

Logica Nu

Sonja Smets
(ILLC, Universiteit van Amsterdam)

Overzicht

Logica in de oude tekstboeken = De studie van het menselijk redeneren

Logica Nu

De Studie van ieder informatieverwerkend proces.

- ▶ Leren, observeren, communicatie, spelen (games), argumentatie, dialoog, informatie-overdracht.
- ▶ Meerdere Actoren ('Agenten')
- ▶ Doelgericht: Klassieke kennisverwerving, geloofsherziening, strategie ontwikkeling, planning.

Bijvoorbeeld: We gebruiken Logica in de studie van epistemisch sociale fenomenen, bijvoorbeeld in de studie van de kennis in sociale netwerken.

Hogere-Orde Redeneren

“... as we know, there are known knowns; there are things we know we know.

We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know”

US Secretary of Defense, D. Rumsfeld, Feb. 12, 2002, Dept. of Defense news briefing



Hogere Orde Redeneren



Hij denkt dat zij denkt dat de kat aan haar denkt.

[bron beeldmateriaal: www.psychologytoday.com]

Epistemische Attitudes

Kennis

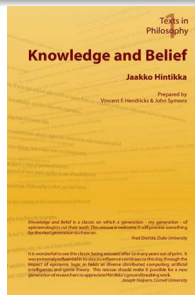
- ▶ We kijken naar een groep van **agenten met bepaalde kennis**. Dus we kennen **epistemische attitudes** toe aan agenten.
- ▶ Agenten kunnen ook nieuwe feiten leren (en dus kennis kan vergroten).

Nota over de Etymologie

- ▶ 'Epistemische logica' bevat het woord 'epistèmè', Grieks voor 'kennis'.
- ▶ 'Doxastische logica' bevat het woord 'doxai', Grieks voor 'Geloof of opinie'.

Oorsprong

Epistemische logica (om te redeneren over 'kennis'), zoals we het vandaag kennen, gaat terug naar het werk van Jaakko Hintikka (1962) en de eerdere ontwikkelingen in de jaren 1948-1950s (door Carnap, Prior, Von Wright).



Verdere ontwikkeling door filosofen (R. Parikh, R. Stalnaker etc.) en informatici (Halpern, Vardi, Fagin etc.)

Representatie van Kennis

In tegenstelling tot het idee om 'knowledge' rechtstreeks te modelleren, representeren we in de **epistemische logica** de **notie van 'epistemische onzekerheid'** (wat we niet weten) en we berekenen wat 'kennis' is.



[bron beeldmateriaal: <https://beursmeesters.nl/binaire-opties/soorten-opties/>]

Epistemische Onzekerheid

In zogenaamde Kripke Models, drukken we **epistemische onzekerheid** uit via een relatie tussen 'mogelijke werelden' of mogelijke situaties zoals een agent denkt dat de wereld zou kunnen zijn.



We nemen alle alternatieve mogelijke werelden in overweging.

Definitie van Kennis

Kennis

“**Kennis van p** ” (uitgedrukt in een symbolische taal via de afkorting Kp) betekent dat:

p waar is in alle werelden die deze agent epistemisch mogelijk acht.



[bron beeldmateriaal: <http://www.webwinkelsucces.nl>]

Symbolische Taal

De taal van (1-agent) epistemische logica, bestaat uit de taal van de propositie logica met de uitbreiding van een epistemische operator K .

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K\varphi$$

- ▶ We lezen $K\varphi$ als “**de agent weet dat φ het geval is**”.
- ▶ We kunnen de volgende afkorting gebruiken: $\langle K \rangle \varphi$ staat voor $\neg K \neg \varphi$ en lezen we als “ φ is consistent met de kennis van de agent”.

Epistemisch Kripke Model for 1 agent

Model

Een epistemisch model

$$(W, \sim, ||\cdot||)$$

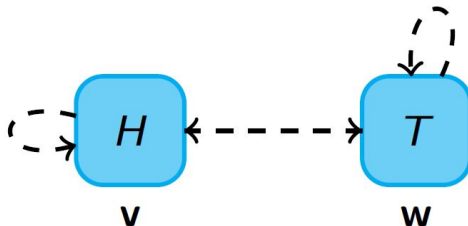
Bestaat uit:

- ▶ een verzameling van mogelijke werelden W (de epistemisch mogelijke situaties)
- ▶ een binaire relatie $\sim$ (voor de impliciete agent)
- ▶ een valuatiefunctie $||\cdot||$, die voor elke atomaire zin p (uit een gegeven verzameling Ω) zegt in welke mogelijke wereld die zin waar is.

Voorbeeld: Muntstuk

Een muntstuk ligt op tafel, het is bedekt, dus de agent acht het mogelijk dat de bovenzijde kop (Heads) is en acht het mogelijk dat het munt (Tails) is. Gegeven de atomaire zinnen H (Heads) en T (Tails), dan tekenen we het model $M = (W, \sim, ||.||)$ als volgt:

(W is de gegeven verzameling van mogelijke werelden, $\sim$ de epistemische relatie en $||.||$ de valuatie-functie)



$M, s \models K\varphi$ als en slechts dan als (asda) $M, t \models \varphi$ voor all t zodanig dat $s \sim t$

Interpretatie van de K -operator

- De epistemische relatie gebruiken we om een interpretatie te geven aan de K -operator:

$$M, s \models K\varphi \text{ asda } M, t \models \varphi \text{ voor alle } t \text{ zodanig dat } s \sim t$$

$$M, s \models \langle K \rangle \varphi \text{ asda}$$

Er bestaat een wereld t zodanig dat $s \sim t$ en $M, t \models \varphi$

Eigenschappen van Kennis

Aangezien we de kennis van agenten willen weergeven, gaan we ervan uit dat de relatie aan de volgende eisen moet voldoen:

voor alle $v, s, t \in W$:

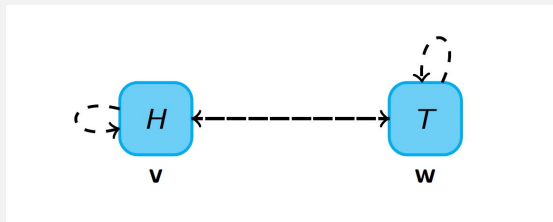
- ▶ **reflexief**, $v \sim v$
- ▶ **transitief**, als $s \sim t$ en $t \sim v$ dan $s \sim v$
- ▶ **euclidisch**, als $s \sim t$ en $s \sim v$ dan $t \sim v$

Oefening: Dit impliceert dat de relatie ook een equivalentie-relatie is (reflexief, transitief en symmetrisch).

Eigenschappen van Kennis

- ▶ Reflexiviteit, correspondeert met de formule $K\varphi \rightarrow \varphi$ en drukt uit dat **alle kennis ook waar is**.
- ▶ Transitiviteit, correspondeert met de formule $K\varphi \rightarrow KK\varphi$ en drukt uit dat agenten **positief introspectief** zijn.
- ▶ Euclidisch, correspondeert met de formule $\langle K \rangle \varphi \rightarrow K \langle K \rangle \varphi$ en drukt uit dat agenten **negatief introspectief** zijn.

Muntstuk voorbeeld



Gegeven dit model M voor de presentatie van kennis, zien we dat de agent het mogelijk acht dat (H) het geval is en dat zij het mogelijk acht dat (T) het geval is.

Ga na dat de volgende zinnen waar zijn in de aangegeven wereld:

- ▶ $M, v \models \neg KH$
- ▶ $M, v \models \neg KT$
- ▶ $M, v \models \neg KH \wedge \neg KT$
- ▶ $M, v \models K(\neg KH \wedge \neg KT)$

Multi-agenten Epistemische Logica: Semantiek

Model

Een multi-agenten epistemisch model

$$(W, \sim_{a \in \mathcal{A}}, ||\cdot||)$$

Bestaat uit:

- ▶ Een verzameling van mogelijke werelden W (voor de weergave van alle situaties die epistemisch mogelijk zijn)
- ▶ Een **binaire epistemische relatie** $\sim_a$, één voor elke agent a
- ▶ Een **valuatie functie** $||\cdot||$, die voor elke atomaire zin p (uit een gegeven verzameling Ω) zegt in welke mogelijke wereld de zin waar is.

De taal van Multi-agenten Epistemische Logica

We voegen aan de taal van de klassieke propositie logica een aantal operatoren toe K_a , eentje voor elke agent in de verzameling van agenten $\mathcal{A}$

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \varphi \mid K_a\varphi$$

- ▶ definitie $\langle K \rangle_a \varphi := \neg K_a \neg \varphi$
- ▶ We werken nu met labels op de epistemische relatie $\sim_a$ om een interpretatie aan K_a te geven:

$$M, s \models K_a\varphi \text{ asda } M, t \models \varphi \text{ voor alle } t \text{ zodanig dat } s \sim_a t$$

- ▶ $K_a\varphi$ lezen we als: **agent a weet dat φ het geval is.**

Hogere Orde Redeneren

In de taal van de epistemische logica kunnen we de operatoren gebruiken om iteraties van kennis over kennis (door dezelfde of door verschillende agenten) weer te geven:

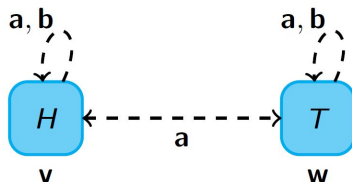
$$K_a K_b \neg K_a \varphi$$

lezen we als:

Alice weet dat Bob weet dat zij niet weet dat φ het geval is



Multi-agenten Muntstuk

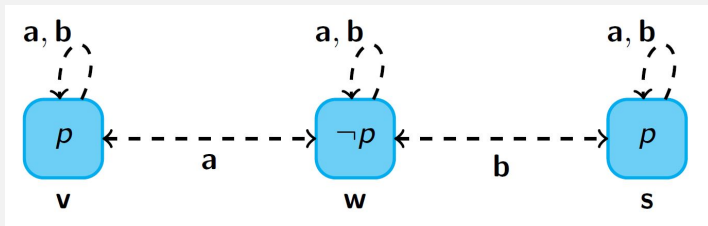


Dit model M gebruiken we voor het scenario waarin Alice (a) niet weet wat de bovenzijde van het muntstuk is maar waarin Bob (b) dat wel weet.

Ga na dat:

- ▶ $M, v \models K_a(K_b H \vee K_b T)$
- ▶ $M, v \models K_b H$
- ▶ $M, v \models \neg K_a K_b H$

Hogere Orde Kennis (Theory of Mind)

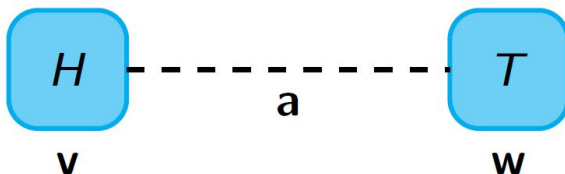


Ga na dat in dit 2-agenten epistemisch model waarin (a) staat voor Alice en (b) voor Bob, dat in wereld v :

- ▶ Bob weet dat p het geval is,
- ▶ Alice weet niet dat Bob weet dat p het geval is.
- ▶ Alice acht het mogelijk dat Bob niet weet dat p het geval is.
- ▶ Bob weet dat Alice niet weet dat hij weet dat p het geval is.

Vereenvoudiging van onze notatie

Wanneer we de epistemische relaties tekenen, doen we dat vanaf nu zonder pijlpunten te tekenen en zonder al de reflexieve en transitieve pijlen expliciet te tekenen (die er wel zijn). Dus het 2-agenten muntstuk-voorbeeld ziet er zo uit:



Gemeenschappelijke Kennis

Wanneer kunnen we spreken over **gemeenschappelijke kennis** (common knowledge)?

Stel dat iedereen de verkeersregels kent in Nederland. Bijvoorbeeld, iedereen weet dat 'rood licht' staat voor 'stop' en 'groen licht' betekent 'doorgaan'. Stel ook dat iedereen deze regels respecteert.

Vraag: Is dit voldoende om je veilig te voelen op de weg?

Kennis in een groep agenten

Wat we moeten veronderstellen is dat de verkeersregels (en het feit dat ze worden gevolgd) ook echt **gemeenschappelijke kennis** zijn:

Iedereen kent (en volgt) de regels, en iedereen weet dat iedereen de regels kent (en volgt), en dat iedereen weet dat iedereen weet dat iedereen weet ... etc.

Verschillende vormen van groep-kennis zijn mogelijk:

- ▶ gemeenschappelijke kennis (common knowledge),
- ▶ gedistribueerde kennis (distributed knowledge),
- ▶ iedereen (individueel) weet dat iets het geval is.

Gemeenschappelijke Kennis

Een groep agenten heeft **gemeenschappelijke kennis** van een propositie p , als iedereen weet dat p het geval is, iedereen weet dat iedereen weet dat p het geval is, iedereen weet dat iedereen weet dat iedereen weet dat p het geval is, etc.

- ▶ Bestudeerd door David Lewis in the context van conventies
- ▶ Beschreven door McCarthy in KI als dat **what 'any fool' knows**.

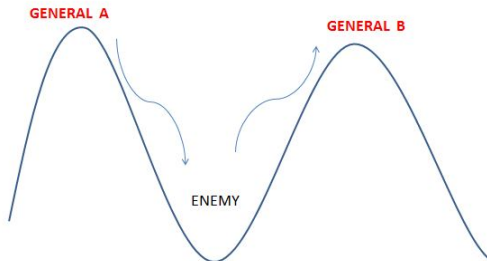
Sillari and Vanderschraaf, Stanford Encyclopedia of Philosophy

“If a married couple is separated in a department store, they stand a good chance of finding one another because their common knowledge of each others’ tastes and experiences ... Since the spouses both love cappuccino, each expects the other to go to the coffee bar, and they find one another.”

Bereiken van Gemeenschappelijke Kennis

Coordinated Attack Problem

Twee legers (elk heeft een generaal), bereiden zich voor om de gemeenschappelijke vijand aan te vallen. Als ze de aanval niet coördineren dan kunnen ze niet winnen. De enige manier waarop ze kunnen coördineren is om een bericht te sturen via een messenger van A naar B (via het kamp van de vijand).



Achieving Common Knowledge: continued

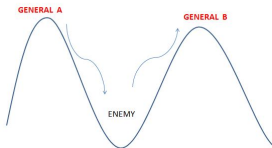
Coordinated Attack Problem

Dus: **A stuurt het bericht:** “aanval morgenochtend”

B ontvangt het bericht en denkt: “A weet niet dat ik dit bericht heb ontvangen”, ik moet een bevestiging van ontvangst sturen.

A ontvangt de bevestiging en denkt: B heeft mijn bericht ontvangen maar B weet niet dat ik zijn bevestiging heb ontvangen. Dus ik moet een bevestiging van de bevestiging versturen.

B ontvangt de bevestiging van de bevestiging en denkt nu: A weet niet dat ik dit bericht heb ontvangen en dus moet ik een bevestiging terugsturen...



Bereiken we gemeenschappelijke kennis

Coordinated Attack Problem

Kunnen deze generalen op deze wijze gemeenschappelijke kennis verkrijgen om volledig zeker te zijn over het tijdstip van de gecoördineerde aanval?

Verkrijgen van Gemeenschappelijke Kennis

Kunnen we ooit gemeenschappelijke kennis over een basis feit verkrijgen?

Ja, bijvoorbeeld door **een ware publieke uitspraak van een basis feit**:

het publiek zal de aankondiging horen en zal het feit daarna kennen; iedereen weet dat iedereen weet dat iedereen weet ... dat iedereen de uitspraak heeft gehoord en het feit kent.



De rol van vraagstellingen en publieke aankondigingen in het verkrijgen van kennis

We kijken naar een beroemde logica puzzel, “the muddy children puzzle” (in een andere versie gaat dit terug tot 1958).

Epistemische Puzzel: De Kinderen met Modder op hun voorhoofd

Stel dat er 4 kinderen buiten spelen, allen zijn excellent in de logica, 3 hebben modder op hun voorhoofd. *Elk kind kan het voorhoofd van de anderen zien maar niet zijn/haar eigen voorhoofd..*

De vader maakt een publieke aankondiging:

“Minstens 1 van jullie heeft modder op het voorhoofd” .

vervolgens

De vader begint nu met het herhalen van dezelfde vraag: **“Weet jij als je wel of geen modder op je voorhoofd hebt? Ja/Nee, en indien Ja welke van de twee?”**

Na elke vraag, geven de kinderen **tegelijkertijd een publiek antwoord, enkel op basis van hun kennis**. Geen andere communicatie is toegestaan en niemand mag liegen.

We kunnen formeel aantonen dat na 2 rondes van vragen en antwoorden, **al de kinderen met modder op hun voorhoofd weten dat ze een vuil voorhoofd hebben!** Dus ze geven dit antwoord in de derde ronde, waarna ook **het kind met het schoon voorhoofd weet dat zij schoon is**, en zij geeft het juiste antwoord in de vierde ronde.

Als we k kinderen met modder op hun voorhoofd hebben, dan is er een bewijs dat aantoonst dat de $k - 1$ keer dat de vader de vraag stelt, al de kinderen 'nee' zeggen, maar als hij de vraag stelt voor de k -ste keer, zullen alle kinderen met modder op hun voorhoofd simultaan antwoorden 'Ja, ik weet dat ik modder op mijn voorhoofd heb'.

Sketch van het bewijs (Fagin et al., 'Reasoning about Knowledge', 1995)

Het bewijs is per inductie over het aantal kinderen met modder op hun voorhoofd k .

- ▶ Stel $k = 1$, het enige kind met modder ziet dat al de andere kinderen een schoon voorhoofd hebben. Zij kan afleiden dat zij het enige kind is met een vuil voorhoofd en zal dit ook onmiddellijk antwoorden.
- ▶ Stel $k = 2$, twee kinderen met modder, a en b . Elk geeft het antwoord 'Nee' in de eerste rond, omdat ze elkaar zien. Maar nadat b 'nee' zegt, weet a dat zij modder op haar voorhoofd heeft, want anders zou b ja gezegd hebben (zie geval $k = 1$) en dus zegt a 'ja' in de tweede rond. Idem voor b .

Vervolg Sketch van het bewijs

- ▶ Stel $k = 3$, drie kinderen met modder, a, b, c . Kind a denkt: als ik schoon ben, dan in het geval $k = 2$ moeten zowel b en c 'Ja' antwoorden in de tweede ronde, als ze dat niet doen dan was haar veronderstelling fout en komt ze tot de conclusie dat ze modder op haar voorhoofd heeft en dus zegt ze 'ja' in de derde ronde. De anderen redeneren op gelijke wijze.
- ▶ Het argument gaat zo verder voor alle k .

Eerste Vraag: *Wat wil de vader bereiken met zijn eerste aankondiging als hij zegt: Ten minste 1 van jullie heeft modder op het voorhoofd?*

Op het eerste zicht zou je denken dat dit bericht niet informatief is. Maar toch zorgt deze zin ervoor dat **er nieuwe informatie bij komt**:

De kinderen leren nu dat wat iedereen individueel reeds wist (namelijk dat ten minste 1 kind vuil is), opeens **gemeenschappelijke kennis is**.

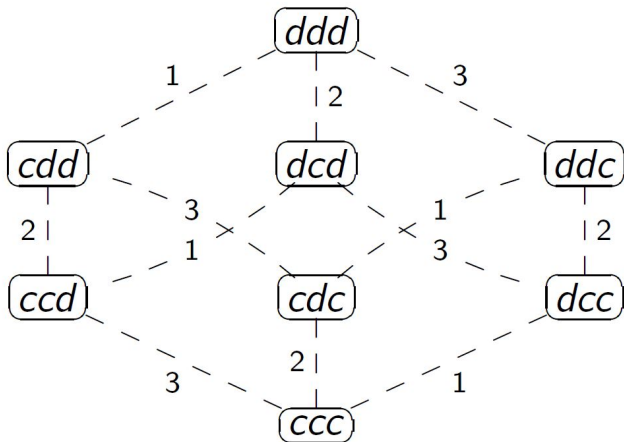
Tweede Vraag: *Wat is de bedoeling van de herhaling van de vragen van de vader?*

De bedoeling is niet om een gebrek aan kennis bij de vader te verhelpen, maar de kinderen leren door de vragen die de vader stelt. Dit is een socratische leer methode.

Derde Vraag: *Hoe kunnen de uitspraken van de kinderen over hun gebrek aan kennis hen tot kennis brengen?* De kinderen leren van de uitspraken van de anderen.

Voorbeeld: “Muddy Children”

Drie kinderen, waarvan kind 1 en kind 2 modder op hun voorhoofd kregen (d , dirty). Oorspronkelijk gaan we ervan uit dat de kinderen geen informatie hebben over hun eigen toestand, dus ze achten zowel (d , dirty) en (c , clean) mogelijk:



Vanuit de positie van de persoon die het model opstelt, kunnen we aangeven welke de echte **reële** wereld is. In dit geval is dat de wereld waarin (*ddc*) waar is, dus de wereld waarin het eerste en tweede kind modder op hun voorhoofd hebben en het derde kind een schoon voorhoofd heeft.

Teken de (serie van) modellen waarin we analyseren wat het resultaat is van de publieke aankondigen van de vader en van de antwoorden van de kinderen op al de vragen die de vader stelt.

Puzzel aangepast: Sneaky Children (door A. Baltag)

Stel dat we 4 kinderen hebben waarvan 3 modder op hun voorhoofd hebben. Stel ook dat kinderen nu een straf krijgen als ze incorrect antwoorden en dus willen ze zo snel mogelijk tot de juiste conclusie komen.

Stel dat, **na de eerste ronde van vragen en antwoorden, twee van de kinderen met modder op hun voorhoofd elkaar een teken geven (tegen de regels van het spel in)**, terwijl de anderen hiervan niets merken. Deze kinderen die zich niet aan de regels houden geven in de tweede rond nu het antwoord: ja, ik weet dat ik modder op mijn voorhoofd heb.

Wat zijn de antwoorden van het derde kind op de vragen van de vader?

Een tweede vorm van groepskennis

Gedistribueerde Kennis

Een propositie p is *gedistribueerde kennis* in een groep G als het kan worden afgeleid van een verzameling premissen, waarbij elk van deze premisse tot de individuele kennis behoort van ten minste 1 lid van de groep.

- ▶ “That what a wise man knows”.
- ▶ Dat wat we weten als we alle kennis van iedereen in de groep samenbrengen en hieruit de logische gevolgen trekken.



Hoe kunnen we gedistribueerde kennis verkrijgen?

In de meeste situaties zien we dat mensen die communiceren of delibereren niet tot het niveau van **gedistribueerde kennis** komen, in veel situaties vallen ze terug op dat wat al gemeenschappelijke kennis was nog voor de deliberatie begon.

Kennis van de leden van de groep

Vraag: Is de kennis in een groep 'meer' dan dat van de individuele leden van de groep?

Meer of minder dan Individuele Kennis?

(+)

Agenten kunnen hun kennis combineren (via logische afleidingen, communicatie, stemmen etc.) en nieuwe informatie in de groep brengen die geen van de individuele leden heeft.

(-)

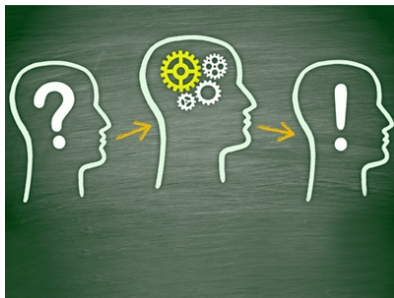
Het tegenovergestelde kan ook gebeuren waarbij sommige individuele kennis niet toegankelijk is tot de groep.



- **Factoren:** topologie van het communicatie (sociale) netwerk, vertrouwensrelaties, agenda/interesse van de groepsleden, mechanismen voor geloofsherziening etc.

Dit Brengt ons tot Verder Onderzoek Over:

- ▶ Individuele kennis versus (verschillende soorten van) groepskennis
- ▶ Hogere-orde redeneren en noties van rationaliteit
- ▶ Verschillende attitudes (kennis, geloof, voorkeuren, intenties etc)
- ▶ Beïnvloeding van opinies en de voorplanting van kennis in verschillende soorten sociale netwerken



Vragen?

- ▶ 'Reasoning about Knowledge' door Ronald Fagin, Joseph Y. Halpern, Yoram Moses en Moshe Vardi, MIT Press 1995
- ▶ 'Logical Dynamics of Information and Interaction' door Johan van Benthem, Cambridge UP, 2011.
- ▶ 'Epistemic logic and information update'. door A. Baltag, H. P. van Ditmarsch en L.S. Moss. In Handbook of Philosophy of Information, Elsevier, 2008.
- ▶ Referenties naar het werk van de leden van de 'Amsterdam Dynamics Group': <http://events.illc.uva.nl/dg/>