Epsilon Calculus I

"In the ε-calculus it is hard to understand anything"[1]

Georg Moser

Department of Computer Science
University of Innsbruck

International Summer School for Proof Theory in First-Order Logic,
August. 22–27, 2017

[1] © Michel Parigot

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator
- if $A(x)$ is a formula, then $\varepsilon_x A(x)$ is an ε-term

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator
- if $A(x)$ is a formula, then $\varepsilon_x A(x)$ is an ε-term
- $\varepsilon_x A(x)$ is an indefinite description:
  $\varepsilon_x A(x)$ is some x for which $A(x)$ is true

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator
- if $A(x)$ is a formula, then $\varepsilon_x A(x)$ is an ε-term
- $\varepsilon_x A(x)$ is an indefinite description:
  $\varepsilon_x A(x)$ is some x for which $A(x)$ is true
- ε can replace $\exists$: $\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x))$

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator
- if $A(x)$ is a formula, then $\varepsilon_x A(x)$ is an ε-term
- $\varepsilon_x A(x)$ is an indefinite description:
  $\varepsilon_x A(x)$ is some x for which $A(x)$ is true
- ε can replace $\exists$: $\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x))$
- axioms of ε-calculus:

  1. propositional tautologies
  2. equality axioms
  3. $A(t) \rightarrow A(\varepsilon_x A(x))$

What is the Epsilon Calculus?

Definition

- the ε-calculus is a formalisation of logic without quantifiers but with the ε-operator
- if $A(x)$ is a formula, then $\varepsilon_x A(x)$ is an ε-term
- $\varepsilon_x A(x)$ is an indefinite description:
  $\varepsilon_x A(x)$ is some x for which $A(x)$ is true
- ε can replace $\exists$: $\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x))$
- axioms of ε-calculus:
  1. propositional tautologies
  2. equality axioms
  3. $A(t) \to A(\varepsilon_x A(x))$

predicate logic can be embedded in the ε-calculus

Why Should You Care?

Why Should You Care?

1. basis of proof theory

Why Should You Care?

1. basis of proof theory
2. interesting logical formalism
   - trade logical structure for term structure
   - formalisation of choice; recognised in its use in proof assistants, like Coq, Isabelle, etc.
   - full potential for linguistics and computer science is unexplored

Why Should You Care?

1. basis of proof theory

2. interesting logical formalism
   - trade logical structure for term structure
   - formalisation of choice; recognised in its use in proof assistants, like Coq, Isabelle, etc.
   - full potential for linguistics and computer science is unexplored

3. foundation of noteworthy proof-theoretic results
   - ε-theorems and Herbrand's theorem (this lecture)
   - ε-substitution method and its connection to learning
   - Kreisel's no-counterexample interpretation
   - foundation of "unsound, but short proofs" (Matthias' lecture)

Why Should You Care?

1. basis of proof theory

2. interesting logical formalism
   - trade logical structure for term structure
   - formalisation of choice; recognised in its use in proof assistants, like Coq, Isabelle, etc.
   - full potential for linguistics and computer science is unexplored

3. foundation of noteworthy proof-theoretic results
   - ε-theorems and Herbrand's theorem (this lecture)
   - ε-substitution method and its connection to learning
   - Kreisel's no-counterexample interpretation
   - foundation of "unsound, but short proofs" (Matthias' lecture)

4. applications and great interest in linguistics
   - choice functions
   - *anaphora*, that is, an expression whose interpretation depends upon another expression in context

Outline

- Historical Remarks
- Axiomatisation
- The Embedding Lemma
- The First Epsilon Theorem

Rough Timeline

- **1922** introduced by Hilbert in 1921, as the basis for his formulation of mathematics for which Hilbert's Program was supposed to be carried out
- **1930s** original work in proof theory (pre-Gentzen) concentrated on ε-calculus and ε-substitution method (Ackermann [Ack25, Ack40], von Neumann [vN27], Bernays [HB39], see also [Zac03, MZ06, Mos06, Zac17])
- **1950s** ε-substitution method used by Kreisel for no-counterexample interpretation [Kre52, Kre58, Koh99] leading to work on proof analysis by Kreisel, Luckhardt, Kohlenbach
- **1990s** use of the ε-substitution method for ordinal analysis by Arai, Avigad, Mints, Tait
- **recent** renewed interest in connection to update procedures and learning [Avi02, Asc11, Pow16] and computational interpretations

Axioms of the Epsilon Calculus

Definitions

- AxEC: all substitution instances of propositional tautologies
- AxEC$_\varepsilon$: AxEC + all substitution instances of

$$\underbrace{A(t) \to A(\varepsilon_x A(x))}_{\text{critical formula}}$$

- AxPC: AxEC + all substitution instances of

$$A(a) \to \exists x\, A(x) \qquad \forall x\, A(x) \to A(a)$$

- AxPC$_\varepsilon$: AxPC + all substitution instances of critical formulas

Definitions

- a proof in EC (EC_ε) is a sequence $A_1, \ldots, A_n$ of formulas such that each A_i is either in AxEC ($AxEC_\varepsilon$) or it follows from formulas preceding it by modus ponens

- a proof in PC (PC_ε) is a sequence $A_1, \ldots, A_n$ of formulas such that each A_i is either in AxPC ($AxPC_\varepsilon$) or follows from formulas preceding it by modus ponens or generalisation

- if A is provable in say EC_ε we write $EC_\varepsilon \vdash_\pi A$

Definitions

- a proof in EC (EC_ε) is a sequence $A_1, \dots, A_n$ of formulas such that each A_i is either in AxEC ($AxEC_\varepsilon$) or it follows from formulas preceding it by modus ponens

- a proof in PC (PC_ε) is a sequence $A_1, \dots, A_n$ of formulas such that each A_i is either in AxPC ($AxPC_\varepsilon$) or follows from formulas preceding it by modus ponens or generalisation

- if A is provable in say EC_ε we write $EC_\varepsilon \vdash_\pi A$

- the size $\mathrm{sz}(\pi)$ of a proof π is the number of steps in π

Definitions

- a proof in EC (EC_ε) is a sequence $A_1, \ldots, A_n$ of formulas such that each A_i is either in AxEC ($\mathrm{AxEC}_\varepsilon$) or it follows from formulas preceding it by modus ponens
- a proof in PC (PC_ε) is a sequence $A_1, \ldots, A_n$ of formulas such that each A_i is either in AxPC ($\mathrm{AxPC}_\varepsilon$) or follows from formulas preceding it by modus ponens or generalisation
- if A is provable in say EC_ε we write $\mathrm{EC}_\varepsilon \vdash_\pi A$
- the size $\mathrm{sz}(\pi)$ of a proof π is the number of steps in π
- the critical count $\mathrm{cc}(\pi)$ of π is the number of distinct critical formulas and quantifier axioms in π (plus 1)

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping $^{\varepsilon}$:

$$f(t_1, \ldots, t_n)^{\varepsilon} = f(t_1^{\varepsilon}, \ldots, t_n^{\varepsilon})$$

$$x^{\varepsilon} = x \qquad\qquad [\varepsilon_x A(x)]^{\varepsilon} = \varepsilon_x A^{\varepsilon}(x)$$

$$a^{\varepsilon} = a$$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping ${}^{\varepsilon}$:

$$f(t_1, \dots, t_n)^{\varepsilon} = f(t_1^{\varepsilon}, \dots, t_n^{\varepsilon}) \qquad P(t_1, \dots, t_n)^{\varepsilon} = P(t_1^{\varepsilon}, \dots, t_n^{\varepsilon})$$

$$x^{\varepsilon} = x \qquad [\varepsilon_x A(x)]^{\varepsilon} = \varepsilon_x A^{\varepsilon}(x)$$

$$a^{\varepsilon} = a$$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping $^\varepsilon$:

$$f(t_1, \dots, t_n)^\varepsilon = f(t_1^\varepsilon, \dots, t_n^\varepsilon) \qquad P(t_1, \dots, t_n)^\varepsilon = P(t_1^\varepsilon, \dots, t_n^\varepsilon)$$

$$x^\varepsilon = x \qquad (A \to B)^\varepsilon = A^\varepsilon \to B^\varepsilon \qquad [\varepsilon_x A(x)]^\varepsilon = \varepsilon_x A^\varepsilon(x)$$

$$a^\varepsilon = a \qquad (A \vee B)^\varepsilon = A^\varepsilon \vee B^\varepsilon$$

$$(\neg A)^\varepsilon = \neg A^\varepsilon \qquad (A \wedge B)^\varepsilon = A^\varepsilon \wedge B^\varepsilon$$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping $^\varepsilon$:

$$f(t_1,\dots,t_n)^\varepsilon = f(t_1^\varepsilon,\dots,t_n^\varepsilon) \qquad P(t_1,\dots,t_n)^\varepsilon = P(t_1^\varepsilon,\dots,t_n^\varepsilon)$$

$$x^\varepsilon = x \qquad (A \to B)^\varepsilon = A^\varepsilon \to B^\varepsilon \qquad [\varepsilon_x A(x)]^\varepsilon = \varepsilon_x A^\varepsilon(x)$$

$$a^\varepsilon = a \qquad (A \vee B)^\varepsilon = A^\varepsilon \vee B^\varepsilon \qquad (\exists x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x A^\varepsilon(x))$$

$$(\neg A)^\varepsilon = \neg A^\varepsilon \qquad (A \wedge B)^\varepsilon = A^\varepsilon \wedge B^\varepsilon \qquad (\forall x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x \neg A^\varepsilon(x))$$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping ${}^\varepsilon$:

$$f(t_1,\dots,t_n)^\varepsilon = f(t_1^\varepsilon,\dots,t_n^\varepsilon) \qquad P(t_1,\dots,t_n)^\varepsilon = P(t_1^\varepsilon,\dots,t_n^\varepsilon)$$

$$x^\varepsilon = x \qquad (A \to B)^\varepsilon = A^\varepsilon \to B^\varepsilon \qquad [\varepsilon_x A(x)]^\varepsilon = \varepsilon_x A^\varepsilon(x)$$

$$a^\varepsilon = a \qquad (A \vee B)^\varepsilon = A^\varepsilon \vee B^\varepsilon \qquad (\exists x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x A^\varepsilon(x))$$

$$(\neg A)^\varepsilon = \neg A^\varepsilon \qquad (A \wedge B)^\varepsilon = A^\varepsilon \wedge B^\varepsilon \qquad (\forall x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x \neg A^\varepsilon(x))$$

Lemma (Embedding Lemma)

if π is a regular PC_ε-proof of A then there is an EC_ε-proof π^ε of A^ε with $\mathrm{sz}(\pi^\varepsilon) \leqslant 3 \cdot \mathrm{sz}(\pi)$ and $\mathrm{cc}(\pi^\varepsilon) \leqslant \mathrm{cc}(\pi)$

Definition (tentative)

quantifiers in a quantifier-free system:

$$\exists x\, A(x) \Leftrightarrow A(\varepsilon_x A(x)) \qquad \forall x\, A(x) \Leftrightarrow A(\varepsilon_x \neg A(x))$$

Definition

define a mapping $^\varepsilon$:

$$f(t_1, \dots, t_n)^\varepsilon = f(t_1^\varepsilon, \dots, t_n^\varepsilon) \qquad P(t_1, \dots, t_n)^\varepsilon = P(t_1^\varepsilon, \dots, t_n^\varepsilon)$$

$$x^\varepsilon = x \qquad (A \to B)^\varepsilon = A^\varepsilon \to B^\varepsilon \qquad [\varepsilon_x A(x)]^\varepsilon = \varepsilon_x A^\varepsilon(x)$$

$$a^\varepsilon = a \qquad (A \lor B)^\varepsilon = A^\varepsilon \lor B^\varepsilon \qquad (\exists x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x A^\varepsilon(x))$$

$$(\neg A)^\varepsilon = \neg A^\varepsilon \qquad (A \land B)^\varepsilon = A^\varepsilon \land B^\varepsilon \qquad (\forall x\, A(x))^\varepsilon = A^\varepsilon(\varepsilon_x \neg A^\varepsilon(x))$$

Lemma (Embedding Lemma)

if π is a regular PC_ε-proof of A then there is an EC_ε-proof π^ε of A^ε with $\mathrm{sz}(\pi^\varepsilon) \leqslant 3 \cdot \mathrm{sz}(\pi)$ and $\mathrm{cc}(\pi^\varepsilon) \leqslant \mathrm{cc}(\pi)$

Example: Epsilon Mapping

Example

$$[\exists x(P(x) \quad \vee \quad \forall y Q(y))]^{\varepsilon} =$$

Example: Epsilon Mapping

Example

$$\begin{aligned} [\exists x(P(x) \quad &\vee \quad \forall y Q(y))]^{\varepsilon} = \\ &= \quad [P(x) \vee \forall y Q(y)]^{\varepsilon} \quad \{x \leftarrow \varepsilon_x[P(x) \vee \forall y Q(y)]^{\varepsilon}\} \end{aligned}$$

Example: Epsilon Mapping

Example

$$[\exists x(P(x) \quad \vee \quad \forall y Q(y))]^{\varepsilon} =$$
$$= \quad [P(x) \vee \forall y Q(y)]^{\varepsilon} \quad \{x \leftarrow \varepsilon_x [P(x) \vee \forall y Q(y)]^{\varepsilon}\}$$
$$[P(x) \vee \forall y Q(y)]^{\varepsilon} = P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})$$

Example: Epsilon Mapping

Example

$$
\begin{aligned}
[\exists x(P(x) \quad \vee \quad \forall y Q(y))]^{\varepsilon} &= \\
&= [P(x) \vee \forall y Q(y)]^{\varepsilon} \quad \{x \leftarrow \varepsilon_x [P(x) \vee \forall y Q(y)]^{\varepsilon}\} \\
&\qquad [P(x) \vee \forall y Q(y)]^{\varepsilon} = P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \\
&= P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \quad \{x \leftarrow \underbrace{\varepsilon_x [P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}\}
\end{aligned}
$$

Example: Epsilon Mapping

Example

$$
\begin{aligned}
[\exists x(P(x) \;\vee\; \forall y Q(y))]^{\varepsilon} &= \\
&= [P(x) \vee \forall y Q(y)]^{\varepsilon} \quad \{x \leftarrow \varepsilon_x[P(x) \vee \forall y Q(y)]^{\varepsilon}\} \\
&\quad [P(x) \vee \forall y Q(y)]^{\varepsilon} = P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \\
&= P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1}) \quad \underbrace{\{x \leftarrow \varepsilon_x[P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]\}}_{e_2} \\
&= P(\underbrace{\varepsilon_x[P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})
\end{aligned}
$$

Drinker's Paradox

Example

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(a) \Rightarrow P(a)}{P(a) \Rightarrow P(a), \forall y P(y)}}{\Rightarrow P(a) \to \forall y P(y), P(a)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), P(a)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), \forall y P(y)}}{P(b) \Rightarrow \exists x(P(x) \to \forall y P(y)), \forall y P(y)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), P(b) \to \forall y P(y)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), \exists x(P(x) \to \forall y P(y))}}{\Rightarrow \exists x(P(x) \to \forall y P(y))}$$

Drinker's Paradox

Example

$$
\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{\cfrac{P(a) \Rightarrow P(a)}{P(a) \Rightarrow P(a), \forall y P(y)}}{\Rightarrow P(a) \to \forall y P(y), P(a)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), P(a)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), \forall y P(y)}}{P(b) \Rightarrow \exists x(P(x) \to \forall y P(y)), \forall y P(y)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), P(b) \to \forall y P(y)}}{\Rightarrow \exists x(P(x) \to \forall y P(y)), \exists x(P(x) \to \forall y P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))}
$$

where we employ

$$
[\forall y P(y)]^\varepsilon = P(\varepsilon_y \neg P(y))
$$

$$
[\exists x(P(x) \to \forall y P(y)]^\varepsilon = P(\underbrace{\varepsilon_x(P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))
$$

Drinker's Paradox

Example

$$\begin{array}{c}
P(a) \Rightarrow P(a) \\ \hline
P(a) \Rightarrow P(a), \forall y P(y) \\ \hline
\Rightarrow P(a) \to \forall y P(y), P(a) \\ \hline
\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(a) \\ \hline
\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), \forall y P(y) \\ \hline
P(b) \Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), \forall y P(y) \\ \hline
\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(b) \to \forall y P(y) \\ \hline
\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y)) \\ \hline
\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))
\end{array}$$

where we employ

$$[\forall y P(y)]^{\varepsilon} = P(\varepsilon_y \neg P(y))$$

$$[\exists x (P(x) \to \forall y P(y)]^{\varepsilon} = P(\underbrace{\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))$$

Drinker's Paradox

Example

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(a) \Rightarrow P(a)}{P(a) \Rightarrow P(a), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(a) \to P(\varepsilon_y \neg P(y)), P(a)}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(a)}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{P(b) \Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(b) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))}$$

where we employ

$$[\forall y P(y)]^{\varepsilon} = P(\varepsilon_y \neg P(y))$$

$$[\exists x (P(x) \to \forall y P(y)]^{\varepsilon} = P(\underbrace{\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))$$

Drinker's Paradox

Example

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y))}{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{P(\varepsilon) \Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))}$$

where we employ

$$[\forall y P(y)]^\varepsilon = P(\varepsilon_y \neg P(y))$$

$$[\exists x (P(x) \to \forall y P(y)]^\varepsilon = P(\underbrace{\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))$$

Drinker's Paradox

Example

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y))}{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{P(\varepsilon) \Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))}$$

where we employ

$$[\forall y P(y)]^\varepsilon = P(\varepsilon_y \neg P(y))$$

$$[\exists x (P(x) \to \forall y P(y)]^\varepsilon = P(\underbrace{\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))$$

Drinker's Paradox

Example

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y))}{P(\varepsilon_y \neg P(y)) \Rightarrow P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{P(\varepsilon) \Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y)), P(\varepsilon) \to P(\varepsilon_y \neg P(y))}}{\Rightarrow P(\varepsilon) \to P(\varepsilon_y \neg P(y))}$$

where we employ

$$[P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y))] \to [P(\underbrace{\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y)))}_{\varepsilon}) \to P(\varepsilon_y \neg P(y))]$$

Drinker's Paradox (II)

Example (cont'd)

1	$P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y))$	TAUT
2	$(P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y))) \to$	
	$\to (P(\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y))))) \to P(\varepsilon_y \neg P(y)))$	critical axiom
3	$P(\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y))))) \to P(\varepsilon_y \neg P(y))$	$1, 2$, *MP*

Drinker's Paradox (II)

Example (cont'd)

1	$P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y))$	TAUT
2	$(P(\varepsilon_y \neg P(y)) \to P(\varepsilon_y \neg P(y))) \to$	
	$\to (P(\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y))))) \to P(\varepsilon_y \neg P(y)))$	critical axiom
3	$P(\varepsilon_x (P(x) \to P(\varepsilon_y \neg P(y))))) \to P(\varepsilon_y \neg P(y))$	$1, 2,$ *MP*

Remarks

- ε-calculus allows proof compression
- propositional inferences and structural rules become irrelevant
- focus on quantifier inferences

Lemma

for any semi-formula A, a fresh, any term $t \in \mathrm{L}(\mathrm{PC}_\varepsilon)$ and any substitution for free variables, we have

1. $(A\{a \mapsto t\})^\varepsilon = A^\varepsilon\{a \mapsto t^\varepsilon\}$
2. $(A\{a \mapsto t\})\sigma = A\sigma\{a \mapsto t\sigma\}$

Proof.

by induction on A ∎

Lemma

for any semi-formula A, a fresh, any term $t \in \mathrm{L}(\mathrm{PC}_\varepsilon)$ and any substitution for free variables, we have

1. $(A\{a \mapsto t\})^\varepsilon = A^\varepsilon\{a \mapsto t^\varepsilon\}$
2. $(A\{a \mapsto t\})\sigma = A\sigma\{a \mapsto t\sigma\}$

Proof.

by induction on A ∎

Corollary

for any formula A and any substitution from free variables into $\mathrm{L}(\mathrm{EC}_\varepsilon)$, we have $(A\sigma)^\varepsilon = A^\varepsilon\sigma$

Lemma

for any semi-formula A, a fresh, any term $t \in \mathrm{L}(\mathrm{PC}_\varepsilon)$ and any substitution for free variables, we have

1. $(A\{a \mapsto t\})^\varepsilon = A^\varepsilon\{a \mapsto t^\varepsilon\}$
2. $(A\{a \mapsto t\})\sigma = A\sigma\{a \mapsto t\sigma\}$

Proof.

by induction on A ∎

Corollary

for any formula A and any substitution from free variables into $\mathrm{L}(\mathrm{EC}_\varepsilon)$, we have $(A\sigma)^\varepsilon = A^\varepsilon\sigma$

Lemma

for any semi-formula A: $\mathsf{Var}(A) = \mathsf{Var}(A^\varepsilon)$

(Corrected) Proof of the Embedding Lemma[1]

Definition

- let $\pi\colon A_1, \ldots, A_n$ be a regular proof in PC_ε
- let $E_1, \ldots, E_p$ be a subsequence of π consisting of conclusions of quantifier rules
- for each $j \leqslant p$, let a_j and $C_j(a_j)$ denote the eigenvariable and formula associated to E_j

[1] Existing proofs in the literature [HB39, MZ06, Zac17] are false, correction due to M. Parigot.

(Corrected) Proof of the Embedding Lemma[1]

Definition

- let $\pi \colon A_1, \dots, A_n$ be a regular proof in PC_ε
- let $E_1, \dots, E_p$ be a subsequence of π consisting of conclusions of quantifier rules
- for each $j \leqslant p$, let a_j and $C_j(a_j)$ denote the eigenvariable and formula associated to E_j

then we define a substitution σ from free variables to $\mathrm{L}(\mathrm{EC}_\varepsilon)$:

$$\sigma = \{a_1 \mapsto t_1\} \circ \cdots \circ \{a_p \mapsto t_p\}$$

$$\text{where } t_j = \begin{cases} \varepsilon_x \neg C_j^\varepsilon(x) & C_j(a) \text{ associated to } \forall i \\ \varepsilon_x C_j^\varepsilon(x) & C_j(a) \text{ associated to } \exists i \end{cases}$$

[1] Existing proofs in the literature [HB39, MZ06, Zac17] are false, correction due to M. Parigot.

Definition

for $q \leqslant p$, define

$$\sigma_{<q} = \{a_1 \mapsto t_1\} \circ \cdots \circ \{a_{q-1} \mapsto t_{q-1}\}$$

define $\sigma_{>q}$ similarly

Definition

for $q \leqslant p$, define

$$\sigma_{<q} = \{a_1 \mapsto t_1\} \circ \cdots \circ \{a_{q-1} \mapsto t_{q-1}\}$$

define $\sigma_{>q}$ similarly

Observations

for $q \leqslant p$ and $j \leqslant q$

- for any formula A: $(A\sigma)^\varepsilon = A^\varepsilon\sigma$
- C_q doesn't contain a_j
- $a_q\sigma = t_q\sigma_{>q}$

Proof.

using the regularity assumption and the previous corollary ∎

Proof of Embedding Lemma.

- we show $\forall$ regular proofs $\pi\colon A_1, \dots, A_n$
  $\exists$ proof π^ε containing $A_1^\varepsilon\sigma, \dots, A_n^\varepsilon\sigma$ (+ extra formulas)
- we use by induction on n

Proof of Embedding Lemma.

- we show $\forall$ regular proofs $\pi\colon A_1, \dots, A_n$
  $\exists$ proof π^ε containing $A_1^\varepsilon\sigma, \dots, A_n^\varepsilon\sigma$ (+ extra formulas)
- we use by induction on n
- base case is trivial and if $A_n =: A$ is a propositional tautology, $A^\varepsilon\sigma$ is also a tautology

Proof of Embedding Lemma.

- we show $\forall$ regular proofs $\pi\colon A_1, \ldots, A_n$
  $\exists$ proof π^ε containing $A_1^\varepsilon\sigma, \ldots, A_n^\varepsilon\sigma$ (+ extra formulas)
- we use by induction on n
- base case is trivial and if $A_n =: A$ is a propositional tautology, $A^\varepsilon\sigma$ is also a tautology
- Case A an instance of a quantifier axiom; suppose
  $A = A(t) \to \exists x A(x)$; hence

$$[A(t) \to \exists x\, A(x)]^\varepsilon\sigma = A^\varepsilon\sigma(t^\varepsilon\sigma) \to A^\varepsilon\sigma(\varepsilon_x A^\varepsilon\sigma(x))$$

  the latter is an instance of a critical axiom

Proof of Embedding Lemma.

- we show $\forall$ regular proofs $\pi \colon A_1, \dots, A_n$
  $\exists$ proof π^ε containing $A_1^\varepsilon\sigma, \dots, A_n^\varepsilon\sigma$ (+ extra formulas)
- we use by induction on n
- base case is trivial and if $A_n =: A$ is a propositional tautology, $A^\varepsilon\sigma$ is also a tautology
- Case A an instance of a quantifier axiom; suppose $A = A(t) \to \exists x A(x)$; hence

$$[A(t) \to \exists x\, A(x)]^\varepsilon\sigma = A^\varepsilon\sigma(t^\varepsilon\sigma) \to A^\varepsilon\sigma(\varepsilon_x A^\varepsilon\sigma(x))$$

  the latter is an instance of a critical axiom
- Case A follows by modus ponens from A_i and $A_j \equiv A_i \to A$
  applying IH there exists a proof π^* containing $A_i^\varepsilon\sigma$ and $A_i^\varepsilon\sigma \to A_j^\varepsilon\sigma$;
  we add $A^\varepsilon\sigma$ to π^*

Proof (cont'd).

- Case A follows by quantifier rule; e.g. $A = B \to \forall x\, C_q(x)$ and there exists $A_i = B \to C_q(a)$; a eigenvariable

  by IH there exists a proof π^* containing $A_i^\varepsilon \sigma \equiv B^\varepsilon \sigma \to C_q^\varepsilon(a)\sigma$; by definition σ replacing the eigenvariable a by $\varepsilon_x \neg A^\varepsilon(x)\sigma$.

Proof (cont'd).

- Case A follows by quantifier rule; e.g. $A = B \to \forall x\, C_q(x)$ and there exists $A_i = B \to C_q(a)$; a eigenvariable

  by IH there exists a proof π^* containing $A_i^\varepsilon\sigma \equiv B^\varepsilon\sigma \to C_q^\varepsilon(a)\sigma$; by definition σ replacing the eigenvariable a by $\varepsilon_x \neg A^\varepsilon(x)\sigma$. Hence

$$\begin{aligned} A_i^\varepsilon\sigma &= B^\varepsilon\sigma \to C_q^\varepsilon\sigma(\varepsilon_x \neg A^\varepsilon(x)\sigma) \\ &= B^\varepsilon\sigma_{>q} \to C_q^\varepsilon\sigma_{>q}(\varepsilon_x \neg A^\varepsilon(x)\sigma_{>q}) \\ &= B^\varepsilon\sigma_{>q} \to [\forall x\ C_q(x)]^\varepsilon\sigma_{>q} \\ &= B^\varepsilon\sigma \to [\forall x\ C_q(x)]^\varepsilon\sigma \\ &= A^\varepsilon\sigma \end{aligned}$$

  and thus we can set $\pi^\varepsilon = \pi^*$

Proof (cont'd).

- Case A follows by quantifier rule; e.g. $A = B \to \forall x\, C_q(x)$ and there exists $A_i = B \to C_q(a)$; a eigenvariable

  by IH there exists a proof π^* containing $A_i^\varepsilon\sigma \equiv B^\varepsilon\sigma \to C_q^\varepsilon(a)\sigma$; by definition σ replacing the eigenvariable a by $\varepsilon_x \neg A^\varepsilon(x)\sigma$. Hence

$$
\begin{aligned}
A_i^\varepsilon\sigma &= B^\varepsilon\sigma \to C_q^\varepsilon\sigma(\varepsilon_x \neg A^\varepsilon(x)\sigma) \\
&= B^\varepsilon\sigma_{>q} \to C_q^\varepsilon\sigma_{>q}(\varepsilon_x \neg A^\varepsilon(x)\sigma_{>q}) \\
&= B^\varepsilon\sigma_{>q} \to [\forall x\, C_q(x)]^\varepsilon\sigma_{>q} \\
&= B^\varepsilon\sigma \to [\forall x\, C_q(x)]^\varepsilon\sigma \\
&= A^\varepsilon\sigma
\end{aligned}
$$

  and thus we can set $\pi^\varepsilon = \pi^*$

The First Epsilon Theorem

Theorem

suppose $E(e_1, \dots, e_m)$ *is a quantifier-free formula containing only the* ε*-terms* $s_1, \dots, s_m$*, and*

$$\mathsf{EC}_\varepsilon \vdash_\pi E(s_1, \dots, s_m)$$

then there are ε*-free terms* t^i_j *such that*

$$\mathsf{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \dots, t^i_m)$$

The First Epsilon Theorem

Theorem

suppose $E(e_1, \dots, e_m)$ *is a quantifier-free formula containing only the* ε*-terms* s_1, …, s_m, *and*

$$\mathrm{EC}_\varepsilon \vdash_\pi E(s_1, \dots, s_m)$$

then there are ε*-free terms* t^i_j *such that*

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \dots, t^i_m)$$

where $n \leqslant 2^{3 \cdot \mathrm{cc}(\pi)}_{2 \cdot \mathrm{cc}(\pi)}$

The First Epsilon Theorem

Theorem

suppose $E(e_1, \dots, e_m)$ is a quantifier-free formula containing only the ε-terms s_1, ..., s_m, and

$$\mathrm{EC}_\varepsilon \vdash_\pi E(s_1, \dots, s_m)$$

then there are ε-free terms t^i_j such that

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \dots, t^i_m)$$

where $n \leqslant 2^{3 \cdot \mathrm{cc}(\pi)}_{2 \cdot \mathrm{cc}(\pi)}$

number of instances independent off # of propositional inferences

Herbrand's Theorem

Theorem

if $\exists x_1 \dots \exists x_m E(x_1, \dots, x_m)$ *is a purely existential formula containing only the bound variables* $x_1, \dots, x_m$*, and*

$$\mathrm{PC} \vdash_\pi \exists x_1 \dots \exists x_m E(x_1, \dots, x_m) \ ,$$

then there are *terms* t^i_j *such that*

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \dots, t^i_m)$$

Herbrand's Theorem

Theorem

if $\exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m)$ *is a purely existential formula containing only the bound variables* $x_1, \ldots, x_m$, *and*

$$\mathrm{PC} \vdash_\pi \exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m) \;,$$

then there are *terms* t^i_j *such that*

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \ldots, t^i_m)$$

where $n \leqslant 2^{3\cdot\mathrm{cc}(\pi)}_{2\cdot\mathrm{cc}(\pi)}$

Herbrand's Theorem

Theorem

if $\exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m)$ *is a purely existential formula containing only the bound variables* x_1, …, x_m, *and*

$$\mathrm{PC} \vdash_\pi \exists x_1 \ldots \exists x_m E(x_1, \ldots, x_m) \ ,$$

then there are *terms* t^i_j *such that*

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \ldots, t^i_m)$$

where $n \leqslant 2^{3 \cdot \mathrm{cc}(\pi)}_{2 \cdot \mathrm{cc}(\pi)}$

length of Herbrand disjunction independent off # of propositional inferences

Herbrand's Theorem

Theorem

if $\exists x_1 \dots \exists x_m E(x_1, \dots, x_m)$ *is a purely existential formula containing only the bound variables* x_1, ..., x_m, *and*

$$\mathrm{PC}_\varepsilon \vdash_\pi \exists x_1 \dots \exists x_m E(x_1, \dots, x_m) \ ,$$

then there are ε*-free terms* t^i_j *such that*

$$\mathrm{EC} \vdash \bigvee_{i=1}^{n} E(t^i_1, \dots, t^i_m)$$

where $n \leqslant 2^{3 \cdot \mathrm{cc}(\pi)}_{2 \cdot \mathrm{cc}(\pi)}$

length of Herbrand disjunction independent off # of propositional inferences

Degree and Rank

Definition (degree)

- $\deg(\varepsilon_x A(x)) = 1$ if $A(x)$ contains no ε-subterms
- If $e_1, \ldots, e_n$ are all immediate ε-subterms of $A(x)$, then $\deg(\varepsilon_x A(x)) = \max\{\deg(e_1), \ldots, \deg(e_n)\} + 1$

Definition (rank)

- An ε-expression e is subordinate to $\varepsilon_x A$ if e is a proper sub-expression of A and contains x
- $\mathrm{rk}(e) = 1$ if no sub-ε-expression of e is subordinate to e
- If $e_1, \ldots, e_n$ are all the ϵ-expressions subordinate to e, then $\mathrm{rk}(e) = \max\{\mathrm{rk}(e_1), \ldots, \mathrm{rk}(e_n)\} + 1$

Example

consider

$$P(\underbrace{\varepsilon_x[P(x) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})]}_{e_2}) \vee Q(\underbrace{\varepsilon_y \neg Q(y)}_{e_1})$$

then

$$\deg(e_1) = 1 \qquad \deg(e_2) = 2 \qquad \mathrm{rk}(e_1) = \mathrm{rk}(e_2) = 1$$

Example

$$A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x,z))}_{e_2}, \underbrace{\varepsilon_y A(\underbrace{\varepsilon_x A(x, \varepsilon_z A(x,z))}_{e_2}, y)}_{e_1(e_2)})$$

then $\deg(e_2) = 1$, $\deg(e_1(e_2)) = 2$, $\mathrm{rk}(e_2) = 2$, $\mathrm{rk}(e_1(e_2)) = 1$

Rank of Critical Formulas and Derivations

Definitions

- rank of a critical formula $A(t) \to A(\varepsilon_x A(x))$ is $\mathrm{rk}(\varepsilon_x A(x))$
- rank of a derivation $\mathrm{rk}(\pi)$: maximum rank of its critical formulas
- critical ϵ-term of a derivation: ε-term e so that $A(t) \to A(e)$ is a critical formula
- degree of a derivation $\deg(\pi)$: maximum degree of its critical ε-terms of maximal rank
- order of a derivation $o(\pi, r)$ wrt. rank r: number of different critical ε-terms of rank r

W. Ackermann.
Begründung des Tertium non datur mittels der Hilbertschen Theorie der Widerspruchsfreiheit.
Mathematische Annalen, 93:1–36, 1925.

W. Ackermann.
Zur Widerspruchsfreiheit der Zahlentheorie.
Mathematische Annalen, 117:162–194, 1940.

F. Aschieri.
Transfinite update procedures for predicative systems of analysis.
In *Proc. 25th CSL*, volume 12 of *LIPIcs*, pages 20–34, 2011.

J. Avigad.
Update procedures and the 1-consistency of arithmetic.
Mathematical Logic Quarterly, 48:3–13, 2002.

David Hilbert and Paul Bernays.
Grundlagen der Mathematik, volume 2.
Springer, Berlin, 1939.

U. Kohlenbach.
On the no-counterexample interpretation.
Journal of Symbolic Logic, 64:1491–1511, 1999.

Georg Kreisel.
Interpretation of non-finitist proofs I.
Journal of Symbolic Logic, 16:241–267, 1952.

G. Kreisel.
Mathematical significance of consistency proof.
Journal of Symbolic Logic, 23:155–182, 1958.

G. Moser.
Ackermann's substitution method (remixed).
Annals of Pure and Applied Logic, 142(1–3):1–18, 2006.

G. Moser and R. Zach.
The Epsilon calculus and Herbrand complexity.
Studia Logica, 82(1):133–155, 2006.

Thomas Powell.
Gödel's functional interpretation and the concept of learning.
In *Proc. 31th LICS*, pages 136–145. ACM, 2016.

John von Neumann.
Zur Hilbertschen Beweistheorie.
Mathematische Zeitschrift, 26:1–46, 1927.

R. Zach.
The practice of finitism. Epsilon calculus and consistency proofs in Hilbert's program.
Synthese, 137:211–259, 2003.

R. Zach.
Semantics and proof theory of the epsilon calculus.
In *Proc. 7th ICLA*, volume 10119 of *LNCS*, pages 27–47, 2017.