theory* van Martin Osborne en Ariel Rubinstein. In de ontwikkeling van deze logica speelt Hintikka ook een belangrijke rol. In het boek *Knowledge and Belief* beschreef hij voor het eerst kennis in termen van mogelijke werelden[1]. Onzekerheid wordt opgevat als een gebrek aan kennis. Het is dus misschien ook beter om van onwetendheid te spreken, omdat onzekerheid vaak ook met waarschijnlijkheidstheorie in verband wordt gebracht. Kennislogica is een modale logica. Als een bepaalde stand van zaken te rijmen is met de informatie die iemand heeft, dan is er een mogelijke wereld toegankelijk voor haar waar die stand van zaken het geval is. De universele modaliteit wordt niet ge nterpreteerd als Ôet is noodzakelijk datÕ maar Ôij of zij weet datÕ Daarom schrijven we ook wel *K* (van Ôknows Õ) in plaats van het gebruikelijke vierkantje.

Laten we eens naar een voorbeeld kijken. In dit voorbeeld wordt ook duidelijk hoe je situaties kunt modelleren waarbij meerdere personen betrokken zijn. Stel je de volgende spelsituatie voor. 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. Laten we aannemen 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. Nu kunnen we met kennislogische formules iets over deze situatie zeggen. Bijvoorbeeld: $K_1 r_1 \wedge \neg K_1 w_2$. Dat wil zeggen dat speler *1* weet dat hij rood heeft en dat hij niet weet dat speler *2* wit heeft. Het model hexa geeft de onzekerheid van de spelers weer.



Figuur 1 hexa

Hierbij representeert rwb de kaartverdeling waarbij speler *1* rood heeft, speler *2* wit, en speler *3* blauw. Deze kaartverdeling staat in een ander lettertype om aan te geven dat dit de echte kaartverdeling is. Deze situatie kan speler *1* niet onderscheiden van de situatie *rbw* waarbij speler *1* nog steeds de rood heeft, maar waarbij speler 2 blauw heeft en speler 3 wit. Dit wordt in het plaatje weergegeven met een lijn met een *1* van rwb naar *rbw*. (Dit wordt ook wel genoteerd als rwb $\sim_1$ *rbw*.)

ZoÕn plaatje als hierboven is de grafische representatie van een *Kripke-model*. Er wordt in aangegeven welke *mogelijke werelden* er zijn; in dit geval de mogelijke kaartverdelingen. Er is een *toegankelijkheidsrelatie* voor iedere speler; deze wordt aangegeven door de lijnen. Bovendien is precies gegeven wie welke kaart heeft in welke mogelijke wereld. Omdat het om kennis gaat zijn de toegankelijkheidsrelaties equivalentierelaties (reflexief, symmetrisch, en transitief). Dit betekent dat we alleen maar met lijnen hoeven aan te geven welke werelden hetzelfde zijn voor een gegeven speler. En als er geen verbindingen zijn (voor een speler), zijn alle werelden verschillend (voor die speler).

[1] Voor een moderne inleiding in de kennislogica zijn *Reasoning about Knowledge* van Fagin et al. en *Epistemic Logic for AI and Computer Science* zeer geschikt. (Zie de referenties aan het eind van het artikel.)

Gegeven een Kripke-model en een mogelijke wereld kun je zeggen wat waar is en wat niet in die wereld. In rwb weet speler *1* bijvoorbeeld dat speler 2 van zichzelf weet dat hij rood niet heeft. (In formules $K_1K_2\neg r_2$). Dit is waar omdat in alle werelden die toegankelijk zijn voor speler *1* (rwb en *rbw*) 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. Zoõ model noemen we een *gepunt* model. De echte wereld is het *punt* van het model.

Kennislogica verschaft tevens deels een invulling van wat het betekent dat iemand rationeel is. Hiermee willen we niet zeggen dat een realistische interpretatie gegeven wordt van rationaliteit. Er wordt een ideaal van rationaliteit verschaft. Er zijn in het bijzonder twee aspecten die niet realistisch zijn voor menselijke kennis: *logische alwetendheid* en *negatieve introspectie*.

Logische alwetendheid is het gevolg van een modale vorm van *modus ponens*: als je weet dat een zin waar is en je weet dat die zin een andere zin impliceert, dan weet je ook dat die andere zin waar is. Bovendien weet je dat iedere tautologie waar is. Dit heeft als gevolg dat je alle logische consequenties kent van wat je weet. Het probleem komt goed tot uitdrukking als je een spel als schaken bekijkt. Als je aanneemt dat iemand alle regels van het spel kent en logisch alwetend is, dan weet zij, gegeven de beginopstelling van alle stukken op het bord, wat alle mogelijke manieren zijn waarop het spel kan verlopen. Dit gaat echter de capaciteiten van de menselijke geest (en ook van een computer) ver te boven. Toch wordt deze aanname in de speltheorie gedaan om dergelijke spelen te analyseren.

Positieve introspectie wil zeggen dat als je iets weet, je dan ook weet dat je dat weet. Dit wordt vaak nog als een realistische interpretatie van menselijke kennis beschouwd. *Negatieve* introspectie is hetzelfde, maar dan voor wat je niet weet: als je iets niet weet, dan weet je dat je dat niet weet. Het belangrijkste bezwaar tegen negatieve introspectie is de mogelijkheid dat er feiten zijn die je niet kent, doordat je er nog nooit over gehoord hebt. Als je een persoon niet kent, weet je ook niet wanneer die jarig is. Het is absurd hieruit te concluderen dat je wel weet dat je niet weet wanneer die persoon jarig is. Toch is het voor spelsituaties waarbij het alleen om relevante proposities gaat al een stuk realistischer.

Met behulp van kennislogica kun je ook een definitie geven van het begrip *gemeenschappelijke kennis* (in het engels *common knowledge*). Dit begrip speelt een belangrijke rol in de speltheorie. Wat het inhoudt komt goed naar voren bij de analyse van het begrip conventie door David Lewis. In het verkeer geldt in Nederland bijvoorbeeld de conventie dat je rechts houdt. Dat wil zeggen dat iedereen weet dat je rechts moet houden. Dat is echter niet het enige. Je zou je niet veilig voelen in het verkeer als je er niet vanuit kon gaan dat iedereen deze conventie kende. Oftewel iedereen weet dat iedereen de conventie kent. Toch moet je ook weten dat iedereen weet dat iedereen de conventie kent, anders voel je je nog onveilig. Enzovoorts. In een Kripke-model is eenvoudig te bepalen of iets gemeenschappelijke kennis voor een groep is of niet. Een propositie *p* is gemeenschappelijke kennis in een mogelijke wereld voor een groep dan en slechts dan als in iedere wereld die bereikt kan worden met een pad dat lijnen volgt met labels voor leden van die groep propositie *p* waar is. Er wordt in situaties die geanalyseerd worden met behulp van speltheorie vaak vanuit gegaan dat bepaalde informatie gemeenschappelijke kennis is. Met behulp van kennislogica kun je nu precies zeggen wat daarmee bedoeld wordt.

3.2 Kennisspelen

In de vorige paragraaf hebben we gezien hoe we met kennislogica de onzekerheid over een gegeven situatie kunnen beschrijven. We lichten dit nu verder toe aan de hand van een type kaartspelen, waarvan we hierboven al een begintoestand zagen: kennisspelen. We beschikken over een aantal kaarten die allemaal van elkaar verschillen. We verdelen de kaarten over een aantal spelers, en nemen aan dat de spelers hun eigen kaarten kunnen inzien. Van andere spelers kunnen ze alleen zien hoeveel kaarten ze hebben, maar niet welke. We hebben nu een typisch voorbeeld van een situatie waarin wel bekend is wat de relevante proposities zijn, maar niet wat hun waarheidswaarde is. Welke speler houdt bijvoorbeeld wit vast? Met vragen en antwoorden kunnen de spelers zich informatie over de kaartverdeling verschaffen. Het doel van het spel is om als eerste duidelijkheid te krijgen over bijvoorbeeld de kaartverdeling, of wie bepaalde kaarten bezit.We beginnen met een eenvoudige versie van zoõ kennisspel: met drie spelers en drie kaarten. In paragraaf 3.2.1 beschrijven we een aantal eenvoudige kennisveranderingen vanuit de beginsituatie van dit spel. In paragraaf 3.2.2 laten we zien hoe we een echt spel kunnen defini ren met dit soort kennisveranderingen.

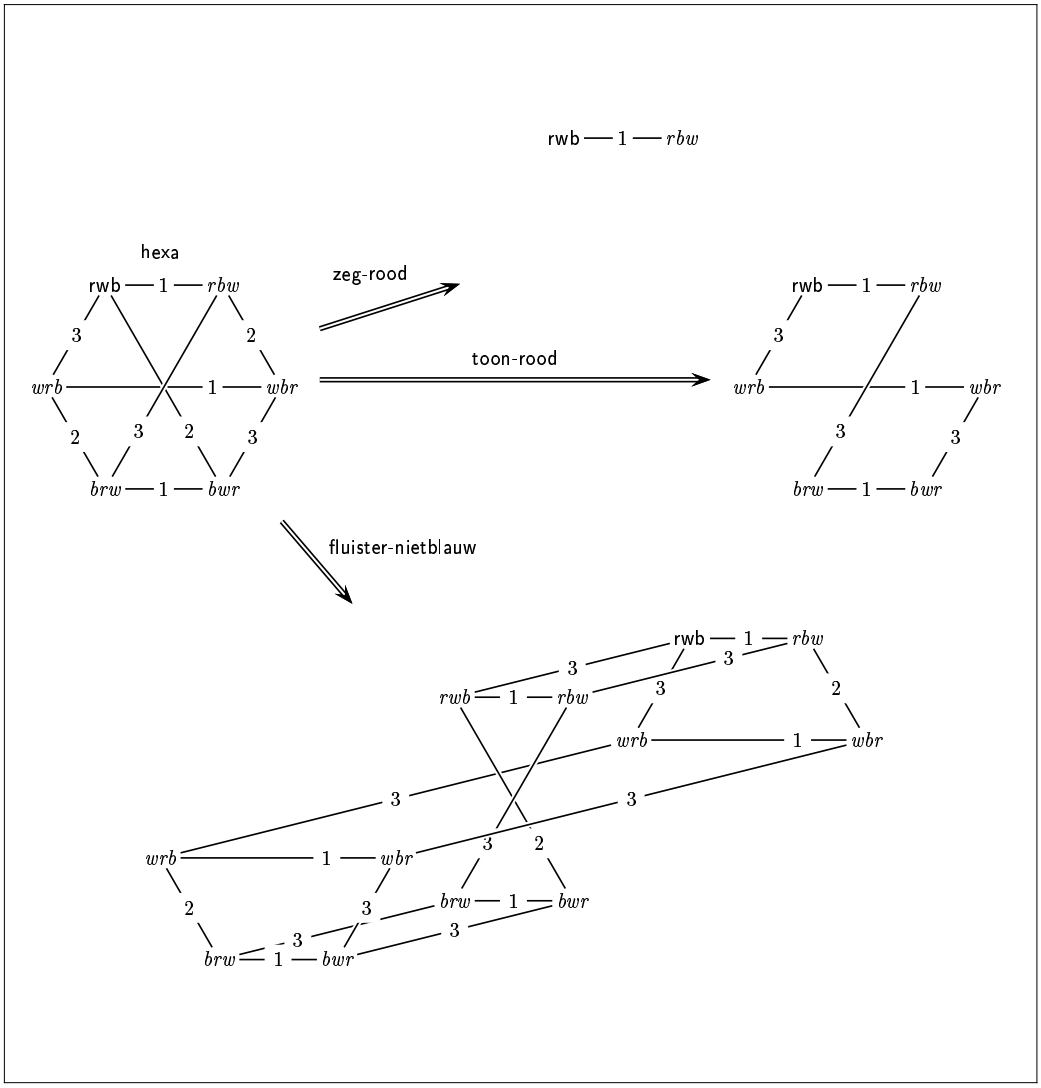
3.2.1 Kennisverandering in kaartspelen

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 equivalentiesklassen 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 drie verschillende manieren waarop speler *2* door een vraag te stellen kan achterhalen welke kaart speler *1* heeft . De vraag moet eerlijk beantwoord worden. Deze manieren zijn:

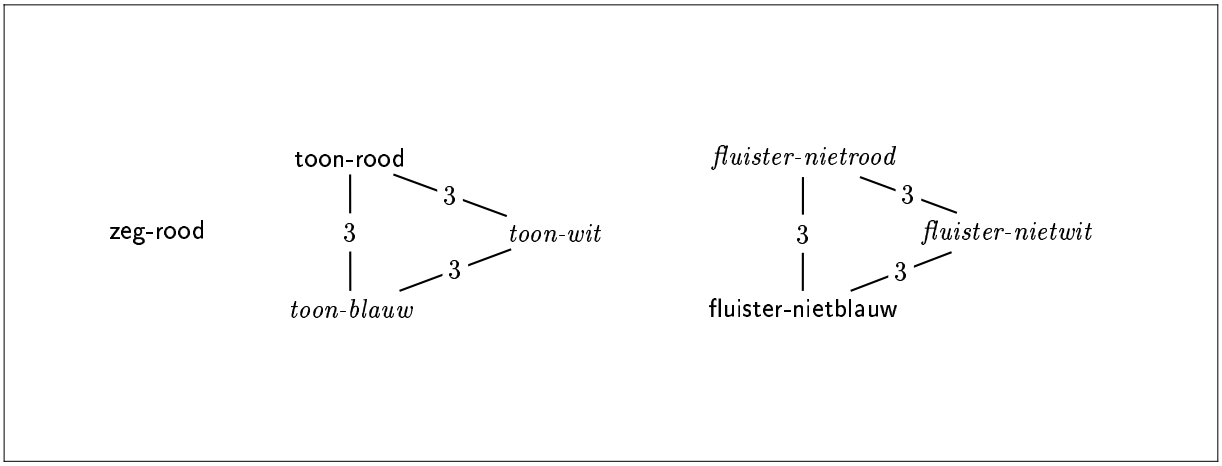
- Speler *2* vraagt speler *1* zijn kaart. Speler *1* legt als antwoord rood op tafel. Gegeven dat de spelers niet liegen is dit dezelfde actie als die waarin speler *1 zegt* dat hij rood heeft. Deze actie noemen we (mede daarom) *zeg-rood*.

- 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*: Ò*k heb blauw niet*ÓDeze 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 alledrie 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!
Een eenvoudige manier om acties te modelleren is met een Kripke-model. Net als een kennistoestand, kunnen we ook een actie weergeven met een soort Ô*Kripke-model*Õ waarbij nu weergegeven wordt welke *acties* (en niet welke werelden) voor de spelers voorstelbaar zijn. We spreken van een Kripke-*frame* en niet van een Kripke-model, omdat de werelden niet gekarakteriseerd worden door een waardering van atomaire proposities. Wel worden ze gekenmerkt door de preconditie van de betreffende actie, zoals bijvoorbeeld dat *1* inderdaad rood heeft bij *zeg-rood.* Net als bij spelmodellen zijn ook bij actieframes de toegankelijkheidsrelaties equivalentierelaties. Gemakshalve identificeren we steeds het punt van het frame met de naam voor het hele frame, net zoals we het model (hexa, rwb) ook wel identificeren met de kaartverdeling rwb. In Figuur 3 zien we welke frames bij de verschillende acties horen.



Figuur 3 Acties als frames.

- *zeg-rood* Als *1* zegt dat hij de rood heeft, is voor alle betrokkenen glashelder wat er zich afspeelt. De gegeven actie is voorstelbaar voor alle spelers, en er zijn geen alternatieve acties voorstelbaar. De preconditie voor het uitvoeren *zeg-rood* is dat *1* inderdaad rood heeft, (r_1 dus). We schrijven dit als $pre(zeg\text{-}rood) = r_1$. De actie is uitvoerbaar, immers in hexa, rwb geldt r_1.
- *toon-rood* Er zijn drie alternatieve acties voorstelbaar: *1* toont wit, *1* toont rood, en *1* toont blauw. Alledrie zijn relevant! Dit kunnen we als volgt toelichten. Als *1* zijn roood aan *2* laat zien zonder dat *3* het ziet, is voor *3* ook voorstelbaar dat wit is geweest. Voor *3* is niet voorstelbaar dat dit blauw was, want die heeft *3* zelf. Maar dit is echter geen algemene kennis: voor *2* is namelijk voorstelbaar dat voor *3* voorstelbaar is dat *1* blauw toonde, want *2* weet niet dat *3* blauw heeft. Alle acties waarvan het geen algemene kennis is dat *3* ze van elkaar kan onderscheiden, zijn voor *3* ÔhetzelfdeÕop het actie-frame voor *toon-rood*. Uiteraard kunnen *1* en *2* de drie acties wel van elkaar onderscheiden. De preconditie voor het uitvoeren van *toon-rood* is dezelfde als bij *zeg-rood*, namelijk r_1. Analoog voor *toon-wit* en *toon-blauw*.
- *fluister-nietblauw* Net als bij *toon-rood* zijn er ook bij *fluister-nietblauw* drie alternatieve acties voorstelbaar: *1* fluistert Ônet roodÕ, Ônet witÕof Ônet blauwÕDe preconditie voor het uitvoeren van de actie *fluister-nietblauw* is dat *1* inderdaad niet blauw heeft, ($\neg\, b_1$ dus). Analoog voor *fluister-nietrood* en *fluister-nietwit*.

Gegeven een Kripke-model $(M,w) = (\langle W, \{\sim_a\}_{a\in A}, V\rangle, w)$ voor een speltoestand, dat de kennis van de actoren weergeeft, en een Kripke-frame $(\langle A, \{\sim_a\}_{a\in A}\rangle, a)$ met precondities $Pre(A) = \{pre(a) \mid a \in A\}$ voor een actie, dat de mogelijke acties weergeeft in relatie tot hun voorstelbaarheid voor de actoren, kunnen we nu heel eenvoudig het Kripke-model $(M',w') = (\langle W', \{\sim_a\}_{a \in A}, V'\rangle, w')$ berekenen dat de volgende toestand weergeeft, dat wil zeggen de kennis van de actoren na het uitvoeren van de actie:

W'	=	$\{(w',a') \mid w' \in W$ *en* $a' \in A$ en in M,w' geldt $pre(a')\}$
$(w_1,a_1) \sim_a' (w_2,a_2)$	≡	$w_1 \sim_a w_2$ en $a_1 \sim_a a_2$
$V'_{(w',a')}$	=	$V_{w'}$
w'	=	$(w,a) \in W'$

Wat betekent deze definitie nu? Uit de verzameling van mogelijke werelden en de verzameling van mogelijke acties maken we alle combinaties van acties en werelden zodanig dat de actie in de wereld uitvoerbaar is. Zo krijgen we de verzameling van volgende toestanden. Twee volgende toestanden zijn ononderscheidbaar voor een speler, dan en slechts dan als zowel de vorige toestanden als de acties niet van elkaar te onderscheiden waren. Omdat er verder niets aan de wereld verandert blijft waardheidswaarde van atomaire proposities hetzelfde.

We passen deze definitie nu toe op een van de drie voorbeeld-acties, namelijk op de actie toon-rood. De drie alternatieve acties zijn toon-rood, *toon-wit* en *toon-blauw*. De actie toon-rood is uitvoerbaar in de werelden *rwb* en *rbw*; de actie *toon-wit* is uitvoerbaar in de werelden *wrb* en *wbr*, de actie *toon-blauw* is uitvoerbaar in de werelden *brw* en *bwr*. De volgende speltoestand bestaat dus uit zes wereld-actie-paren (rwb, toon-rood), (*rbw, toon-rood*), (*wrb, toon-wit*), (*wbr, toon-wit*), (*brw, toon-blauw*) en (*bwr, toon-blauw*), waarbij de eerste de werkelijke is. De waardering van atomen blijft hetzelfde, dus in (rwb, toon-rood) bijvoorbeeld geldt r_1 nog steeds. De toegankelijkheid voor de verschillende spelers kunnen we nu ook berekenen, bijvoorbeeld: (rwb, toon-rood)$\sim_1$ (*rbw*, toon-rood) want rwb $\sim_1$ *rbw* en toon-rood $\sim_1$ toon-rood; (rwb, toon-rood) $\sim_3$ (*wrb, toon-wit*) want rwb $\sim_3$ *wrb* en *toon-rood* $\sim_3$ *toon-wit*, etc.Merk ook op dat alle zes de wereld-actie-paren reflexief zijn voor zowel *1*, *2* als *3*, omdat zowel wereld-modellen als actie-frames zelf ook reflexief zijn. Het resultaat van de andere acties, *zeg-rood* en *fluister-nietblauw*, kunnen we net zo berekenen. Vergelijk de resultaten nogmaals met Figuur 2.

De Kripke-modellen voor de kaartverdelingen en de kennis van de spelers konden we beschrijven met kennislogica. Op dezelfde manier kunnen we de actie-frames beschrijven met een logische taal voor ‘kennisacties’ (knowledge actions). Dit gebeurt in het proefschrift *Knowledge Games* van Hans van Ditmarsch. 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 ‘leren’ en de nummers in subscript bij deze operator staat 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 $[toon\text{-}rood]K_2r_1$. In van Ditmarsch’ proefschrift (zie de referenties aan het eind van het artikel) worden dergelijke acties overigens in eerste instantie geïnterpreteerd als een binaire relatie tussen Kripke-modellen en niet als frames. Dit is vergelijkbaar met de relationele manier waarop in het werk van Jelle Gerbrandy (zie de referenties) updates worden geïnterpreteerd. De interpretatie als actie-frame is een alternatieve interpretatie gedefinieerd in het werk van Alexandru Baltag (zie de referenties).

3.2.2 Het spelen van een kennisspel

We hebben nu modellen en acties, maar nog geen spel. Een actie zoals *toon-rood* zouden we kunnen zien als een zet in een spel van vragen en antwoorden. We hebben verder nog weinig gezegd over de regels. Zoals we al eerder zeiden is het doel van het spel is om de kaartverdeling als eerste te kennen. De beurt gaat met de klok mee (de spelers zitten rond een tafel), waarbij de eerste vragensteller door het lot bepaald wordt. 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: ‘ja’ of het tonen van een kaart. De vraag naar een van drie kaarten komt dus neer op de vraag ‘toon me je kaart’. Nadat de vraag beantwoord is, mag een speler zeggen dat hij weet wat de kaartverdeling is (dan moet hij het natuurlijk wel weten). De eerste speler die dat doet, wint het spel. Hiermee hebben we een kennisspel gedefinieerd. Voor 3 spelers en 3 kaarten, en vragen naar één van drie kaarten zijn de strategieën 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 één 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 ‘nee’, want speler *2* heeft zelf wit. Speler *2* beëindigt 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. Uiteraard had speler *2* het spel ook kunnen winnen, namelijk door een andere strategie te kiezen, dat wil zeggen een andere eerste vraag te stellen. Zowel de strategie ‘heb je wit?’ als ‘heb je blauw?’ leidt in dit geval tot winst. Hier kiest speler *2* natuurlijk voor.

Merk op dat ook de acties waarin een speler een vraag met ‘nee’ beantwoordt, of waarin een speler wint, of waarin een speler zijn zet beëindigt (dat wil zeggen ‘niet wint’), acties zijn die geanalyseerd kunnen worden met de methoden uit de vorige paragraaf. 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* ‘nee’ antwoordt op de vraag van *2* of hij wit bezit. In het tweede en in het derde geval is het een test op een bewering over de kennis van spelers. Als in het op-een-na-laatste spelvoorbeeld speler *2* zijn zet beëindigt 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 δ_{rwb} voor de atomaire beschrijving van deze kaartverdeling, d.w.z. $\delta_{rwb} = r_1 \wedge w_2 \wedge b_3 \wedge \neg r_2 \wedge \neg r_3 \wedge \neg w_1 \wedge \neg w_3 \wedge \neg b_1 \wedge \neg b_2$, etc. voor de overige vijf kaartverdelingen. Het bijbehorende actieframe bestaat uit één enkele mogelijke actie, reflexief voor alle spelers, met preconditie $\neg(K_2\,\delta_{rwb} \vee K_2\,\delta_{rbw} \vee ...)$. Al dit soort publieke bekendmakingen (dus ook ‘nee’ zeggen) zijn frames die uit één wereld bestaan.

3.3 Cluedo

Als er maar 3 spelers en 3 kaarten zijn, is het dus eigenlijk niet zoõ spannend 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. Tot slot van deze verhandeling over kennislogica en spelen lichten we toe hoe het spel Cluedo in zijn werk gaat, en wat het verband met kennislogica is.

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 een 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 al iets over hoe de moord gepleegd kan 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 ÔgeeÕ ofwel met het tonen van een van deze drie gevraagde kaarten. Dit is dus hetzelfde soort actie als *toon-rood*. Zolang het antwoord ÔgeeÕ 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. 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? Het is ook de vraag hoe logische methoden en technieken zouden kunnen bijdragen aan het bepalen van optimale strategie n. Het uitbreiden van een logische taal zodat ook uitspraken over waarschijnlijkheden kunnen worden gedaan, zoals in recent werk van Kooi, lijkt een stap in de goede richting. De logische modellering van spelsituaties en zetten is echter onmisbaar, voordat je over strategie n kunt gaan nadenken.

Hiermee is de cirkel van speltheorie naar kennislogica weer rond: binnen speltheorie is de rol van kennislogica om onzekerheid precies uit te drukken. Met dynamische kennislogica kunnen we tevens kennisveranderingen uitdrukken: acties. Op basis van dit soort kennisveranderingen kunnen we dan weer spelen defini ren. En tot slot moeten we ons dan opnieuw afvragen wat de speltheorie van dit soort spelen is. We zijn weer terug bij speltheorie. De toekomst zal ons leren hoe deze vragen beantwoord moeten worden.

4 Conclusie

Zo zien we dat er op het raakvlak van speltheorie en logica veel te beleven valt. Hoewel sommigen het nut betwijfelen van het toepassen van speltheoretische middelen in de logica enerzijds en het toepassen van logische middelen in de speltheorie anderzijds, zijn wij ervan overtuigd dat er sprake is van een vredig samenzijn. Sterker nog van een samenwerking die beide vakgebieden doet bloeien, waarvan ook dit themanummer getuigt.

Referenties

- Alexandru Baltag, *A logic of epistemic actions*, manuscript, 1999.
- Johan van Benthem, Hintikka self-applied, an essay on the epistemic logic of imperfect information games, te verschijnen in Lewis Kahn (red.) Hintikka Volume, *Library of Living Philosophers*.
- John Barwise en Jon Etchemendy. *The Language of First-order Logic*, volume 34 van *CSLI lecture notes*. Center for the Study of Language and Information, Stanford, California, third edition, revised and expanded, 1992.
- Ken Binmore. *Fun and Games, A Text on Game Theory*, D.C. Heath & Company, Lexington, Massachusetts, 1992
- Hans van Ditmarsch. *Knowledge Games*, ILLC Dissertation Series, Amsterdam, 2000.
- Ronald Fagin, Joseph Y. Halpern, Yoram Moses, Moshe Y. Vardi. *Reasoning about Knowledge*, MIT Press, Cambridge, Massachusetts, 1995.
- Jelle Gerbrandy. *Bisimulations on Planet Kripke*, ILLC Dissertation Series, Amsterdam, 1999.
- Jaakko Hintikka. *Knowledge and Belief, An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca & London, 1962.
- Jaakko Hintikka. *The Principles of Mathematics Revisited*, Cambridge University Press, Cambridge, 1996.
- Jaakko Hintikka en Gabriel Sandu. Game-theoretical semantics, Chapter 6 in *Handbook of Logic and Language*, Johan van Benthem en A. ter Meulens (red.), Elsevier Science B.V., 1997.

- Barteld Kooi. Probability in Dynamic Epistemic Logic, ingediend voor het *Journal of Logic, Language, and Information*.
- David Lewis. *Convention*, Harvard University Press, Cambridge, Massachusetts, 1969.
- John-Jules Meyer en Wiebe van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, Cambridge, 1995.
- Martin J. Osborne en Ariel Rubinstein. *A Course in Game Theory*, MIT Press, Cambridge, Massachusetts, 1994.