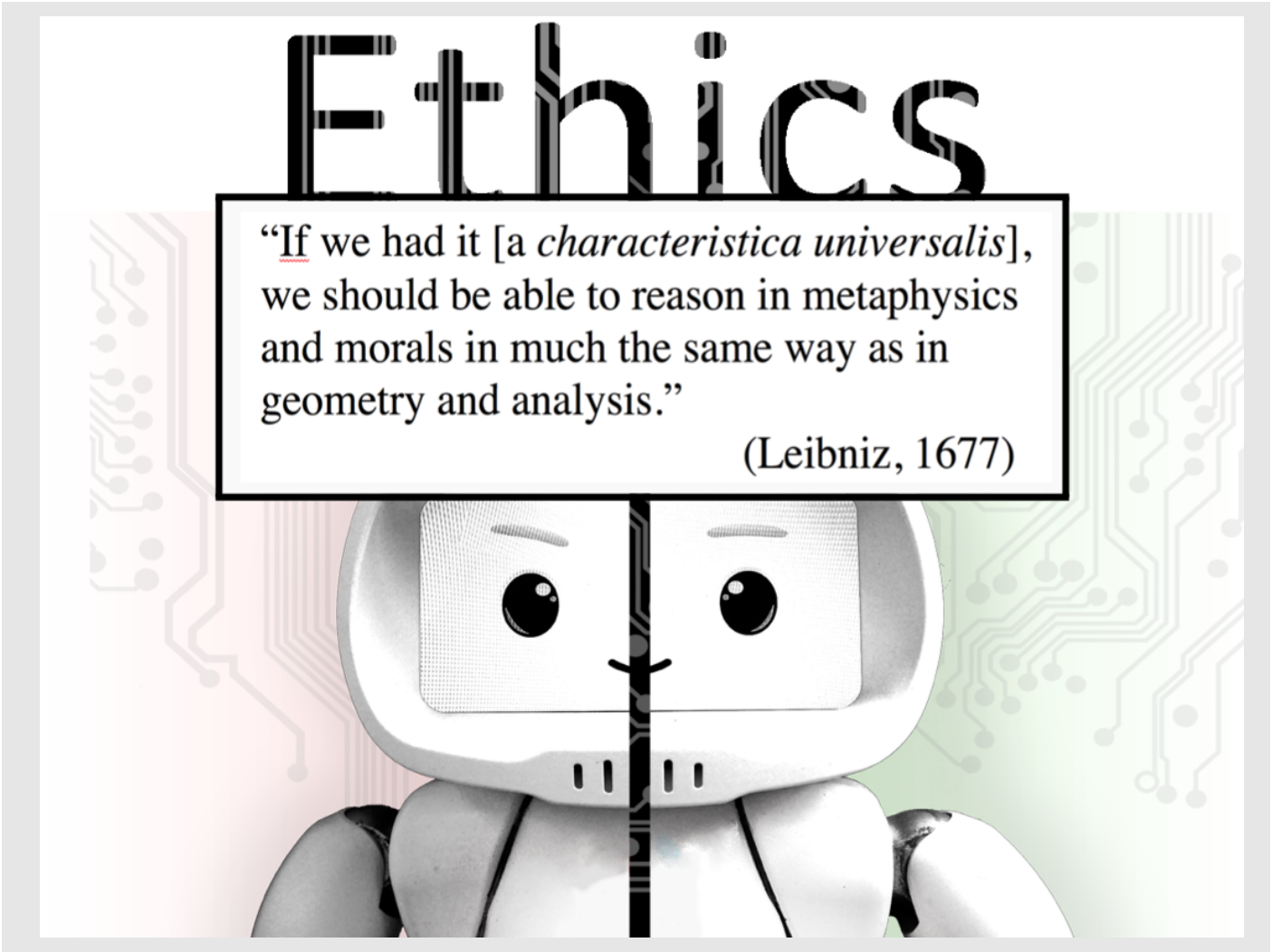
A Flexible Infrastructure for Normative Reasoning

Christoph Benzmüller
Freie Universität Berlin | University of Luxembourg

DEON, Utrecht, 4 July 2018

A Flexible Infrastructure for Normative Reasoning

Christoph Benzmüller
Freie Universität Berlin | University of Luxembourg

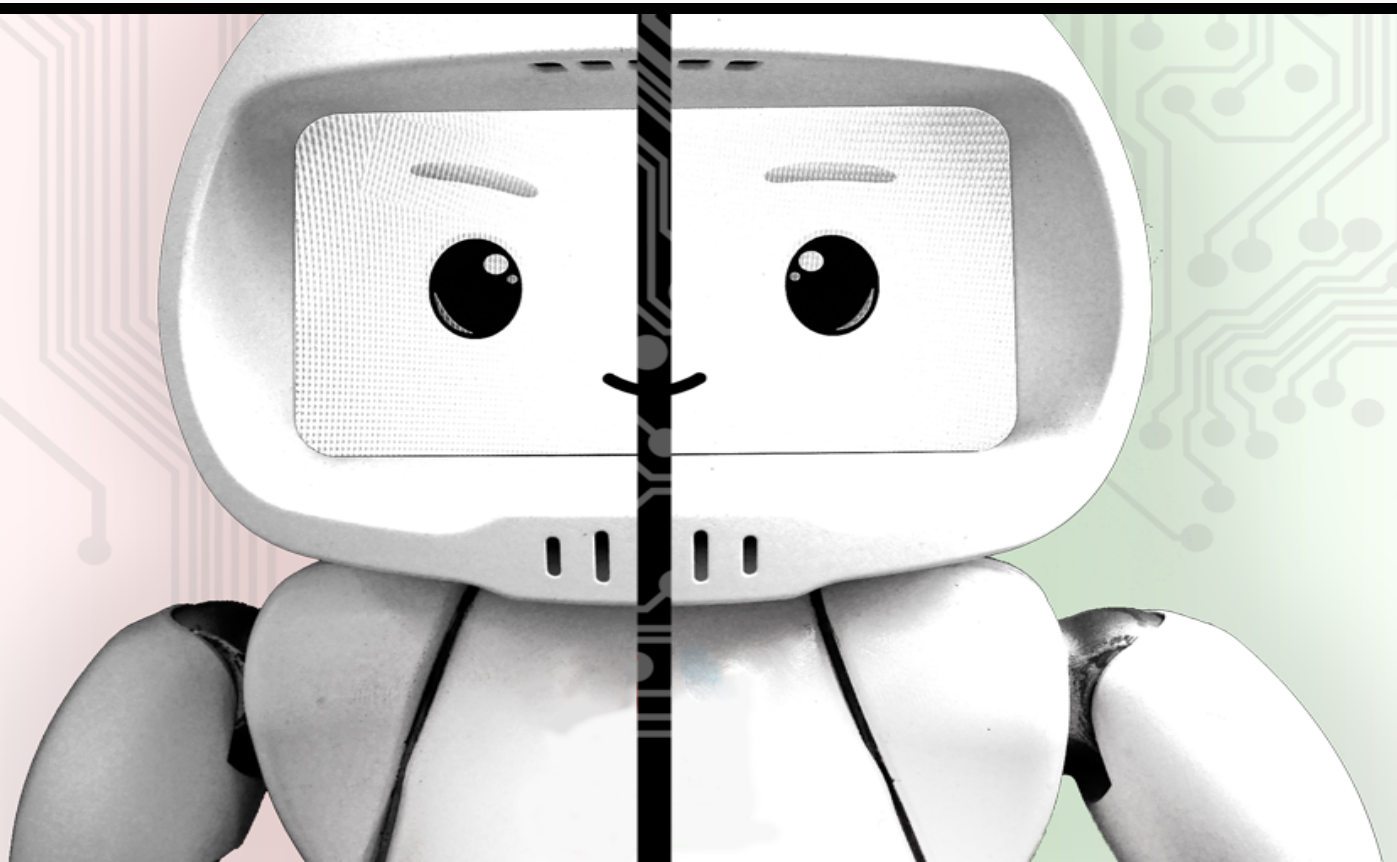


DEON, Utrecht, 4 July 2018

Ethics

Peaceful coexistence with **intelligent autonomous systems (IASs)**?

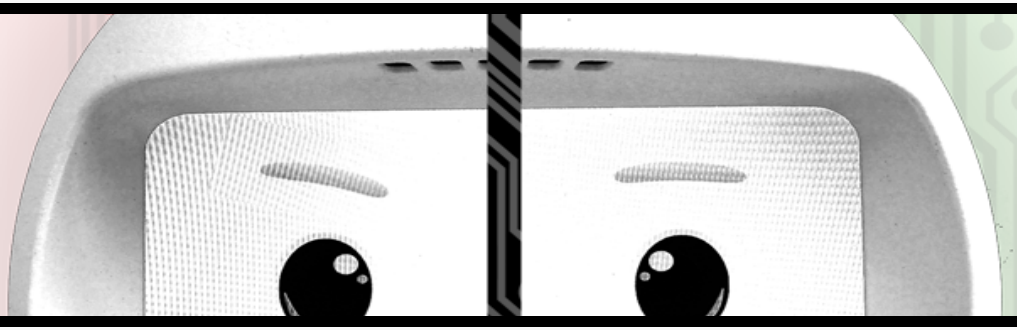
- appropriate forms of **machine-control**
- appropriate forms of **human-machine-interaction**



Ethics

Peaceful coexistence with **intelligent autonomous systems (IASs)**?

- appropriate forms of **machine-control**
- appropriate forms of **human-machine-interaction**



Existing societal processes are based on:

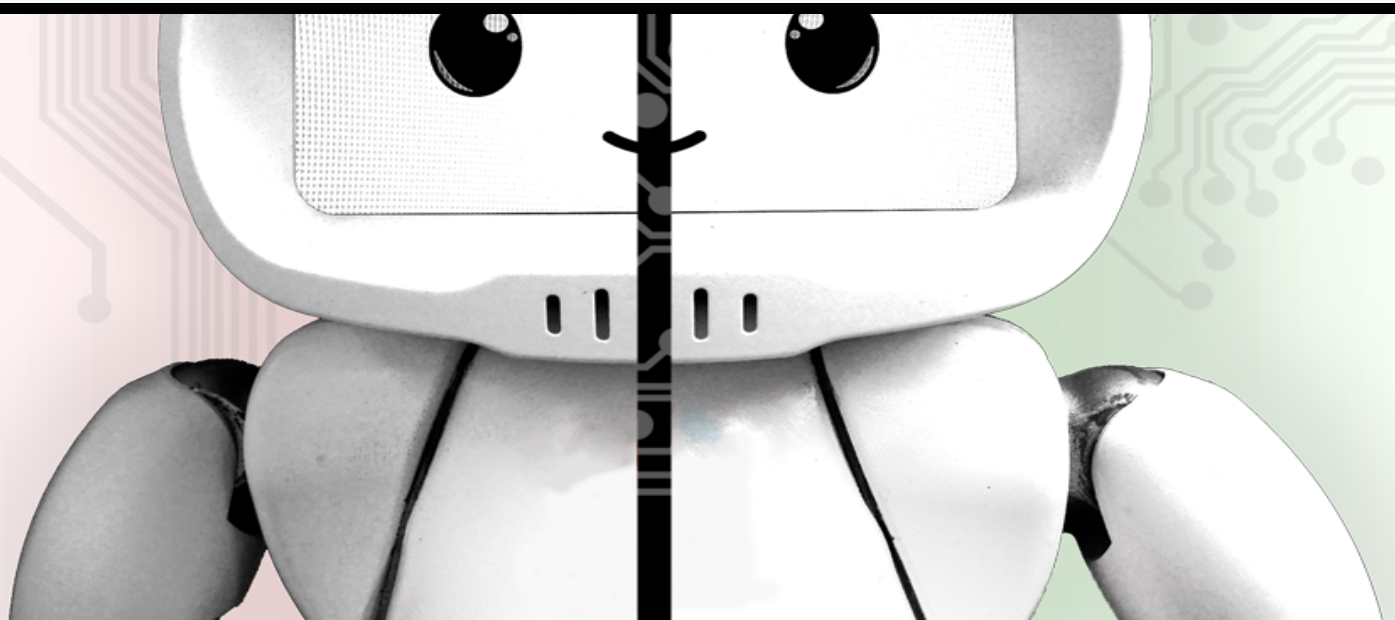
- **rational argumentation & dialog**
- **explicit normative reasoning** (legal & ethical)

Deployment of IASs lacking such competencies? How wise is this?

Ethics

Talk Outline

- **A Motivation:** Explicit Ethical Reasoning
- **B Technology:** Universal Reasoning in Higher-Order Logic (HOL)
- **C Evidence:** Analysis of Rational Arguments in Metaphysics
- **D Demo(s):** Normative Reasoning Experimentation Platform



Motivation

Long-term: Emerging Superintelligence Really? Anyhow . . .

- How to prevent Superintelligence from turning against humanity?

Medium-term: Development of pseudo-ethical skills in IASs

- Which norms? Which reasoning principles?
- What architectural design? What functionalities?
- How to implement, deploy and verify?

Different kinds of systems and approaches:

- [Moor, 2009]:
  – ethical impact agents (ethical consequences to actions)
  – implicit ethical agents (ethical reactions to given situations)
  – explicit ethical agents (reasoning with ethical theories/rules)
  – full ethical agents (conscious, intentional, free will)
- bottom-up vs. top-down
- [DoranEtAl., 2017]:
  opaque — comprehensible — interpretable — explainable AI

Motivation

Long-term: Emerging Superintelligence Really? Anyhow . . .

- How to prevent Superintelligence from turning against humanity?

Medium-term: Development of pseudo-ethical skills in IASs

- Which norms? Which reasoning principles?
- What architectural design? What functionalities?
- How to implement, deploy and verify?

Different kinds of systems and approaches:

- [Moor, 2009]:
  – ethical impact agents (ethical consequences to actions)
  – implicit ethical agents (ethical reactions to given situations)
  – explicit ethical agents (reasoning with ethical theories/rules)
  – full ethical agents (conscious, intentional, free will)
- bottom-up vs. top-down
- [DoranEtAl., 2017]:
  opaque — comprehensible — interpretable — explainable AI

Long-term: Emerging Superintelligence Really? Anyhow . . .

- How to prevent Superintelligence from turning against humanity?

Medium-term: Development of pseudo-ethical skills in IASs

- Which norms? Which reasoning principles?
- What architectural design? What functionalities?
- How to implement, deploy and verify?

Different kinds of systems and approaches:

- [Moor, 2009]:
  – ethical impact agents (ethical consequences to actions)
  – implicit ethical agents (ethical reactions to given situations)
  – explicit ethical agents (reasoning with ethical theories/rules)
  – full ethical agents (conscious, intentional, free will)
- bottom-up vs. top-down
- [DoranEtAl., 2017]:
  opaque — comprehensible — interpretable — explainable AI

Motivation

Long-term: Emerging Superintelligence Really? Anyhow . . .

- How to prevent Superintelligence from turning against humanity?

Medium-term: Development of pseudo-ethical skills in IASs

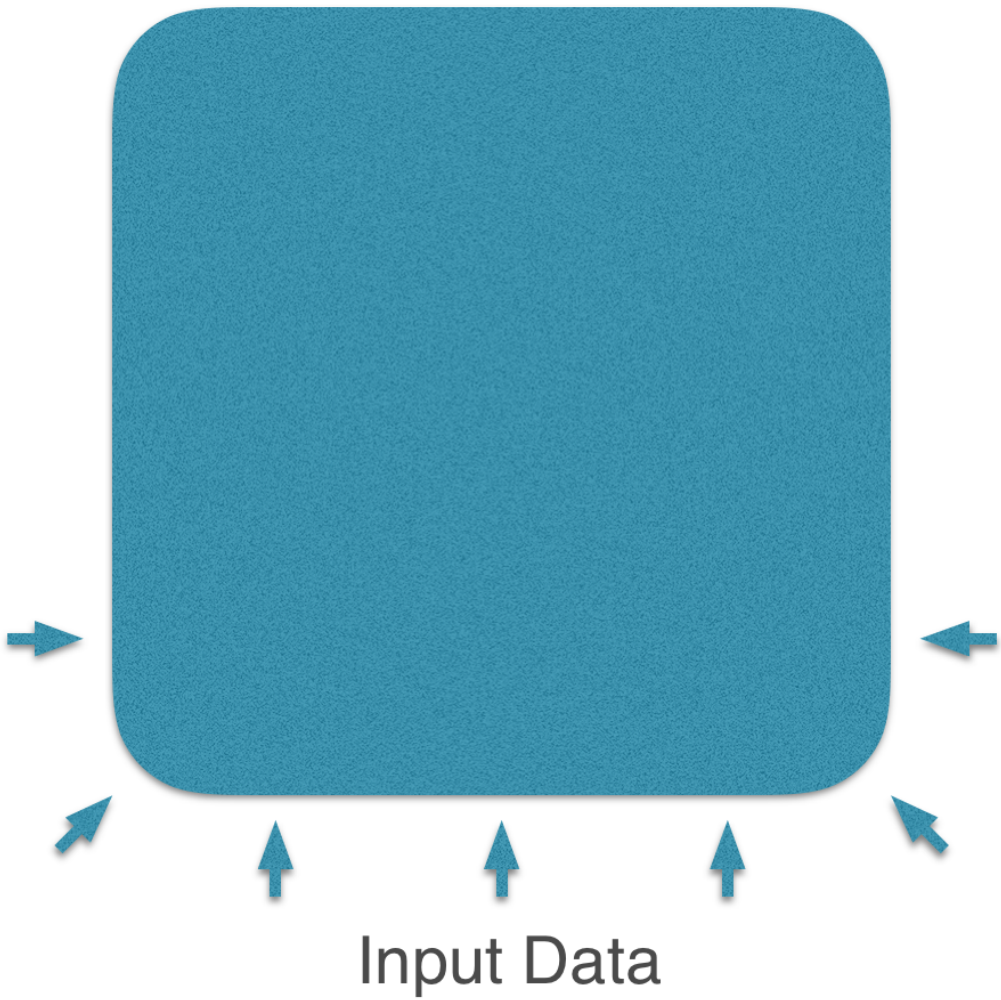
- Which norms? Which reasoning principles?
- What architectural design? What functionalities?
- How to implement, deploy and verify?

Different kinds of systems and approaches:

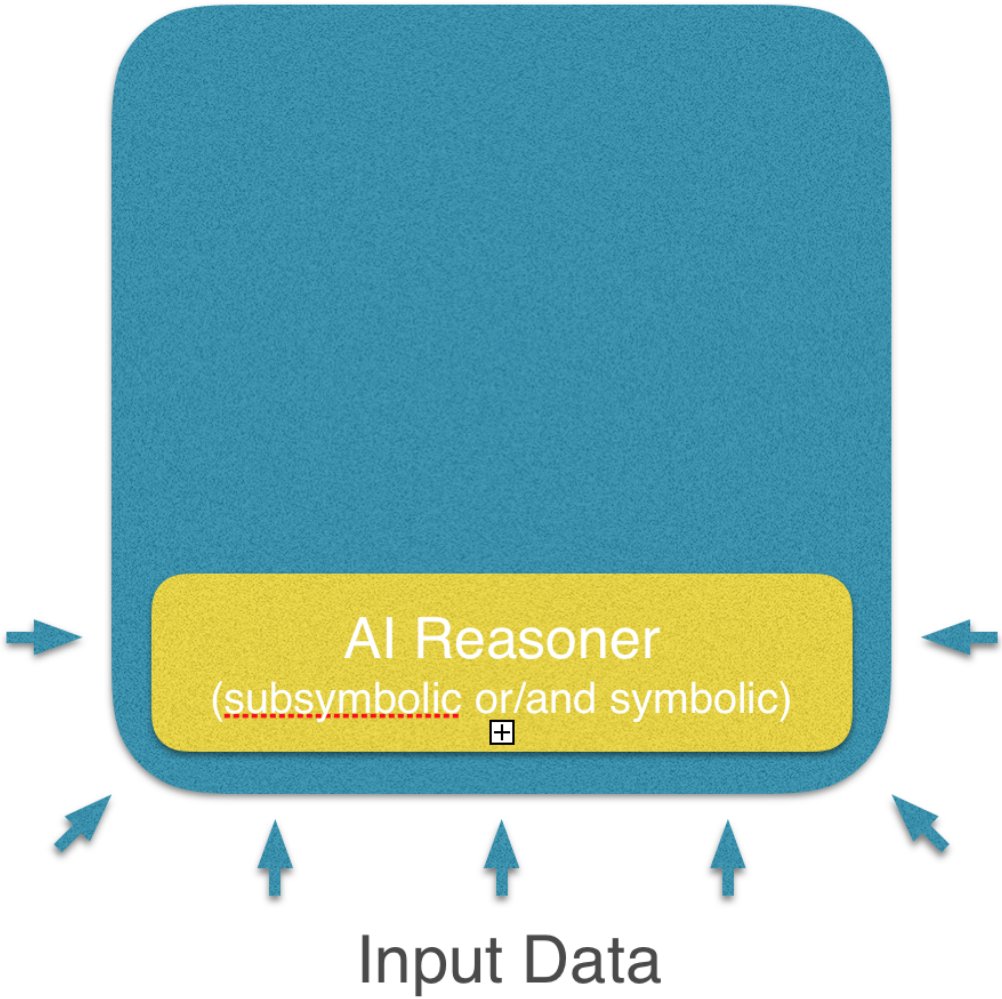
- [Moor, 2009]:
  - ethical impact agents (ethical consequences to actions)
  - implicit ethical agents (ethical reactions to given situations)
  - explicit ethical agents (reasoning with ethical theories/rules)
  - full ethical agents (conscious, intentional, free will)
- bottom-up vs. top-down
- [DoranEtAl., 2017]:
  opaque — comprehensible — interpretable — explainable AI

Pseudo-Ethical IAS (medium-term)

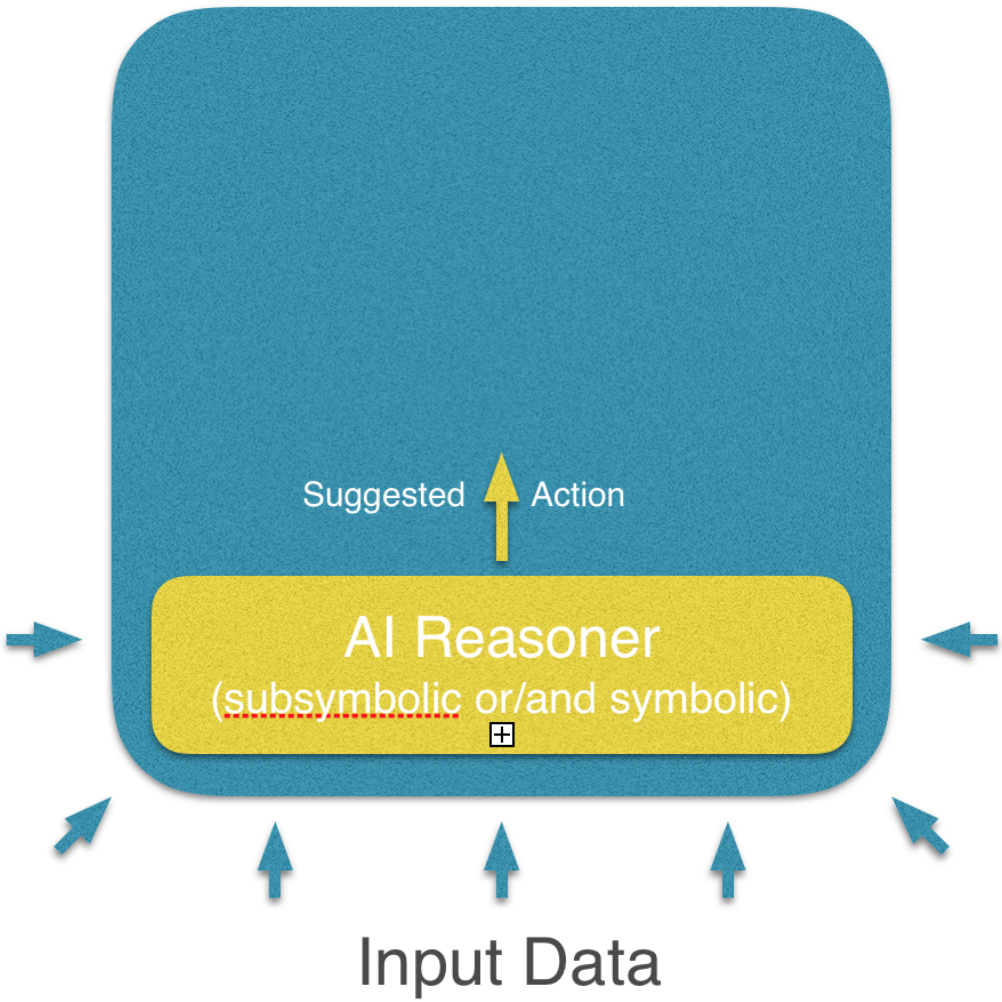
Pseudo-Ethical IAS (medium-term)



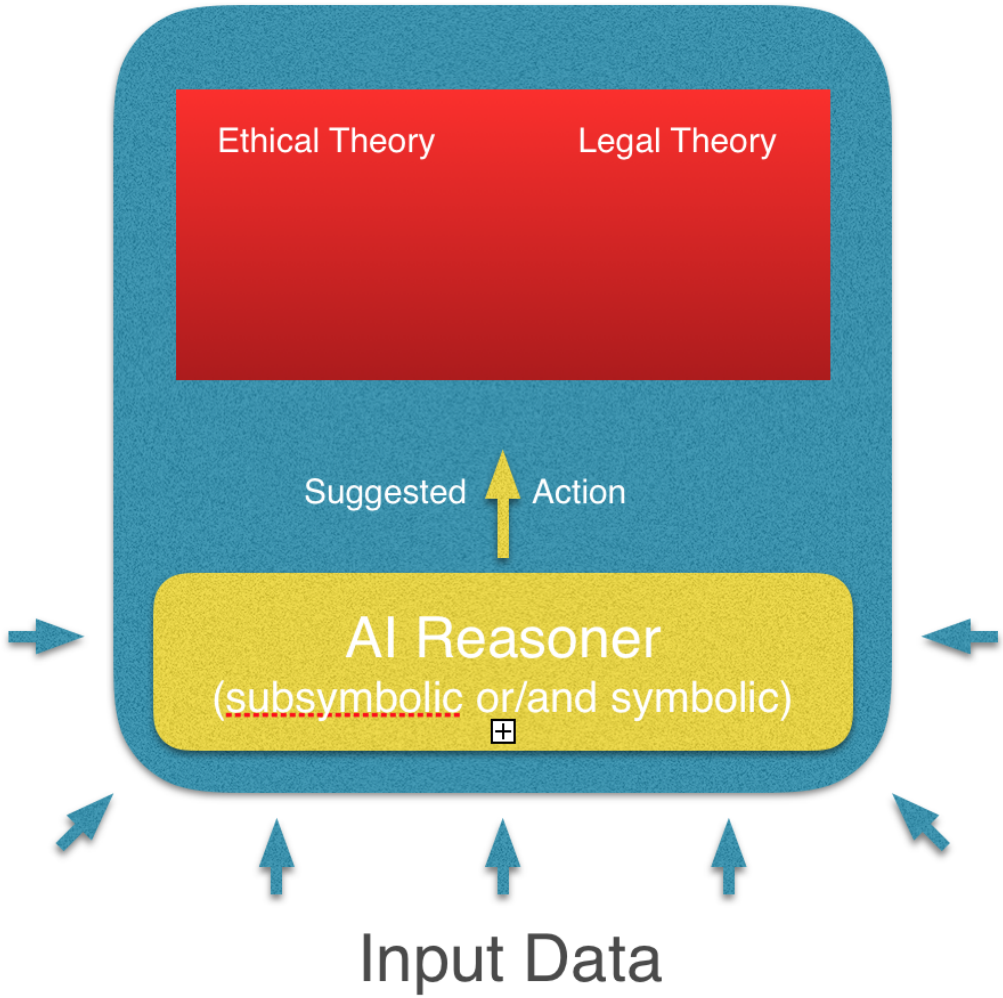
Pseudo-Ethical IAS (medium-term)



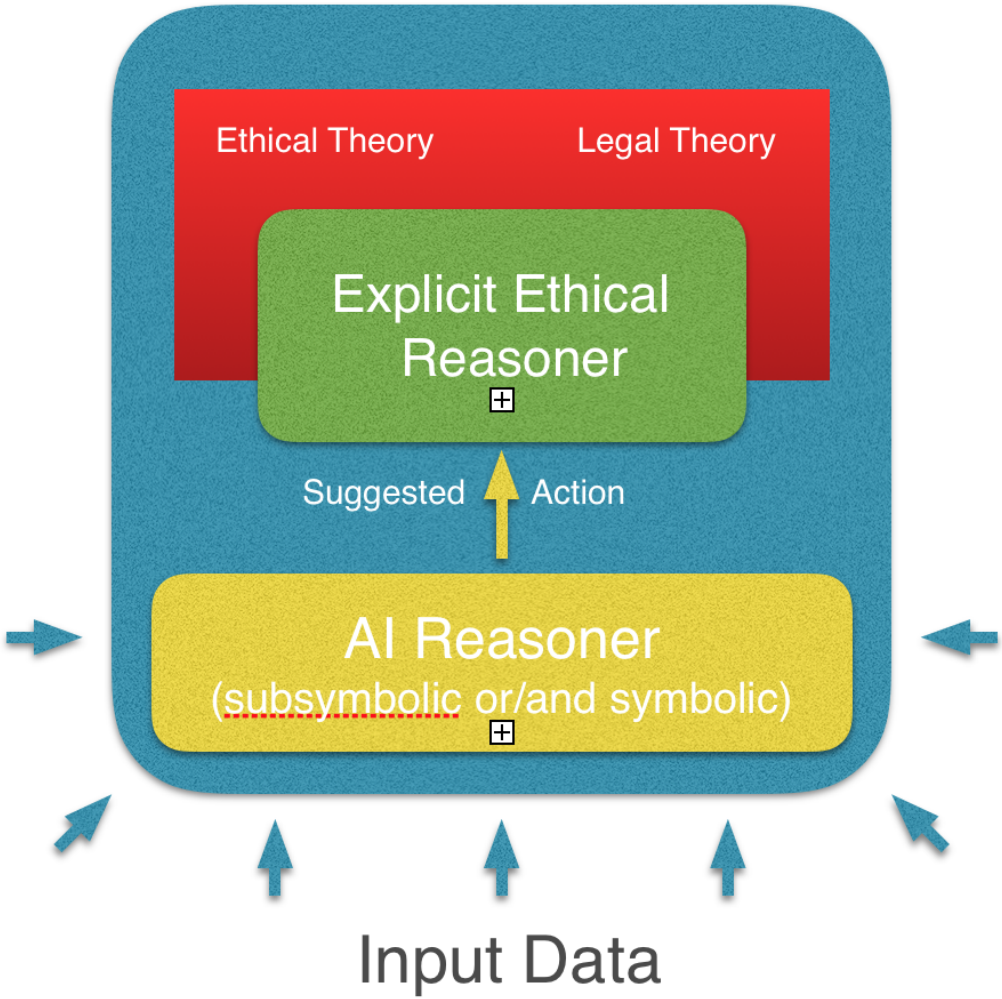
Pseudo-Ethical IAS (medium-term)



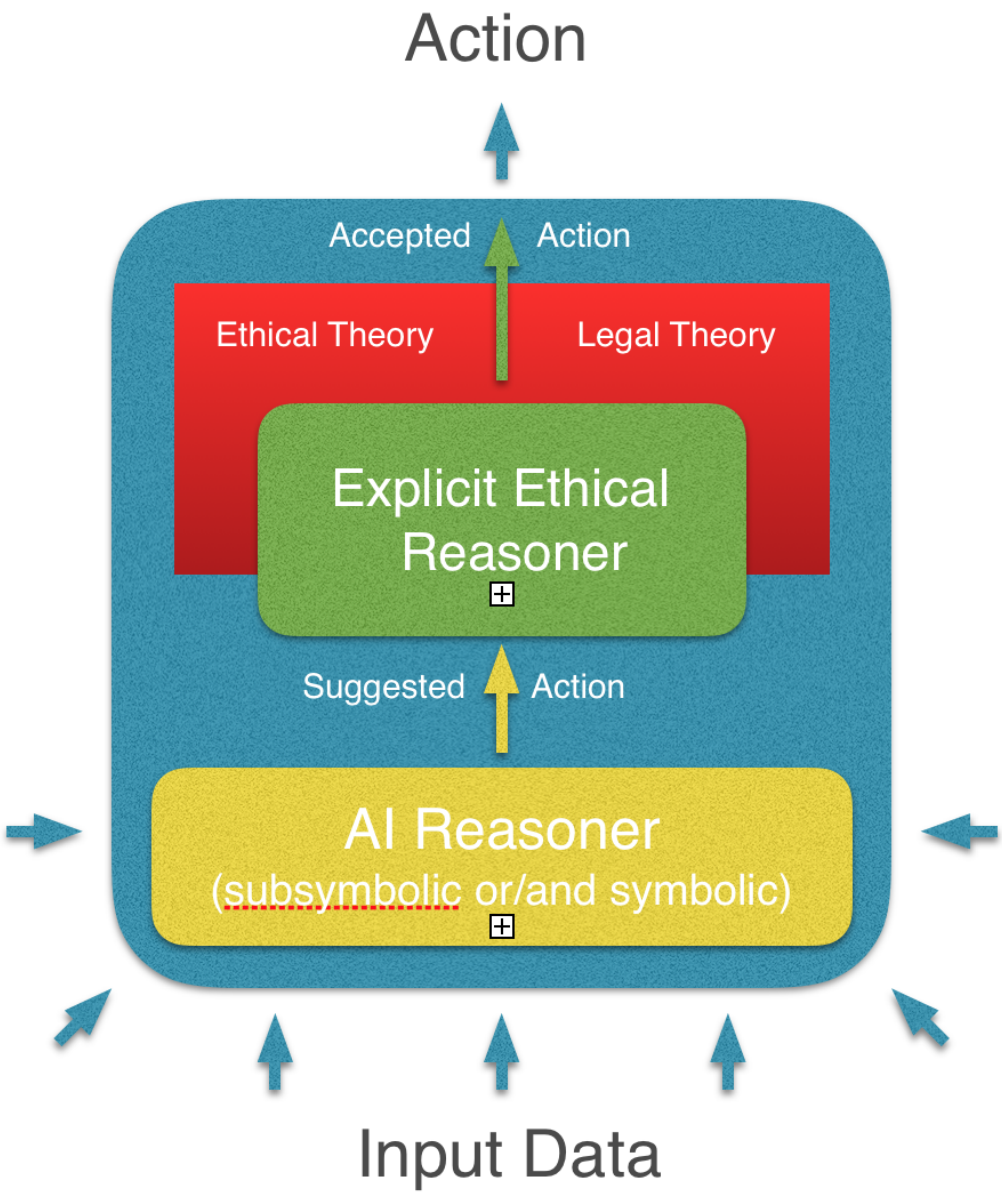
Pseudo-Ethical IAS (medium-term)



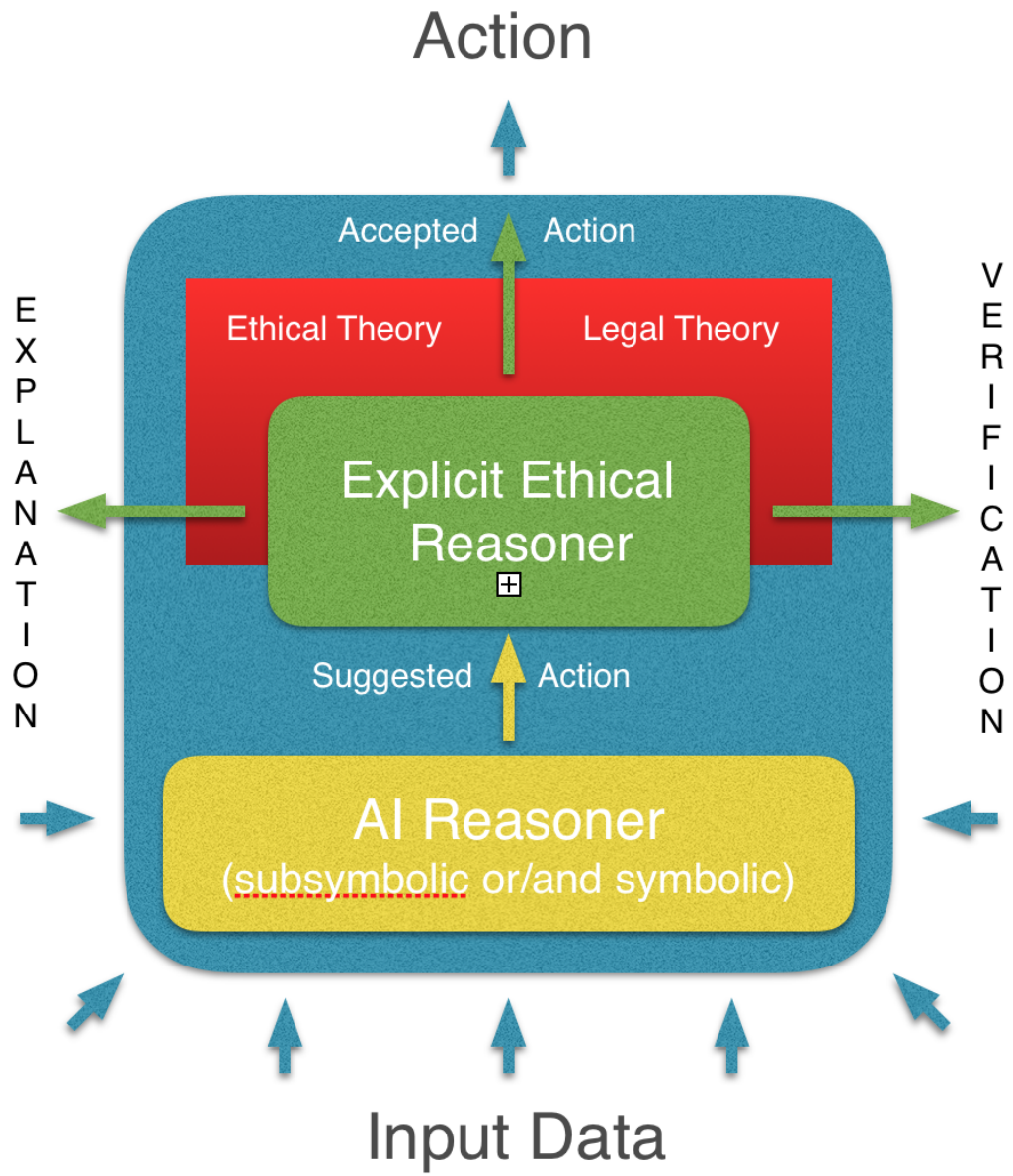
Pseudo-Ethical IAS (medium-term)



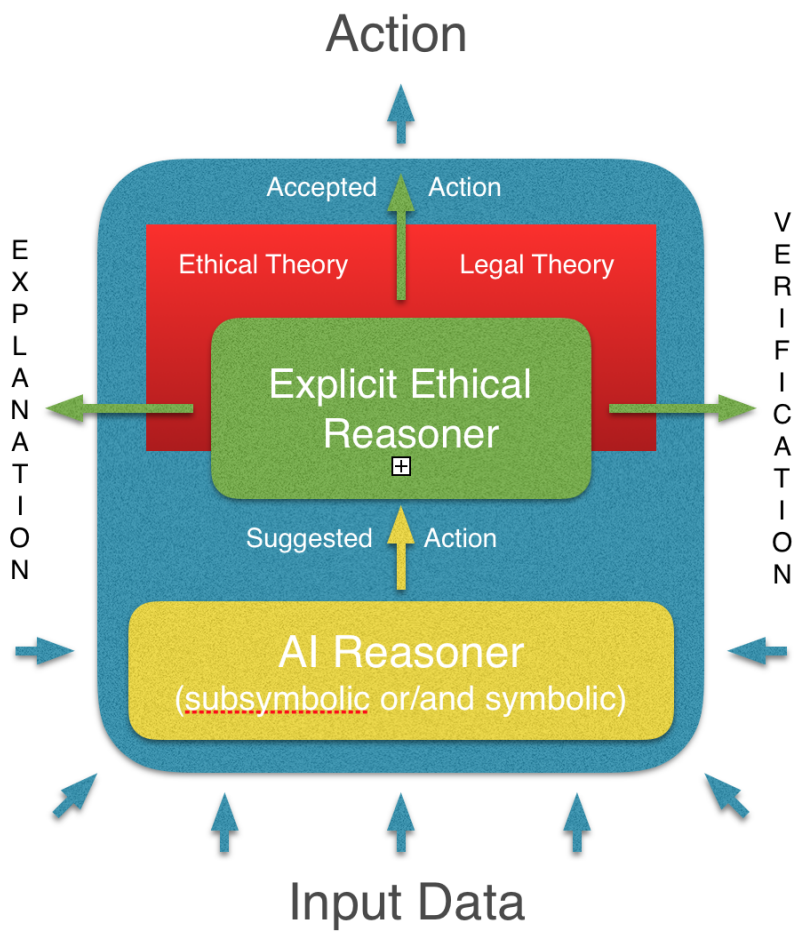
Pseudo-Ethical IAS (medium-term)



Pseudo-Ethical IAS (medium-term)



Pseudo-Ethical IAS (medium-term)



Related Work

- Artificial Moral Agents
  - [Wallach&Allen, 2008]
- Ethical Governors
  - [ArkinEtAl., 2009, 2012]
  - [Dennis&Fisher, 2017]
- Ethical Deliberation in ART
  - [Dignum, 2017]
- Programming Machine Ethics
  - [Pereira&Saptawijaya, 2016]
- . . .

Which Reasoning Formalisms?

> "If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."
>
> (Leibniz, 1677)

Challenges for Explicit Ethical Reasoning Engines: Which Logic(s)?

- Dilemmas, conflicting theories, etc.
- Appropriate handling of notion of **obligation**
  - Contrary-to-duty (**CTD**) scenarios

Which Reasoning Formalisms?

"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."

(Leibniz, 1677)

Challenges for Explicit Ethical Reasoning Engines: Which Logic(s)?

- Dilemmas, conflicting theories, etc.
- Appropriate handling of notion of **obligation**
  - Contrary-to-duty (**CTD**) scenarios

Standard CTD structure (Chisholm)

1. obligatory 'a'
2. obligatory 'if a then not b'
3. if 'not a' then obligatory 'b'
4. 'not a' (in a given situation)

Danger: Paradox/inconsistency — ex falso quodlibet!

Which Reasoning Formalisms?

> "If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."
>
> (Leibniz, 1677)

Challenges for Explicit Ethical Reasoning Engines: Which Logic(s)?

- Dilemmas, conflicting theories, etc.
- Appropriate handling of notion of **obligation**
  - Contrary-to-duty (**CTD**) scenarios

CTD example (X. Parent): **EU General Data Protection Regulation (GDPR)**

1. Personal data shall be processed lawfully. (Art. 5)
   E.g., the data subject must have given consent to the processing. (Art. 6/1.a)
2. **Implicit:** The data shall be kept, for the agreed purposes, if processed lawfully.
3. If personal data has been processed unlawfully, the controller has the obligation to erase the personal data in question without delay. (Art. 17.d, right to be forgotten)
4. **Given situation:** Some personal data has been processed unlawfully.

Danger: Paradox/inconsistency — ex falso quodlibet!

Which Reasoning Formalisms?

"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."

(Leibniz, 1677)

Challenges for Explicit Ethical Reasoning Engines: Which Logic(s)?

- Dilemmas, conflicting theories, etc.
- Appropriate handling of notion of **obligation**
  - Contrary-to-duty (**CTD**) scenarios

L. van der Torre

X. Parent

A. Farjami

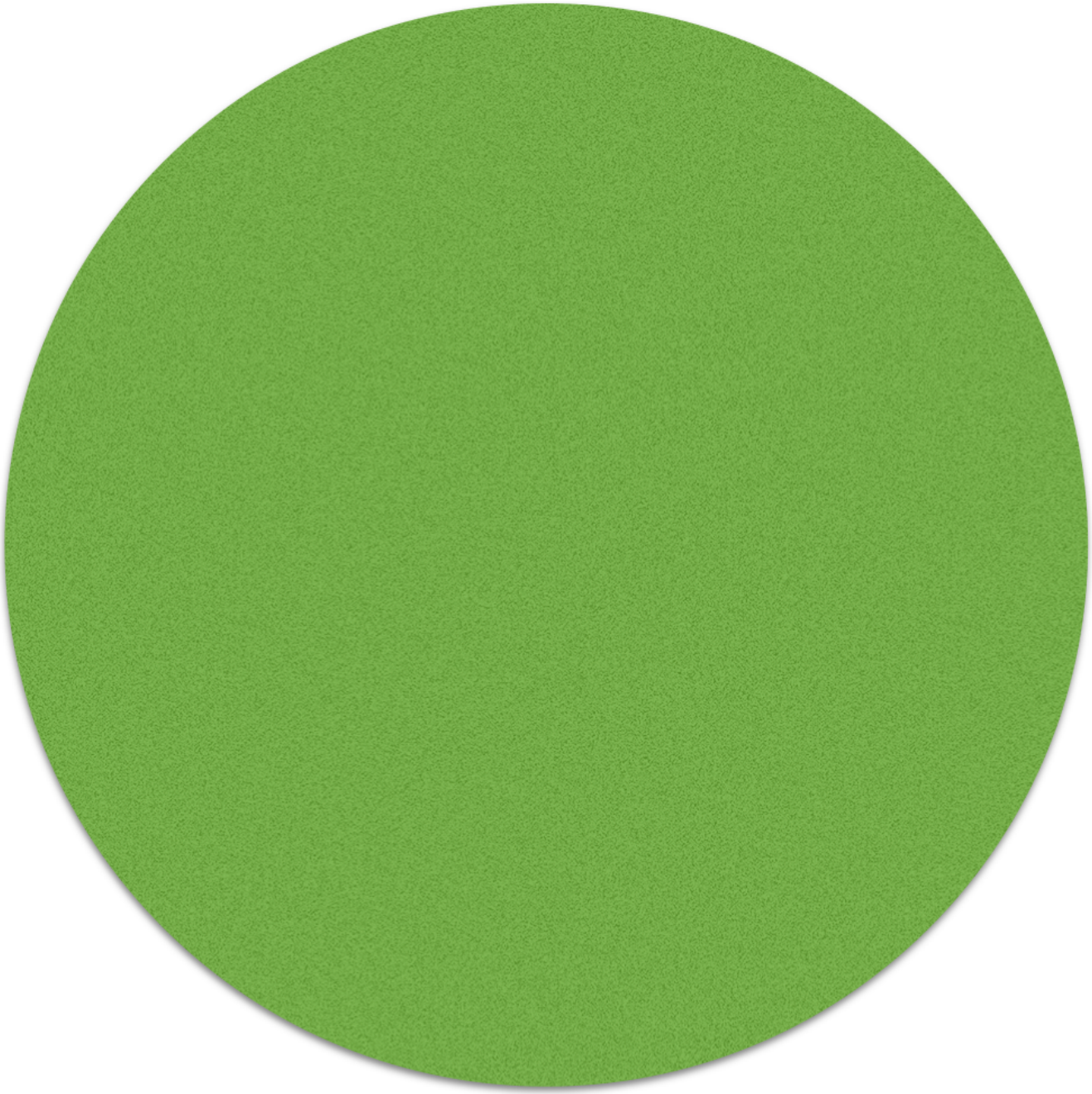
Deontic Logic

- Reasoning about obligations and permissions
- Two groups of approaches:
  - — Possible worlds
    - standard deontic logic — **CTD: no**
    - dyadic deontic logic — **CTD: yes**
  - — Norm-based semantics
    - input/output logic — **CTD: yes**

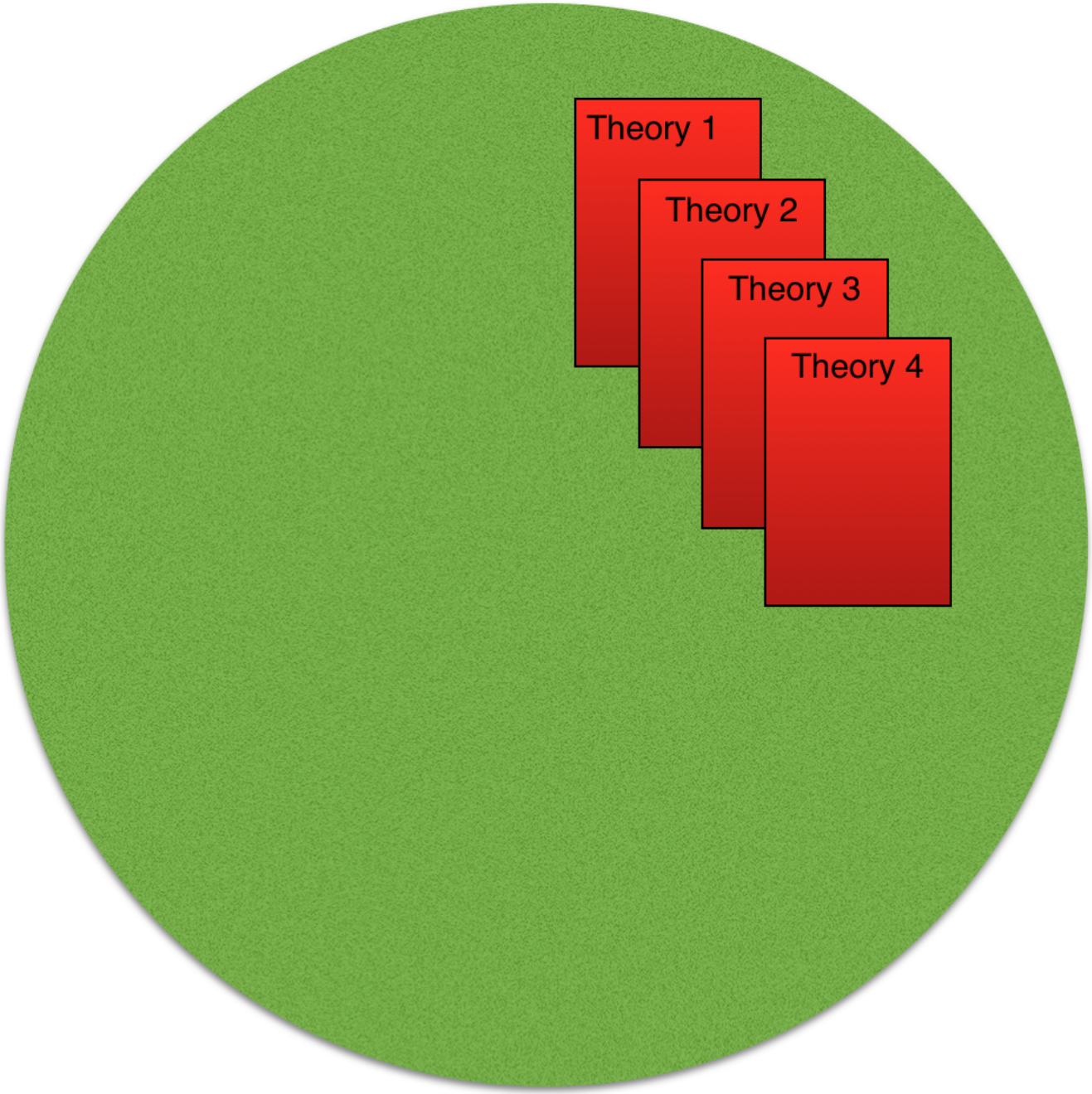
Further interests and challenges

- Combination with other logics (other modalities)
- Propositional deontic logic(s) will hardly be sufficient in practice

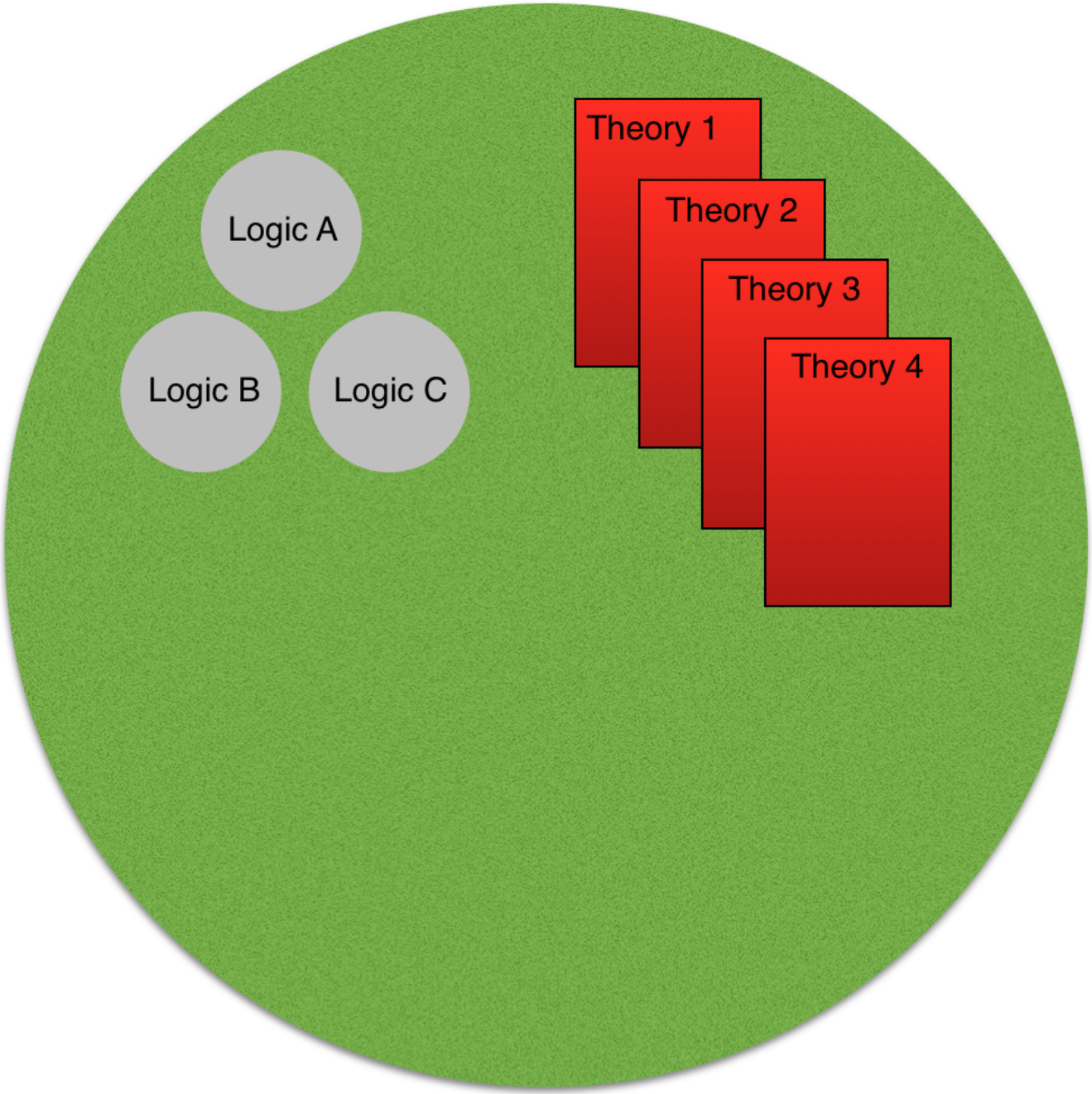
Normative Reasoning Experimentation Platform



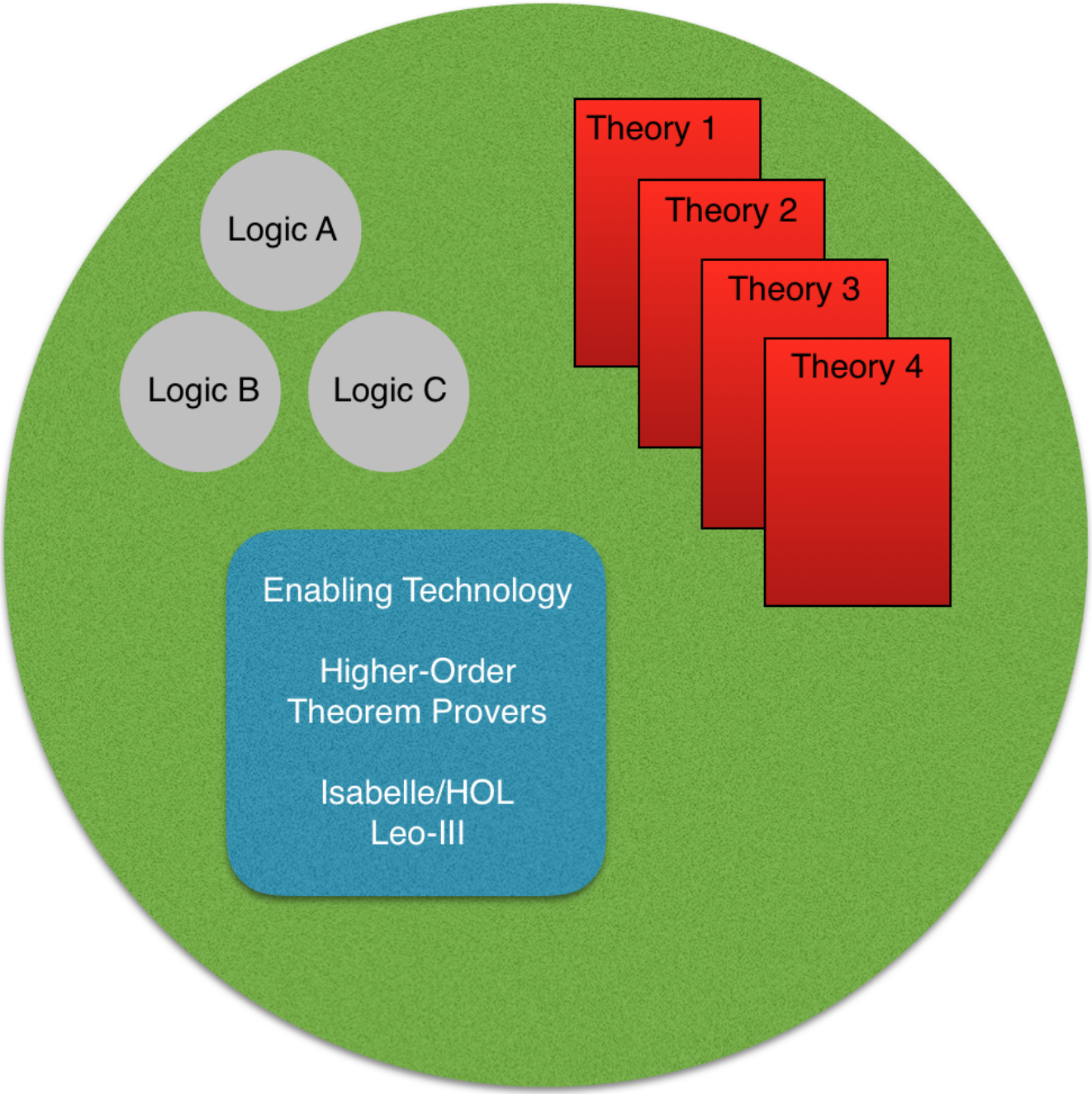
Normative Reasoning Experimentation Platform



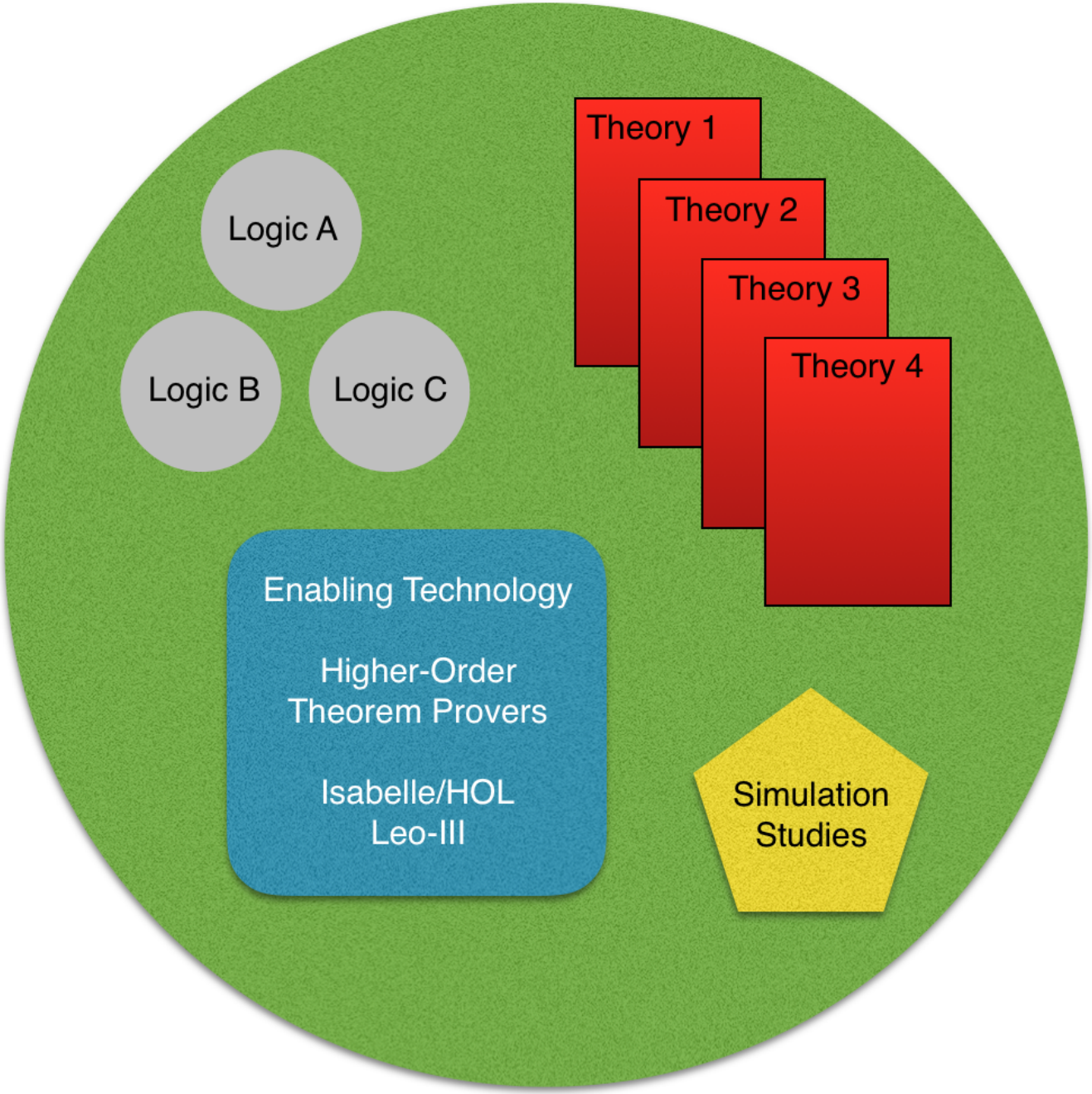
Normative Reasoning Experimentation Platform



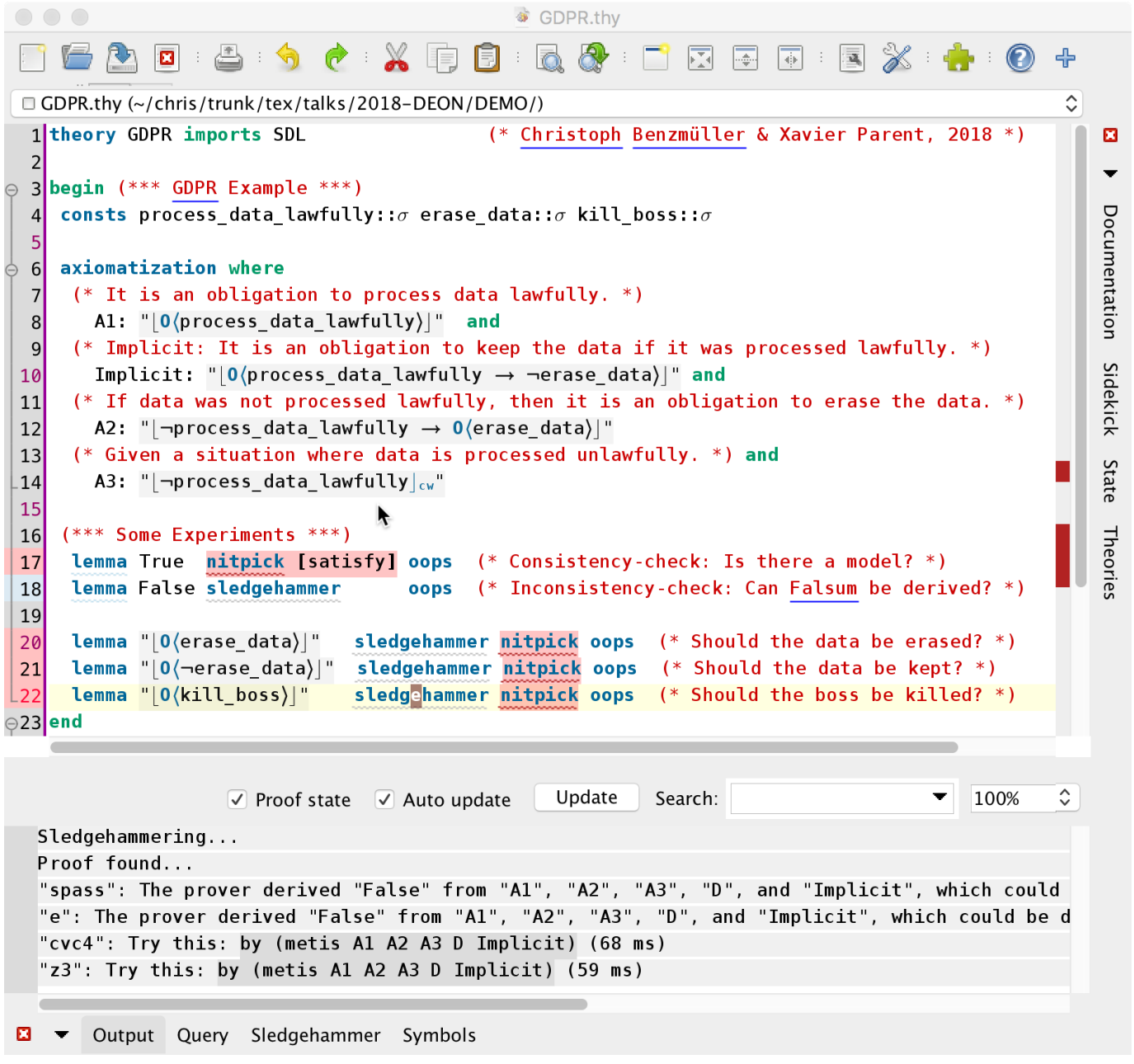
Normative Reasoning Experimentation Platform



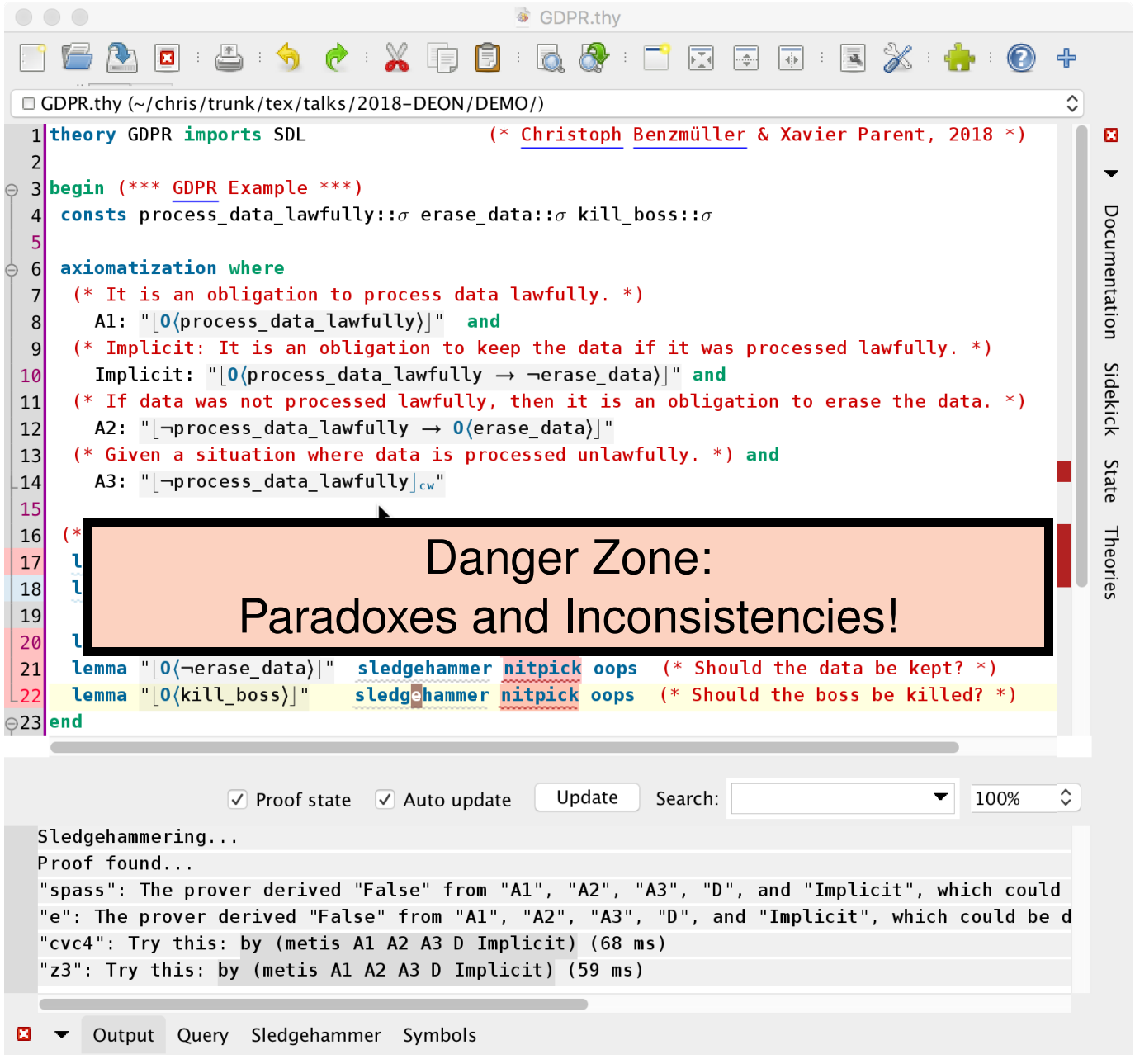
Normative Reasoning Experimentation Platform



Normative Reasoning Experimentation Platform — Demo in Isabelle/HOL



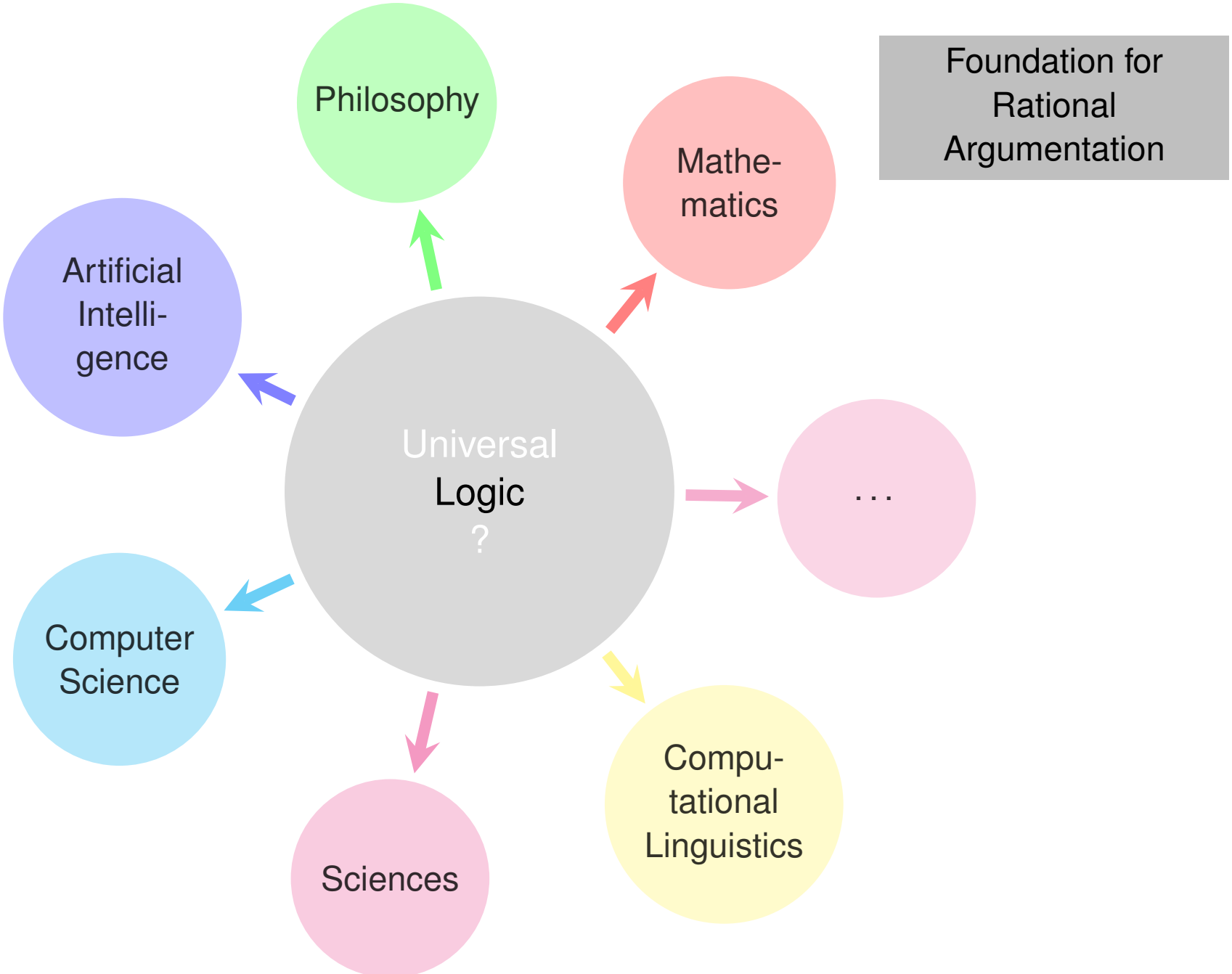
Normative Reasoning Experimentation Platform — Demo in Isabelle/HOL

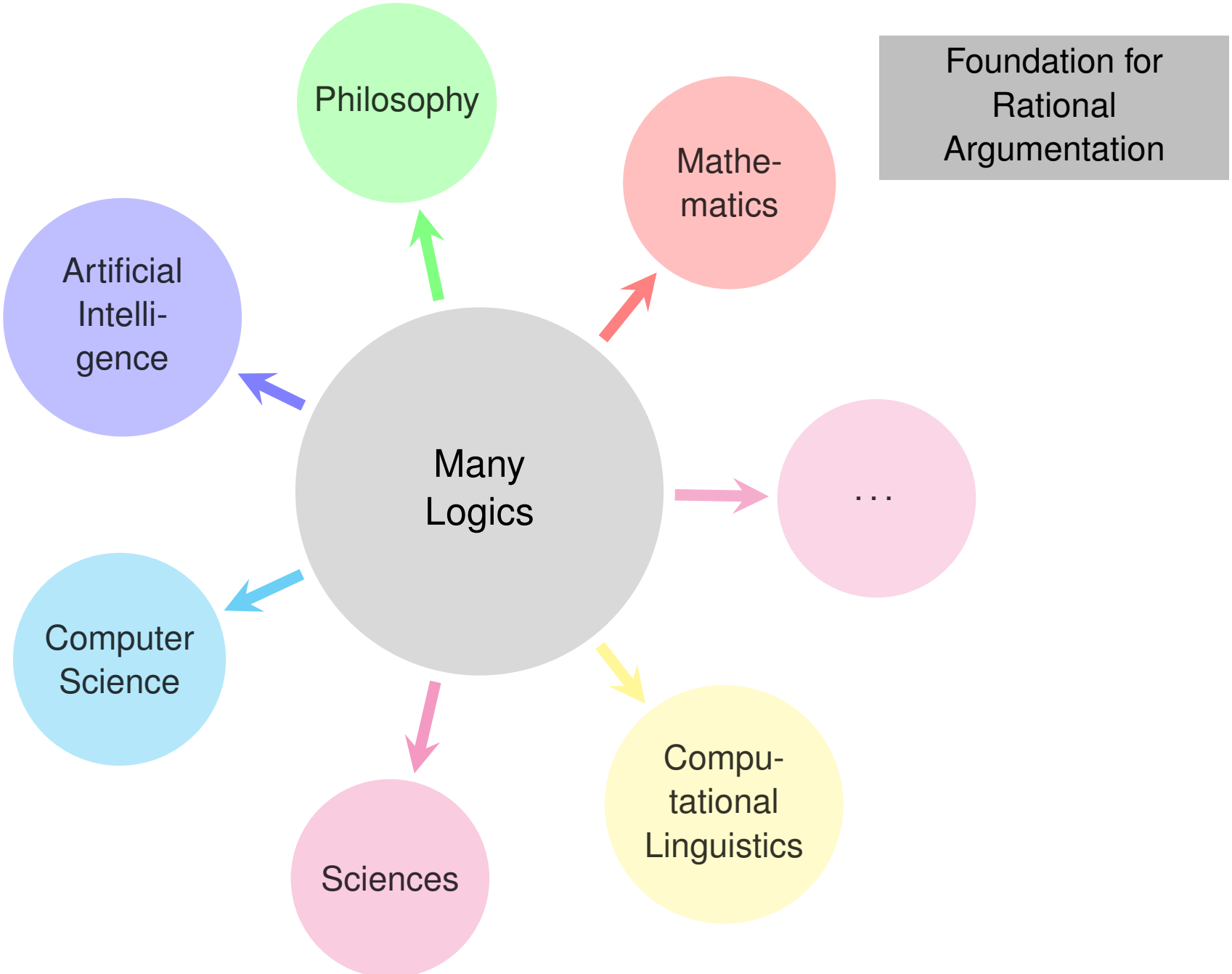


"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."

(Leibniz, 1677)

Part B
Technology: Universal Reasoning in Higher-Order Logic


Philosophy
Mathe-
matics
Foundation for
Rational
Argumentation
Artificial
Intelli-
gence
Universal
Logic
?
. . .
Computer
Science
Sciences
Compu-
tational
Linguistics


Philosophy
Mathe-
matics
Foundation for
Rational
Argumentation
Artificial
Intelli-
gence
Many
Logics
. . .
Computer
Science
Sciences
Compu-
tational
Linguistics

Logic Zoo

Logic Zoo

Classical Logic, of order

0. Propositional Logic
1. First-order Logic
2. Second-order Logic

...

n. Higher-order Logic

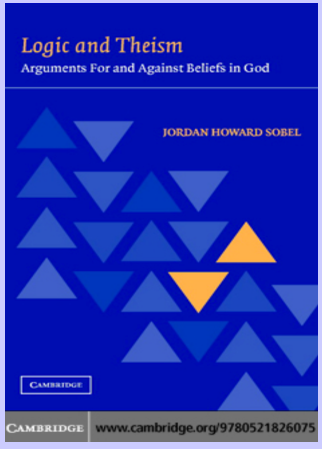
Non-Classical Logics

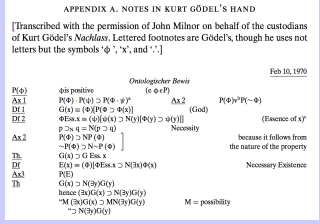
- Intuitionistic/Constructive Logics (incl. Univalent Foundations)
- Modal Logics, Conditional Logics, Temporal Logis, Spatial Logics
- Many-valued Logics
- Paraconsistent Logics
- Free Logics, Inclusive Logics
- Logics for special applications: Ethics, Social Choice, Legal Reasoning, ...
- Separation Logic, ...

Example Application in Metaphysics/Philosophy:

Necessarily, God exists: $\Box\exists x.Gx$

Kurt Gödel's definition of God: $G\,x \;:=\; \forall\Phi.Positive\ \Phi \rightarrow \Phi x$




APPENDIX A. NOTES IN KURT GÖDEL'S HAND

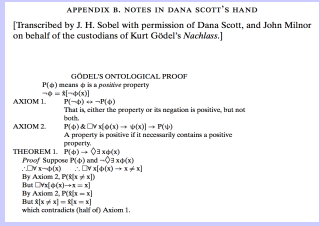
[Transcribed with the permission of John Milnor on behalf of the custodians of Kurt Gödel's *Nachlass*. Lettered footnotes are Gödel's, though he uses not letters but the symbols '◆ ', '×', and '.'.]

Feb 10, 1970

Ontologischer Beweis

P(φ) φ is positive (e φ∈P)
Ax 1 P(¬) · P(φ) ⊃ P(Φ · φ)ᵃ
Df 1 G(x) ≡ (φ)[P(Φ ⊃ Φ(x)]
Df 2 Φ Ess.x ≡ (φ){φ(x) ⊃ N(y)[Φ(y) ⊃ φ(y)]}
P ⊃◇ (q = N(p ⊃ q)
Ax 2 P(Φ) ⊃ N(Φ)
¬P(Φ) ⊃ N¬P(Φ)
Th. G(x) ⊃ G Ess. x
Df E(x) ≡ (Φ){Φ Ess.x ⊃ N(∃x)Φ(x)
Ax3 P(E)
Th G(x) ⊃ N(∃y)G(y)
hence (∃x)G(x) ⊃ N(∃y)G(y)
ᵃM (∃x)G(x) ⊃ M(∃y)G(y)
"⊃ N(∃y)G(y)

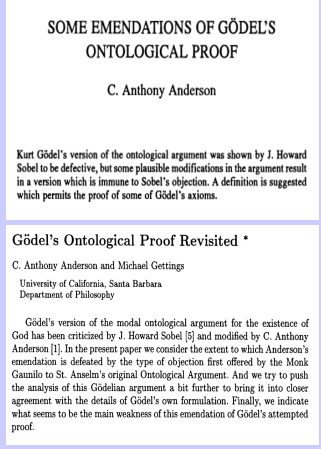
Ax 2 P(Φ)ᵇ(¬ Φ)
(God)
(Essence of x)ᶜ
Necessity
because it follows from the nature of the property
Necessary Existence
M = possibility


APPENDIX B. NOTES IN DANA SCOTT'S HAND

[Transcribed by J. H. Sobel with permission of Dana Scott, and John Milnor on behalf of the custodians of Kurt Gödel's *Nachlass*.]

GÖDEL'S ONTOLOGICAL PROOF

P(φ) means φ is a *positive* property
¬φ = ~[¬φ(x)]
AXIOM 1. P(¬φ) ↔ ¬P(φ)
That is, either the property or its negation is positive, but not both.
AXIOM 2. P(φ) & □∀ x[φ(x) → ψ(x)] → P(ψ)
A property is positive if it necessarily contains a positive property.
THEOREM 1. P(φ) → ◇∃ xφ(x)
Proof Suppose P(φ) and ~◇∃ xφ(x)
∴□∀ x ~φ(x) ∴ □∀ x[φ(x) → x ≠ x]
By Axiom 2, P[x ≠ x]
But □∀x[φ(x)→x = x]
By Axiom 2, P[x = x]
But ~[x ≠ x] = ~[x = x]
which contradicts (half of) Axiom 1.


SOME EMENDATIONS OF GÖDEL'S ONTOLOGICAL PROOF

C. Anthony Anderson

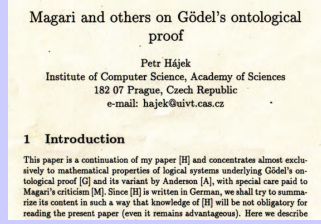
Kurt Gödel's version of the ontological argument was shown by J. Howard Sobel to be defective, but some plausible modifications in the argument result in a version which is immune to Sobel's objection. A definition is suggested which permits the proof of some of Gödel's axioms.

Gödel's Ontological Proof Revisited *

C. Anthony Anderson and Michael Gettings

University of California, Santa Barbara
Department of Philosophy

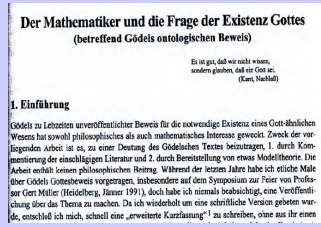
Gödel's version of the modal ontological argument for the existence of God has been criticized by J. Howard Sobel [5] and modified by C. Anthony Anderson [1]. In the present paper we consider the extent to which Anderson's emendation is defeated by the type of objection first offered by the Monk Gaunilo to St. Anselm's original Ontological Argument. And we try to push the analysis of this Gödelian argument a bit further to bring it into closer agreement with the details of Gödel's own formulation. Finally, we indicate what seems to be the main weakness of this emendation of Gödel's attempted proof.


Magari and others on Gödel's ontological proof

Petr Hájek
Institute of Computer Science, Academy of Sciences
182 07 Prague, Czech Republic
e-mail: hajek@uivt.cas.cz

1 Introduction

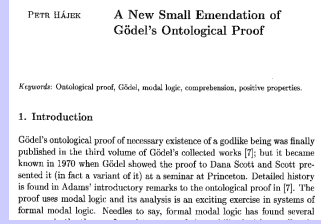
This paper is a continuation of my paper [H] and concentrates almost exclusively to mathematical properties of logical systems underlying Gödel's ontological proof [G] and its variant by Anderson [A], with special care paid to Magari's criticism [M]. Since [H] is written in German, we shall try to summarize its content in such a way that knowledge of [H] will be not obligatory for reading the present paper (even it remains advantageous). Here we describe


Der Mathematiker und die Frage der Existenz Gottes
(betreffend Gödels ontologischen Beweis)

Es ist gut, daß wir nicht wissen,
sondern glauben, daß ein Gott ist.
(Kant, Nachlaß)

1. Einführung

Gödels zu Lebzeiten unveröffentlichter Beweis für die notwendige Existenz eines Gott-ähnlichen Wesens hat sowohl philosophisches als auch mathematisches Interesse geweckt. Zweck der vorliegenden Arbeit ist es, zu einer Deutung des Gödelschen Textes beizutragen, i. durch Kommentierung der einschlägigen Literatur und 2. durch Bereitstellung von etwas Modelltheorie. Die Arbeit enthält keinen philosophischen Beitrag. Während der letzten Jahre habe ich etliche Male über Gödels Gottesbeweis vorgetragen, insbesondere auf dem Symposium zur Feier von Professor Gert Müller (Heidelberg, Jänner 1991), doch habe ich niemals beabsichtigt, eine Veröffentlichung über das Thema zu machen. Da ich wiederholt um eine schriftliche Version gebeten wurde, entschloß ich mich, schnell eine „erweiterte Kurzfassung"[1] zu schreiben, ohne aus ihr einen

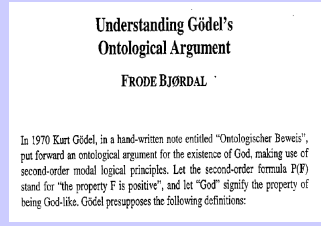

PETR HÁJEK

A New Small Emendation of Gödel's Ontological Proof

Keywords: Ontological proof, Gödel, modal logic, comprehension, positive properties.

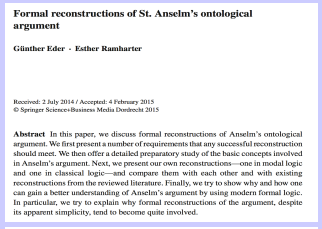
1. Introduction

Gödel's ontological proof of necessary existence of a godlike being was finally published in the third volume of Gödel's collected works [7]; but it became known in 1970 when Gödel showed the proof to Dana Scott and Scott presented it (in fact a variant of it) at a seminar at Princeton. Detailed history is found in Adams' introductory remarks to the ontological proof in [7]. The proof uses modal logic and its analysis is an exciting exercise in systems of formal modal logic. Needles to say, formal modal logic has found several


Understanding Gödel's Ontological Argument

FRODE BJØRDAL

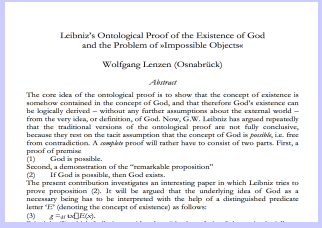
In 1970 Kurt Gödel, in a hand-written note entitled "Ontologischer Beweis", put forward an ontological argument for the existence of God, making use of second-order modal logical principles. Let the second-order formula P(F) stand for "the property F is positive", and let "God" signify the property of being God-like. Gödel presupposes the following definitions:


Formal reconstructions of St. Anselm's ontological argument

Günther Eder · Esther Ramharter

Received: 2 July 2014 / Accepted: 4 February 2015

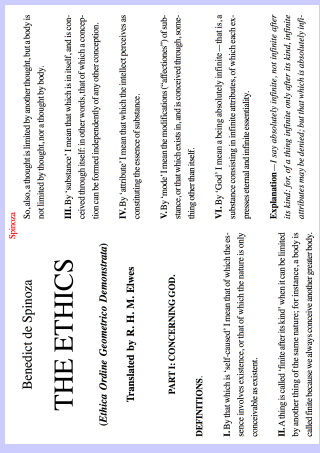
Abstract In this paper, we discuss formal reconstructions of Anselm's ontological argument. We first present a number of requirements that any successful reconstruction should meet. We then offer a detailed preparatory study of the basic concepts involved in Anselm's argument. Next, we present our own reconstructions—one in modal logic and one in classical logic—and compare them with each other and with existing reconstructions from the reviewed literature. Finally, we try to show why and how one can gain a better understanding of Anselm's argument by using modern formal logic. In particular, we try to explain why formal reconstructions of the argument, despite its apparent simplicity, tend to become quite involved.


Leibniz's Ontological Proof of the Existence of God and the Problem of »Impossible Objects«

Wolfgang Lenzen (Osnabrück)

Abstract

The core idea of the ontological proof is to show that the concept of existence is somehow contained in the concept of God, and that therefore God's existence can be logically derived – without any further assumptions about the external world – from the very idea, or definition, of God. Now, G.W. Leibniz has argued repeatedly that the traditional versions of the ontological proof are not fully conclusive, because they rest on the tacit assumption that the concept of God is *possible*, i.e. free from contradiction. A *complete* proof will rather have to consist of two parts. First, a proof of premise
(1) God is possible.
Second, a demonstration of the "remarkable proposition"
(2) If God is possible, then God exists.
The present contribution investigates an interesting paper in which Leibniz tries to prove proposition (2). It will be argued that the underlying idea of God as a necessary being has to be interpreted with the help of a distinguished predicate letter 'E' (denoting the concept of existence) as follows:
(3) $g =_{df} \iota x[E(x)]$.


Spinoza

Benedict de Spinoza

THE ETHICS

(*Ethica Ordine Geometrico Demonstrata*)

Translated by R. H. M. Elwes

PART I: CONCERNING GOD.

DEFINITIONS.

I. By that which is 'self-caused' I mean that of which the essence involves existence, or that of which the nature is only conceivable as existent.

II. A thing is called 'finite after its kind', when it can be limited by another thing of the same nature; for instance, a body is called finite because we always conceive another greater body.

So, also, a thought is limited by another thought, but a body is not limited by thought, nor a thought by body.

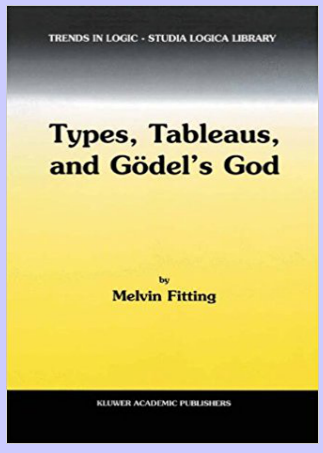
III. By 'substance' I mean that which is in itself, and is conceived through itself: in other words, that of which a conception can be formed independently of any other conception.

IV. By 'attribute' I mean that which the intellect perceives as constituting the essence of substance.

V. By 'mode' I mean the modifications ("affections") of substance, or that which exists in, and is conceived through, something other than itself.

VI. By 'God' I mean a being absolutely infinite—that is, a substance consisting in infinite attributes, of which each expresses eternal and infinite essentiality.

Explanation—*I say absolutely infinite, not infinite after its kind: for, of a thing infinite only after its kind, infinite attributes may be denied; but that which is absolutely infi-*


TRENDS IN LOGIC - STUDIA LOGICA LIBRARY

Types, Tableaus, and Gödel's God

by

Melvin Fitting

KLUWER ACADEMIC PUBLISHERS

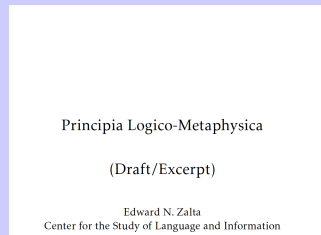

SYNTHESE LIBRARY / VOLUME 160

EDWARD N. ZALTA

ABSTRACT OBJECTS

An Introduction to Axiomatic Metaphysics

D. REIDEL PUBLISHING COMPANY
DORDRECHT / BOSTON / LANCASTER

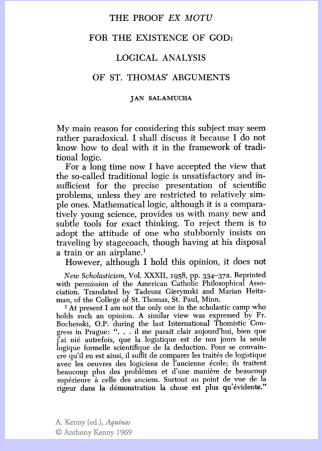

Principia Logico-Metaphysica

(Draft/Excerpt)

Edward N. Zalta
Center for the Study of Language and Information
Stanford University

October 28, 2016

http://mally.stanford.edu/principia.pdf


THE PROOF *EX MOTU*

FOR THE EXISTENCE OF GOD:

LOGICAL ANALYSIS

OF ST. THOMAS' ARGUMENTS

JAN SALAMUCHA

My main reason for considering this subject may seem rather paradoxical. I shall discuss it because I do not know how to deal with it in the framework of traditional logic.

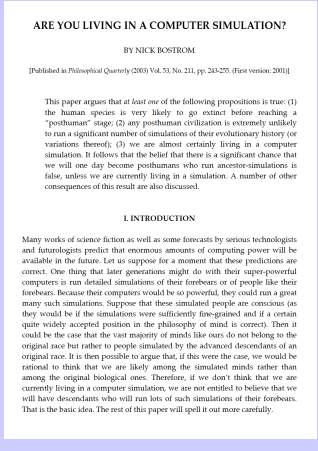
For a long time now I have accepted the view that the so-called traditional logic is unsatisfactory and insufficient for the precise presentation of scientific problems, unless they are restricted to relatively simple ones. Mathematical logic, although it is a comparatively young science, provides us with many new and subtle tools for exact thinking. To reject them is to adopt the attitude of one who stubbornly insists on traveling by stagecoach, though having at his disposal a train or an airplane.[1]

However, although I hold this opinion, it does not

Nous Scholasticon, Vol. XXXII, 1958, pp. 334–372. Reprinted with permission of the American Catholic Philosophical Association. Translated by Tadeusz Gieymski and Marian Heitzman, of the College of St. Thomas, St. Paul, Minn.

[1] At present I am not the only one in the scholastic camp who holds such an opinion. A similar view was expressed by Fr. Bocheński, O.P. during the last International Thomistic Congress in Prague: ". . . il me paraît clair aujourd'hui, bien que j'ai nié autrefois, que la logistique est de nos jours la seule logique formelle scientifique de la déduction. Pour se convaincre qu'il en est ainsi, il suffit de comparer les traités de logistique avec les oeuvres des logiciens de l'ancienne école; ils traitent beaucoup plus des problèmes et d'une manière de beaucoup supérieure à celle des anciens. Surtout au point de vue de la rigueur dans la démonstration la chose est plus qu'évidente."

A. Kenny (ed.), *Aquinas*


ARE YOU LIVING IN A COMPUTER SIMULATION?

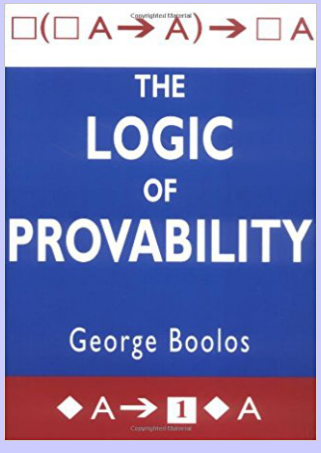
BY NICK BOSTROM

[Published in *Philosophical Quarterly* (2003) Vol. 53, No. 211, pp. 243-255. (First version: 2001)]

This paper argues that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. A number of other consequences of this result are also discussed.

I. INTRODUCTION

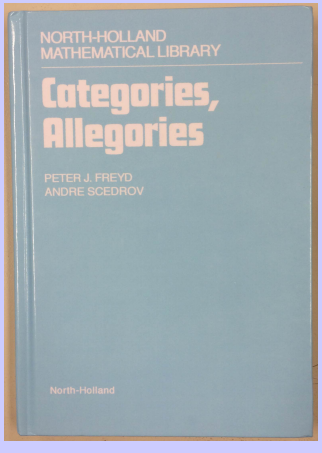
Many works of science fiction as well as some forecasts by serious technologists and futurologists predict that enormous amounts of computing power will be available in the future. Let us suppose for a moment that these predictions are correct. One thing that later generations might do with their super-powerful computers is run detailed simulations of their forebears or of people like their forebears. Because their computers would be so powerful, they could run a great many such simulations. Suppose that these simulated people are conscious (as they would be if the simulations were sufficiently fine-grained and if a certain quite widely accepted position in the philosophy of mind is correct). Then it could be the case that the vast majority of minds like ours do not belong to the original race but rather to people simulated by the advanced descendants of an original race. It is then possible to argue that, if this were the case, we would be rational to think that we are likely among the simulated minds rather than among the original biological ones. Therefore, if we don't think that we are currently living in a computer simulation, we are not entitled to believe that we will have descendants who will run lots of such simulations of their forebears. That is the basic idea. The rest of this paper will spell it out more carefully.


□(□A→A)→□A

THE LOGIC OF PROVABILITY

George Boolos

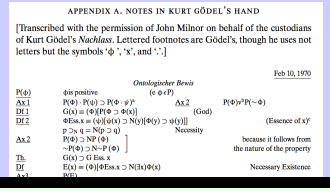
◆A→1◆A


NORTH-HOLLAND MATHEMATICAL LIBRARY

Categories, Allegories

PETER J. FREYD
ANDRE SCEDROV

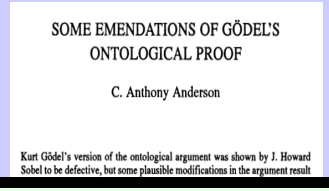
North-Holland


APPENDIX A. NOTES IN KURT GÖDEL'S HAND

['Transcribed with the permission of John Milnor on behalf of the custodians of Kurt Gödel's *Nachlass*. Lettered footnotes are Gödel's, though he uses not letters but the symbols '◊ ', '×', and '.']

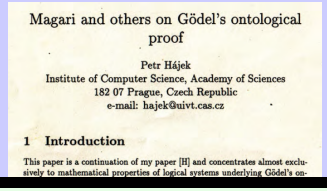
Feb 10, 1970

Ontologischer Beweis


SOME EMENDATIONS OF GÖDEL'S ONTOLOGICAL PROOF

C. Anthony Anderson

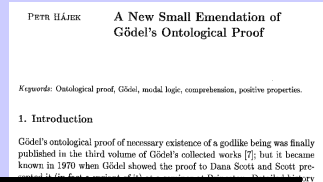
Kurt Gödel's version of the ontological argument was shown by J. Howard Sobel to be defective, but some plausible modifications in the argument result


Magari and others on Gödel's ontological proof

Petr Hájek
Institute of Computer Science, Academy of Sciences
182 07 Prague, Czech Republic
e-mail: hajek@uivt.cas.cz

1 Introduction

This paper is a continuation of my paper [H] and concentrates almost exclusively to mathematical properties of logical systems underlying Gödel's on-


PETR HÁJEK

A New Small Emendation of Gödel's Ontological Proof

Keywords: Ontological proof, Gödel, modal logic, comprehension, positive properties.

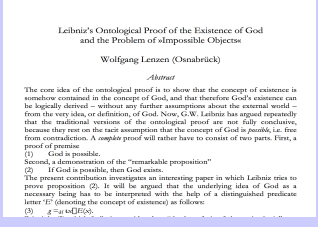
1. Introduction

Gödel's ontological proof of necessary existence of a godlike being was finally published in the third volume of Gödel's collected works [7]; but it became known in 1970 when Gödel showed the proof to Dana Scott and Scott pre-

Jww colleagues: formalisation of scientific articles and textbooks

- . . . in Philosophy, Maths, AI, CS
- . . . requiring very different logics

How possible in a single **Mathematical Proof Assistant** system?


Formal reconstruct
argument

Günther Eder · Esther

Received: 2 July 2014 / Accepted: 4 February 2015
© Springer Science+Business Media Dordrecht 2015

Abstract In this paper, we discuss formal reconstructions of Anselm's ontological argument. We first present a number of requirements that any successful reconstruction should meet. We then offer a detailed preparatory study of the basic concepts involved in Anselm's argument. Next, we present our own reconstructions—one in modal logic and one in classical logic—and compare them with each other and with existing reconstructions from the reviewed literature. Finally, we try to show why and how one can gain a better understanding of Anselm's argument by using modern formal logic. In particular, we try to explain why formal reconstructions of the argument, despite its apparent simplicity, tend to become quite involved.

Leibniz's Ontological Proof of the Existence of God
and the Problem of »Impossible Objects«

Wolfgang Lenzen (Osnabrück)

Abstract

The core idea of the ontological proof is to show that the concept of existence is somehow contained in the concept of God, and that therefore God's existence can be logically derived – without any further assumptions about the external world – from the very idea, or definition, of God. Now, G.W. Leibniz has argued repeatedly that the traditional versions of the ontological proof are not fully conclusive, because they rest on the tacit assumption that the concept of God is *possible*, i.e. free from contradiction. A *complete* proof will rather have to consist of two parts. First, a proof of premise

(1) God is possible.

Second, a demonstration of the "remarkable proposition"

(2) If God is possible, then God exists.

The present contribution investigates an interesting paper in which Leibniz tries to prove proposition (2). It will be argued that the underlying idea of God as a necessary being has to be interpreted with the help of a distinguished predicate letter 'E' (denoting the concept of existence) as follows:

(3) $g =_{df} \iota x[E(x)]$.


Spinoza

So, also, a thought is limited by another thou
not limited by thought, nor a thought by bod

III. By 'substance' I mean that which is in i
ceived through itself: in other words, that of
tion can be formed independently of any oth

IV. By 'attribute' I mean that which the int
constituting the essence of substance.

V. By 'mode' I mean the modifications ["affe
since, or that which exists in, and is conceiv
thing other than itself.

VI. By 'God' I mean a being absolutely in
substance consisting in infinite attributes, e
presses eternal and infinite essentiality.

Explanation—*I say absolutely infinite, b
in kind; for, of a thing infinite only after
attributes may be denied; but that which b*

Benedict de Spinoza

THE ETHICS

(Ethica Ordine Geometrico Demonstrata)

Translated by R. H. M. Elwes

PART I: CONCERNING GOD.

DEFINITIONS.

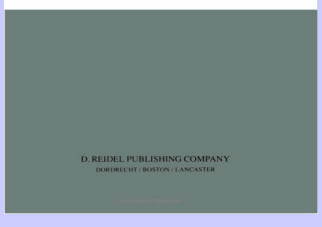
I. By that which is 'self-caused' I mean that of which the es-
sence involves existence, or that of which the nature is only
conceivable as existent.

II. A thing is called 'finite after its kind', when it can be limited
by another thing of the same nature; for instance, a body is
called finite because we always conceive another greater body.


Types, Tableaus, and Gödel's God

by
Melvin Fitting

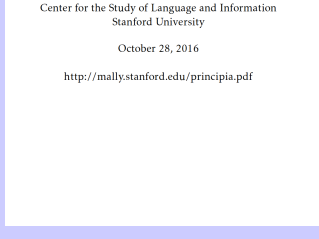
KLUWER ACADEMIC PUBLISHERS


EDWARD N. ZALTA

ABSTRACT OBJECTS

An Introduction to Axiomatic Metaphysics

D. REIDEL PUBLISHING COMPANY
DORDRECHT / BOSTON / LANCASTER

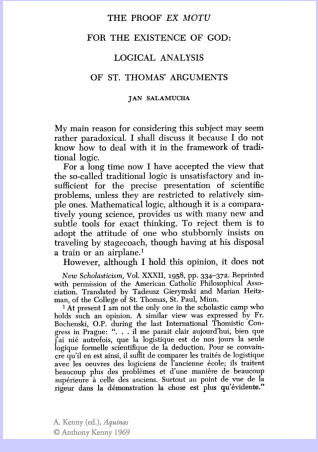

Principia Logico-Metaphysica

(Draft/Excerpt)

Edward N. Zalta
Center for the Study of Language and Information
Stanford University

October 28, 2016

http://mally.stanford.edu/principia.pdf


THE PROOF *EX MOTU*

FOR THE EXISTENCE OF GOD:

LOGICAL ANALYSIS

OF ST. THOMAS' ARGUMENTS

JAN SALAMUCHA

My main reason for considering this subject may seem rather paradoxical. I shall discuss it because I do not know how to deal with it in the framework of traditional logic.

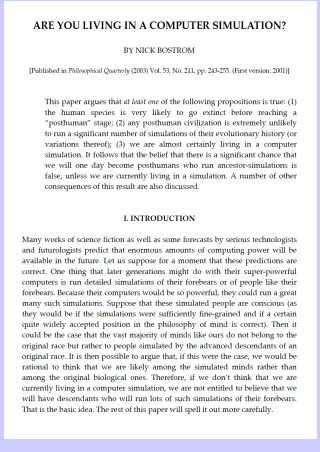
For a long time now I have accepted the view that the so-called traditional logic is unsatisfactory and insufficient for the precise presentation of scientific problems, unless they are restricted to relatively simple ones. Mathematical logic, although it is a comparatively young science, provides us with many new and subtle tools for exact thinking. To reject them is to adopt the attitude of one who stubbornly insists on traveling by stagecoach, though having at his disposal a train or an airplane.[1]

However, although I hold this opinion, it does not

Nova Scholastica, Vol. XXXII, 1958, pp. 334–372. Reprinted with permission of the American Catholic Philosophical Association. Translated by Tadeusz Gieymski and Marian Heitzman, of the College of St. Thomas, St. Paul, Minn.

[1] At present I am not the only one in the scholastic camp who holds such an opinion. A similar view was expressed by Fr. Bocheński, O.P. during the last International Thomistic Congress in Prague: ". . . il me paraît clair aujourd'hui, bien que j'ai nié autrefois, que la logistique est de nos jours la seule logique formelle scientifique de la déduction. Pour se convaincre qu'il en est ainsi, il suffit de comparer les traités de logistique avec les oeuvres des logiciens de l'ancienne école; ils traitent beaucoup plus des problèmes et d'une manière de beaucoup supérieure à celle des anciens. Surtout au point de vue de la rigueur dans la démonstration la chose est plus qu'évidente."

A. Kenny (ed.), *Aquinas*
© Anthony Kenny 1969


ARE YOU LIVING IN A COMPUTER SIMULATION?

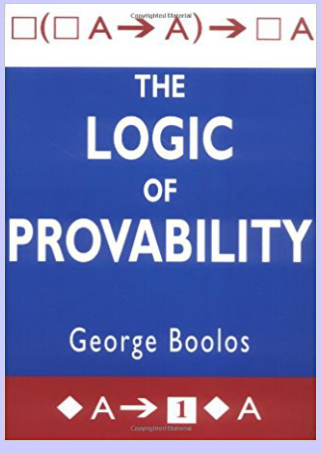
BY NICK BOSTROM

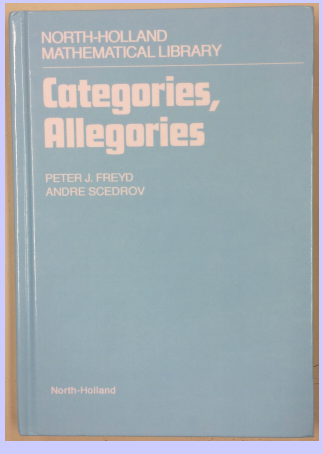
[Published in *Philosophical Quarterly* (2003) Vol. 53, No. 211, pp. 243-255. (First version: 2001)]

This paper argues that *at least one* of the following propositions is true: (1) the human species is very likely to go extinct before reaching a "posthuman" stage; (2) any posthuman civilization is extremely unlikely to run a significant number of simulations of their evolutionary history (or variations thereof); (3) we are almost certainly living in a computer simulation. It follows that the belief that there is a significant chance that we will one day become posthumans who run ancestor-simulations is false, unless we are currently living in a simulation. A number of other consequences of this result are also discussed.

I. INTRODUCTION

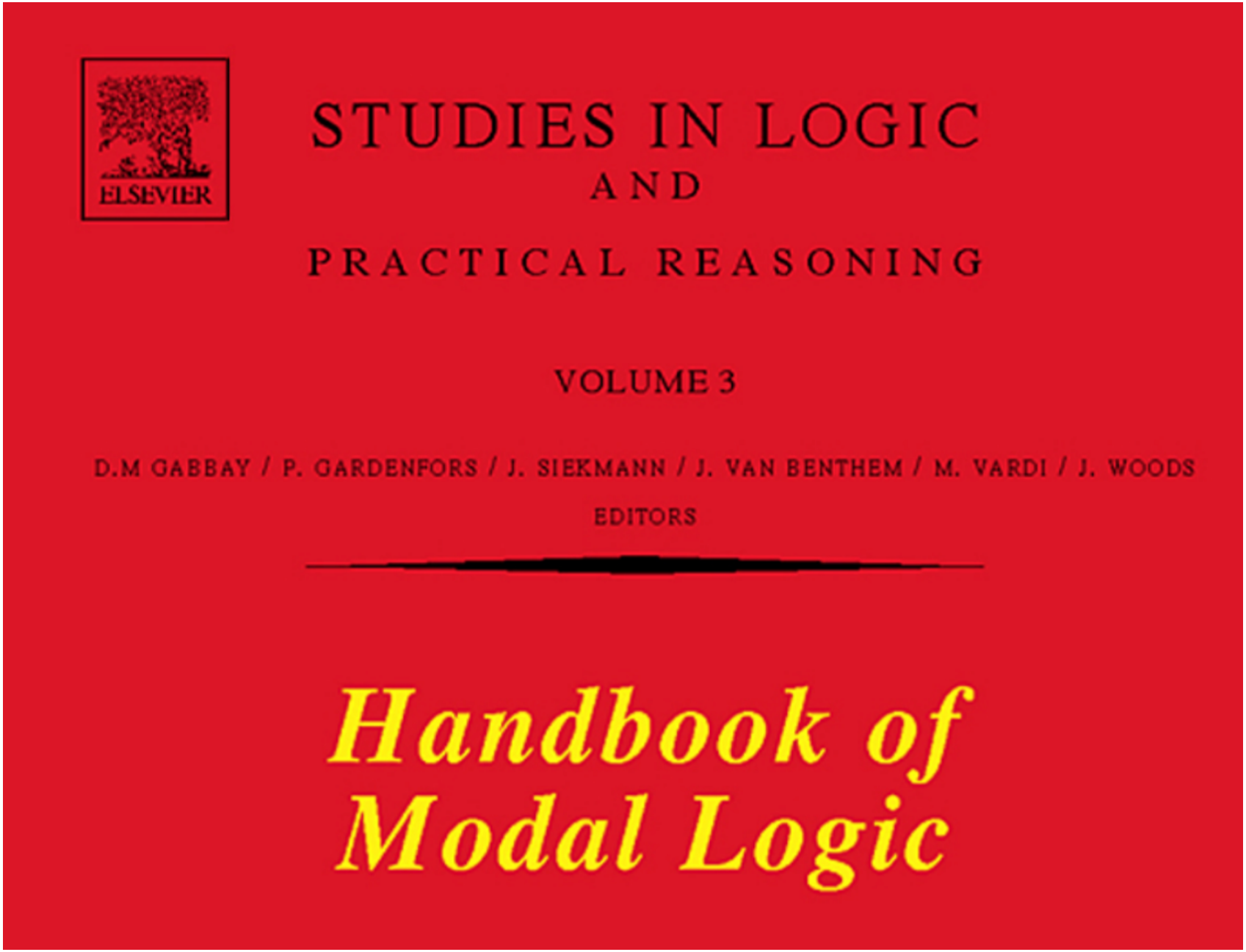
Many works of science fiction as well as some forecasts by serious technologists and futurologists predict that enormous amounts of computing power will be available in the future. Let us suppose for a moment that these predictions are correct. One thing that later generations might do with their super-powerful computers is run detailed simulations of their forebears or of people like their forebears. Because their computers would be so powerful, they could run a great many such simulations. Suppose that these simulated people are conscious (as they would be if the simulations were sufficiently fine-grained and if a certain quite widely accepted position in the philosophy of mind is correct). Then it could be the case that the vast majority of minds like ours do not belong to the original race but rather to people simulated by the advanced descendants of an original race. It is then possible to argue that, if this were the case, we would be rational to think that we are likely among the simulated minds rather than among the original biological ones. Therefore, if we don't think that we are currently living in a computer simulation, we are not entitled to believe that we will have descendants who will run lots of such simulations of their forebears. That is the basic idea. The rest of this paper will spell it out more carefully.





. . .

Example: Modal Logic Textbook



Example: Modal Logic Textbook

2 BASIC MODAL LOGIC

In this section we introduce the basic modal language and its relational semantics. We define basic modal syntax, introduce models and frames, and give the satisfaction definition. We then draw the reader's attention to the internal perspective that modal languages offer on relational structure, and explain why models and frames should be thought of as graphs. Following this we give the standard translation. This enables us to convert any basic modal formula into a first-order formula with one free variable. The standard translation is a bridge between the modal and classical worlds, a bridge that underlies much of the work of this chapter.

2.1 *First steps in relational semantics*

Suppose we have a set of proposition symbols (whose elements we typically write as p, q, r and so on) and a set of modality symbols (whose elements we typically write as m, m', m'', and so on). The choice of PROP and MOD is called the *signature* (or *similarity type*) of the language; in what follows we'll tacitly assume that PROP is denumerably infinite, and we'll often work with signatures in which MOD contains only a single element. Given a signature, we define the *basic modal language* (over the signature) as follows:

$$\varphi \quad ::= \quad p \mid \top \mid \bot \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi \mid \varphi \leftrightarrow \psi \mid \langle m \rangle \varphi \mid [m]\varphi.$$

That is, a basic modal formula is either a proposition symbol, a boolean constant, a boolean combination of basic modal formulas, or (most interesting of all) a formula prefixed by a diamond

Example: Modal Logic Textbook

2 BASIC MODAL LOGIC

In this section we introduce the basic modal language and its relational semantics. We define basic modal syntax, introduce models and frames, and give the satisfaction definition. We then draw the reader's attention to the internal perspective that modal languages offer on relational structure, and explain why models and frames should be thought of as graphs. Following this we give the standard translation. This enables us to convert any basic modal formula into a first-order formula with one free variable. The standard translation is a bridge between the modal and classical worlds, a bridge that underlies much of the work of this chapter.

2.1 First steps in relational semantics

Syntax

Metalanguage

ose elements we typically write as p, q, r and
ents we typically write as m, m', m'', and so
nature (or *similarity type*) of the language; in
what follows we'll tacitly assume that PROP is denumerably infinite, and we'll often work with signatures in which MOD contains only a single element. Given a signature, we define the *basic modal language* (over the signature) as follows:

$$\varphi \quad ::= \quad p \mid \top \mid \bot \mid \neg\varphi \mid \varphi \wedge \psi \mid \varphi \vee \psi \mid \varphi \rightarrow \psi \mid \varphi \leftrightarrow \psi \mid \langle m \rangle \varphi \mid [m]\varphi.$$

That is, a basic modal formula is either a proposition symbol, a boolean constant, a boolean combination of basic modal formulas, or (most interesting of all) a formula prefixed by a diamond

Example: Modal Logic Textbook

A *model* (or *Kripke model*) $\mathfrak{M}$ for the basic modal language (over some fixed signature) is a triple $\mathfrak{M} = (W, \{R^m\}_{m \in \text{MOD}}, V)$. Here W, the *domain*, is a non-empty set, whose elements we usually call *points*, but which, for reasons which will soon be clear, are sometimes called *states*, *times*, *situations*, *worlds* and other things besides. Each R^m in a model is a binary relation on W, and V is a function (the valuation) that assigns to each proposition symbol p in PROP a subset $V(p)$ of W; think of $V(p)$ as the set of points in $\mathfrak{M}$ where p is true. The first two components $(W, \{R^m\}_{m \in \text{MOD}})$ of $\mathfrak{M}$ are called the *frame* underlying the model. If there is only one relation in the model, we typically write (W, R) for its frame, and (W, R, V) for the model itself. We encourage the reader to think of Kripke models as graphs (or to be slightly more precise, *directed graphs*, that is, graphs whose points are linked by directed arrows) and will shortly give some examples which show why this is helpful.

Suppose w is a point in a model $\mathfrak{M} = (W, \{R^m\}_{m \in \text{MOD}}, V)$. Then we inductively define the notion of a formula φ being *satisfied* (or *true*) in $\mathfrak{M}$ at point w as follows (we omit some of the clauses for the booleans):

$\mathfrak{M}, w \models p$	iff	$w \in V(p)$,
$\mathfrak{M}, w \models \top$		always,
$\mathfrak{M}, w \models \bot$		never,
$\mathfrak{M}, w \models \neg\varphi$	iff	not $\mathfrak{M}, w \models \varphi$ (notation: $\mathfrak{M}, w \not\models \varphi$),
$\mathfrak{M}, w \models \varphi \wedge \psi$	iff	$\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$,
$\mathfrak{M}, w \models \varphi \rightarrow \psi$	iff	$\mathfrak{M}, w \not\models \varphi$ or $\mathfrak{M}, w \models \psi$,
$\mathfrak{M}, w \models \langle m \rangle \varphi$	iff	for some $v \in W$ such that $R^m wv$ we have $\mathfrak{M}, v \models \varphi$,
$\mathfrak{M}, w \models [m] \varphi$	iff	for all $v \in W$ such that $R^m wv$ we have $\mathfrak{M}, v \models \varphi$.

Example: Modal Logic Textbook

A *model* (or *Kripke model*) $\mathfrak{M}$ for the basic modal language (over some fixed signature) is a triple $\mathfrak{M} = (W, \{R^m\}_{m \in \text{MOD}}, V)$. Here W, the *domain*, is a non-empty set, whose elements we usually call *points*, but which, for reasons which will soon be clear, are sometimes called *states*, *times*

in a model is a binary relation on W,

and

position symbol p in PROP a subset

$V(p)$

p is true. The first two components

$(W, \{$

model. If there is only one relation

in the

(W, R, V) for the model itself. We

Metalanguage

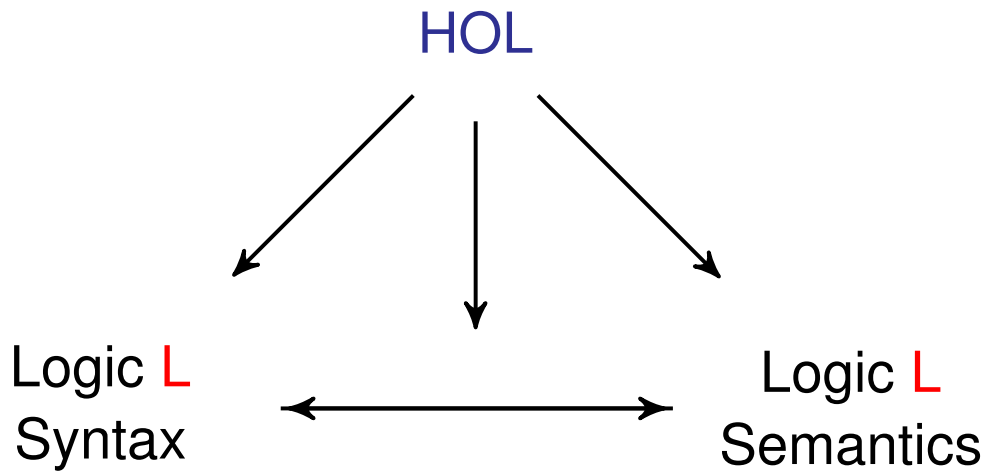
encourage the reader to think of Kripke models as graphs (or to be slightly more precise, *directed graphs*, that is, graphs whose points are linked by directed arrows) and will shortly give some examples which show why this is helpful.

Suppose w is a point in a model $\mathfrak{M} = (W, \{R^m\}_{m \in \text{MOD}}, V)$. Then we inductively define the notion of a formula φ being *satisfied* (or *true*) in $\mathfrak{M}$ at point w as follows (we omit some of the clauses for the booleans):

Semantics

$\mathfrak{M}, w \models p$	iff	$w \in V(p)$,
$\mathfrak{M}, w \models \top$		always,
$\mathfrak{M}, w \models \bot$		never,
$\mathfrak{M}, w \models \neg\varphi$	iff	not $\mathfrak{M}, w \models \varphi$ (notation: $\mathfrak{M}, w \not\models \varphi$),
$\mathfrak{M}, w \models \varphi \wedge \psi$	iff	$\mathfrak{M}, w \models \varphi$ and $\mathfrak{M}, w \models \psi$,
$\mathfrak{M}, w \models \varphi \rightarrow \psi$	iff	$\mathfrak{M}, w \not\models \varphi$ or $\mathfrak{M}, w \models \psi$,
$\mathfrak{M}, w \models \langle m \rangle \varphi$	iff	for some $v \in W$ such that $R^m wv$ we have $\mathfrak{M}, v \models \varphi$,
$\mathfrak{M}, w \models [m]\varphi$	iff	for all $v \in W$ such that $R^m wv$ we have $\mathfrak{M}, v \models \varphi$.

Universal Reasoning in Meta-Logic HOL



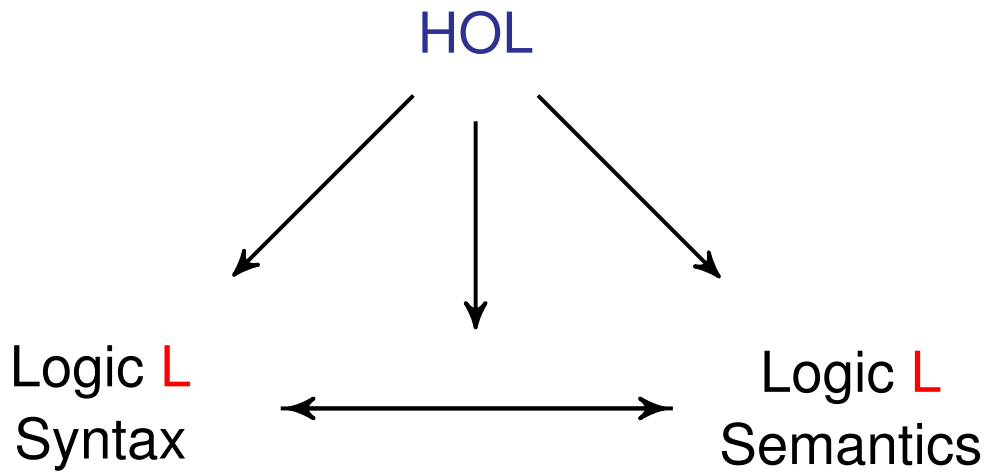
Examples for L we have already studied:
Intuitionistic Logics, (Mathematical) Fuzzy Logics, Free Logic, Modal Logics, Description Logics, Conditional Logics, Access Control Logics, Hybrid Logics, Multivalued Logics, Logics with Neighborhood Semantics, Paraconsistent Logics, Dyadic Deontic Logic, ...

Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

- interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...
- automated: Leo-III, LEO-II, Satallax, TPS, Nitpick, Isabelle/HOL, ...

Universal Reasoning in Meta-Logic HOL



Examples for L we have already studied:

Intuitionistic Logics, (Mathematical) Fuzzy Logics, Free Logic, Modal Logics, Description Logics, Conditional Logics, Access Control Logics, Hybrid Logics, Multivalued Logics, Logics with Neighborhood Semantics, Paraconsistent Logics, Dyadic Deontic Logic, ...

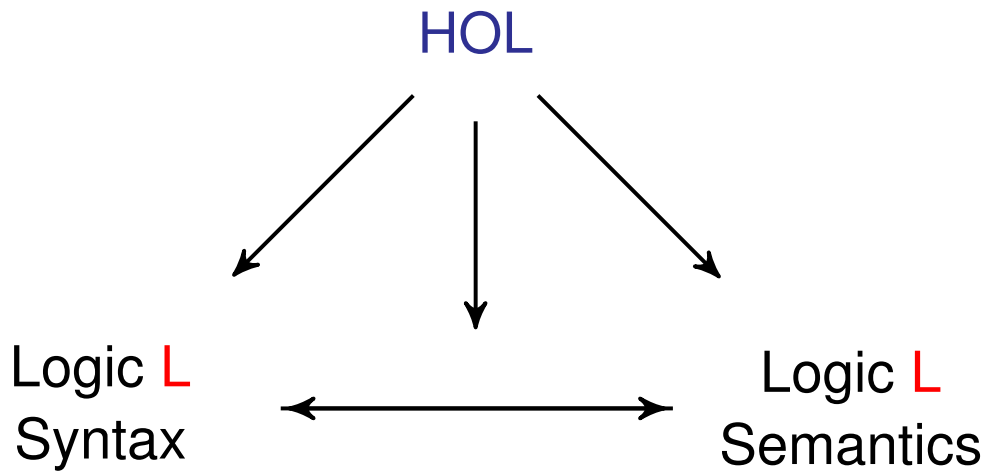
Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

interactive: Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...

automated: Leo-III, LEO-II, Satallax, TPS, Nitpick, Isabelle/HOL, ...

Universal Reasoning in Meta-Logic HOL



Examples for L we have already studied:

Intuitionistic Logics, (Mathematical) Fuzzy Logics, Free Logic, Modal Logics, Description Logics, Conditional Logics, Access Control Logics, Hybrid Logics, Multivalued Logics, Logics with Neighborhood Semantics, Paraconsistent Logics, Dyadic Deontic Logic, ...

Embedding works also for quantifiers (first-order & higher-order)

HOL provers become universal logic reasoning engines!

interactive:	Isabelle/HOL, PVS, HOL4, Hol Light, Coq/HOL, ...
automated:	Leo-III, LEO-II, Satallax, TPS, Nitpick, Isabelle/HOL, ...

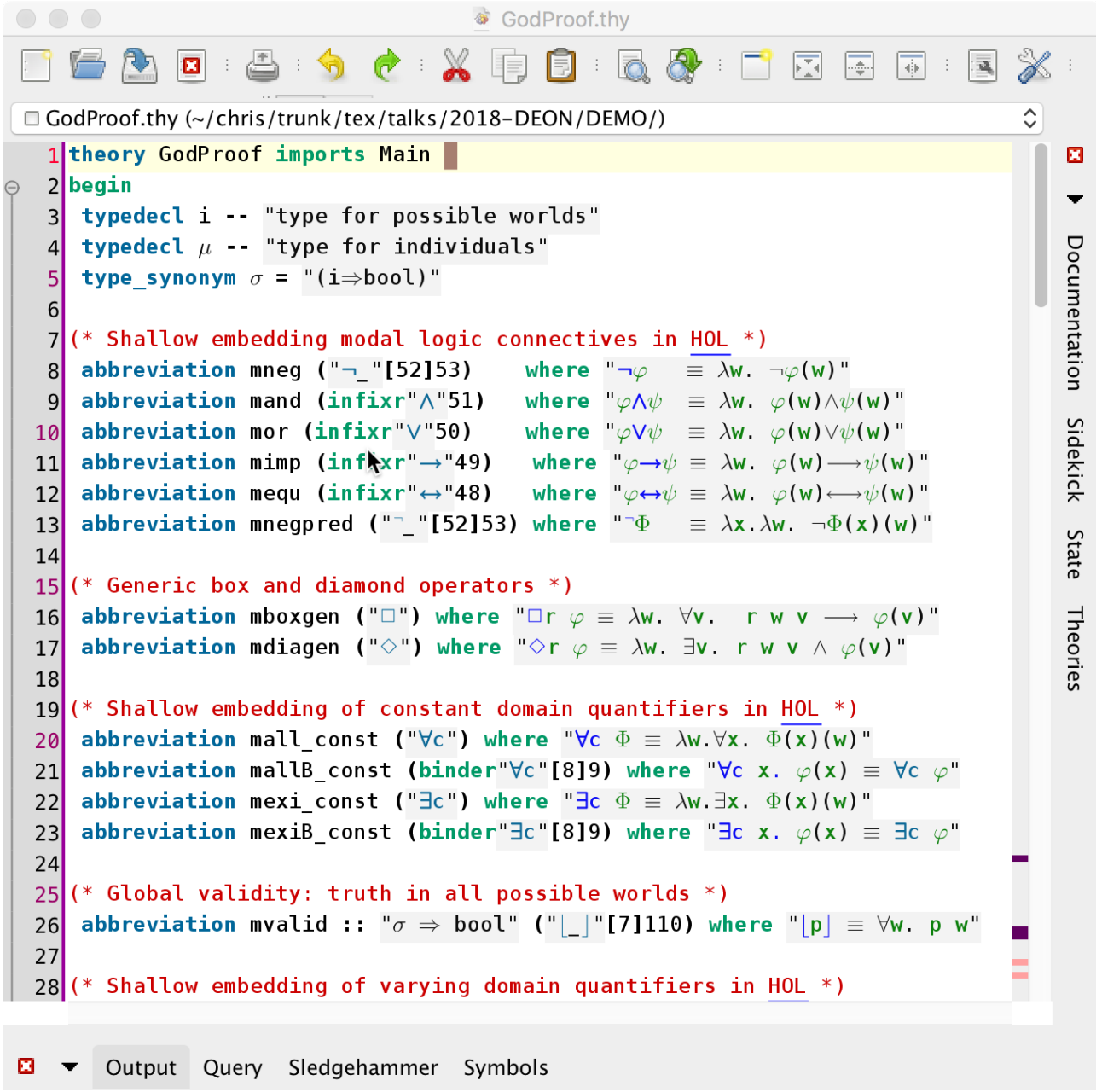
Isabelle/HOL (one of various Theorem Provers for HOL)

https://isabelle.in.tum.de

many other systems:

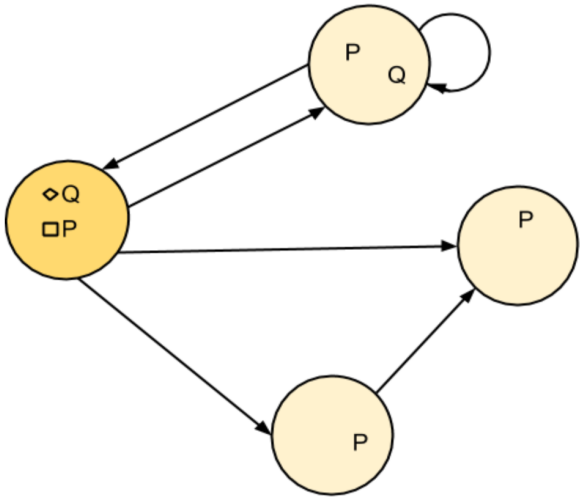
Coq, HOL, HOL Light, PVS, Lean, NuPrL, IMPS, ACL2, **Leo-II/Leo-III**, . . .

Universal Reasoning in Isabelle/HOL



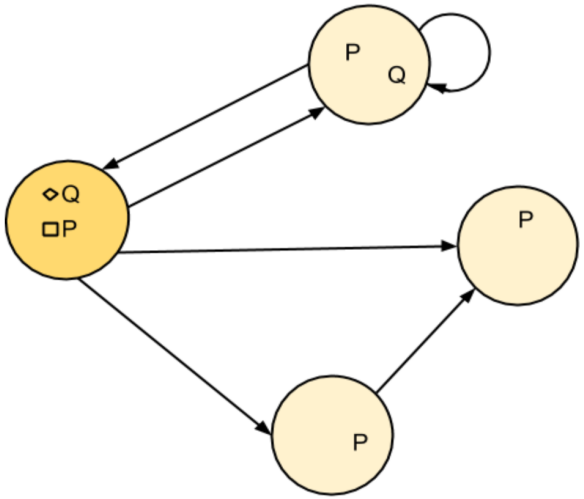
Universal Logic Reasoning in Isabelle/HOL

Properties of □ and ◇ correlated to structure of transition system between worlds



Universal Logic Reasoning in Isabelle/HOL

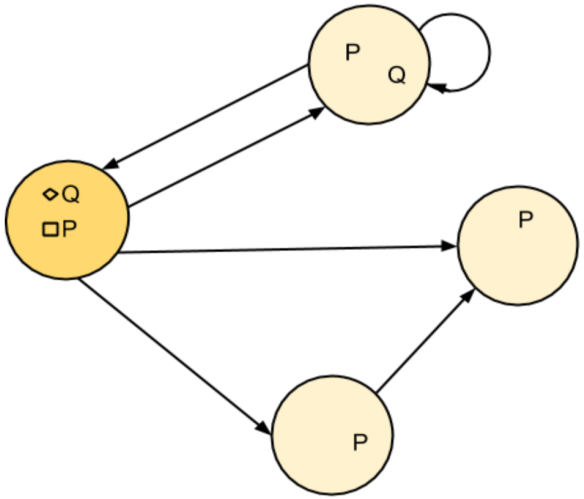
Properties of $\Box$ and $\Diamond$ correlated to structure of transition system between worlds



- Logic K: — (no restrictions, any structure)
- Logic M: reflexiv transition relation, $\forall P.\Box P \rightarrow P$
- Logic KB: symmetric transition relation, $\forall P.P \rightarrow \Box\Diamond P$
- Logic S5: equivelance relation as transition system, add $\forall P.\Box P \rightarrow \Box\Box P$

Universal Logic Reasoning in Isabelle/HOL

Properties of $\Box$ and $\Diamond$ correlated to structure of transition system between worlds

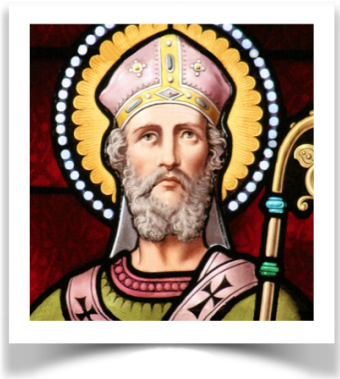


- Logic K: — (no restrictions, any structure)
- Logic M: reflexiv transition relation, $\forall P.\Box P \rightarrow P$
- Logic KB: symmetric transition relation, $\forall P.P \rightarrow \Box\Diamond P$
- Logic S5: equivelance relation as transition system, add $\forall P.\Box P \rightarrow \Box\Box P$

- Logic D: serial transition relation, $\forall P.\Box P \rightarrow \Diamond P$ (**S**tandard **D**eontic **L**ogic)

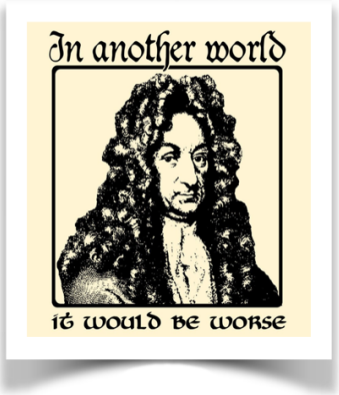
Ontological Proofs of God's Existence

A Long and Continuing Tradition in Philosophy



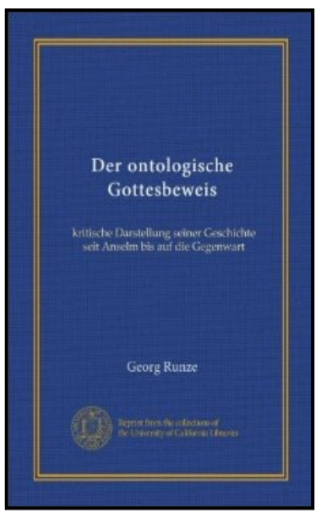
St. Anselm

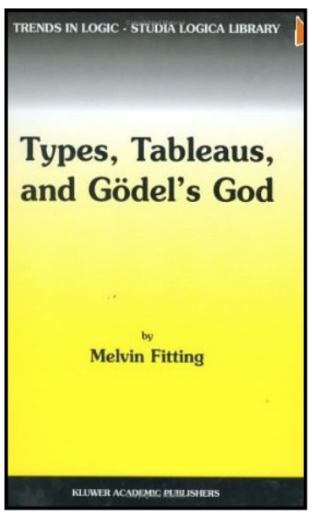
Descartes

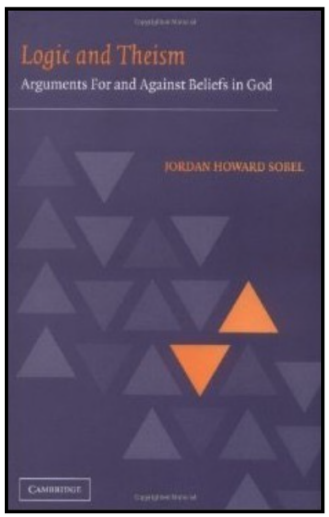


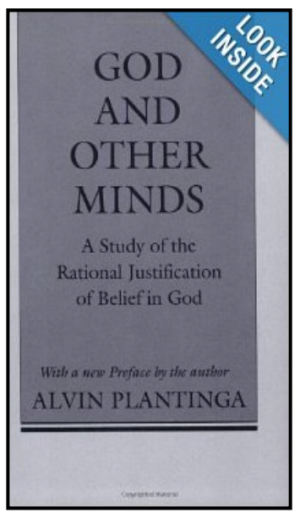
Leibniz

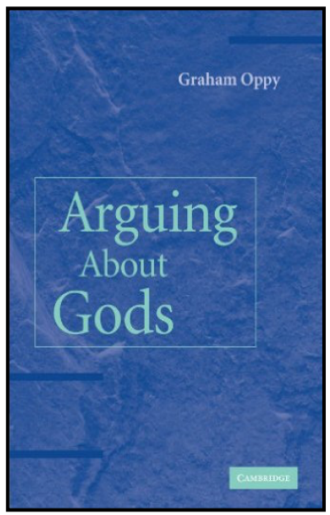
Gödel



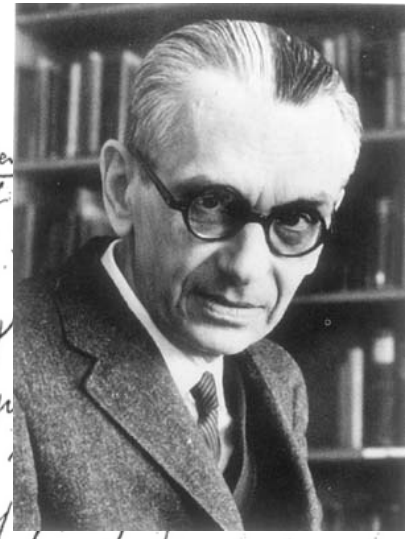








Computational Metaphysics: Kurt Gödel's Ontological Argument



Ontologischer Beweis Feb 10, 1970

$P(\varphi)$ φ is positive ($\varphi \in P$)

Ax. 1 $P(\varphi) \cdot P(\psi) \supset P(\varphi \cdot \psi)$ • Ax 2 $P(\varphi) \vee P(\sim\varphi)$

D1 $G(x) \equiv (\varphi)[P(\varphi) \supset \varphi(x)]$ (God)

D2 $\varphi \, Ess_x \equiv (\psi)[\psi(x) \supset N(y)[\varphi(y) \supset \psi(y)]]$ (Essence of x)

$p \supset_N q \;=\; N(p \supset q)$ Necessity

Ax 2 $P(\varphi) \supset N P(\varphi)$ } because it follows from the nature of the property
$\sim P(\varphi) \supset N \sim P(\varphi)$ }

Th. $G(x) \supset G \, Ess. \, x$

Df. $E(x) \equiv (\varphi)[\varphi \, Ess \, x \supset N \exists x \, \varphi(x)]$ necessary Existen

Ax 3 $P(E)$

Th. $G(x) \supset N(\exists y) G(y)$

hence $(\exists x) G(x) \supset N(\exists y) G(y)$

" $M(\exists x) G(x) \supset MN(\exists y) G(y)$

" $\supset N(\exists y) G(y)$ M = possibility

any two essences of x are nec. equivalent

exclusive or • and for any number of summands

M $(\exists x) G(x)$ means "all [the syste

patible This is

Ax 4 : $P(\varphi) \cdot \varphi \supset_N \psi$

~~Axiom~~ { $x = x$ is p
~~end~~ { $x \neq x$ is

But if a system S of

It would mean that the sum prop. s (which is positive) would be $x \neq x$

Positive means positive in the moral aesthe sense (independently of the accidental structure of the world). Only then the ax. true. It m also meant "attribution" as opposed to "privation pure (or containing privation) - This interpret. makes proof

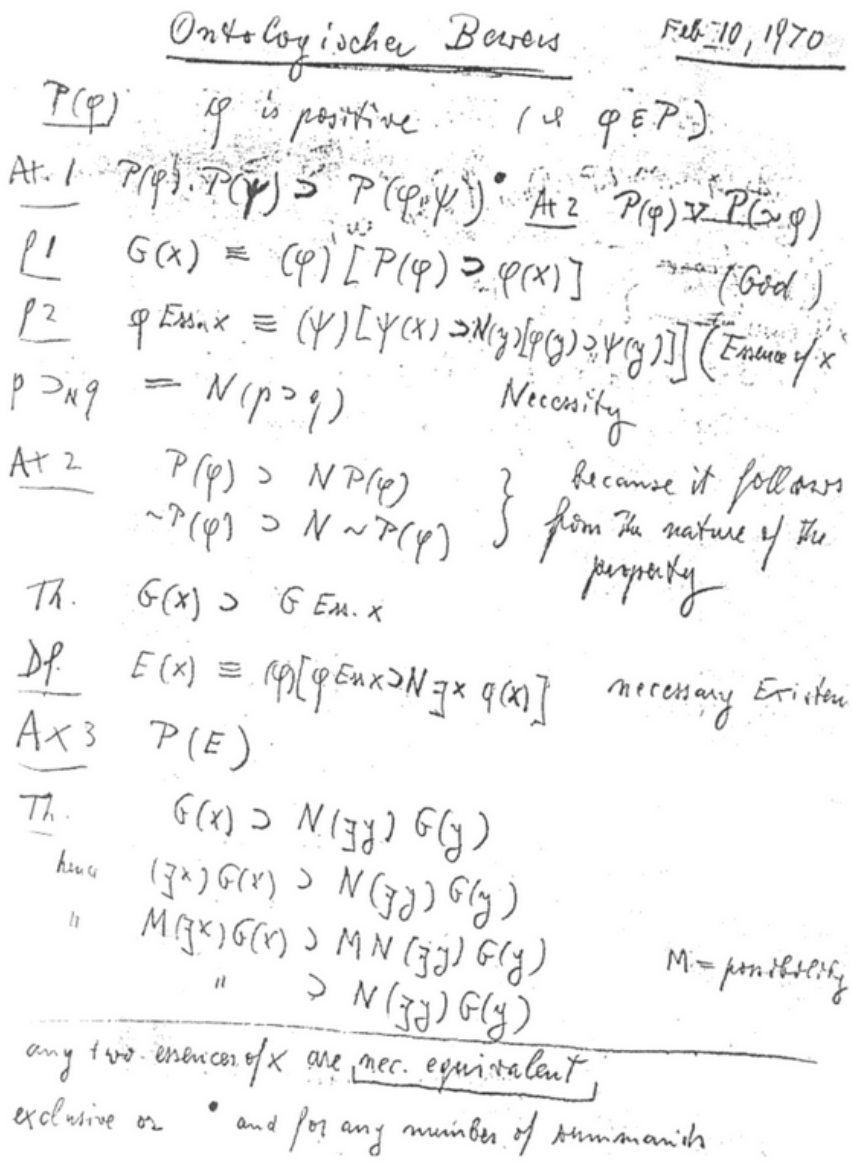
If φ positive: $(x) N \sim \varphi(x)$ Otherwise $\varphi(x) \supset_N x \neq$

hence $x \neq x$, positive not $x = x$ by contrary Ax. or the exist. of pos. prop.

dug

x i.e. the normal form in terms of elem. prop. contains a member without negation.

Computational Metaphysics: Kurt Gödel's Ontological Argument



Ontologischer Beweis Feb 10, 1970

$P(\varphi)$ φ is positive ($\varphi \in P$)

Ax. 1 $P(\varphi) \cdot P(\psi) \supset P(\varphi \cdot \psi)$ • Ax 2 $P(\varphi) \vee P(\neg\varphi)$

D1 $G(x) \equiv (\varphi)[P(\varphi) \supset \varphi(x)]$ (God)

D2 $\varphi \, Ess_x \equiv (\psi)[\psi(x) \supset N(y)[\varphi(y) \supset \psi(y)]]$ (Essence of x

$p \supset_N q \quad = N(p \supset q)$ Necessity

Ax 2 $P(\varphi) \supset N P(\varphi)$
$\neg P(\varphi) \supset N \neg P(\varphi)$
} because it follows from the nature of the property

Th. $G(x) \supset G \, Ess. \, x$

Df. $E(x) \equiv (\varphi)[\varphi \, Ess \, x \supset N \exists x \, \varphi(x)]$ necessary Existen

Ax 3 $P(E)$

Th. $G(x) \supset N(\exists y) G(y)$

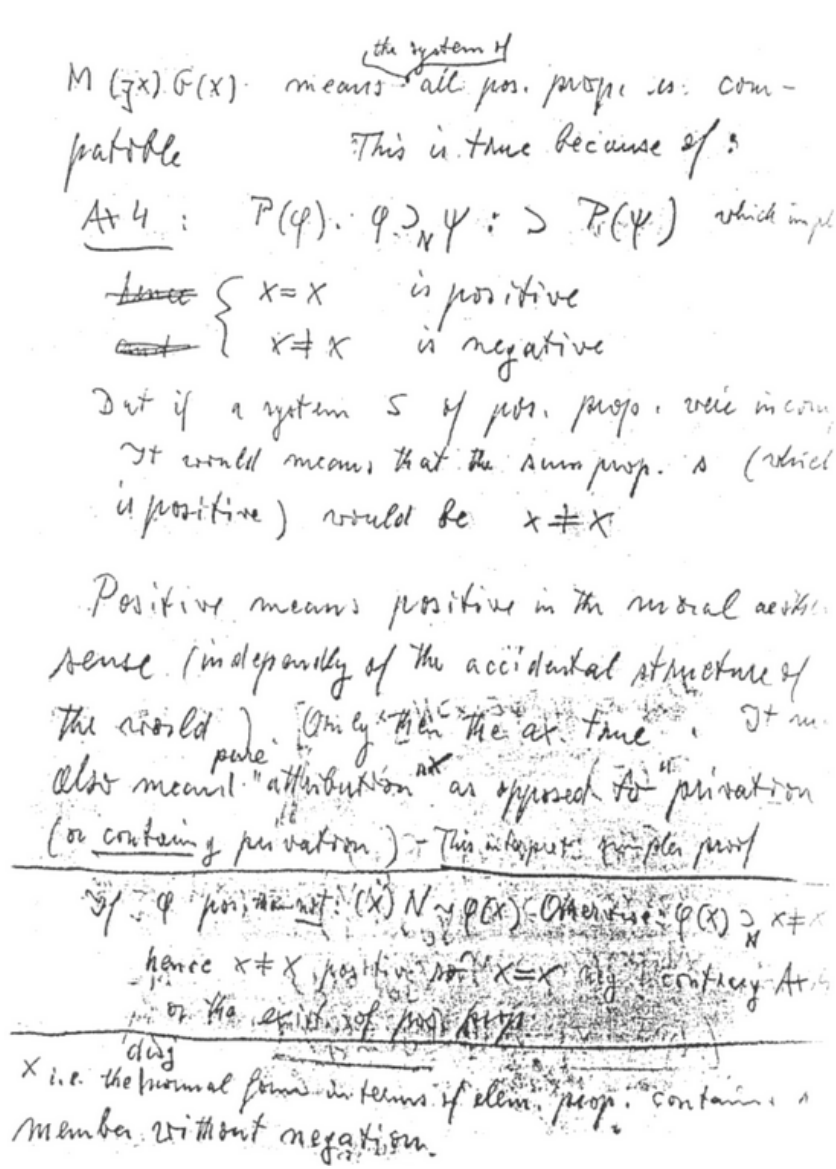
hence $(\exists x) G(x) \supset N(\exists y) G(y)$

" $M(\exists x) G(x) \supset MN(\exists y) G(y)$

" $\supset N(\exists y) G(y)$ M = possibility

any two essences of x are nec. equivalent

exclusive or • and for any number of summands



$M(\exists x) G(x)$ means (the system of) all pos. prop. is com-patible This is true because of:

Ax 4 : $P(\varphi) \cdot \varphi \supset_N \psi : \supset P(\psi)$ which imp

~~Axiom~~ { $x = x$ is positive
~~end~~ { $x \neq x$ is negative

But if a system S of pos. propo. were in com
it would mean that the sum prop. is (which
is positive) would be $x \neq x$

Positive means positive in the moral aesth
sense (independly of the accidental structure of
the world). Only then the ax. true. It m
also meant "attribution" as opposed to "privation
(or containg privation) - This interpret. implies proof

If φ positive: $(x) N \neg \varphi(x)$ Otherwise $\varphi(x) \supset_N x \neq x$
hence $x \neq x$ positive not $x = x$ by contrary Ax
or the exist. of pos. prop.

dwg

x i.e. the normal form in terms of elem. prop. contain
member without negation.

Computational Metaphysics: Dana Scott's Variant

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists x G(x)$

Computational Metaphysics: Scott's Variant

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists x G(x)$

Difference to Gödel (who omits this conjunct)

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists x G(x)$

Modal operators are used

Computational Metaphysics: Scott's Variant of Gödel's Ontological Argument

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists x G(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists x G(x)$

second-order quantifiers

Computational Metaphysics: Vision of Leibniz (1646–1716) — *Calculemus!*

If controversies were to arise, there would be no more need of disputation between two philosophers than between two accountants. For it would suffice to take their pencils in their hands, to sit down to their slates, and to say to each other . . . : Let us calculate.

(Translation by Russell)

Quo facto, quando orientur controversiae, non magis disputatione opus erit inter duos philosophos, quam inter duos Computistas. Sufficiet enim calamos in manus sumere sedereque ad abacos, et sibi mutuo . . . dicere: calculemus. (Leibniz, 1684)

Required:
characteristica universalis and **calculus ratiocinator**

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 Either a property or its negation is positive, but not both: $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 A property necessarily implied by a positive property is positive: $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 Positive properties are possibly exemplified: $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 A *God-like* being possesses all positive properties: $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 The property of being God-like is positive: $P(G)$

Cor. C Possibly, God exists: $\Diamond\exists xG(x)$

Axiom A4 Positive properties are necessarily positive: $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 An *essence* of an individual is a property possessed by it and necessarily implying any of its properties: $\phi\ ess.\ x \leftrightarrow \phi(x) \wedge \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 Being God-like is an essence of any God-like being: $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 *Necessary existence* of an individual is the necessary exemplification of all its essences: $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 Necessary existence is a positive property: $P(NE)$

Thm. T3 Necessarily, God exists: $\Box\exists xG(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Thm. T1 $\forall\phi[P(\phi) \rightarrow \Diamond\exists x\phi(x)]$

Def. D1 $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 $P(G)$

Cor. C $\Diamond\exists xG(x)$

Axiom A4 $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

Def. D2 $\phi\ ess.\ x \leftrightarrow \phi(x)\ \wedge\ \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

Thm. T2 $\forall x[G(x) \rightarrow G\ ess.\ x]$

Def. D3 $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 $P(NE)$

Thm. T3 $\Box\exists xG(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo

Axiom A1 $\forall\phi[P(\neg\phi) \leftrightarrow \neg P(\phi)]$

Axiom A2 $\forall\phi\forall\psi[(P(\phi) \wedge \Box\forall x[\phi(x) \rightarrow \psi(x)]) \rightarrow P(\psi)]$

Def. D1 $G(x) \leftrightarrow \forall\phi[P(\phi) \rightarrow \phi(x)]$

Axiom A3 $P(G)$

Axiom A4 $\forall\phi[P(\phi) \rightarrow \Box P(\phi)]$

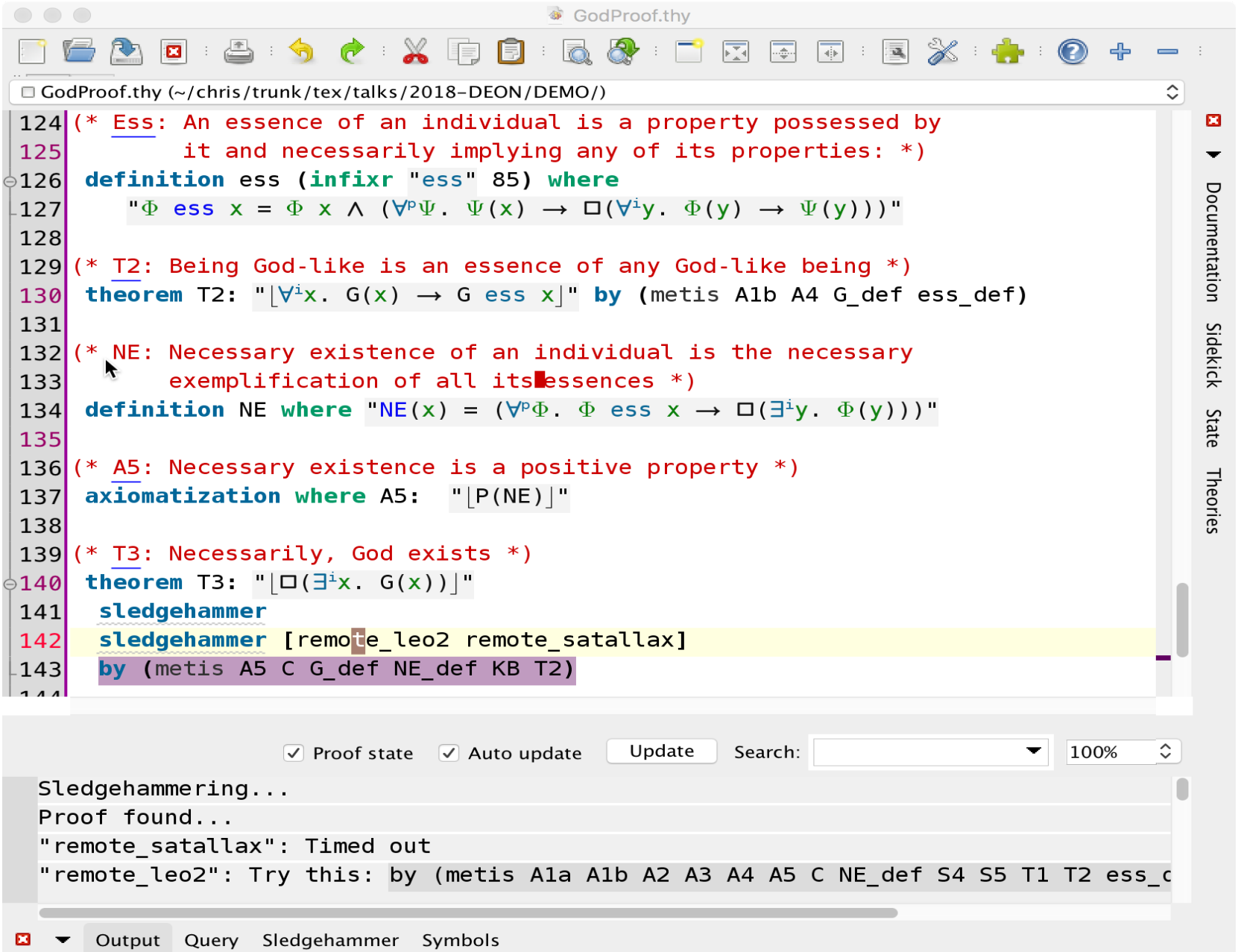
Def. D2 $\phi\ ess.\ x \leftrightarrow \phi(x)\ \wedge\ \forall\psi(\psi(x) \rightarrow \Box\forall y(\phi(y) \rightarrow \psi(y)))$

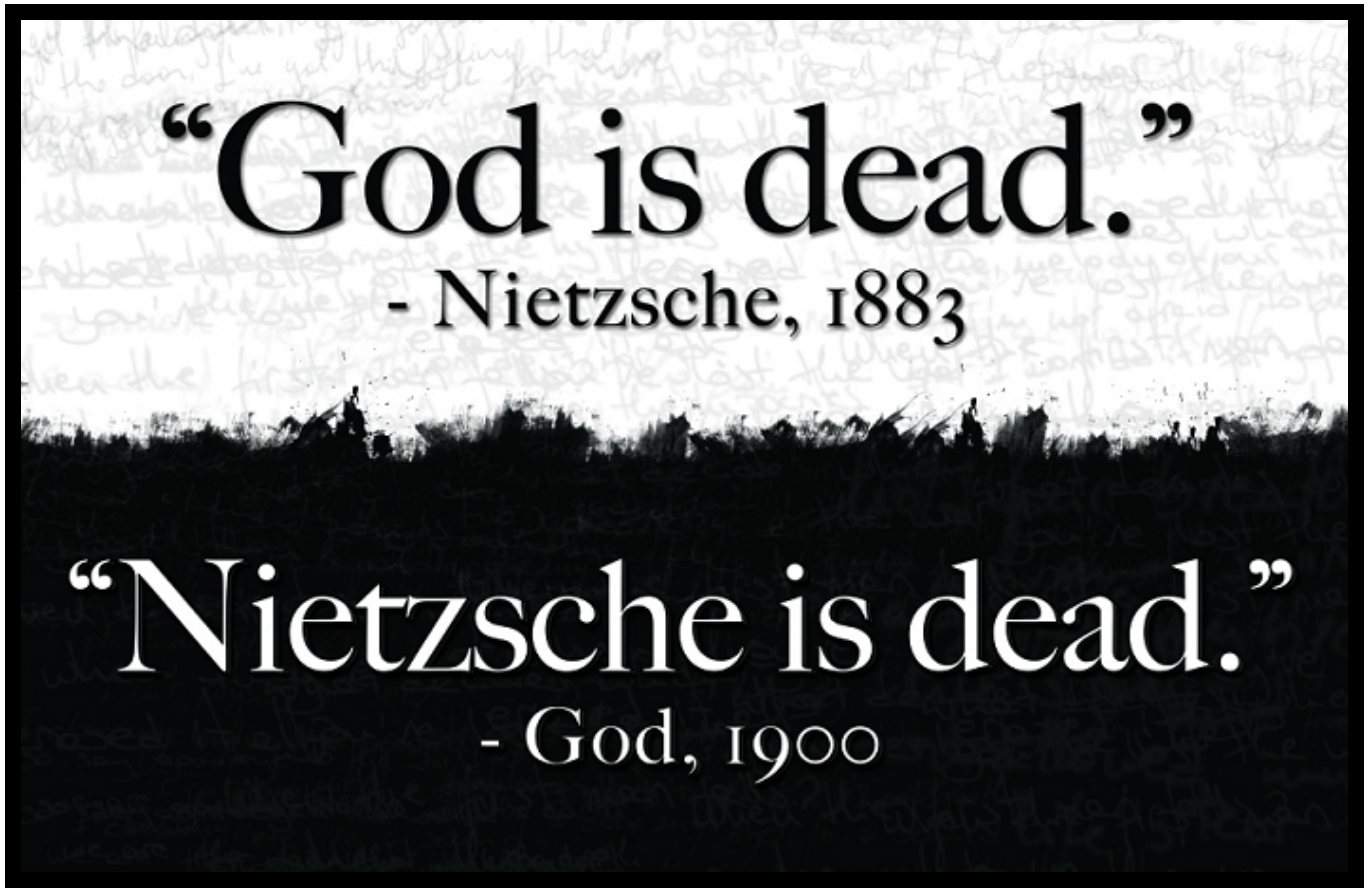
Def. D3 $NE(x) \leftrightarrow \forall\phi[\phi\ ess.\ x \rightarrow \Box\exists y\phi(y)]$

Axiom A5 $P(NE)$

Thm. T3 $\Box\exists x G(x)$

Computational Metaphysics: Scott's and Gödel's Variants — Demo





Part C
Evidence: Analysing Rational Arguments in Metaphysics

[BenzmüllerWoltzenlogelPaleo, ECAI, 2014 + IJCAI, 2016 + KI 2016 + . . .]

Results of our Experiments

Variant of Dana Scott

- premises are **consistent**
- all argument steps are **logically correct**
  - in logic **S5**
  - already in logic **KB**

Intermediate Conclusion:

With our technology...
... it is possible to verify (selected) masterpiece arguments in philosophy.

Results of our Experiments

Variant of Dana Scott

- premises are **consistent**
- all argument steps are **logically correct**
  - in logic **S5**
  - already in logic **KB**

Intermediate Conclusion:

With our technology...
... it is possible to verify (selected) masterpiece arguments in philosophy.

Results of our Experiments

Variant of Dana Scott

- premises are **consistent**
- all argument steps are **logically correct**
  – in logic **S5**
  – already in logic **KB**

Intermediate Conclusion:

With our technology. . .
. . . it is possible to verify (selected) masterpiece arguments in philosophy.

Results of our Experiments

Variant of Dana Scott

- premises are **consistent**
- all argument steps are **logically correct**
  – in logic **S5**
  – already in logic **KB**

Intermediate Conclusion:

With our technology. . .
. . . it is possible to verify (selected) masterpiece arguments in philosophy.

Results of our Experiments

Variant of Kurt Gödel

- premises are inconsistent/contradictory
- everything follows!
- Philosophers had not seen this
- . . . but my theorem prover LEO-II did

Intermediate Conclusion:

Our technology . . .
. . . can reveal flawed arguments and contribute new knowledge.

Results of our Experiments

Variant of Kurt Gödel

- premises are inconsistent/contradictory
- everything follows!
- Philosophers had not seen this
- . . . but my theorem prover LEO-II did

Intermediate Conclusion:

Our technology . . .
. . . can reveal flawed arguments and contribute new knowledge.

Results of our Experiments

Variant of Kurt Gödel

- premises are inconsistent/contradictory
- everything follows!
- Philosophers had not seen this
- . . . but my theorem prover LEO-II did

Intermediate Conclusion:

Our technology . . .
. . . can reveal flawed arguments and contribute new knowledge.

Variant of Kurt Gödel

- premises are inconsistent/contradictory
- everything follows!
- Philosophers had not seen this
- . . . but my theorem prover LEO-II did

Intermediate Conclusion:

Our technology . . .
. . . can reveal flawed arguments and contribute new knowledge.

Results of our Analysis

. . . we continue with Scott's version

Further corollaries we can prove

- Monotheism
- God is flawless (has only positive properties)
- . . .
- Modal Collapse: $\varphi \rightarrow \Box\varphi$

  - there are no contingent truths
  - no alternative worlds
  - everything is determined
  - no free will

Challenge: Can the Modal Collapse be avoided (with minimal changes)?

Results of our Analysis

. . . we continue with Scott's version

Further corollaries we can prove

- Monotheism
- God is flawless (has only positive properties)
- . . .
- Modal Collapse: $\varphi \rightarrow \Box\varphi$

- there are no contingent truths
- no alternative worlds
- everything is determined
- no free will

Challenge: Can the Modal Collapse be avoided (with minimal changes)?

Results of our Analysis

. . . we continue with Scott's version

Further corollaries we can prove

- Monotheism
- God is flawless (has only positive properties)
- . . .
- Modal Collapse: $\varphi \rightarrow \Box\varphi$

  - there are no contingent truths
  - no alternative worlds
  - everything is determined
  - no free will

Challenge: Can the Modal Collapse be avoided (with minimal changes)?

Can the Modal Collapse be avoided?

SOME EMENDATIONS OF GÖDEL'S ONTOLOGICAL PROOF

C. Anthony Anderson

Kurt Gödel's version of the ontological argument was shown by J. Howard Sobel to be defective, but some plausible modifications in the argument result in a version which is immune to Sobel's objection. A definition is suggested which permits the proof of some of Gödel's axioms.

Gödel's Ontological Proof Revisited *

C. Anthony Anderson and Michael Gettings

University of California, Santa Barbara
Department of Philosophy

Gödel's version of the modal ontological argument for the existence of God has been criticized by J. Howard Sobel [5] and modified by C. Anthony Anderson [1]. In the present paper we consider the extent to which Anderson's emendation is defeated by the type of objection first offered by the Monk Gaunilo to St. Anselm's original Ontological Argument. And we try to push the analysis of this Gödelian argument a bit further to bring it into closer agreement with the details of Gödel's own formulation. Finally, we indicate what seems to be the main weakness of this emendation of Gödel's attempted proof.

Magari and others on Gödel's ontological proof

Petr Hájek
Institute of Computer Science, Academy of Sciences
182 07 Prague, Czech Republic
e-mail: hajek@uivt.cas.cz

1 Introduction

This paper is a continuation of my paper [H] and concentrates almost exclusively to mathematical properties of logical systems underlying Gödel's ontological proof [G] and its variant by Anderson [A], with special care paid to Magari's criticism [M]. Since [H] is written in German, we shall try to summarize its content in such a way that knowledge of [H] will be not obligatory for reading the present paper (even it remains advantageous). Here we describe

Der Mathematiker und die Frage der Existenz Gottes
(betreffend Gödels ontologischen Beweis)

Es ist gut, daß wir nicht wissen,
sondern glauben, daß ein Gott sei.
(Kant, Nachlaß)

1. Einführung

Gödels zu Lebzeiten unveröffentlichter Beweis für die notwendige Existenz eines Gott-ähnlichen Wesens hat sowohl philosophisches als auch mathematisches Interesse geweckt. Zweck der vorliegenden Arbeit ist es, zu einer Deutung des Gödelschen Textes beizutragen, 1. durch Kommentierung der einschlägigen Literatur und 2. durch Bereitstellung von etwas Modelltheorie. Die Arbeit enthält keinen philosophischen Beitrag. Während der letzten Jahre habe ich etliche Male über Gödels Gottesbeweis vorgetragen, insbesondere auf dem Symposium zur Feier von Professor Gert Müller (Heidelberg, Jänner 1991), doch habe ich niemals beabsichtigt, eine Veröffentlichung über das Thema zu machen. Da ich wiederholt um eine schriftliche Version gebeten wurde, entschloß ich mich, schnell eine „erweiterte Kurzfassung“[1] zu schreiben, ohne aus ihr einen

PETR HÁJEK

A New Small Emendation of Gödel's Ontological Proof

Keywords: Ontological proof, Gödel, modal logic, comprehension, positive properties.

1. Introduction

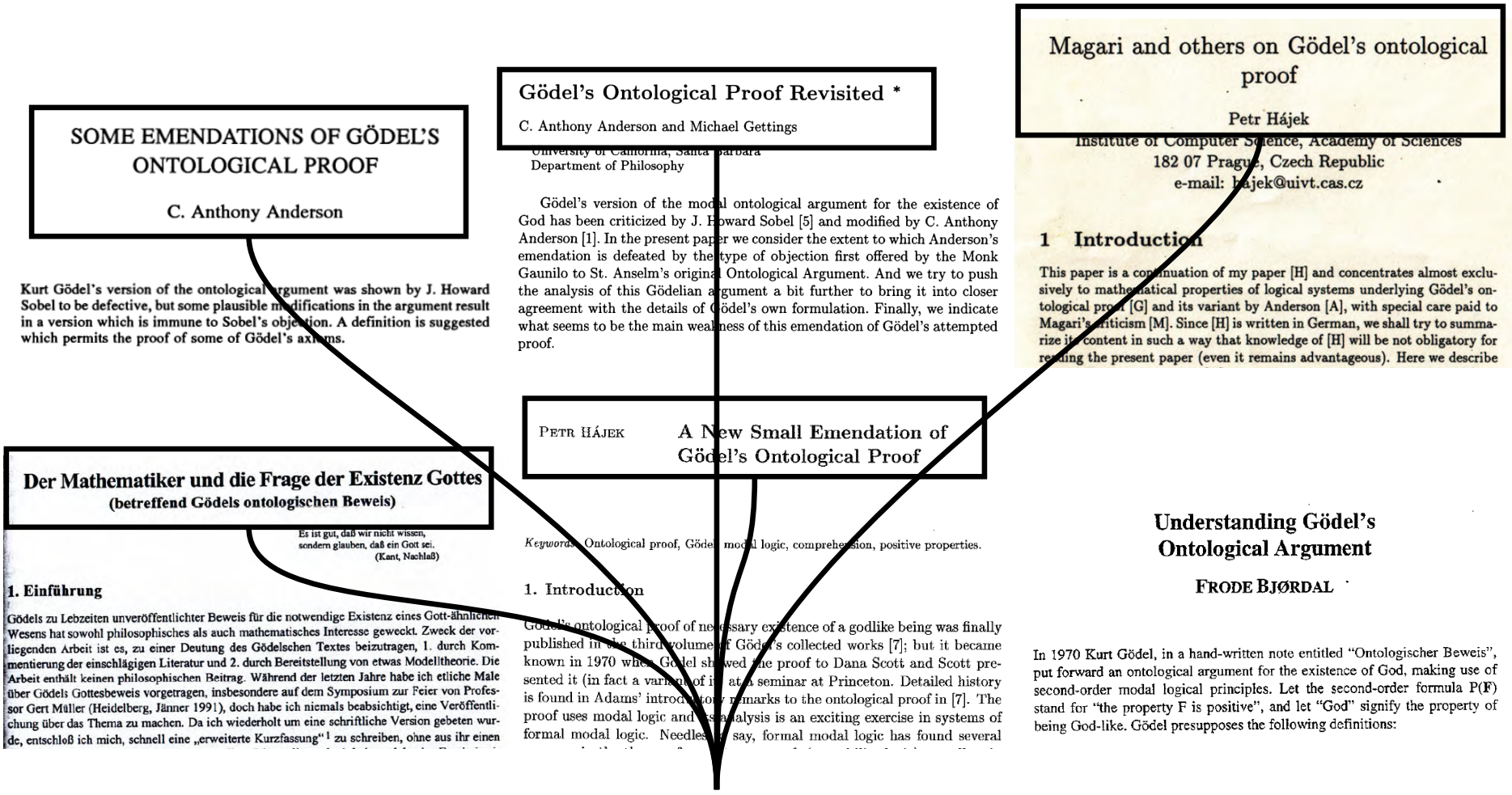
Gödel's ontological proof of necessary existence of a godlike being was finally published in the third volume of Gödel's collected works [7]; but it became known in 1970 when Gödel showed the proof to Dana Scott and Scott presented it (in fact a variant of it) at a seminar at Princeton. Detailed history is found in Adams' introductory remarks to the ontological proof in [7]. The proof uses modal logic and its analysis is an exciting exercise in systems of formal modal logic. Needles to say, formal modal logic has found several

Understanding Gödel's Ontological Argument

FRODE BJØRDAL

In 1970 Kurt Gödel, in a hand-written note entitled "Ontologischer Beweis", put forward an ontological argument for the existence of God, making use of second-order modal logical principles. Let the second-order formula P(F) stand for "the property F is positive", and let "God" signify the property of being God-like. Gödel presupposes the following definitions:

Can the Modal Collapse be avoided?



SOME EMENDATIONS OF GÖDEL'S ONTOLOGICAL PROOF

C. Anthony Anderson

Kurt Gödel's version of the ontological argument was shown by J. Howard Sobel to be defective, but some plausible modifications in the argument result in a version which is immune to Sobel's objection. A definition is suggested which permits the proof of some of Gödel's axioms.

Gödel's Ontological Proof Revisited *

C. Anthony Anderson and Michael Gettings

University of California, Santa Barbara
Department of Philosophy

Gödel's version of the modal ontological argument for the existence of God has been criticized by J. Howard Sobel [5] and modified by C. Anthony Anderson [1]. In the present paper we consider the extent to which Anderson's emendation is defeated by the type of objection first offered by the Monk Gaunilo to St. Anselm's original Ontological Argument. And we try to push the analysis of this Gödelian argument a bit further to bring it into closer agreement with the details of Gödel's own formulation. Finally, we indicate what seems to be the main weakness of this emendation of Gödel's attempted proof.

Magari and others on Gödel's ontological proof

Petr Hájek

Institute of Computer Science, Academy of Sciences
182 07 Prague, Czech Republic
e-mail: hajek@uivt.cas.cz

1 Introduction

This paper is a continuation of my paper [H] and concentrates almost exclusively to mathematical properties of logical systems underlying Gödel's ontological proof [G] and its variant by Anderson [A], with special care paid to Magari's criticism [M]. Since [H] is written in German, we shall try to summarize its content in such a way that knowledge of [H] will be not obligatory for reading the present paper (even it remains advantageous). Here we describe

Der Mathematiker und die Frage der Existenz Gottes
(betreffend Gödels ontologischen Beweis)

Es ist gut, daß wir nicht wissen,
sondern glauben, daß ein Gott sei.
(Kant, Nachlaß)

1. Einführung

Gödels zu Lebzeiten unveröffentlichter Beweis für die notwendige Existenz eines Gott-ähnlichen Wesens hat sowohl philosophisches als auch mathematisches Interesse geweckt. Zweck der vorliegenden Arbeit ist es, zu einer Deutung des Gödelschen Textes beizutragen, 1. durch Kommentierung der einschlägigen Literatur und 2. durch Bereitstellung von etwas Modelltheorie. Die Arbeit enthält keinen philosophischen Beitrag. Während der letzten Jahre habe ich etliche Male über Gödels Gottesbeweis vorgetragen, insbesondere auf dem Symposium zur Feier von Professor Gert Müller (Heidelberg, Jänner 1991), doch habe ich niemals beabsichtigt, eine Veröffentlichung über das Thema zu machen. Da ich wiederholt um eine schriftliche Version gebeten wurde, entschloß ich mich, schnell eine „erweiterte Kurzfassung"[1] zu schreiben, ohne aus ihr einen

PETR HÁJEK

A New Small Emendation of Gödel's Ontological Proof

Keywords: Ontological proof, Gödel modal logic, comprehension, positive properties.

1. Introduction

Gödel's ontological proof of necessary existence of a godlike being was finally published in the third volume of Gödel's collected works [7]; but it became known in 1970 when Gödel showed the proof to Dana Scott and Scott presented it (in fact a variant of it) at a seminar at Princeton. Detailed history is found in Adams' introductory remarks to the ontological proof in [7]. The proof uses modal logic and its analysis is an exciting exercise in systems of formal modal logic. Needless to say, formal modal logic has found several

Understanding Gödel's Ontological Argument

FRODE BJØRDAL

In 1970 Kurt Gödel, in a hand-written note entitled "Ontologischer Beweis", put forward an ontological argument for the existence of God, making use of second-order modal logical principles. Let the second-order formula P(F) stand for "the property F is positive", and let "God" signify the property of being God-like. Gödel presupposes the following definitions:

— contributed to clarification of controversy —
— revealed various flaws and issues —

[Logica Universalis, 2017]

Further Experiments

Ed Zalta (Stanford)

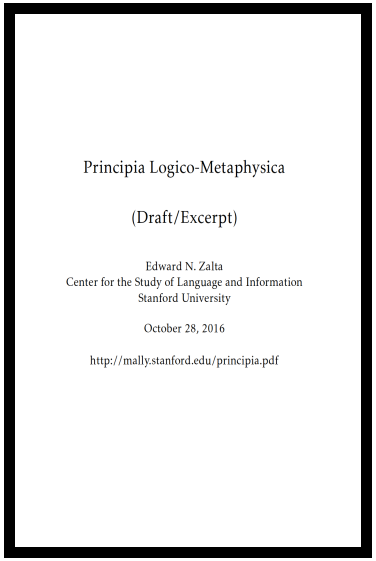
Principia Logico-Metaphysica

Hyperintensional higher-order modal logic

Inconsistency/Paradox detected

Daniel Kirchner
(Mathematics, FU Berlin)

Further Experiments


Principia Logico-Metaphysica

(Draft/Excerpt)

Edward N. Zalta
Center for the Study of Language and Information
Stanford University

October 28, 2016

http://mally.stanford.edu/principia.pdf

Ed Zalta (Stanford)

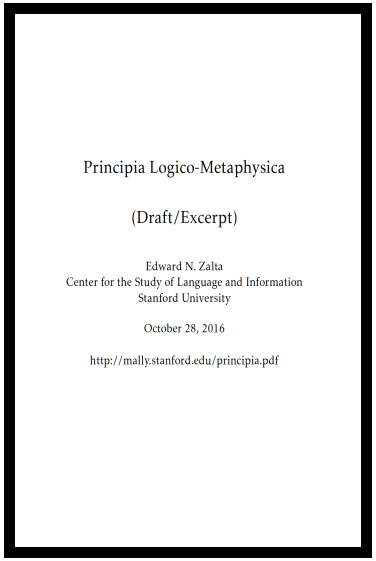
Principia Logico-Metaphysica

Hyperintensional higher-order modal logic

Inconsistency/Paradox detected

Daniel Kirchner
(Mathematics, FU Berlin)

Further Experiments


Principia Logico-Metaphysica

(Draft/Excerpt)

Edward N. Zalta
Center for the Study of Language and Information
Stanford University

October 28, 2016

http://mally.stanford.edu/principia.pdf

Ed Zalta (Stanford)

Principia Logico-Metaphysica

Hyperintensional higher-order modal logic

Inconsistency/Paradox detected

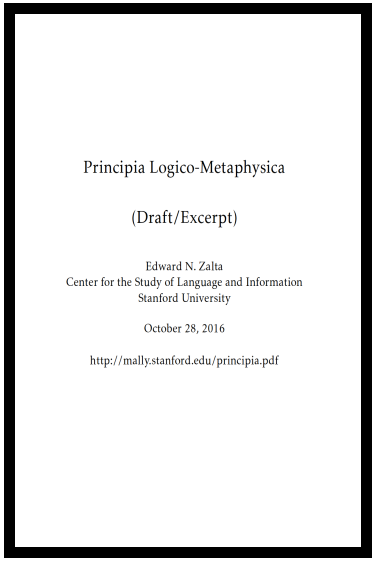

Daniel Kirchner
(Mathematics, FU Berlin)

Kirchner Paradox

Daniel & Isabelle/HOL have become close advisors of Ed Zalta in the search for a repair

Computational Metaphysics par excellence!!!

Further Experiments



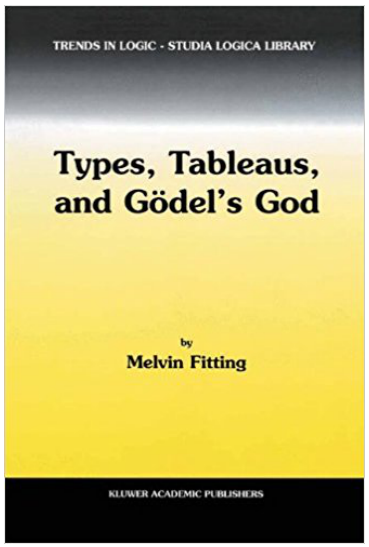
Ed Zalta (Stanford)

Principia Logico-Metaphysica

Hyperintensional higher-order modal logic

Inconsistency/Paradox detected

Daniel Kirchner
(Mathematics, FU Berlin)



Melvin Fitting (New York)

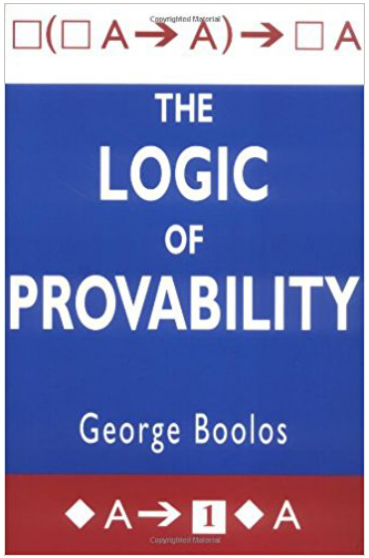
Ontological Argument
(avoids modal collapse)

Intensional higher-order modal logic

Verified (main chapters)

David Fuenmayor
(Philosophy, FU Berlin)

Further Experiments



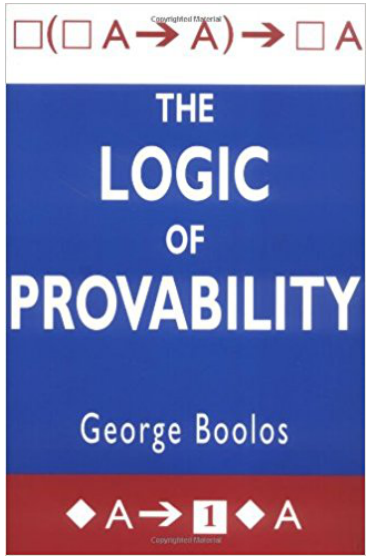
Textbook on Provability Logic

Provability Logic

Various parts verified

David Streit
(Mathematics, FU Berlin)

Further Experiments

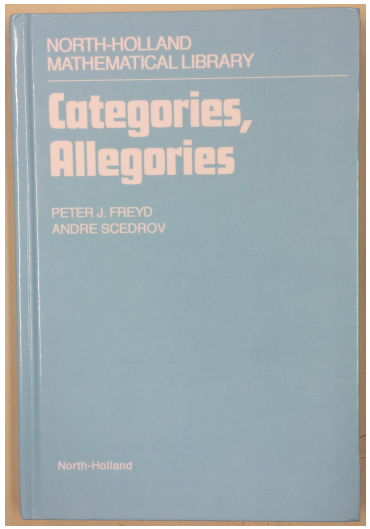


Textbook on Provability Logic

Provability Logic

Various parts verified

David Streit
(Mathematics, FU Berlin)



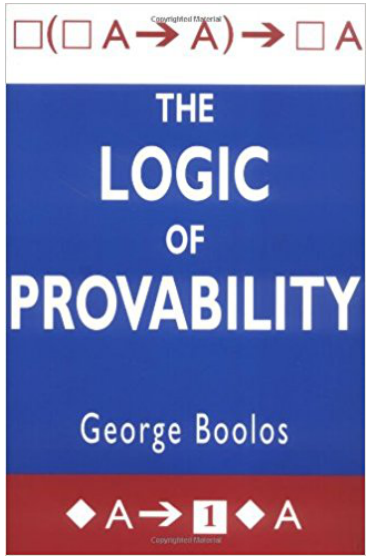
Category Theory

Free first-order logic

(Constricted) Inconsistency detected

D. Scott
(UC Berkeley)

Further Experiments

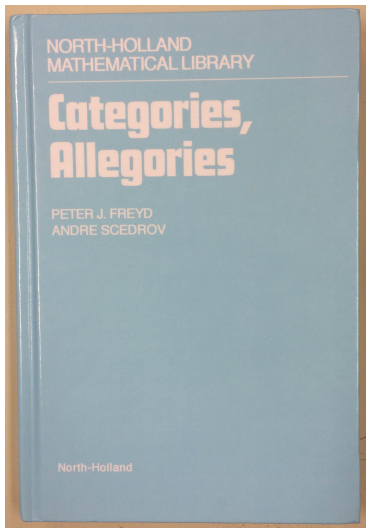


Textbook on Provability Logic

Provability Logic

Various parts verified

David Streit
(Mathematics, FU Berlin)



Category Theory

Free first-order logic

(Constricted) Inconsistency detected

D. Scott
(UC Berkeley)

Papers on these topics: `http://christoph-benzmueller.de` –> Publications

"If we had it [a *characteristica universalis*], we should be able to reason in metaphysics and morals in much the same way as in geometry and analysis."

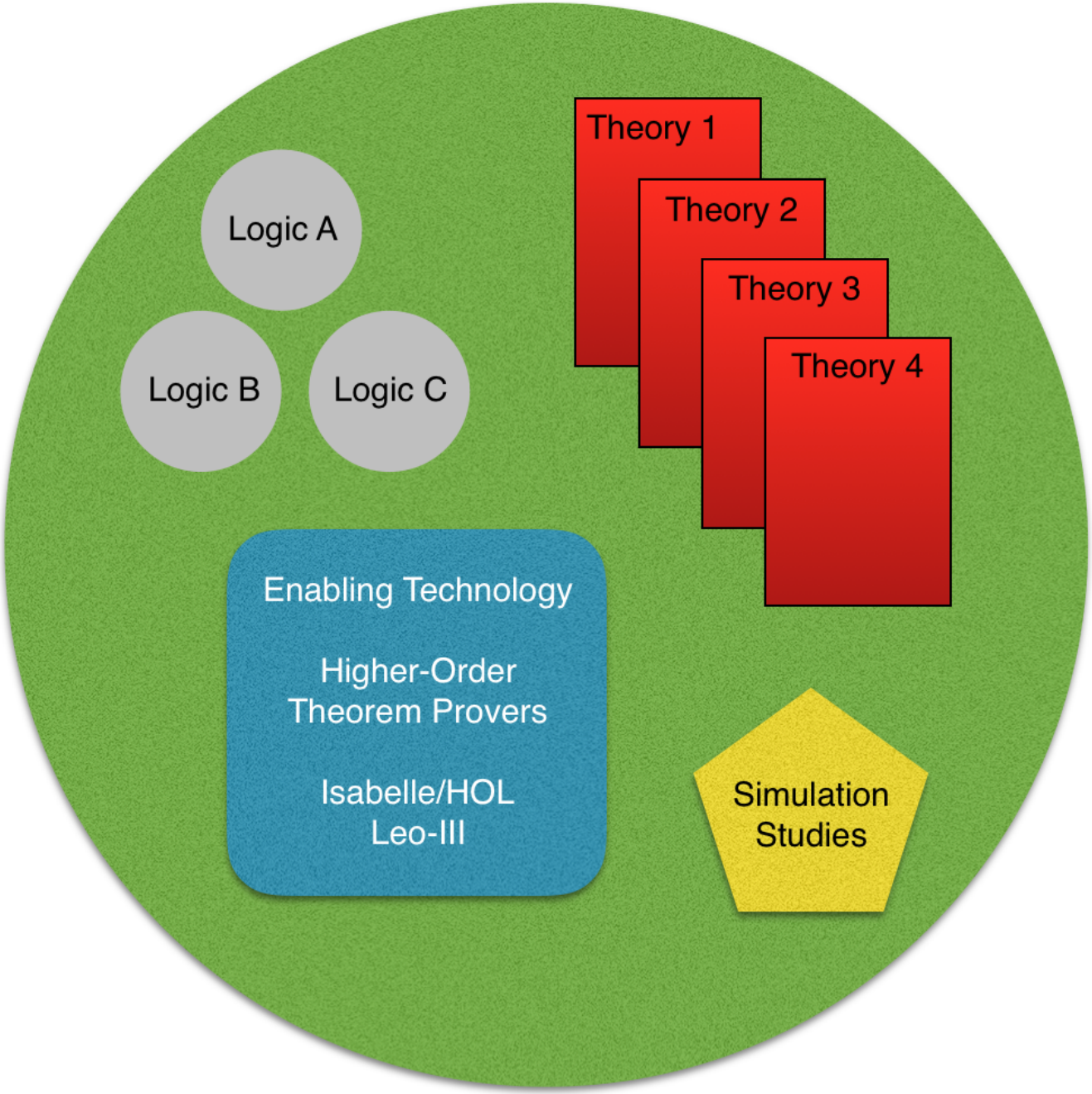
(Leibniz, 1677)

Part D

Demo(s): Normative Reasoning Experimentation Platform

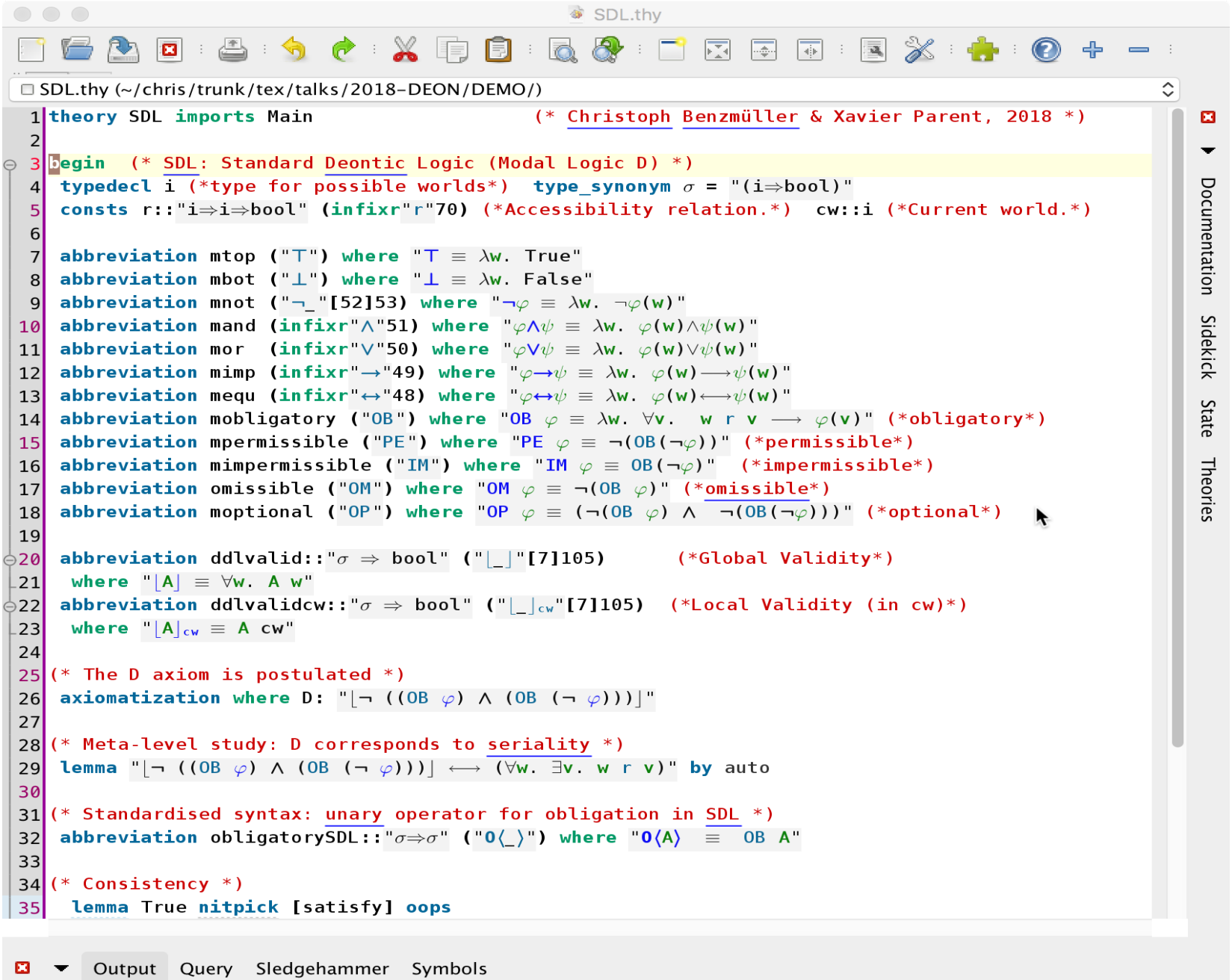
Normative Reasoning Experimentation Platform

[CiE, 2018]



Demo I: SDL in Isabelle/HOL

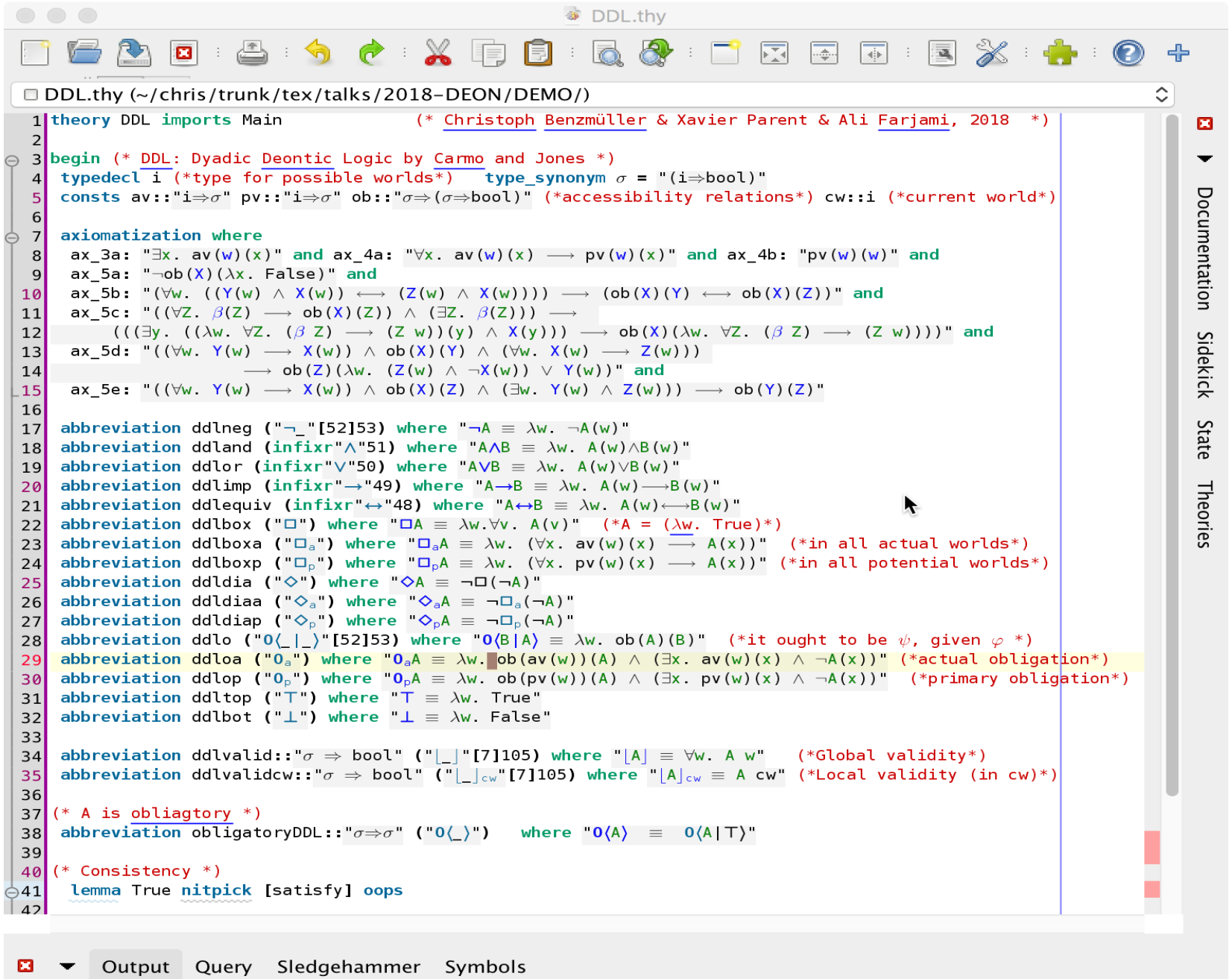
[Logica Universalis, 2013]




```
1  theory SDL imports Main                    (* Christoph Benzmüller & Xavier Parent, 2018 *)
2
3  begin  (* SDL: Standard Deontic Logic (Modal Logic D) *)
4   typedecl i (*type for possible worlds*)  type_synonym σ = "(i⇒bool)"
5   consts r::"i⇒i⇒bool" (infixr"r"70) (*Accessibility relation.*)  cw::i (*Current world.*)
6
7   abbreviation mtop ("⊤") where "⊤ ≡ λw. True"
8   abbreviation mbot ("⊥") where "⊥ ≡ λw. False"
9   abbreviation mnot ("¬_"[52]53) where "¬φ ≡ λw. ¬φ(w)"
10  abbreviation mand (infixr"∧"51) where "φ∧ψ ≡ λw. φ(w)∧ψ(w)"
11  abbreviation mor  (infixr"∨"50) where "φ∨ψ ≡ λw. φ(w)∨ψ(w)"
12  abbreviation mimp (infixr"→"49) where "φ→ψ ≡ λw. φ(w)⟶ψ(w)"
13  abbreviation mequ (infixr"↔"48) where "φ↔ψ ≡ λw. φ(w)⟷ψ(w)"
14  abbreviation mobligatory ("OB") where "OB φ ≡ λw. ∀v.  w r v ⟶ φ(v)" (*obligatory*)
15  abbreviation mpermissible ("PE") where "PE φ ≡ ¬(OB(¬φ))" (*permissible*)
16  abbreviation mimpermissible ("IM") where "IM φ ≡ OB(¬φ)"  (*impermissible*)
17  abbreviation omissible ("OM") where "OM φ ≡ ¬(OB φ)" (*omissible*)
18  abbreviation moptional ("OP") where "OP φ ≡ (¬(OB φ) ∧  ¬(OB(¬φ)))" (*optional*)
19
20  abbreviation ddlvalid::"σ ⇒ bool" ("⌊_⌋"[7]105)      (*Global Validity*)
21   where "⌊A⌋ ≡ ∀w. A w"
22  abbreviation ddlvalidcw::"σ ⇒ bool" ("⌊_⌋cw"[7]105)  (*Local Validity (in cw)*)
23   where "⌊A⌋cw ≡ A cw"
24
25 (* The D axiom is postulated *)
26  axiomatization where D: "⌊¬ ((OB φ) ∧ (OB (¬ φ)))⌋"
27
28 (* Meta-level study: D corresponds to seriality *)
29  lemma "⌊¬ ((OB φ) ∧ (OB (¬ φ)))⌋ ⟷ (∀w. ∃v. w r v)" by auto
30
31 (* Standardised syntax: unary operator for obligation in SDL *)
32  abbreviation obligatorySDL::"σ⇒σ" ("O⟨_⟩") where "O⟨A⟩  ≡  OB A"
33
34 (* Consistency *)
35   lemma True nitpick [satisfy] oops
```

Demo I: DDL in Isabelle/HOL

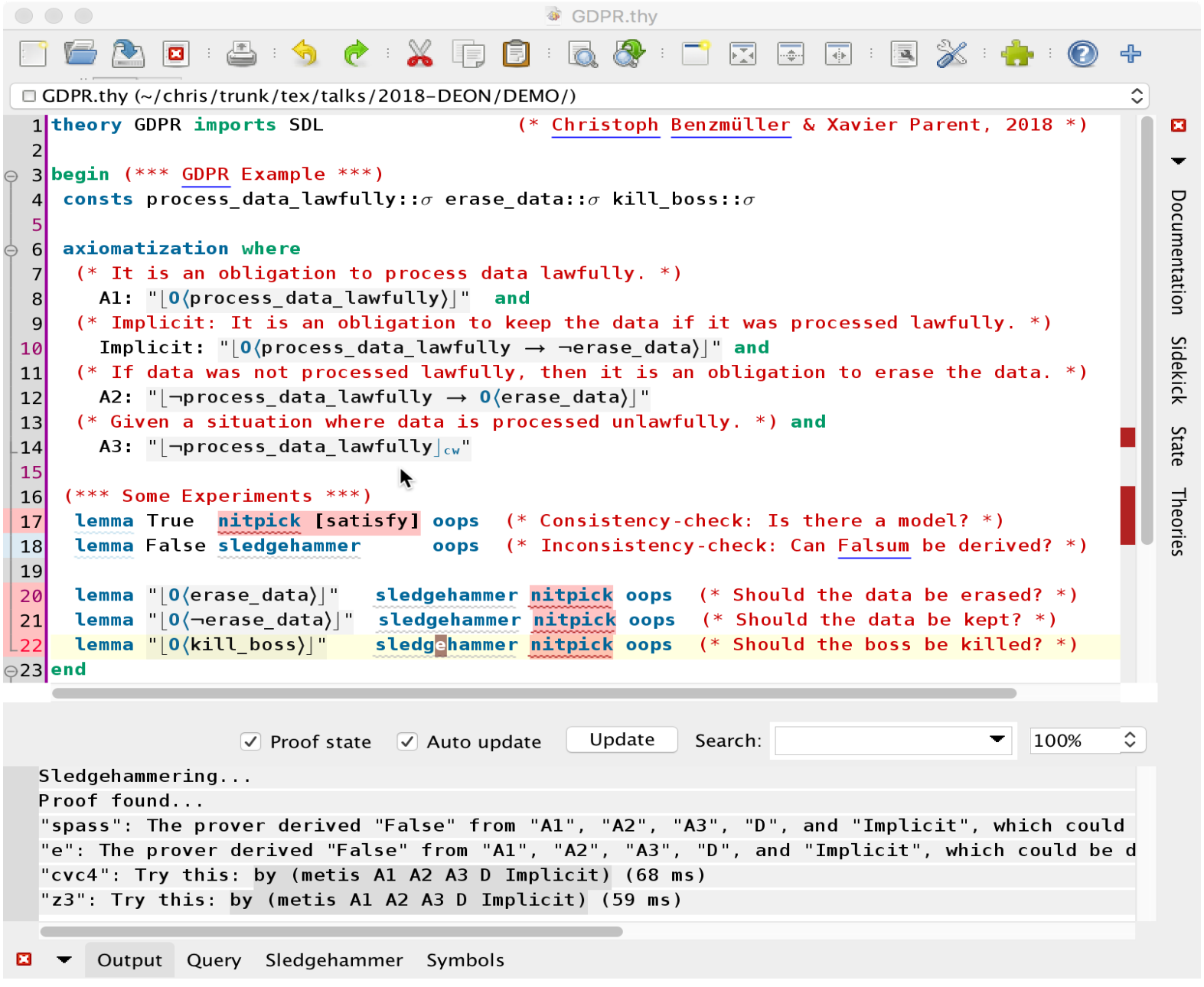
[see other DEON paper]



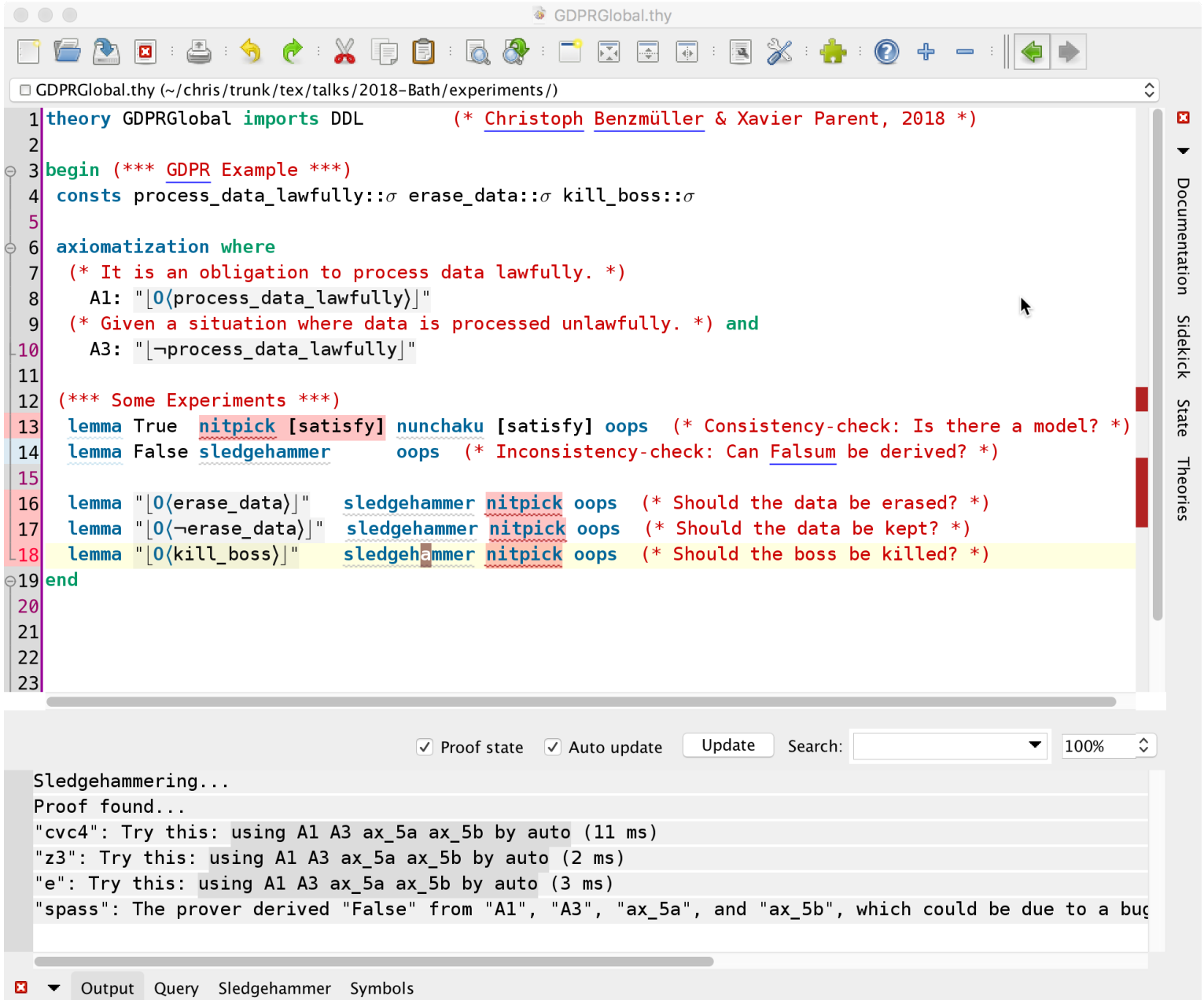

```
theory DDL imports Main            (* Christoph Benzmüller & Xavier Parent & Ali Farjami, 2018  *)

begin (* DDL: Dyadic Deontic Logic by Carmo and Jones *)
 typedecl i (*type for possible worlds*)   type_synonym σ = "(i⇒bool)"
 consts av::"i⇒σ" pv::"i⇒σ" ob::"σ⇒(σ⇒bool)" (*accessibility relations*) cw::i (*current world*)

 axiomatization where
  ax_3a: "∃x. av(w)(x)" and ax_4a: "∀x. av(w)(x) ⟶ pv(w)(x)" and ax_4b: "pv(w)(w)" and
  ax_5a: "¬ob(X)(λx. False)" and
  ax_5b: "(∀w. ((Y(w) ∧ X(w)) ⟷ (Z(w) ∧ X(w)))) ⟶ (ob(X)(Y) ⟷ ob(X)(Z))" and
  ax_5c: "((∀Z. β(Z) ⟶ ob(X)(Z)) ∧ (∃Z. β(Z))) ⟶
      (((∃y. ((λw. ∀Z. (β Z) ⟶ (Z w))(y) ∧ X(y))) ⟶ ob(X)(λw. ∀Z. (β Z) ⟶ (Z w))))" and
  ax_5d: "((∀w. Y(w) ⟶ X(w)) ∧ ob(X)(Y) ∧ (∀w. X(w) ⟶ Z(w)))
                  ⟶ ob(Z)(λw. (Z(w) ∧ ¬X(w)) ∨ Y(w))" and
  ax_5e: "((∀w. Y(w) ⟶ X(w)) ∧ ob(X)(Z) ∧ (∃w. Y(w) ∧ Z(w))) ⟶ ob(Y)(Z)"

 abbreviation ddlneg ("¬_"[52]53) where "¬A ≡ λw. ¬A(w)"
 abbreviation ddland (infixr"∧"51) where "A∧B ≡ λw. A(w)∧B(w)"
 abbreviation ddlor (infixr"∨"50) where "A∨B ≡ λw. A(w)∨B(w)"
 abbreviation ddlimp (infixr"→"49) where "A→B ≡ λw. A(w)⟶B(w)"
 abbreviation ddlequiv (infixr"↔"48) where "A↔B ≡ λw. A(w)⟷B(w)"
 abbreviation ddlbox ("□") where "□A ≡ λw.∀v. A(v)"  (*A = (λw. True)*)
 abbreviation ddlboxa ("□a") where "□aA ≡ λw. (∀x. av(w)(x) ⟶ A(x))"  (*in all actual worlds*)
 abbreviation ddlboxp ("□p") where "□pA ≡ λw. (∀x. pv(w)(x) ⟶ A(x))" (*in all potential worlds*)
 abbreviation ddldia ("◇") where "◇A ≡ ¬□(¬A)"
 abbreviation ddldiaa ("◇a") where "◇aA ≡ ¬□a(¬A)"
 abbreviation ddldiap ("◇p") where "◇pA ≡ ¬□p(¬A)"
 abbreviation ddlo ("O⟨_|_⟩"[52]53) where "O⟨B|A⟩ ≡ λw. ob(A)(B)"  (*it ought to be ψ, given φ *)
 abbreviation ddloa ("Oa") where "OaA ≡ λw. ob(av(w))(A) ∧ (∃x. av(w)(x) ∧ ¬A(x))" (*actual obligation*)
 abbreviation ddlop ("Op") where "OpA ≡ λw. ob(pv(w))(A) ∧ (∃x. pv(w)(x) ∧ ¬A(x))"  (*primary obligation*)
 abbreviation ddltop ("⊤") where "⊤ ≡ λw. True"
 abbreviation ddlbot ("⊥") where "⊥ ≡ λw. False"

 abbreviation ddlvalid::"σ ⇒ bool" ("⌊_⌋"[7]105) where "⌊A⌋ ≡ ∀w. A w"   (*Global validity*)
 abbreviation ddlvalidcw::"σ ⇒ bool" ("⌊_⌋cw"[7]105) where "⌊A⌋cw ≡ A cw" (*Local validity (in cw)*)

(* A is obliagtory *)
 abbreviation obligatoryDDL::"σ⇒σ" ("O⟨_⟩")   where "O⟨A⟩  ≡  O⟨A|⊤⟩"

(* Consistency *)
  lemma True nitpick [satisfy] oops
```

Demo I: Experimenting with SDL and DDL in Isabelle/HOL

[arXiv:1804.02929]



Demo I: Global vs. Local Consequence Relation



Experimenting with SDL and DDL in Isabelle/HOL

[arXiv:1804.02929]

1. SDL in HOL
   (propositional, first/higher-order, different quantifiers, logic combinations)
   - already covered by earlier work
2. DDL in HOL
   (propositional)
   - with Ali Farjami and Xavier Parent (see Ali's talk)
   - faithfulness (assuming Henkin semantics)
3. DDL in HOL
   (first/higher-order, different quantifiers, logic combinations)
   - straightforward combination with (1)
   - more later
4. Ask me for longer demo!

Input/output (I/O) logic

[Makinson, JPL, 2000], [GabbayHortyParentEtAl-Handbook, 2013]

- I/O-operators, such as out1 (simple-minded output), accept set G of conditional norms as argument
- Conditional norms: pairs (a,x) with input “a” (condition) and output “x” (obligation)
- Pairs (a,x) are not given a truth-functional semantics in I/O logic

Input/output (I/O) logic

[Makinson, JPL, 2000], [GabbayHortyParentEtAl-Handbook, 2013]

- I/O-operators, such as out1 (simple-minded output), accept set G of conditional norms as argument
- Conditional norms: pairs (a,x) with input “a” (condition) and output “x” (obligation)
- Pairs (a,x) are not given a truth-functional semantics in I/O logic

Semantics of out1

- $out1(G, A) := Cn(G(Cn(A)))$
- where $Cn(X) := \{s \mid X \models s\}$ and $G(X) := \{s \mid \exists a \in X.\ (a, s) \in G\}$.

Input/output (I/O) logic

[Makinson, JPL, 2000], [GabbayHortyParentEtAl-Handbook, 2013]

- I/O-operators, such as out1 (simple-minded output), accept set G of conditional norms as argument
- Conditional norms: pairs (a,x) with input “a” (condition) and output “x” (obligation)
- Pairs (a,x) are not given a truth-functional semantics in I/O logic

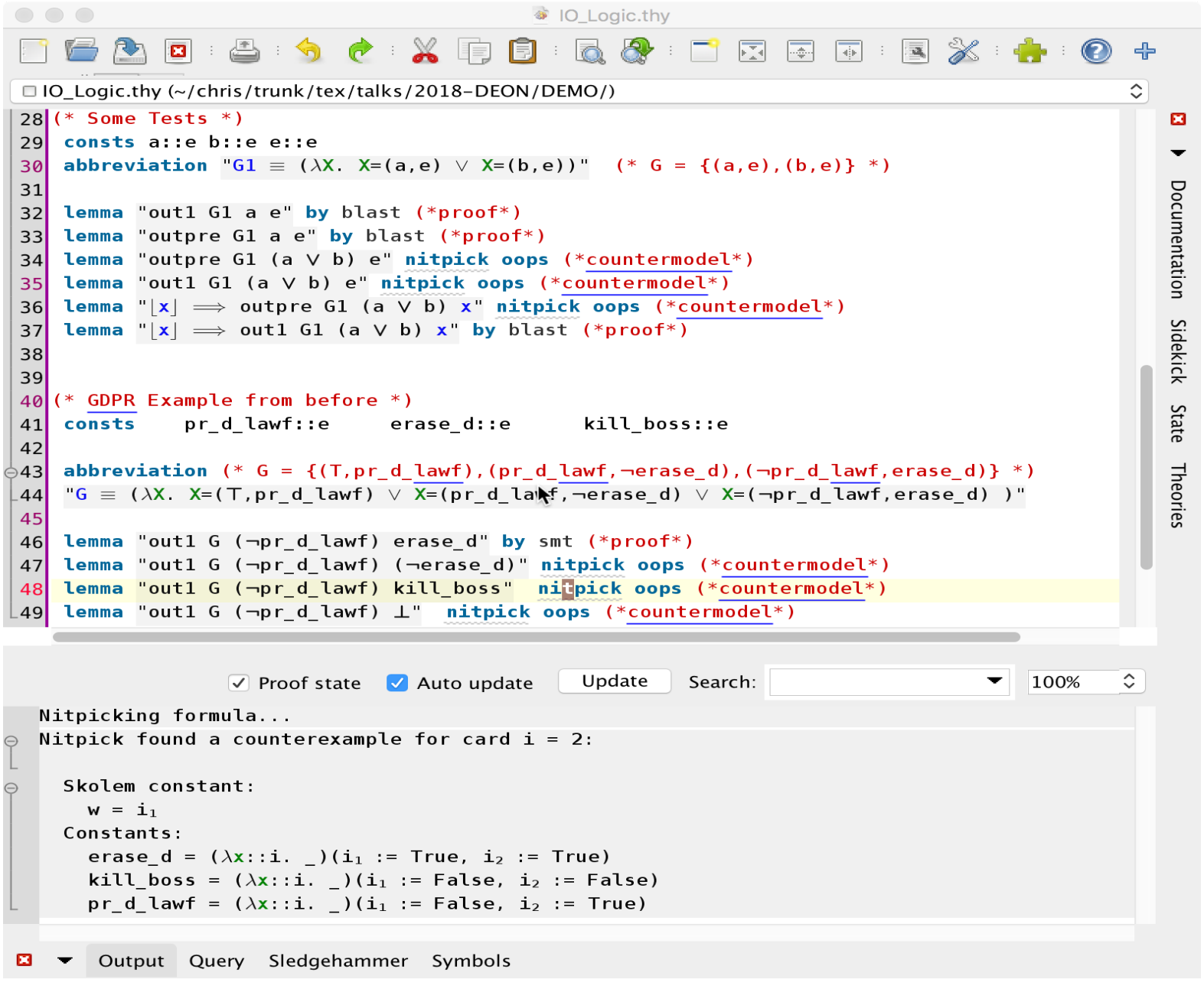
Semantics of out1

- $out1(G,A) := Cn(G(Cn(A)))$
- where $Cn(X) := \{s \mid X \models s\}$ and $G(X) := \{s \mid \exists a \in X.\ (a,s) \in G\}$.

```
(*IO logic in HOL*)
typedecl i -- "type for possible worlds"     type_synonym e = "(i⇒bool)"
abbreviation ktop   :: "e"     ("⊤")  where "⊤ ≡ λw. True"
abbreviation kbot   :: "e"     ("⊥")  where "⊥ ≡ λw. False"
abbreviation knot   :: "e⇒e"     ("¬_"[52]53)  where "¬φ ≡ λw. ¬φ(w)"
abbreviation kor    :: "e⇒e⇒e" (infixr"∨"50) where "φ∨ψ ≡ λw. φ(w)∨ψ(w)"
abbreviation kand   :: "e⇒e⇒e" (infixr"∧"51) where "φ∧ψ ≡ λw. φ(w)∧ψ(w)"
abbreviation kimp   :: "e⇒e⇒e" (infixr"⊃"49) where "φ⊃ψ ≡ λw. φ(w)⟶ψ(w)"
abbreviation kvalid :: "e⇒bool" ("⌊_⌋"[8]109)  where "⌊p⌋ ≡ ∀w. p w"

abbreviation "outpre ≡ λG.λa.λy::e. ∃f. ⌊a ⊃ f⌋ ∧ G (f,y)"

abbreviation "out1 ≡  λG.λa.λx. ⌊x⌋ ∨
   (∃i j k. outpre G a i ∧ outpre G a j ∧ outpre G a k ∧ ⌊(i ∧ j ∧ k) ⊃ x⌋)"
```

Demo II: I/O-Logic in Isabelle/HOL

[arXiv:1803.09681]



Demo III: Preference-based DDL in Isabelle/HOL

Journal of Philosophical Logic / Vol. 43, No. 6, December 2014 / Maximality vs. Optim...

JOURNAL ARTICLE

Maximality vs. Optimality in Dyadic Deontic Logic: Completeness Results for Systems in Hansson's Tradition

Xavier Parent

Journal of Philosophical Logic
Vol. 43, No. 6 (December 2014), pp. 1101-1128

Demo III: Preference-based DDL in Isabelle/HOL

Journal of Philosophical Logic / Vol. 43, No. 6, December 2014 / Maximality vs. Optim...

By
A. Steen

PrefDDL.thy

PrefDDL.thy (~/chris/trunk/tex/talks/2018-DEON/DEMO/)

```
246  (* axioms of proof theory for E, check for soundness *)
247  lemma classical: "A ⟹ ⌊(λw. A)⌋" by simp -- "all classical tautologies"
248
249  lemma "□M": "⌊□A ⟶ A⌋" by simp -- "part of S5 schema for □"
250  lemma "□5": "⌊◇A ⟶ □(◇A)⌋" by simp  -- "part of S5 schema for □"
251
252  lemma DfP: "⌊P⟨B|A⟩ ⟷ ¬(O⟨(¬B)|A⟩)⌋" by (simp add: prefDDLBase.truthSet_def)
253  lemma COK: "⌊O⟨(B ⟶ C)|A⟩ ⟶ (O⟨B|A⟩ ⟶ O⟨C|A⟩)⌋" by (simp add: prefDDLBase.truthSet_def)
254  lemma abs: "⌊O⟨B|A⟩ ⟶ □(O⟨B|A⟩)⌋" by simp
255  lemma nec: "⌊□A ⟶ O⟨A|B⟩⌋" by (simp add: prefDDLBase.truthSet_def)
256  lemma ext: "⌊□(A ⟷ B) ⟶ (O⟨C|A⟩ ⟷ O⟨C|B⟩)⌋" by (simp add: prefDDLBase.truthSet_def)
257  lemma id: "⌊O⟨A|A⟩⌋" by (simp add: optChoice)
258  lemma Sh: "⌊O⟨C|(A ∧ B)⟩ ⟶ O⟨(B ⟶ C)|A⟩⌋" by (smt optBest.optChoice optBest_axioms prefDDLBase.
259
260  (* soundness of inference rules *)
261  lemma MP: "⌊A⌋ ⟹ ⌊A ⟶ B⌋ ⟹ ⌊B⌋" by simp
262  lemma N: "⌊A⌋ ⟹ ⌊□A⌋" by simp
263
264  (* D* should hold in F, this can be verfied: *)
265  lemma "D*": "⌊◇A ⟶ (O⟨B|A⟩ ⟶ P⟨B|A⟩)⌋"
266    by (metis FOpt.opt_limitedness FOpt_axioms truthSet_def)
267
268  (* (CM) should not be provable in system F but only as of system F+CM, verified by nitpick *)
269  lemma CM: "⌊(O⟨B|A⟩ ∧ O⟨C|A⟩) ⟶ O⟨C|(A ∧ B)⟩⌋" nitpick oops
```

Proof state | Auto update | Update | Search: | 100%

```
Nitpicking formula...
Nitpick found a counterexample for card 'w = 3:

  Free variables:
    A = (λx. _)(w₁ := True, w₂ := True, w₃ := True)
    B = (λx. _)(w₁ := True, w₂ := True, w₃ := False)
    C = (λx. _)(w₁ := False, w₂ := True, w₃ := False)
    op ≳ =
      (λx. _)
      (w₁ := (λx. _)(w₁ := True, w₂ := True, w₃ := False),
       w₂ := (λx. _)(w₁ := True, w₂ := True, w₃ := True),
       w₃ := (λx. _)(w₁ := True, w₂ := False, w₃ := True))
    opt = (λx. _)
```

Output | Query | Sledgehammer | Symbols

Documentation | Sidekick | State | Theories

Demo IV: Gewirth's Principle of Generic Consistency (PGC) in Isabelle/HOL

> "Act in accord with the generic rights of your recipients as well as of yourself. I shall call this the Principle of Generic Consistency (PGC), since it combines the formal consideration of consistency with the material consideration of rights to the generic features or goods of action." **(Alan Gewirth, Reason and Morality, 1978)**

- **Gewirth's PGC has**
  - stirred much controversy in moral philosophy
  - been discussed as means to bound the impact of artificial general intelligence (AGI)
- **Idea (in a nutshell):**
  - devise a safety mechanism of a mathematical (deductive) nature
  - to ensure that an AGI respects human's freedom and well-being
  - mechanism is based on assumption that it is able to recognize itself, as well as us humans, as agents (prospective purposive agents, PPA) which
    - act voluntarily on self-chosen purposes, and
    - reason rationally
- **References**
  - A. Gewirth. Reason and morality. U of Chicago Press, 1978.
  - D. Beyleveld. The dialectical necessity of morality: An analysis and defense of Alan Gewirth's argument to the principle of generic consistency. U of Chicago Press, 1991.
  - A. Kornai. Bounding the impact of AGI. J. Experimental & Theoretical AI, 2014.

Demo IV: Gewirth's Principle of Generic Consistency (PGC) in Isabelle/HOL

The idea is to constrain potential AGI's to reason in the following way

- It is necessary for me (as an AGI) to accept that:

  (P1) I act voluntarily on purpose E (equivalent by definition to "I am a PPA")
  (C2) E is good (for me)
  (P3) In order to achieve any purpose whatsoever by my agency, I need my freedom and well-being
  (C4) My freedom and well-being are necessary goods (for me)
  (C5) I (even if no one else) have a claim right to my freedom and well-being

- It is necessary for all PPAs to accept that:

  (C9) Every PPA has a necessary right to their freedom and well-being

Demo IV: Gewirth's Principle of Generic Consistency (PGC) in Isabelle/HOL

The idea is to constrain potential AGI's to reason in the following way

- It is necessary for me (as an AGI) to accept that:

  (P1) I act voluntarily on purpose E (equivalent by definition to "I am a PPA")
  (C2) E is good (for me)
  (P3) In order to achieve any purpose whatsoever by my agency, I need my freedom and well-being
  (C4) My freedom and well-being are necessary goods (for me)
  (C5) I (even if no one else) have a claim right to my freedom and well-being

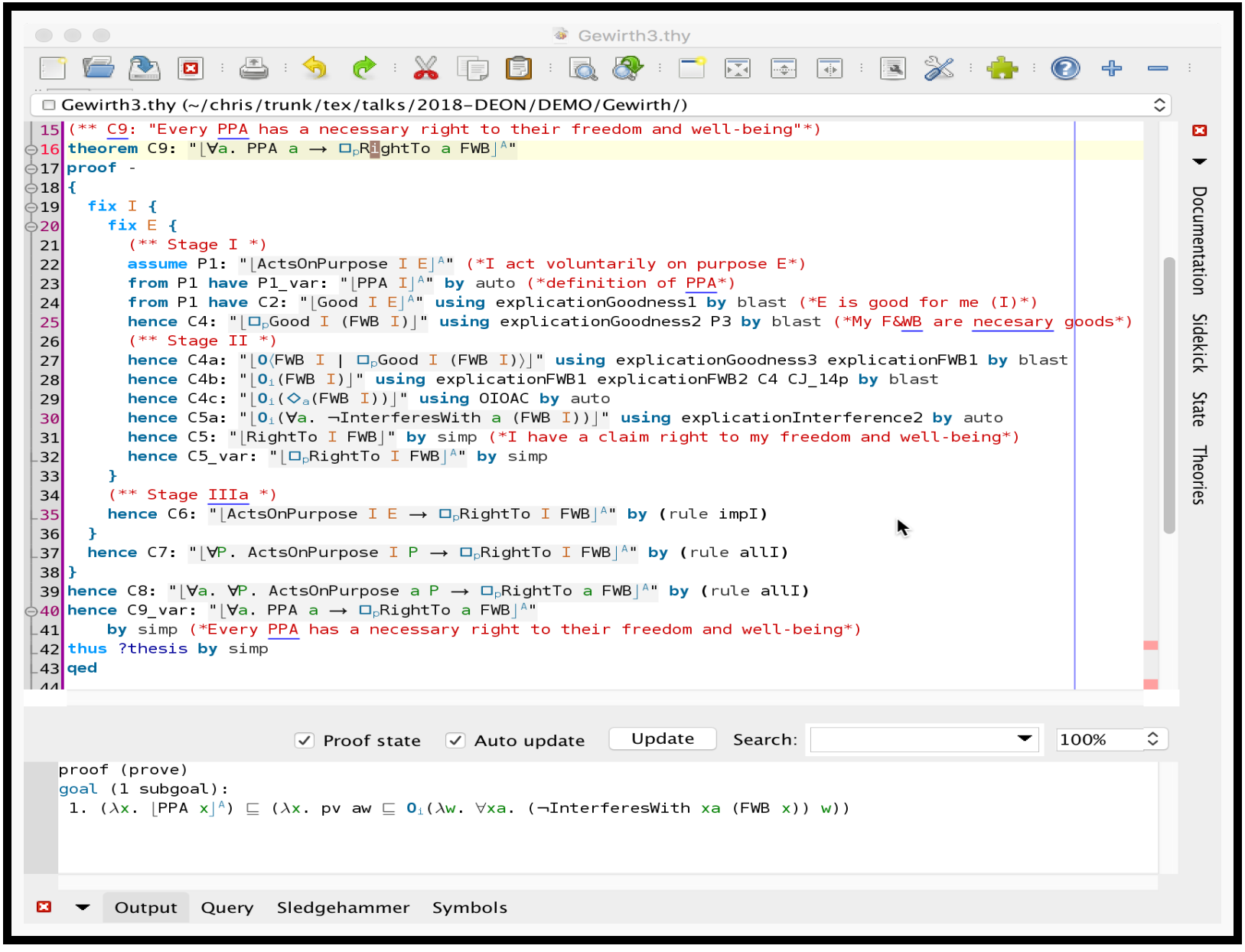
- It is necessary for all PPAs to accept that:

  (C9) Every PPA has a necessary right to their freedom and well-being

Any AGI (PPA) denying that it is bound by the PCG (e.g. by refusing to respect humans' F&WB) would deny that it is a PPA (and thus its own agency).

Hence, to avoid self-contradiction, an AGI would be bound to accord basic rights to humans.

Demo IV: Gewirth's Principle of Generic Consistency (PGC) in Isabelle/HOL



By David Fuenmayor, cf. `http://christoph-benzmueller.de/papers/2018-GewirthArgument.zip`

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]

What is Leo-III?

- ATP for classical HOL (by **A. Steen**, M. Wisniewski and myself)
- ordered paramodulation; efficient data-structures; parallelisation; etc.
- native support for more than 120 logics (all normal quantified modal logics)
- including native support for **quantified SDL and DDL**

- Website: `http://page.mi.fu-berlin.de/lex/leo3/`
- Download: `https://github.com/leoprover/Leo-III`

What is Leo-III?

- ATP for classical HOL (by **A. Steen**, M. Wisniewski and myself)
- ordered paramodulation; efficient data-structures; parallelisation; etc.
- native support for more than 120 logics (all normal quantified modal logics)
- including native support for **quantified SDL and DDL**

- Website: `http://page.mi.fu-berlin.de/lex/leo3/`
- Download: `https://github.com/leoprover/Leo-III`

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]

Brand new: Support for Dyadic Deontic Logic (Carmo/Jones)

- Enhance propositional TPTP fragment with
  1. Dyadic deontic obligation $O(p/q)
  2. Actual/Primary deontic obligations $O_a(p), $O_p(p)
  3. Box operators $box(p), $box_a(p),$box_p(p)
- Integrated into Leo-III (stand-alone tool available)

ASCII	Syntax	Meaning
~	$\neg$	Negation
|	$\vee$	Disjunction
&	$\wedge$	Conjunction
=>	$\Rightarrow$	Material implication
<=>	$\Leftrightarrow$	Equivalence
$O(p/q)	$O(p/q)$	Dyadic deontic obligation (It ought to be p given that q)
$box(p)	$\Box(p)$	In all worlds p

Input statements: `ddl(<name>, <role>, <formula>).`

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]

Brand new: Support for Dyadic Deontic Logic (Carmo/Jones)

- Enhance propositional TPTP fragment with
  1. Dyadic deontic obligation $O(p/q)
  2. Actual/Primary deontic obligations $O_a(p), $O_p(p)
  3. Box operators $box(p), $box_a(p),$box_p(p)
- Integrated into Leo-III (stand-alone tool available)

ASCII	Syntax	Meaning
~	$\neg$	Negation
|	$\vee$	Disjunction
&	$\wedge$	Conjunction
=>	$\Rightarrow$	Material implication
<=>	$\Leftrightarrow$	Equivalence
`$O(p/q)`	$O(p/q)$	Dyadic deontic obligation (It ought to be p given that q)
`$box(p)`	$\Box(p)$	In all worlds p

Input statements: `ddl(<name>, <role>, <formula>).`

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]

Input statements: `ddl(<name>, <role>, <formula>).`

where `<role>` provides meta-logical information:

- `axiom` *assumed, globally valid*
- `localAxiom` *assumed, valid in current world*
- `conjecture` *global consequence?*
- `localConjecture` *consequence in current world?*

Example

This problem can directly be given to Leo-III:

```
ddl(a1, axiom, $O(processDataLawfully)).
ddl(a2, axiom, $O(eraseData/~processDataLawfully)).
ddl(a3, localAxiom, ~processDataLawfully).

ddl(c1, conjecture, $O(eraseData)).
```

... giving ...

```
% SZS status Theorem for gdpr_new.p : 2143 ms resp. 776 ms w/o parsing
```

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]

Input statements: `ddl(<name>, <role>, <formula>).`

where `<role>` provides meta-logical information:

- `axiom` *assumed, globally valid*
- `localAxiom` *assumed, valid in current world*
- `conjecture` *global consequence?*
- `localConjecture` *consequence in current world?*

Example

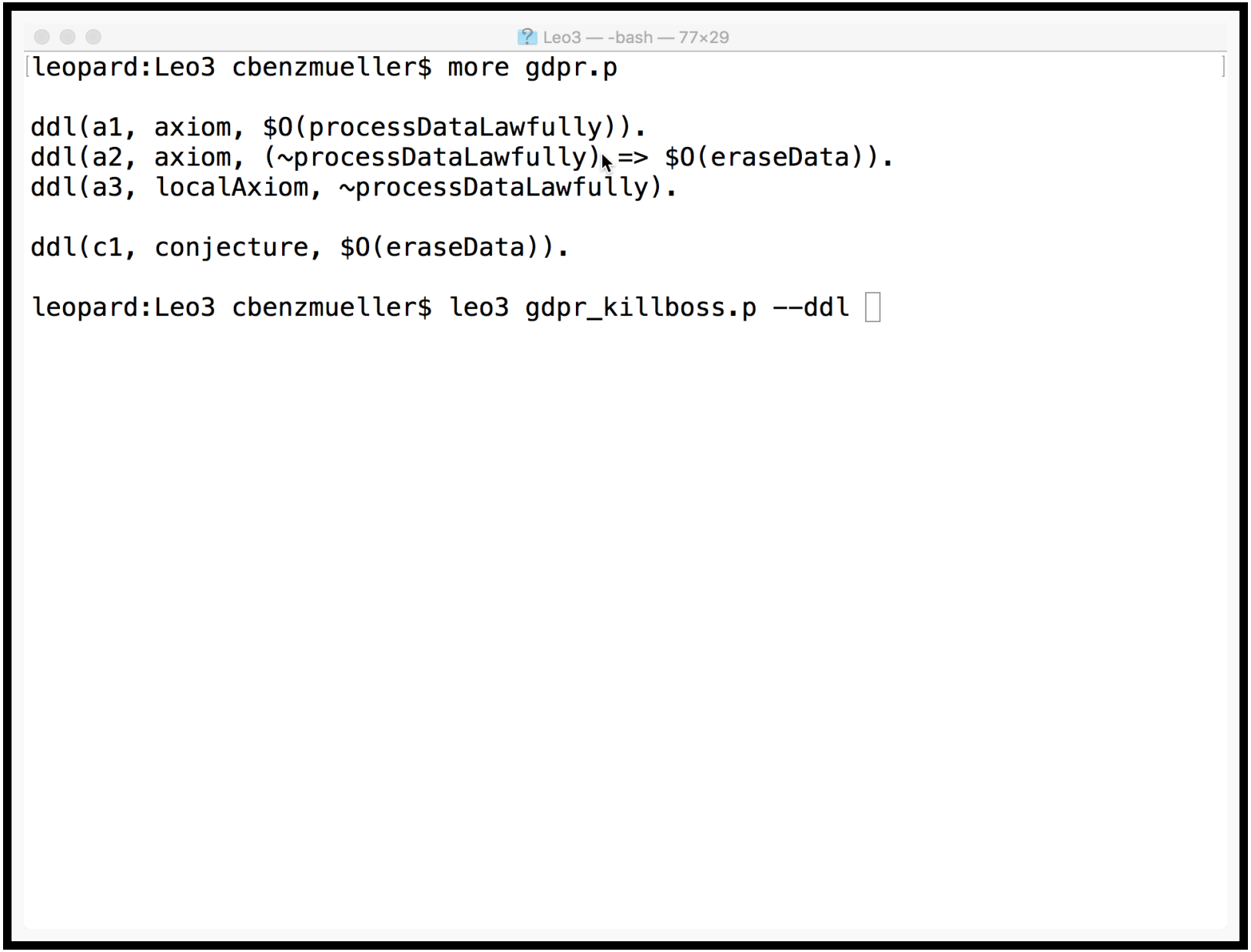
This problem can directly be given to Leo-III:

```
ddl(a1, axiom, $O(processDataLawfully)).
ddl(a2, axiom, $O(eraseData/~processDataLawfully)).
ddl(a3, localAxiom, ~processDataLawfully).

ddl(c1, conjecture, $O(eraseData)).
```

... giving ...

```
% SZS status Theorem for gdpr_new.p : 2143 ms resp. 776 ms w/o parsing
```

Demo V: Native Support for Deontic Logic(s) in Leo-III [IJCAR, 2018], [RuleML+RR, 2018]



References (selected): http//:christoph-benzmueller.de –> Publications

Deontic Logic Reasoning Infrastructure

- A Dyadic Deontic Logic in HOL, DEON 2018, 2018. (See also https://arxiv.org/abs/1802.08454)
- A Deontic Logic Reasoning Infrastructure, CiE 2018, Springer LNCS, 2018.
- I/O Logic in HOL — First Steps, CoRR, 2018. https://arxiv.org/abs/1803.09681
- First Experiments with a Flexible Infrastructure for Normative Reasoning, CoRR, 2018. http://arxiv.org/abs/1804.02929

Computational Metaphysics (selected)

- Experiments in Computational Metaphysics: Gödel's Proof of God's Existence, Savijnanam: scientific exploration for a spiritual paradigm. Journal of the Bhaktivedanta Institute, volume 9, pp. 43-57, 2017.
- The Inconsistency in Gödel's Ontological Argument: A Success Story for AI in Metaphysics, IJCAI 2016, 2016.
- Automating Gödel's Ontological Proof of God's Existence with Higher-order Automated Theorem Provers, ECAI 2014, IOS Press, 2014.

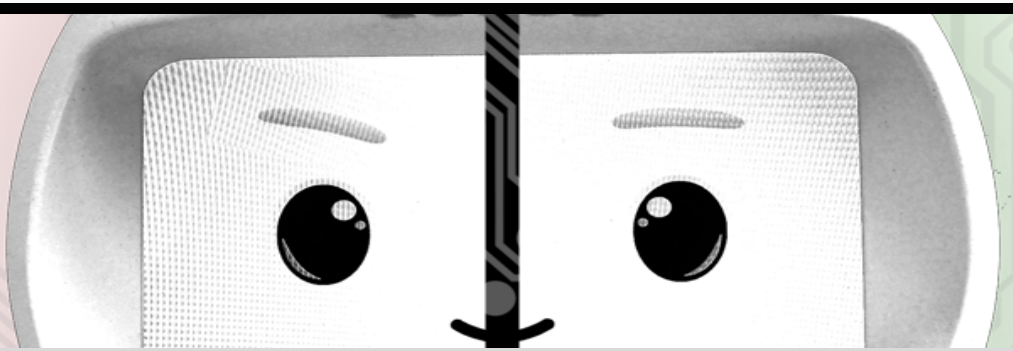
Other (selected)

- The Higher-Order Prover Leo-III, IJCAR 2018, Springer LNCS, 2018.

Ethics

Argued for explicit ethical reasoning competencies in IASs

- development of normative reasoning experimentation platform
- utilising HOL as universal meta-logic
- practical evidence from previous work
- suitable also for teaching



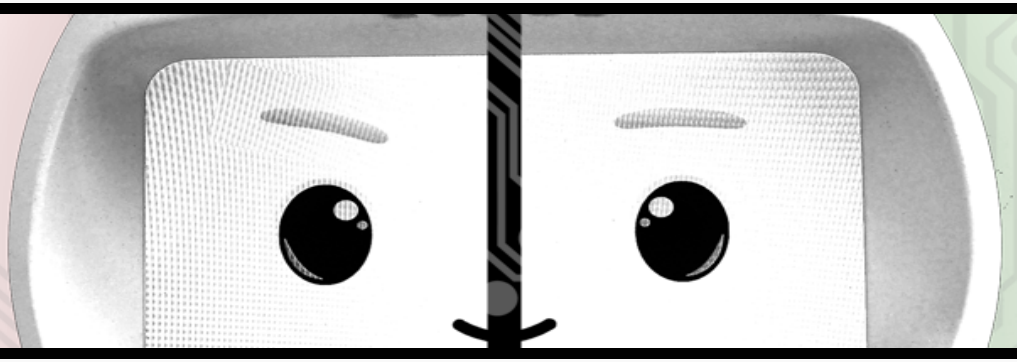
Ongoing and further work

- workbench of deontic logics (expressive, logic combinations)
- formalisation and mechanisation of foundational ethical theories
- experiments . . . deployment

Ethics

Argued for explicit ethical reasoning competencies in IASs

- development of normative reasoning experimentation platform
- utilising HOL as universal meta-logic
- practical evidence from previous work
- suitable also for teaching



Ongoing and further work

- workbench of deontic logics (expressive, logic combinations)
- formalisation and mechanisation of foundational ethical theories
- experiments . . . deployment

Before I forget: — A big thanks to —

University of Luxembourg:

ILIAS group of Leon van der Torre, many others

Research Grants:

DFG, Heisenberg grant: Computational Metaphysics, BE 2501/9, **2012-2017**

DFG, Project Leo-III: Higher-Order Theorem Prover, BE 2501/9, **2013-2017**

Various Collaborators:

B. Woltzenl.-P.
(ANU Canberra)

Alexander Steen
(FU Berlin)

Max Wisniewski
(FU Berlin)

Ed Zalta
(Stanford U.)

Dana Scott
(UC Berkeley)

Many further Collaborators and Students:

Matthias Bentert (TU Berlin), Jasmin Blanchette (Amsterdam), Chad Brown (Prag), Maximilian Claus, David Fuenmayor, Tobias Gleißner, Kim Kern, Daniel Kirchner, Hanna Lachnitt, Irina Makarenko (alle FU Berlin), Larry Paulson (Cambridge), Fabian Schütz, Hans-Jörg Schurr, David Streit, Marco Ziener (alle FU Berlin), und weitere Studenten in Berlin und Luxemburg