INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

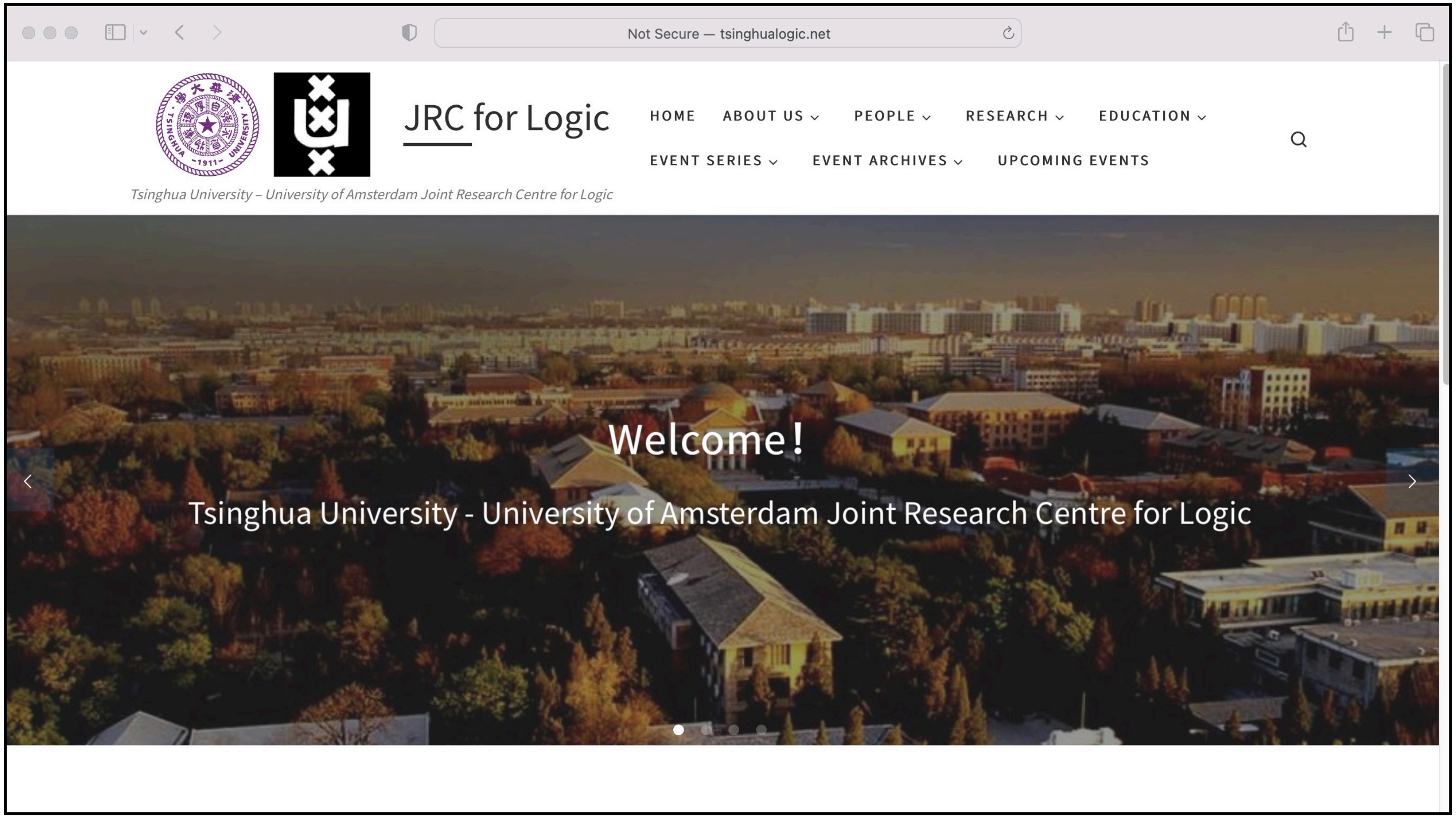
FROM COMPUTATION TO AGENCY:
WHAT, HOW, AND WHO

Johan van Benthem

14 December 2022

Logic and Thinking course, Tsinghua University

Institute for Logic,
Language and Computation



JRC for Logic

Tsinghua University – University of Amsterdam Joint Research Centre for Logic

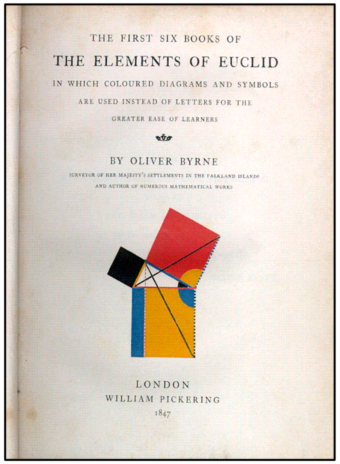
Welcome!

Tsinghua University - University of Amsterdam Joint Research Centre for Logic

Since Antiquity: Philosophy, Mathematics

Aristotle Logic, knowledge, argumentation

Euclid Proofs and constructions/algorithms

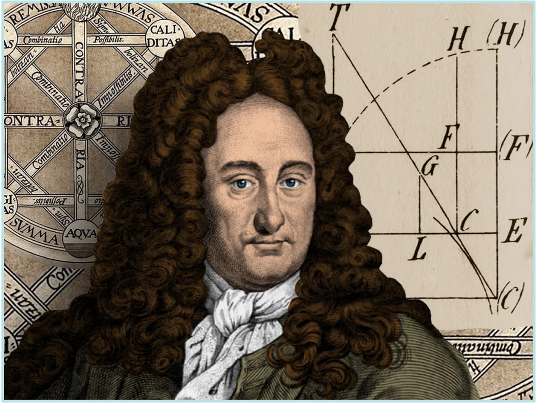


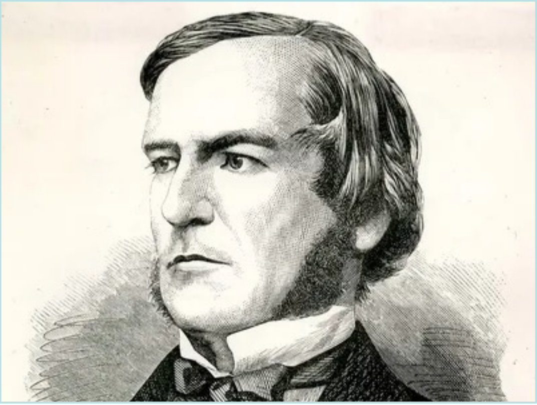
THE FIRST SIX BOOKS OF
THE ELEMENTS OF EUCLID
IN WHICH COLOURED DIAGRAMS AND SYMBOLS
ARE USED INSTEAD OF LETTERS FOR THE
GREATER EASE OF LEARNERS

BY OLIVER BYRNE
SURVEYOR OF HER MAJESTY'S SETTLEMENTS IN THE FALKLAND ISLANDS
AND AUTHOR OF NUMEROUS MATHEMATICAL WORKS

LONDON
WILLIAM PICKERING
1847

Add Mechanical Computation



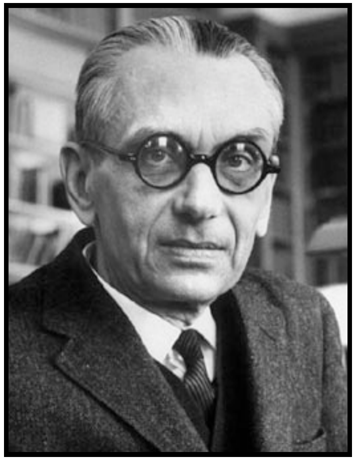


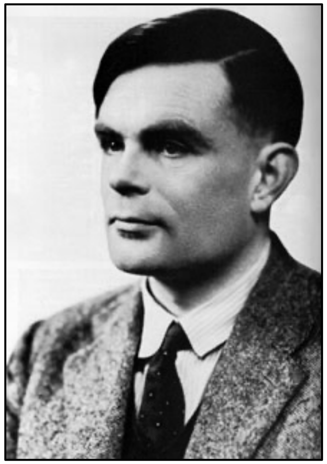
I

The Foundational Era

Institute for Logic,
Language and Computation

Logic, Proof, Computation





1930s The Grand Foundational Results

Kurt Gödel

What can be proved, but: Incompleteness theorems

Alan Turing

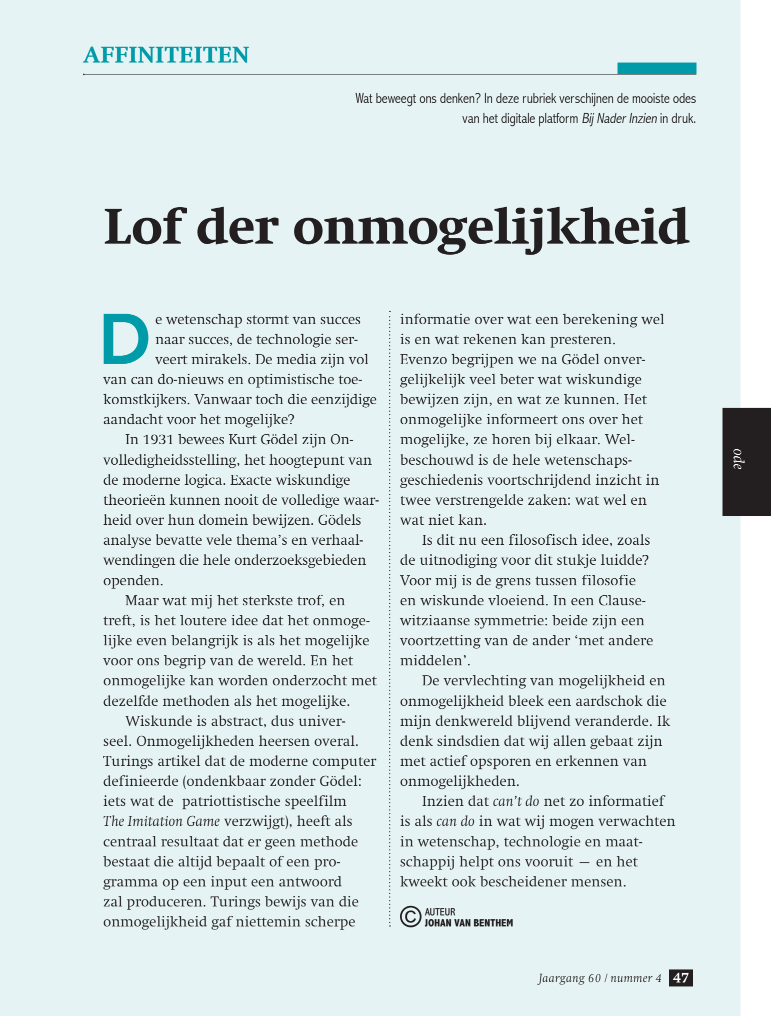
What can be computed, but: Uncomputable problems

Theoretical, yet turned out to be highly practical

Impossibilities requires deep study of possibilities

Institute for Logic, Language and Computation

In Praise of Impossibility



AFFINITEITEN

Wat beweegt ons denken? In deze rubriek verschijnen de mooiste odes van het digitale platform *Bij Nader Inzien* in druk.

Lof der onmogelijkheid

De wetenschap stormt van succes naar succes, de technologie serveert mirakels. De media zijn vol van can do-nieuws en optimistische toekomstkijkers. Vanwaar toch die eenzijdige aandacht voor het mogelijke?

In 1931 bewees Kurt Gödel zijn Onvolledigheidsstelling, het hoogtepunt van de moderne logica. Exacte wiskundige theorieën kunnen nooit de volledige waarheid over hun domein bewijzen. Gödels analyse bevatte vele thema's en verhaalwendingen die hele onderzoeksgebieden openden.

Maar wat mij het sterkste trof, en treft, is het loutere idee dat het onmogelijke even belangrijk is als het mogelijke voor ons begrip van de wereld. En het onmogelijke kan worden onderzocht met dezelfde methoden als het mogelijke.

Wiskunde is abstract, dus universeel. Onmogelijkheden heersen overal. Turings artikel dat de moderne computer definieerde (ondenkbaar zonder Gödel: iets wat de patriottistische speelfilm *The Imitation Game* verzwijgt), heeft als centraal resultaat dat er geen methode bestaat die altijd bepaalt of een programma op een input een antwoord zal produceren. Turings bewijs van die onmogelijkheid gaf niettemin scherpe informatie over wat een berekening wel is en wat rekenen kan presteren. Evenzo begrijpen we na Gödel onvergelijkelijk veel beter wat wiskundige bewijzen zijn, en wat ze kunnen. Het onmogelijke informeert ons over het mogelijke, ze horen bij elkaar. Welbeschouwd is de hele wetenschapsgeschiedenis voortschrijdend inzicht in twee verstrengelde zaken: wat wel en wat niet kan.

Is dit nu een filosofisch idee, zoals de uitnodiging voor dit stukje luidde? Voor mij is de grens tussen filosofie en wiskunde vloeiend. In een Clausewitziaanse symmetrie: beide zijn een voortzetting van de ander 'met andere middelen'.

De vervlechting van mogelijkheid en onmogelijkheid bleek een aardschok die mijn denkwereld blijvend veranderde. Ik denk sindsdien dat wij allen gebaat zijn met actief opsporen en erkennen van onmogelijkheden.

Inzien dat *can't do* net zo informatief is als *can do* in wat wij mogen verwachten in wetenschap, technologie en maatschappij helpt ons vooruit — en het kweekt ook bescheidener mensen.

© AUTEUR
JOHAN VAN BENTHEM

ode

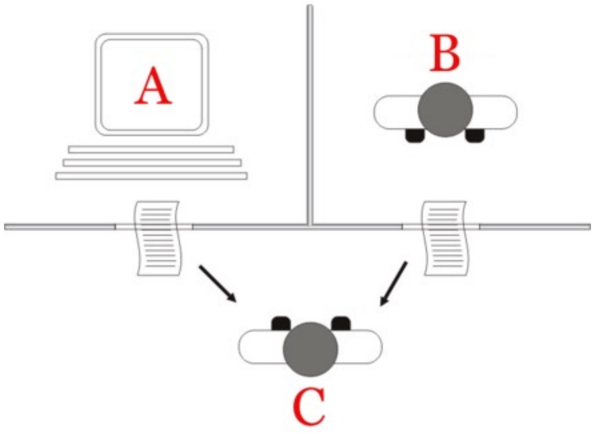
Jaargang 60 / nummer 4 47

The Turing Test (1950)

Turing machines well-chosen abstraction of human capabilities

In principle applicable to many intelligent tasks

Scenario Can we tell a computer apart from a human in natural conversation?



Benchmark for broader field of Artificial Intelligence

Replacement as a goal, or mixed systems?

II

Computer Science Themes

How to Make It Work: General Themes in CS

Machine building tradition 19th century to 1930s (Zuse)

von Neumann architecture 1940s

programming languages 1950s

program correctness 1960s

How you compute, ever more important theme

Computing produces not just output, but **behavior**

Distributed computing: societies 1980s

Finite and **infinite computation** 1990s

many of these themes are entangled with logic

III

Case Study:

Logics of Programs and Actions

1960s - 1970s Correctness and Program Logics

Correctness concerns:

how can we be sure a program

does what it is supposed to do?

prove for given program in suitable logic system

many approaches to achieving such goals

do one particular example here:

dynamic logics for imperative programs

Program Semantics: State Transitions

Imperative program

usual logical formulas describe what **is**

Here: instructions to **do**: $x := 3$, $x := y + 1$, $x: = x – 2$

variables: addresses, registers

computer **memory**: assignment of values to variables

just like assignments for standard **first-order logic**

Semantics: model **M** with state transitions

$\mathbf{M}, \mathbf{s}, \mathbf{t} \models \pi$ some successful execution of program π in **M** starts in **s** and ends in **t**

Program Syntax

atomic instructions work as follows

M, s, t |= x := T iff **t** = **s** $[x := [[T]]^{M, s}]$

program constructions

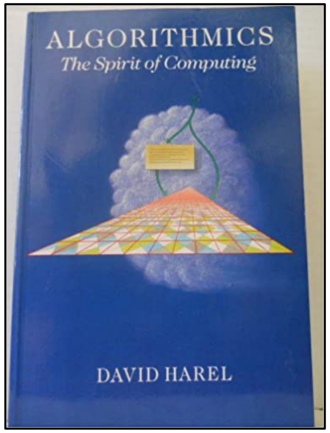
sequential **composition** $\pi 1$; $\pi 2$

conditional choice IF φ THEN $\pi 1$ ELSE $\pi 2$

iteration WHILE φ DO π

Also general structures in any action:

program structure in **cooking**! Also ||



Hoare Calculus

{P} S {Q} if program S starts in a state where **precondition** P is true,

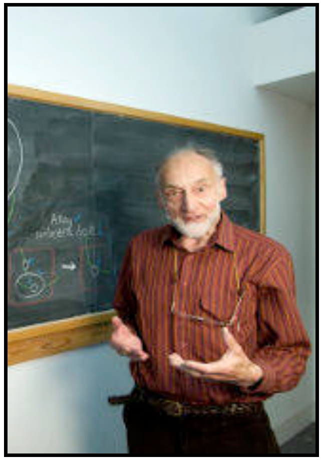
each succesful execution ends in a state with **postcondition** Q true

$$\frac{\{P\}\,S\,\{A\} \qquad\qquad \{A\}\,T\,\{R\}}{\{P\}\,S\,;T\,\{R\}}$$

$$\frac{\{P\,\&\,E\}\,S\,\{Q\} \qquad \{P\,\&\,\neg E\}\,T\,\{Q\}}{\{P\}\,IF\;E\;THEN\;S\;ELSE\;T\,\{Q\}}$$

$$\frac{\{I\}\,S\,\{I\}}{\{I\}\,WHILE\;E\;DO\;S\,\{I\,\&\,\neg E\}}$$

$$\frac{\{P\}\,S\,\{Q\}}{\{P\,\&\,A\}\,S\,\{Q\vee B\}}$$



Institute for Logic,
Language and Computation

General Dynamic Logic of Structured Action

PDL language with two components:

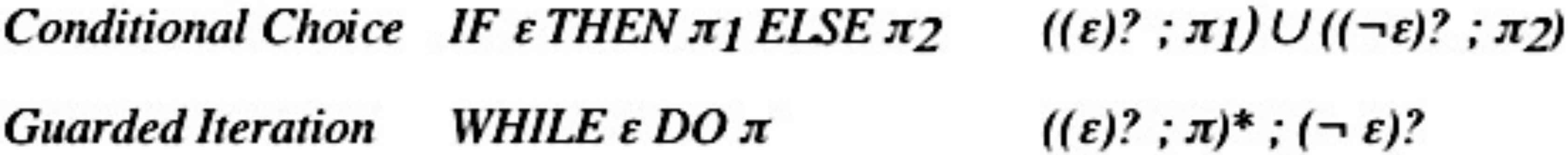
statements $p \mid \neg\varphi \mid \varphi \wedge \psi \mid \langle\pi\rangle\varphi$

programs $a \mid \pi1 \; ; \; \pi2 \mid \pi1 \cup \pi2 \mid \pi^* \mid (\varphi)?$

Can write regular expressions for actions

such as $(a \cup b)^*$, $(a^* \; ; \; b^*)^*$

Can define all the Hoare operations



Conditional Choice	*IF* ε *THEN* π_1 *ELSE* π_2	$((\varepsilon)? \; ; \; \pi_1) \cup ((\neg\varepsilon)? \; ; \; \pi_2)$
Guarded Iteration	*WHILE* ε *DO* π	$((\varepsilon)? \; ; \; \pi)^* \; ; \; (\neg\, \varepsilon)?$

partly inspired by a philosophical book on modal logic

Institute for Logic, Language and Computation

Some Key Reasoning Principles

- All principles of the minimal modal logic for all modalities $[\pi]$
- Computation rules for decomposing program structure:

$$<\pi_1;\pi_2>\phi \leftrightarrow <\pi_1><\pi_2>\phi$$

$$<\pi_1 \cup \pi_2>\phi \leftrightarrow <\pi_1>\phi \vee <\pi_2>\phi$$

$$<\phi?>\psi \leftrightarrow \phi \wedge \psi$$

$$<\pi^*>\phi \leftrightarrow \phi \vee <\pi><\pi^*>\phi$$

- The Induction Axiom $(\phi \,\&\, [\pi^*](\phi \rightarrow [\pi]\phi)) \rightarrow [\pi^*]\phi$.

Uses of PDL: Broader Influences

Program semantics

but challenge of parallel computation ||

Mathematics: unify **abstract algebra** of regular expressions and **modal logic**

Abstract computability theory (on any data structures, not just the natural numbers)

but also crossed to other fields

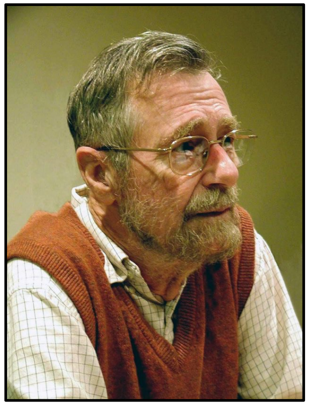
Philosophy of action

Epistemic notions: common knowledge is $[\ (\cup_{i \in G} \sim_i)^*\ ]\ \varphi$

Further Views of Logic and Programs

‘Structured programming’ methodology not prove correctness after the programming has been done, but work in tandem:

design programs together with logical correctness statements



Design programming languages closer to logic:

functional programming (LISP, Haskell, etc.)

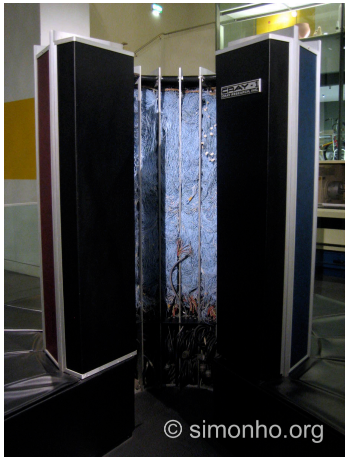
logic programming (PROLOG, etc.)

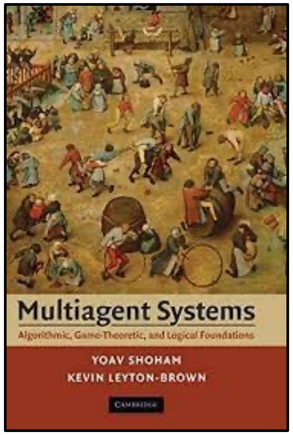
IV

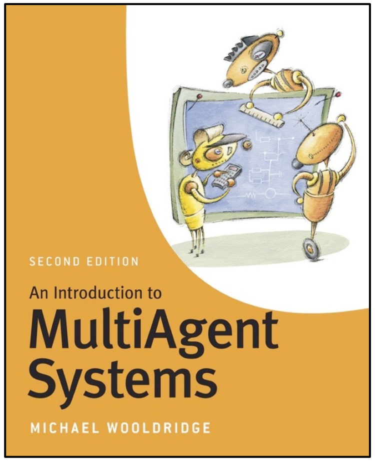
From Action to Agency

Institute for Logic,
Language and Computation

Modern Computation: From Action to Agency

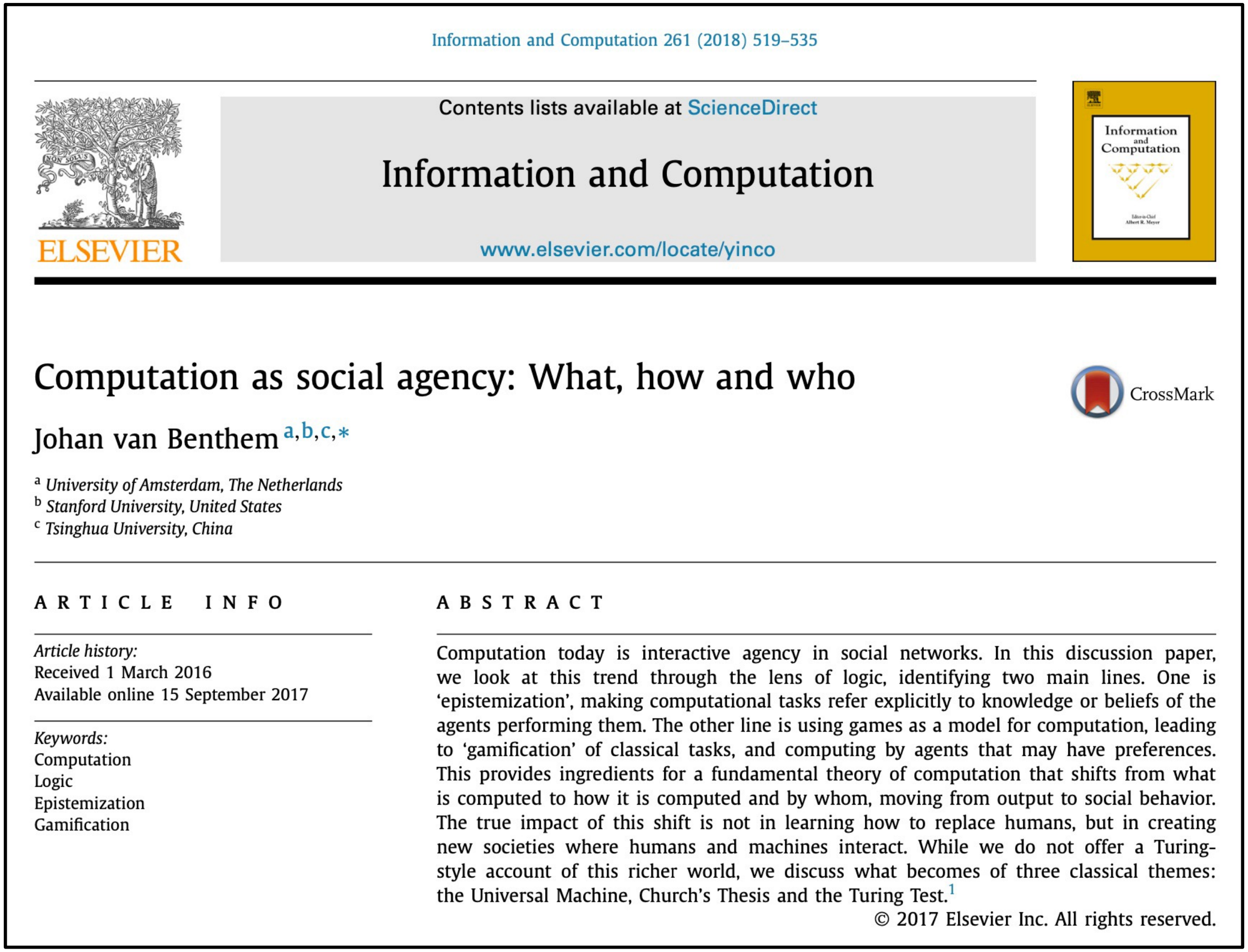



Multiagent Systems
Algorithmic, Game-Theoretic, and Logical Foundations
YOAV SHOHAM
KEVIN LEYTON-BROWN
CAMBRIDGE


SECOND EDITION
An Introduction to
MultiAgent Systems
MICHAEL WOOLDRIDGE

Institute for Logic, Language and Computation

What, How, Who



Information and Computation 261 (2018) 519–535

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Computation as social agency: What, how and who

Johan van Benthem[a,b,c,*]

[a] University of Amsterdam, The Netherlands
[b] Stanford University, United States
[c] Tsinghua University, China

ARTICLE INFO

Article history:
Received 1 March 2016
Available online 15 September 2017

Keywords:
Computation
Logic
Epistemization
Gamification

ABSTRACT

Computation today is interactive agency in social networks. In this discussion paper, we look at this trend through the lens of logic, identifying two main lines. One is ‘epistemization’, making computational tasks refer explicitly to knowledge or beliefs of the agents performing them. The other line is using games as a model for computation, leading to ‘gamification’ of classical tasks, and computing by agents that may have preferences. This provides ingredients for a fundamental theory of computation that shifts from what is computed to how it is computed and by whom, moving from output to social behavior. The true impact of this shift is not in learning how to replace humans, but in creating new societies where humans and machines interact. While we do not offer a Turing-style account of this richer world, we discuss what becomes of three classical themes: the Universal Machine, Church’s Thesis and the Turing Test.[1]

V

Information Dynamics

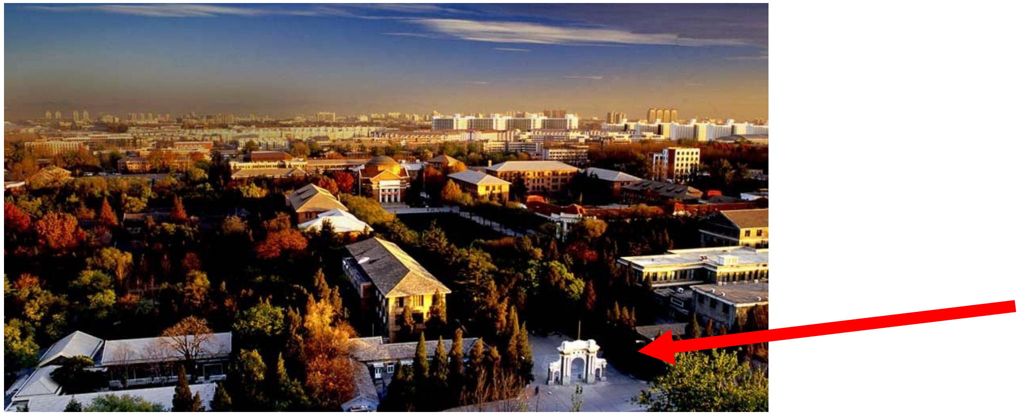
Information Dynamics

The Restaurant Three people order drinks: water, beer, wine. A new waiter comes with 3 glasses. There are 6 ways these could be distributed. Solved by **2 questions, 1 inference.**

知 闻 说 亲 *zhi wen shuo qin*

Knowledge comes from: hearing from others, proof, or experience.

Question and Answer



Q “Is this the Tsinghua Old Gate?

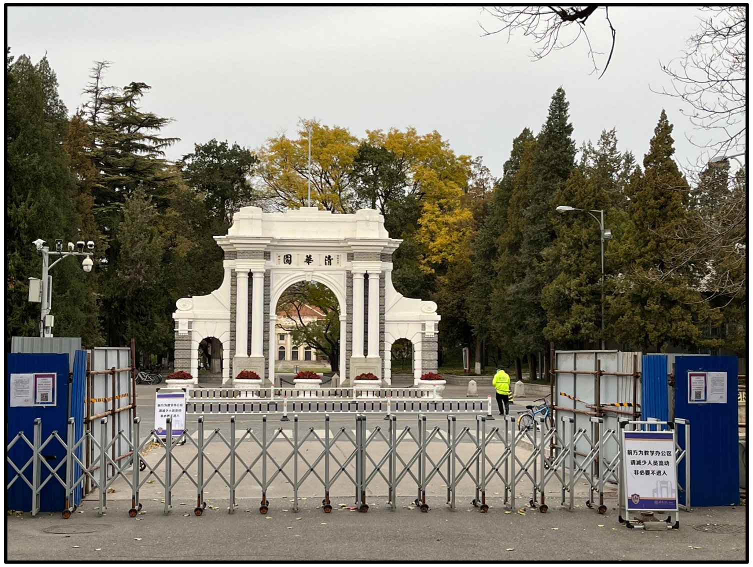
A “Yes”

Q conveys: does not know, thinks the other knows, wants to know

A creates knowledge of facts, about others, common knowledge.

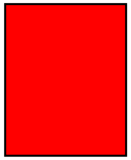
Institute for Logic, Language and Computation

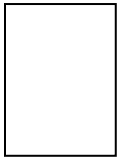
The Tsinghua Old Gate These Days

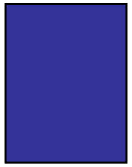


Three Cards

John, Mary, Paul get one card each







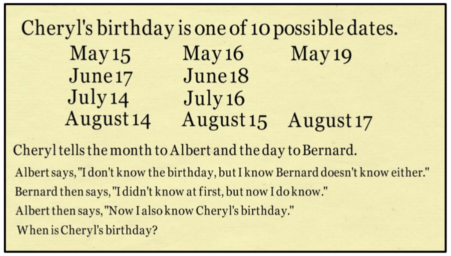
John **Red** Mary **White** Paul **Blue**

Mary asks John: **Do you have the blue card?**

Who knows the deal of the cards now?

John answers: **No**.

Who knows what now?



Cheryl's birthday is one of 10 possible dates.

May 15	May 16	May 19
June 17	June 18	
July 14	July 16	
August 14	August 15	August 17

Cheryl tells the month to Albert and the day to Bernard.

Albert says, "I don't know the birthday, but I know Bernard doesn't know either."

Bernard then says, "I didn't know at first, but now I do know."

Albert then says, "Now I also know Cheryl's birthday."

When is Cheryl's birthday?

Major Themes in One Small Scenario

information

knowledge

information update

knowledge change

multi-agent

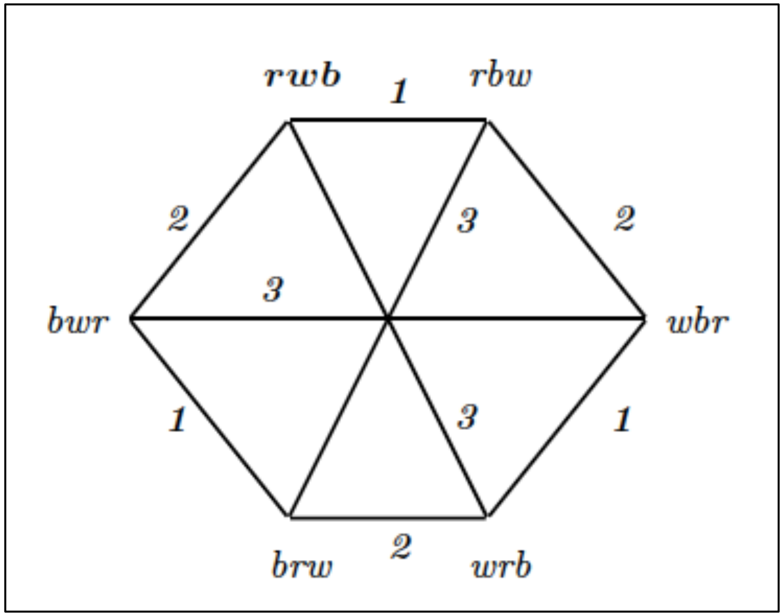
knowledge about facts, about others

communication

interaction, games

social groups

Information Diagram



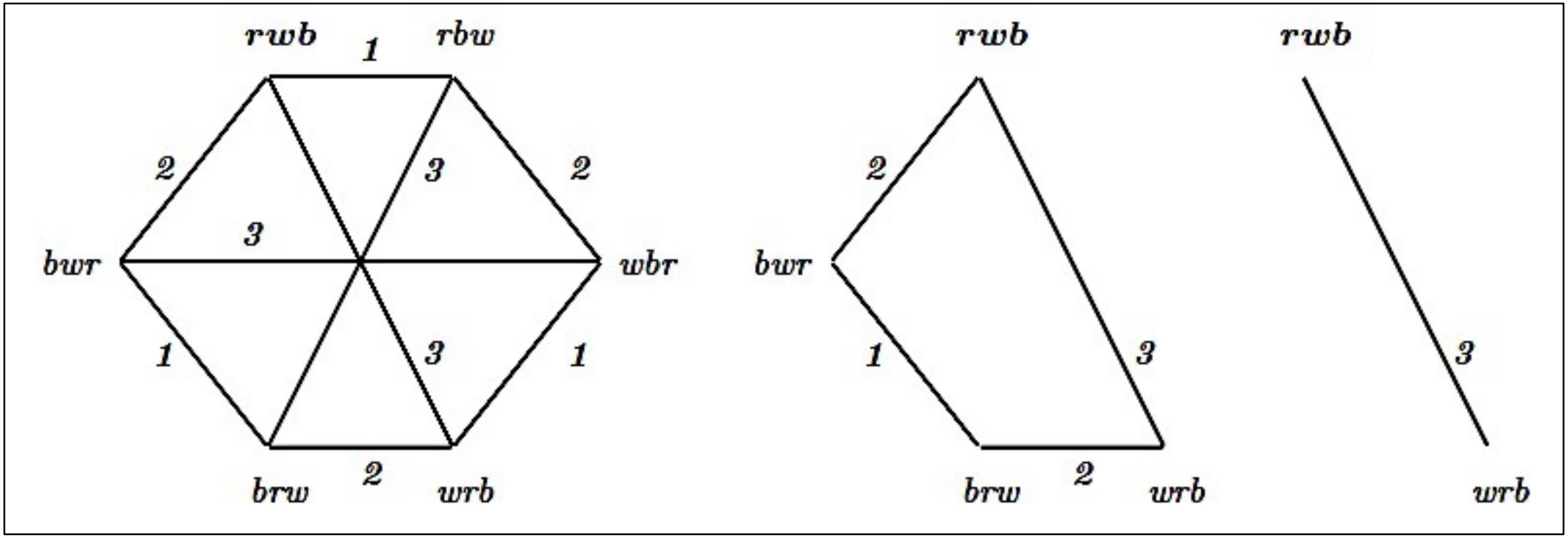
points ~ possible deals of cards

uncertainty lines ~ what agents can(not) observe

knowledge: what is true in all your current options

Institute for Logic, Language and Computation

CS Influence: Information Change and Updates



Final state John and Mary know the cards, Paul does not. But Paul knows that the others know, and in fact, this is common knowledge in the group

VI

Ideas From Philosophy Enter

Modeling Knowledge in Epistemic Logic

Enter the Philosophers: Modal Logic

Once the tools of modern mathematical logic were available philosophers started using them to study the laws of reasoning for fundamental notions in their field like necessity, possibility, time, knowledge, belief, obligation, and so on

Basic technical paradigm: **Modal Logic**

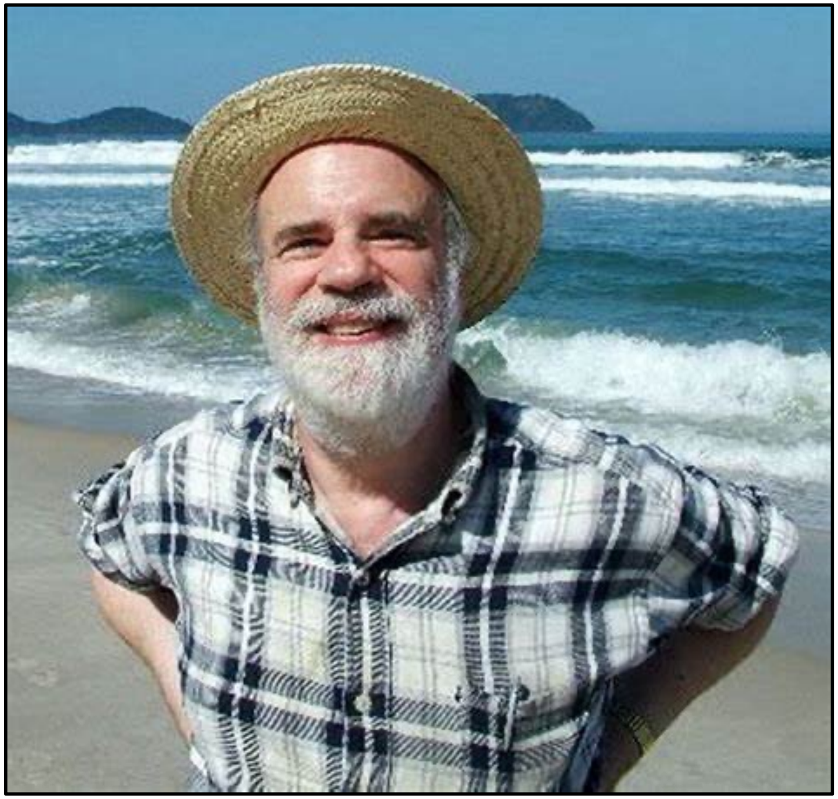
Considers truth of propositions in different **'possible worlds'** that are suitably connected by **'accessibility relations'**

In what follows we look at one such case: **knowledge**

But the earlier dynamic logic of **action** is another instance

Institute for Logic, Language and Computation

Saul Kripke



INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Knowledge, Formal and Natural Language

Language $\mathbf{p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_i\varphi}$

defined $\mathbf{<i>}\varphi := \neg\mathbf{K_i}\neg\varphi$ 'consider possible'

Question: "φ?" Answer: "Yes".

$\neg\mathbf{K_Q}\varphi$	**Q** does not know that φ
$\neg\mathbf{K_Q}\varphi \wedge \neg\mathbf{K_Q}\neg\varphi$	**Q** does not know whether φ
$\mathbf{K_Q(K_A}\varphi \vee \mathbf{K_A}\neg\varphi)$	**Q** knows that **A** knows whether φ
$\mathbf{<Q>(K_A}\varphi \vee \mathbf{K_A}\neg\varphi)$	…

afterwards: **common knowledge** $\mathbf{C_{\{Q, A\}}}\varphi$

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

Epistemic Models and Semantics

semantic information: range of options for the real world

(other important kinds of information also exist in logic)

epistemic model **M** = (W, $\{\sim_i\}_i$, V)

W **worlds/points**, epistemic **accessibility relations** $\sim_i$,

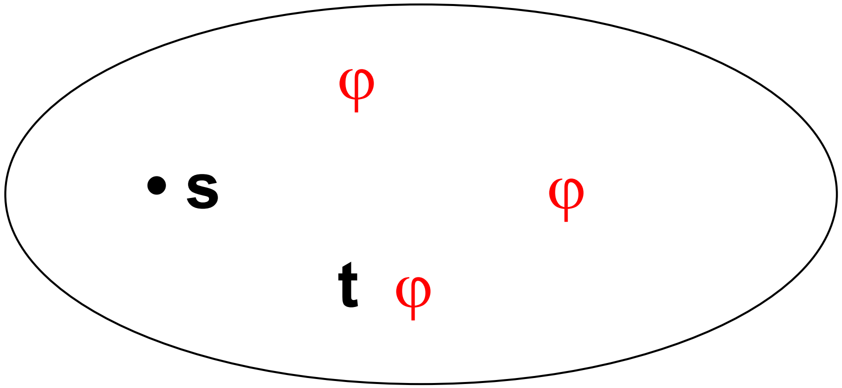
(for now: equivalence relation: reflexive, symmetric, transitive)

valuation V: truth values for proposition letters at worlds

truth definition **M**, s |= $K_i\varphi$ iff for all t $\sim_i$ s: **M**, t |= φ

Semantics Supports Helpful Pictures

knowledge: true according to my **semantic information**



Old Gate

for you, | not: for me

not Old Gate

truth definition $\mathbf{M}, s \models \mathbf{K_i}\varphi$ iff for all $t \sim_i s$: $\mathbf{M}, t \models \varphi$

Valid Principles

$\models \varphi$ φ is **valid**: φ true in all models at all points

logical consequence $\Phi \models \psi$: validity of implication $\&\Phi \rightarrow \psi$

valid

$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$ distribution

$K(\varphi \wedge \psi) \leftrightarrow (K\varphi \wedge K\psi)$

$K\varphi \rightarrow \varphi$ veridicality

$K\varphi \rightarrow KK\varphi$ positive introspection

$\neg K\varphi \rightarrow K\neg K\varphi$ negative introspection

Complete Multi-S5 Axiom System

1 all valid principles and inference rules of **propositional logic**

e.g., Modus Ponens from $\varphi, \varphi \rightarrow \psi$ infer ψ

2 principles of the **minimal modal logic** (K)

$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$, $<>\varphi \leftrightarrow \neg K\neg\varphi$

From already proved φ infer $K\varphi$ **Necessitation Rule**

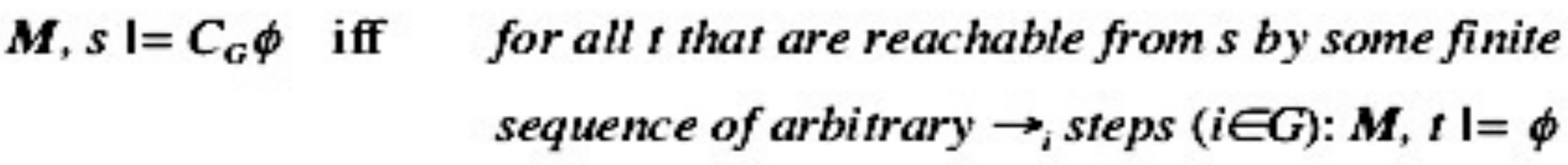
3 S5 axioms

$K\varphi \rightarrow \varphi$ (reflexivity) $K\varphi \rightarrow KK\varphi$ (transitivity)

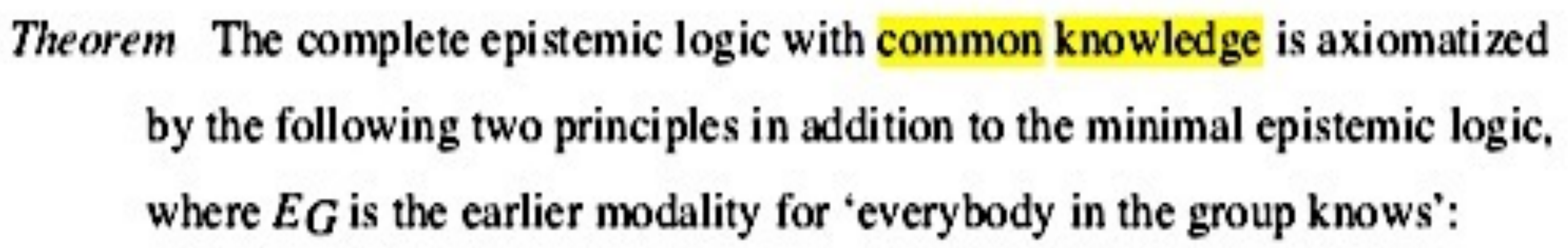
$\neg\varphi \rightarrow K\neg K\varphi$ (symmetry)

Digression: Common Knowledge in Groups

Notation everybody knows $E_G\varphi \neq C_G\varphi$ **common knowledge**



$M, s \models C_G\phi$ iff *for all t that are reachable from s by some finite sequence of arbitrary $\rightarrow_i$ steps ($i \in G$): $M, t \models \phi$*



Theorem The complete epistemic logic with common knowledge is axiomatized by the following two principles in addition to the minimal epistemic logic, where E_G is the earlier modality for 'everybody in the group knows':

$C_G\,\phi \leftrightarrow (\phi \wedge E_G\,C_G\,\phi)$ *Fixed-Point Axiom*

$(\phi \wedge C_G\,(\phi \rightarrow E_G\,\phi)) \rightarrow C_G\,\phi$ *Induction Axiom*

Outlook, Philosophy

knowledge is more than information

S5 axioms unreasonable idealization for knowledge?

omniscience, (positive, negative) **introspection**

Richer views of knowledge proposed by philosophers

justified true belief (Plato)

information in the (most) **relevant** worlds (Dretske)

stable belief under true new information …

all these richer notions are studied in logic today

Criticisms of Logical Omniscience

$K(\varphi \to \psi) \to (K\varphi \to K\psi)$ Do we know all known consequences of what we know? Assuming we know the basic laws of logic, we would then know all logical consequences of what we know.

Unlikely!

Still, there are some defenses:

The modality K models ascribed\implicit knowledge

Current topic: model the more fine-grained information provided by logical inference, and by computation …

Student Question

About the Three Cards:

Is not there a cognitive or computational **cost** to Paul of arriving at the knowledge he gets from the scenario?

Yes, but this requires us to modeling acts of inference and other information updates more finely, maybe bringing in notions like (working) **memory** and **attention** from the psychological literature

So we get to a connection with **empirical sciences**

Outlook, Cognitive Science

Indeed, **epistemic logic meets cognitive psychology**

Our logical system is an idealized model

What happens in reality?

reasoning about what others know:

Theory of Mind

growing ability in children

how many levels of iteration can humans achieve?

experiments by logicians, psychologists, game theorists

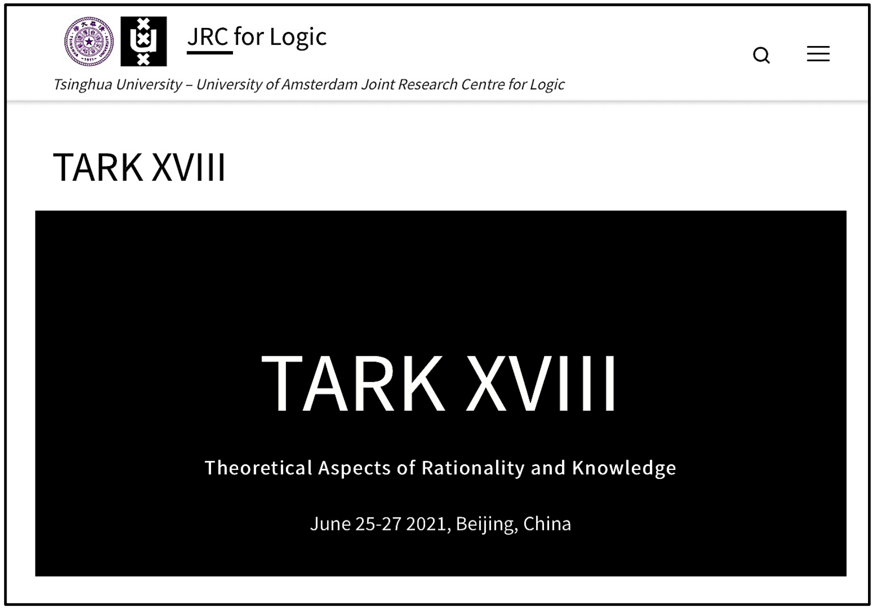
Institute for Logic, Language and Computation

Epistemic Logic: Emerged Across Disciplines

Philosophy 1960s Sociology 1960s

Economics 1970s Computer science 1980s

Cognitive psychology 1990s Back to philosophy 200s



JRC for Logic

Tsinghua University – University of Amsterdam Joint Research Centre for Logic

TARK XVIII

TARK XVIII

Theoretical Aspects of Rationality and Knowledge

June 25-27 2021, Beijing, China

Challenging But Exciting Triangle

Logic, Computer Science, Cognitive Science

Logic produces largely **theoretical analyses** that **we construct**

[Although this has led to concrete computing machines!]

Even CS is still about things we can **design** to suit our purposes:

programming languages and computing devices

Natural language and ordinary reasoning **emerged** in evolutionary **history**, not designed by us, not easy to change/’improve’

but this is also exciting, and generates many research questions

VII

Dynamic-Epistemic Logic

Mixing Philosophy and CS Themes

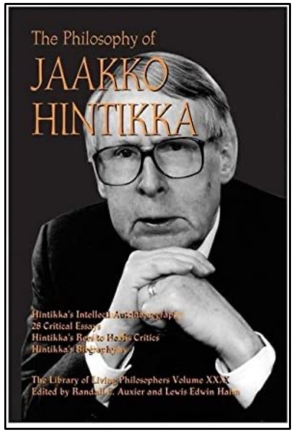
Institute for Logic, Language and Computation

Dynamic Logics of Information

Informational acts satisfy precise logical axioms describing what happens to one's existing knowledge through an informational event:

$$[!\varphi]K_i\psi \leftrightarrow (\varphi \rightarrow K_i(\varphi \rightarrow [!\varphi]\psi))$$

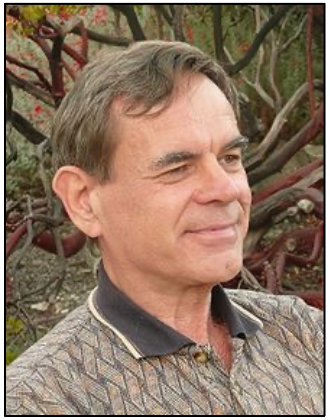
Cooperation of computer science, logic and philosophy.



The Philosophy of JAAKKO HINTIKKA

Hintikka's Intellectual Autobiography
28 Critical Essays
Hintikka's Replies to His Critics
Hintikka's Bibliography

The Library of Living Philosophers Volume XXX
Edited by Randall E. Auxier and Lewis Edwin Hahn



STUDIES IN LOGIC AND PRACTICAL REASONING

Handbook of Modal Logic

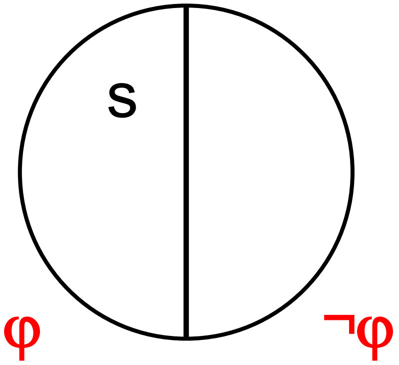
Update with True Information, Picture

epistemic model M, s ~ group information state

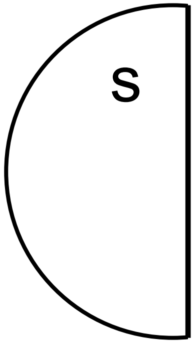
actual world s (seen as such by the external modeler)

learning that φ is true **eliminates all ¬φ-worlds**

from **M**



to **M**|φ



hard information

Update with True Information, Words

event $!\varphi$ of receiving true new information φ

φ is true at actual world, perhaps at others

public **announcement** (or public silence …),

or just as well public **observation** (needs no words)

update to **submodel** $\mathbf{M}|\varphi$ with domain $\{ t \in \mathbf{M} \mid \mathbf{M}, t \models \varphi \}$

just simplest case: many other informational events

Public Announcement Language & Semantics

simple pilot system for richer dynamic-epistemic logics

PAL **language**

grammar of multi-agent epistemic logic plus **$[!\varphi]\psi$**

PAL **semantics**

$\mathbf{M}, s \models [!\varphi]\psi$ iff if $\mathbf{M}, s \models \varphi$, then $\mathbf{M}|\varphi, s \models \psi$

PAL Axiom System

In technical settings, we often write modal box □ for **K**

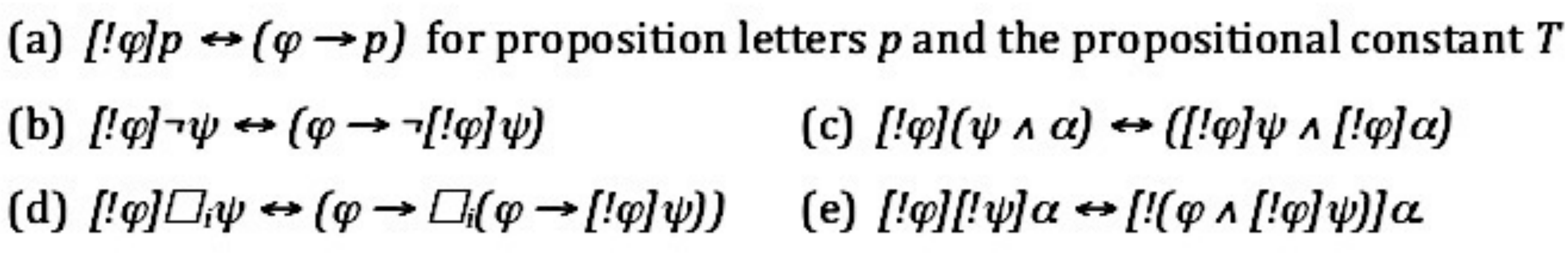
axiom system for PAL

1 all proof principles of multi-agent S5

2 all proof principles of basic modal logic for **$[!\varphi]$**

plus RE: if $\vdash \varphi \leftrightarrow \psi$, then $\vdash \alpha(\varphi) \leftrightarrow \alpha(\psi)$

3 recursion axioms for postconditions ψ in **$[!\varphi]\psi$**:



(a) $[!\varphi]p \leftrightarrow (\varphi \to p)$ for proposition letters p and the propositional constant T

(b) $[!\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[!\varphi]\psi)$

(c) $[!\varphi](\psi \wedge \alpha) \leftrightarrow ([!\varphi]\psi \wedge [!\varphi]\alpha)$

(d) $[!\varphi]\Box_i\psi \leftrightarrow (\varphi \to \Box_i(\varphi \to [!\varphi]\psi))$

(e) $[!\varphi][!\psi]\alpha \leftrightarrow [!(\varphi \wedge [!\varphi]\psi)]\alpha$

Institute for Logic,
Language and Computation

Quick Look at Two Axioms

(d) $[!\varphi]\Box_i\psi \leftrightarrow (\varphi \rightarrow \Box_i(\varphi \rightarrow [!\varphi]\psi))$ (e) $[!\varphi][!\psi]\alpha \leftrightarrow [!(\varphi \wedge [!\varphi]\psi)]\alpha$

(d) Knowledge that results from true new information can be analyzed as conditional knowledge that the agent has beforehand about the effect of the new information

(e) The dynamic effect of two consecutive information updates can be simulated by one update with respect to a suitable conjunction involving a dynamic modality [this law explains many subtle phenomena from the earlier literature]

VIII

Epistemization

Epistemization: A Few Examples

Epistemic specifications

Robot stops: when it knows it has reached the goal.

Knowledge programs

Act when you know that the condition holds.

Know your program/plan

More abstract notion, not fully explained

More general: ethics

Must help when (I know that) you are sick.

Must know whether you are sick?

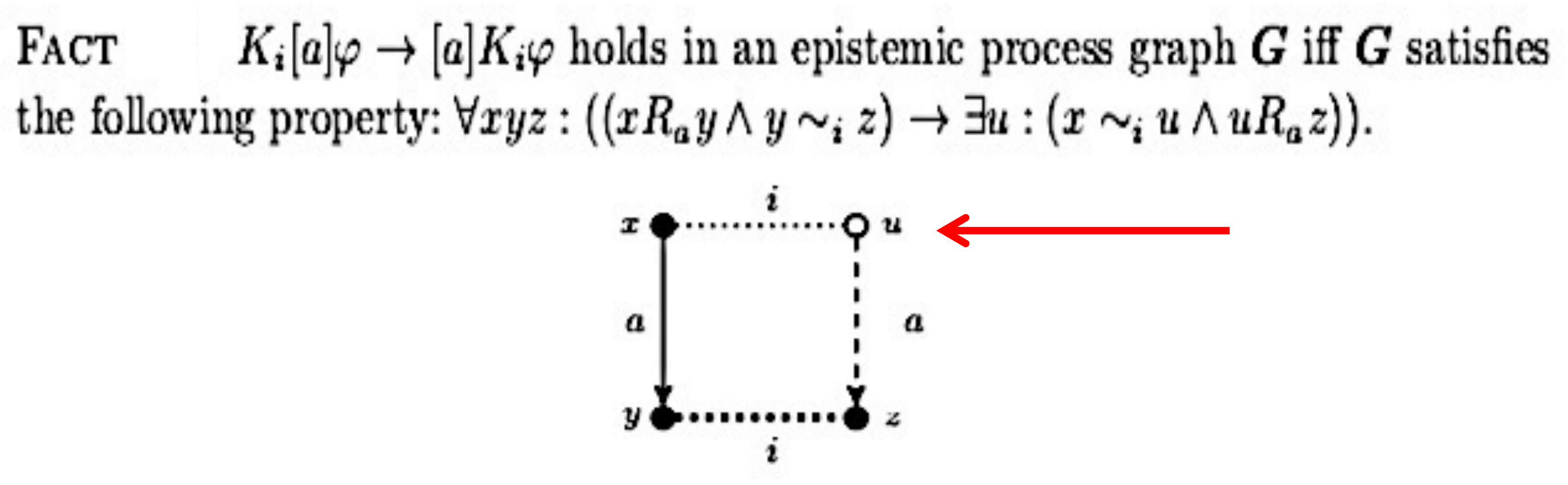
General issues Then: Action and Knowledge

Standard logical principles for knowledge and action separately

$$\mathbf{K[a]\varphi \rightarrow [a]K\varphi}$$

only valid for **epistemically transparent** actions

(I know that, after drinking, I behave stupidly, but unfortunately…)



FACT $K_i[a]\varphi \to [a]K_i\varphi$ holds in an epistemic process graph $\boldsymbol{G}$ iff $\boldsymbol{G}$ satisfies the following property: $\forall xyz : ((xR_a y \wedge y \sim_i z) \to \exists u : (x \sim_i u \wedge uR_a z))$.

Laws Embody Assumptions About Agents!

$K[a]\varphi \to [a]K\varphi$ alternative interpretation in game theory **Perfect Recall**

uncertainty after action can only come from uncertainty you had before

Attention to agency again: **what sort of agents** are we studying?

Also wıtn the putative converse **$[a]K\varphi \to K[a]\varphi$.** Is this valid?

Expresses: **No Learning, No Miracles**

one can only gain new knowledge by

observing new events with known preconditions

holds for simple observable actions, not epistemic **$!\varphi$**

IX

Gamification

Institute for Logic, Language and Computation

Interactive Computing

Much like a game with different actors with different information, different abilities (think: humans, machines), and different goals

if goals are aligned, we get **cooperative games**

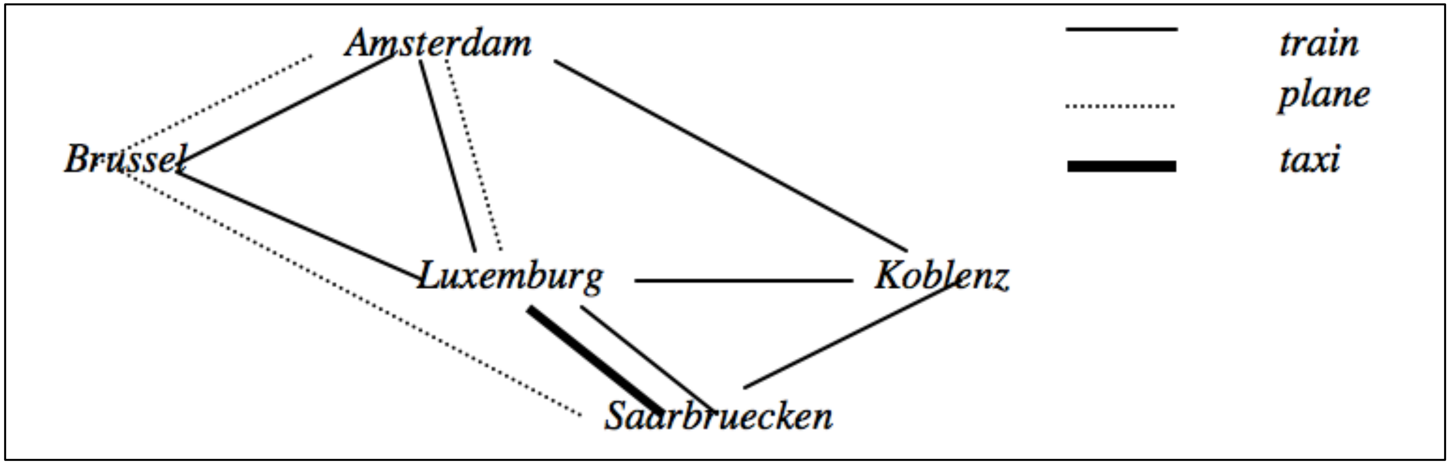
If goals are at odds, we get **competitive games**

Now in addition to **actions**, we must think about **strategies**

Here is one simple scenario testing the robustness of standard computing tasks:

Gamification: Changing Environments

Sabotaging algorithmic tasks. Agents should cope.



Strategies are the interactive solution that we need.

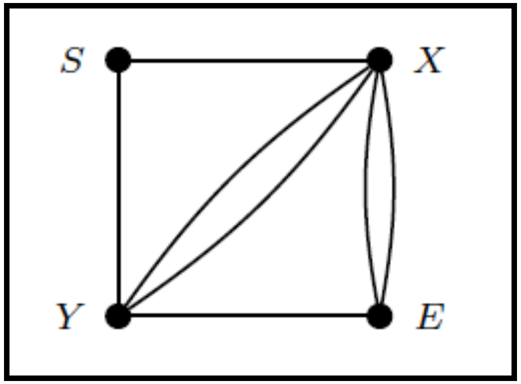
Also get: **new logic** with model changing in evaluation.

Complexity jump:

model checking *Pspace-complete*, logic *undecidable*.

Gamification: Learning

Teaching Game



Try to escape

Zermelo

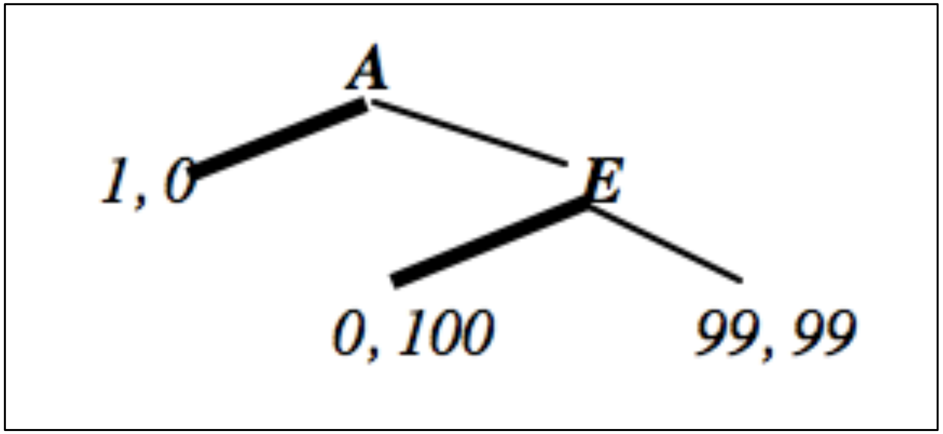
NEMO children

Background: **Zermelo’s Theorem**

Current uses: cognitive experiments.

The Surplus of Games: Goals, Preference

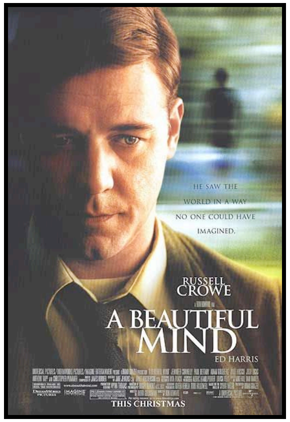
Agency and games involve much more than information processing, and pure strategic interaction:

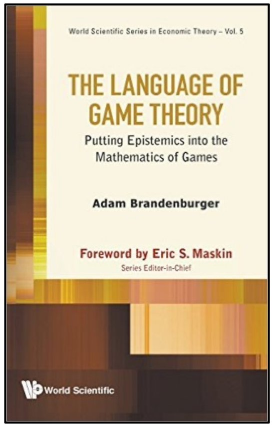


Reasoning to the bold-face (or any other) strategy involves knowledge, action, but also **goals**, **preferences** and **beliefs**.

Institute for Logic, Language and Computation

Epistemic Game Theory





World Scientific Series in Economic Theory – Vol. 5

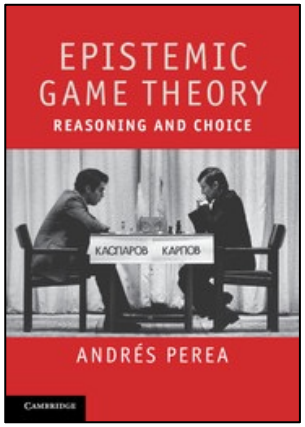
THE LANGUAGE OF GAME THEORY

Putting Epistemics into the Mathematics of Games

Adam Brandenburger

Foreword by Eric S. Maskin
Series Editor-in-Chief

World Scientific



EPISTEMIC GAME THEORY

REASONING AND CHOICE

ANDRÉS PEREA

CAMBRIDGE

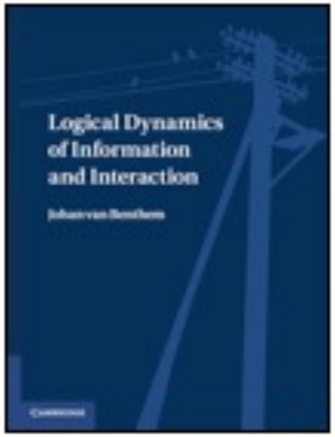
New look at many game-theoretic themes, such as information-dynamic scenarios for game solution.

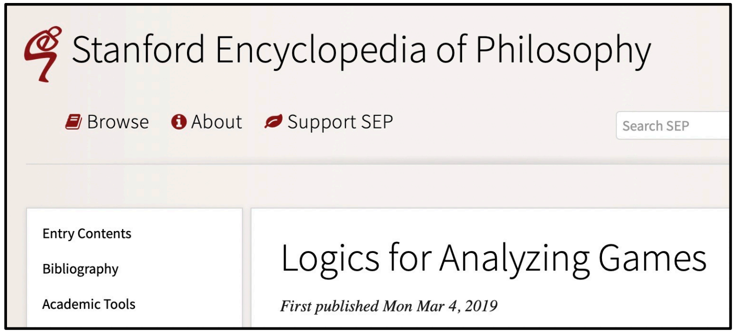
Even the very notion of game identity becomes **agent-dependent**, because of players’ preferences, beliefs, abilities, and this also shows in the logic.

INSTITUTE FOR LOGIC,
LANGUAGE AND COMPUTATION

Logical Studies of Agency in Games

Logic + Games + Agency = Theory of Play



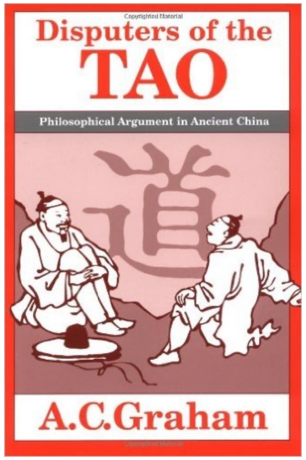

Stanford Encyclopedia of Philosophy

Logics for Analyzing Games

First published Mon Mar 4, 2019

Argumentation Games

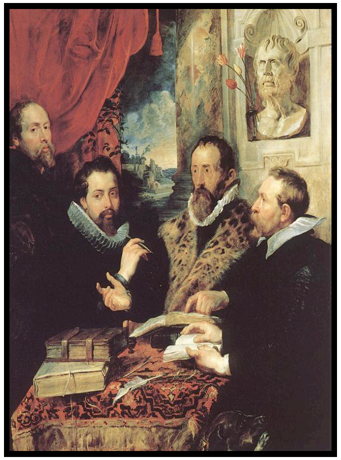
Argumentation is also a game


Copyrighted Material

Disputers of the
TAO

Philosophical Argument in Ancient China

A.C.Graham



Lorenzen dialogue games

Game semantics for computing

Back to where it all began for logic?

X

The Temporal Long Term

From Termination to Infinite Behavior

Single informational actions form longer **histories** of games.

Strategies are plans enforcing histories, obeying logical laws:

$$\{G, i\}\varphi \vee \{G, j\}\ \mathit{Always}\ \neg\{G, i\}\varphi$$

Histories, finite or infinite, show new structure of their own: logics of limit behavior that can only be seen over time.

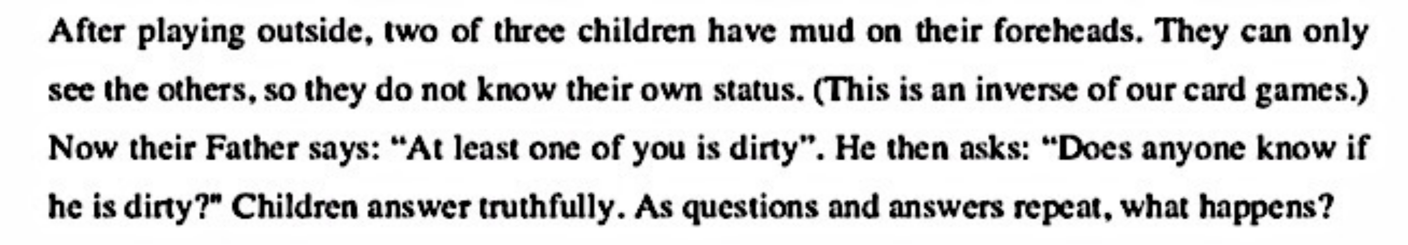
Surprising scenarios: self-fulfilling, self-refuting assertions.

Same in formal learning theory with limit learning.

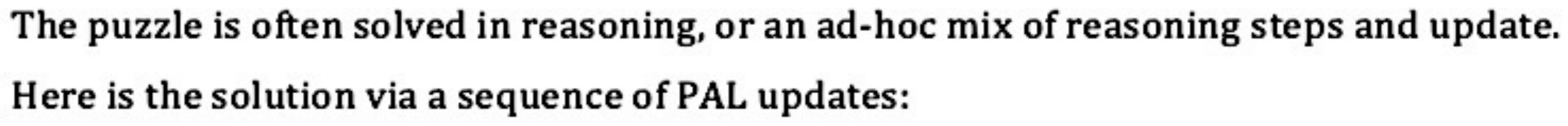
CS: infinite streams and co-algebra || Evolutionary game theory.

Current topic: Interface logics of agency - dynamical systems

Muddy Children, PAL Programs

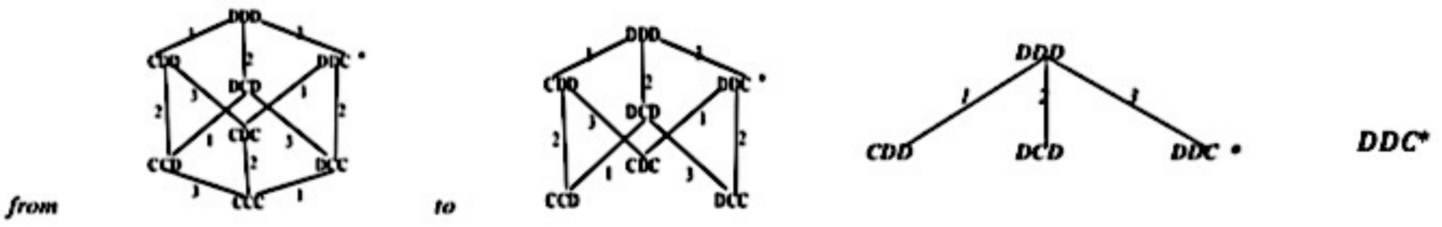


After playing outside, two of three children have mud on their foreheads. They can only see the others, so they do not know their own status. (This is an inverse of our card games.) Now their Father says: "At least one of you is dirty". He then asks: "Does anyone know if he is dirty?" Children answer truthfully. As questions and answers repeat, what happens?



The puzzle is often solved in reasoning, or an ad-hoc mix of reasoning steps and update.

Here is the solution via a sequence of PAL updates:



program structure in communication

WHILE you don't know your status **!**"don't know" ; **!**"I know"

; **IF THEN ELSE** **WHILE DO** even **||**

PAL*, Logic of Programmed Conversation

add **regular program algebra** to PAL

$!\varphi; !\psi$ (composition) $!\varphi \cup !\psi$ (choice) $(!\varphi)^*$ (iteration)

Theorem Validity in PAL* is non-axiomatizable.

Proof SAT for PAL* can encode the Recurrent Tiling Problem.

Open problem May be wrong modeling, better use

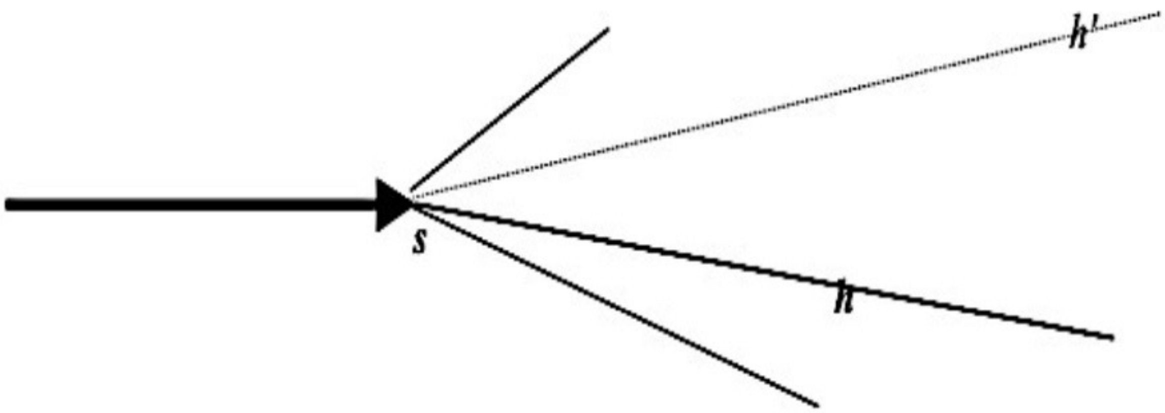
protocol models restricting available update histories:

simple logics of conversation analysis/planning?

Long-Term Dynamics and Limit Behavior

Infinite histories of (update) computation

can be described in epistemic-temporal logics



Both Ways Again: From Logic to CS

Protocol analysis in computer science (since 1980s)

Agents exchanging information, while also maintaining privacy

studied using epistemic-temporal logics

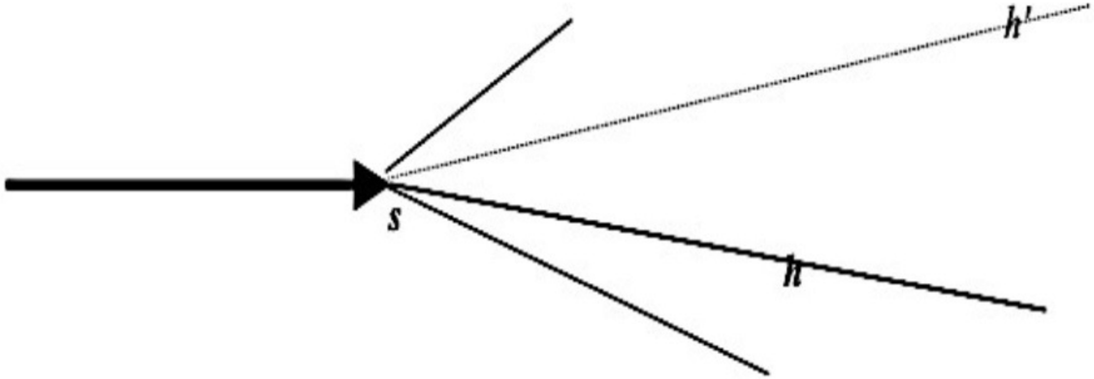
Ronald Fagin
Joseph Y. Halpern
Yoram Moses
Moshe Y. Vardi

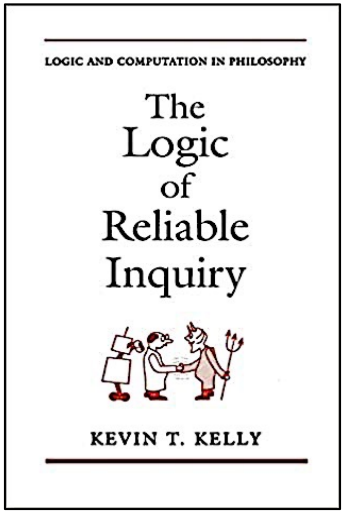
Reasoning About Knowledge

Philosophy: Formal Learning Theory

Modeling inquiry in science, but also daily practice cases

Infinite histories of (update) events, methods from Topology




LOGIC AND COMPUTATION IN PHILOSOPHY

The Logic of Reliable Inquiry

KEVIN T. KELLY

Formal learning theory

What we come to know/believe over time

needs temporal logics that extend our dynamic-epistemic logics

XI

AI Today

Classical AI

Many interactions with logic and philosophy

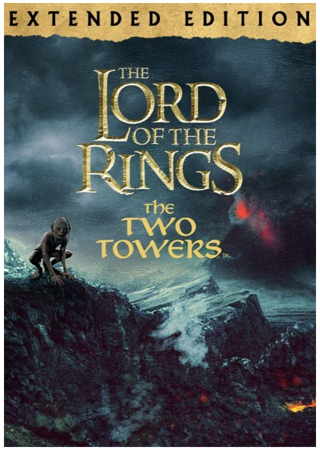
Automated theorem proving

Expert systems

Knowledge bases

Multi-agent systems

AI = ML? The Challenge of Machine Learning



Just blind emergence of conclusions?

no representation or inference

Logic and judgment: not necessary, not even possible?

The mind has lost out to the brain?

Beyond Shallow Scare Rhetoric

Leitgeb, van Lambalgen neural nets ~ nonmonotonic logics

Icard et al. conditional logics match the causal hierarchy

Grohe et al. modal logics classify types of learning systems

recent findings State spaces of ML systems reduce to well-known logical models under natural information reductions

lots of exciting research at Logic - ML interface today

too early for an integration, but will happen once the dust settles

XII

Conclusion

Summary

Logic and computation form a natural historical unity,

both in classical style and in modern agency styles

The classical foundational results from the 1930s

still govern what is (im-)possible

The interface of computing as agency and human behavior is

a constant source of new logical perspectives and theories

logic, philosophy, computer science, AI and cognitive science meet

PS Student Question: Probability

How does all the above logical analysis fit with **probability**?

Dynamic-epistemic logics interface with **Bayesian epistemology**

Combined update logics that integrate qualitative logic parts

with quantitative probabilistic parts

But much more generally: fast-growing current

literature on many **Interfaces of logic and probability**

May be just what we need when studying the role of rational individual agents

in a society with largely statistical phenomena like flows of public opinion