Numbers, constructive truth, and the Kreisel-Goodman Paradox

Göteborg
19-21 October 2012

Walter Dean
Department of Philosophy
University of Warwick

w.h.dean@warwick.ac.uk

Hidenori Kurkokawa
Department of Philosophy
City University of New York

hkurokawa@gc.cuny.edu

General problematics

- What is the status of the semantic paradoxes relative to the constructive understanding of truth?
- More specifically:
  1) Are the combinatory principles implicit in the BHK interpretation of the intuitionistic connectives *consistent*?
  2) And can they be used to interpret arithmetic?
- What does this tell us about constructive provability?
- How does self-reference emanating from combinatory logic bear on the semantic paradoxes?

Outline

1) Constructive truth and the BHK interpretation
2) Montague's Theorem (aka "The Provability Liar") as an antinomy for Constructivism?
3) Kreisel's Theory of Constructions ($\mathcal{C}$) and the Kreisel-Goodman Paradox
4) A tentative diagnosis: quantification over proofs, reflection, and the role of Internalization (aka "Necessitation")
5) Interpreting HA in $\mathcal{C}^+$

The constructive view of truth

$$\text{The P-schema:} \quad \varphi \text{ is true} \iff \varphi \text{ is constructively provable}$$
$$\iff \exists \mathfrak{p} \text{ s.t. } \mathfrak{p} \text{ verifies } \varphi \quad [\mathfrak{p} : \varphi]$$

The constructive view of truth

The P-schema: φ is true $\iff$ φ is constructively provable
$\iff \exists \mathfrak{p}$ s.t. $\mathfrak{p}$ verifies φ $\quad [\mathfrak{p} : \varphi]$

- Dummett (1982): "A sentence may then be said to be true if and only if there exists a construction that constitutes a proof of it."

The constructive view of truth

The P-schema: φ is true $\iff$ φ is constructively provable
$\iff \exists\mathfrak{p}$ s.t. $\mathfrak{p}$ verifies φ $[\mathfrak{p} : \varphi]$

- Dummett (1982): "A sentence may then be said to be true if and only if there exists a construction that constitutes a proof of it."
- Martin-Löf (1995): "[T]he notion of truth is not taken as a primitive notion, like a truth conditional theory of meaning, but is rather defined in terms of an underlying notion of verification by the principle that A is true if there exists a proof of A."

The constructive view of truth

The P-schema: φ is true $\iff$ φ is constructively provable
$\iff \exists \mathfrak{p}$ s.t. $\mathfrak{p}$ verifies φ $[\mathfrak{p} : \varphi]$

- Dummett (1982): "A sentence may then be said to be true if and only if there exists a construction that constitutes a proof of it."
- Martin-Löf (1995): "[T]he notion of truth is not taken as a primitive notion, like a truth conditional theory of meaning, but is rather defined in terms of an underlying notion of verification by the principle that A is true if there exists a proof of A."
- Prawitz (1998): "[I]t is hardly controversial within verificationism that the truth of a proposition is to be identified with provability or existence of proofs ..."

What are constructive proofs?

- Heyting: we can given an account of the constructive meaning of φ in terms of the *proof conditions* of its constituents.

  (BHK$_\wedge$) A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.

  (BHK$_\neg$) A proof of $\neg\varphi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \bot$.

  (BHK$_\rightarrow$) A proof of $\varphi \rightarrow \psi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \psi$.

What are constructive proofs?

- Heyting: we can given an account of the constructive meaning of φ in terms of the *proof conditions* of its constituents.

  $(\mathrm{BHK}_\wedge)$ A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.

  $(\mathrm{BHK}_\neg)$ A proof of $\neg\varphi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \bot$.

  $(\mathrm{BHK}_\rightarrow)$ A proof of $\varphi \rightarrow \psi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \psi$.

- *Approximates* an implicit definition of "constructive proof".

What are constructive proofs?

- Heyting: we can given an account of the constructive meaning of φ in terms of the *proof conditions* of its constituents.

  $(\mathrm{BHK}_{\wedge})$ A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.

  $(\mathrm{BHK}_{\neg})$ A proof of $\neg\varphi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \bot$.

  $(\mathrm{BHK}_{\rightarrow})$ A proof of $\varphi \rightarrow \psi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \psi$.
- *Approximates* an implicit definition of "constructive proof".
- Gödel, Kreisel: the characterization is *impredicative* due to the quantifier over **all** constructive proofs in $\mathrm{BHK}_{\neg,\rightarrow}$.

What are constructive proofs?

- Heyting: we can given an account of the constructive meaning of φ in terms of the *proof conditions* of its constituents.

  $(\mathrm{BHK}_\wedge)$ A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.

  $(\mathrm{BHK}_\neg)$ A proof of $\neg\varphi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \bot$.

  $(\mathrm{BHK}_\rightarrow)$ A proof of $\varphi \rightarrow \psi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \psi$.

- *Approximates* an implicit definition of “constructive proof”.
- Gödel, Kreisel: the characterization is *impredicative* due to the quantifier over **all** constructive proofs in $\mathrm{BHK}_{\neg,\rightarrow}$.
- Sundholm, van Atten: an *explication* rather than an *analysis*.

What are constructive proofs?

- Heyting: we can given an account of the constructive meaning of φ in terms of the *proof conditions* of its constituents.

  (BHK$_\wedge$) A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.

  (BHK$_\neg$) A proof of $\neg\varphi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \bot$.

  (BHK$_\to$) A proof of $\varphi \to \psi$ is a construction $\mathfrak{f}$ s.t. for all proofs $\mathfrak{p}$, if $\mathfrak{p} : \varphi$, then $\mathfrak{f}(\mathfrak{p}) : \psi$.

- *Approximates* an implicit definition of "constructive proof".
- Gödel, Kreisel: the characterization is *impredicative* due to the quantifier over **all** constructive proofs in $\mathrm{BHK}_{\neg,\to}$.
- Sundholm, van Atten: an *explication* rather than an *analysis*.
- But there have still been many attempts to *formalize* BHK ...

A naive question

$$\text{P-schema:} \quad \varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$$

A naive question

$$\text{P-schema:} \quad \varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$$

- The Diagonal Lemma is intuitionistically derivable – e.g.

$$\mathrm{HA} \vdash \delta \leftrightarrow \neg \exists x \mathsf{Proof}_{\mathrm{HA}}(x, \ulcorner \delta \urcorner)$$

A naive question

$$\text{P-schema:} \quad \varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$$

- The Diagonal Lemma is intuitionistically derivable – e.g.

$$\mathrm{HA} \vdash \delta \leftrightarrow \neg \exists x \mathsf{Proof}_{\mathrm{HA}}(x, \ulcorner \delta \urcorner)$$

- So doesn't the P-schema get us into the same kind of trouble as the Liar?

A naive question

$$\text{P-schema:} \quad \varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$$

- The Diagonal Lemma is intuitionistically derivable – e.g.

$$\mathrm{HA} \vdash \delta \leftrightarrow \neg\exists x \mathsf{Proof}_{\mathrm{HA}}(x, \ulcorner\delta\urcorner)$$

- So doesn't the P-schema get us into the same kind of trouble as the Liar?
- This depends ...

Constructive proof, quantification, indefinite extensibility

1) "φ is true iff φ is constructively provable"
2) P-schema: $\varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$

Beeson (1985) asks the following about 1) - 2):

Constructive proof, quantification, indefinite extensibility

1) "φ is true iff φ is constructively provable"
2) P-schema: $\varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$

Beeson (1985) asks the following about 1) - 2):

i) Is "$\mathfrak{p}$ is a proof of φ" a proposition to be treated on an equal basis with φ itself? Or is it a proposition "of another level"?

Constructive proof, quantification, indefinite extensibility

1) "φ is true iff φ is constructively provable"
2) P-schema: $\varphi \Leftrightarrow (\exists \mathfrak{p})\ \mathfrak{p} : \varphi$

Beeson (1985) asks the following about 1) - 2):

i) Is "$\mathfrak{p}$ is a proof of φ" a proposition to be treated on an equal basis with φ itself? Or is it a proposition "of another level"?
ii) Is there a "universe" to which "everything" belongs, so that the quantifier $\exists \mathfrak{p}$ makes sense?

Plausible principles about constructive proof

- The assumption that there is a fixed universe of constructive proofs calls $\varphi \Rightarrow (\exists \mathfrak{p})\mathfrak{p} : \varphi$ into question.

Plausible principles about constructive proof

- The assumption that there is a fixed universe of constructive proofs calls $\varphi \Rightarrow (\exists \mathfrak{p})\mathfrak{p} : \varphi$ into question.
- However, $(\exists \mathfrak{p})\mathfrak{p} : \varphi \Rightarrow \varphi$ seems unobjectionable.

Plausible principles about constructive proof

- The assumption that there is a fixed universe of constructive proofs calls $\varphi \Rightarrow (\exists \mathfrak{p})\mathfrak{p} : \varphi$ into question.
- However, $(\exists \mathfrak{p})\mathfrak{p} : \varphi \Rightarrow \varphi$ seems unobjectionable.
- The many attempts to formalize the BHK interpretation suggest that $\mathfrak{p} : \varphi$ **is a mathematical proposition**.

Plausible principles about constructive proof

- The assumption that there is a fixed universe of constructive proofs calls $\varphi \Rightarrow (\exists \mathfrak{p})\mathfrak{p} : \varphi$ into question.
- However, $(\exists \mathfrak{p})\mathfrak{p} : \varphi \Rightarrow \varphi$ seems unobjectionable.
- The many attempts to formalize the BHK interpretation suggest that $\mathfrak{p} : \varphi$ **is a mathematical proposition**.
- Arithmetizing:
  - Constructively acceptable background theory Z (e.g. HA).
  - Primitive predicate $P(\ulcorner \varphi \urcorner)$ intended to express "φ is constructively provable".
  - Principles about $P(x)$:

(Rfn) $P(\ulcorner \varphi \urcorner) \rightarrow \varphi$
(Nec) $\mathrm{T} \vdash \varphi \quad \therefore \quad \mathrm{T} \vdash P(\ulcorner \varphi \urcorner)$ where $\mathrm{T} = \mathrm{Z} + \mathrm{Rfn}$.

Plausible principles about constructive proof

- The assumption that there is a fixed universe of constructive proofs calls $\varphi \Rightarrow (\exists \mathfrak{p})\mathfrak{p} : \varphi$ into question.
- However, $(\exists \mathfrak{p})\mathfrak{p} : \varphi \Rightarrow \varphi$ seems unobjectionable.
- The many attempts to formalize the BHK interpretation suggest that $\mathfrak{p} : \varphi$ **is a mathematical proposition**.
- Arithmetizing:
  - Constructively acceptable background theory Z (e.g. HA).
  - Primitive predicate $P(\ulcorner\varphi\urcorner)$ intended to express "φ is constructively provable".
  - Principles about $P(x)$:

(Rfn) $P(\ulcorner\varphi\urcorner) \rightarrow \varphi$

(Nec) $\mathrm{T} \vdash \varphi \quad \therefore \quad \mathrm{T} \vdash P(\ulcorner\varphi\urcorner)$ where $\mathrm{T} = \mathrm{Z} + \mathrm{Rfn}$.

- Motivations:
  - Rfn is "analytic" of the notion of constructive proof. (?)
  - If Z embodies "safe" constructive principles, then derivability in Z is at least *sufficient* for constructive provability.

Montague's Theorem

T ($=$ closure of $\mathrm{Z} + \mathrm{Rfn}$ under Nec) is **inconsistent**:

i)	$\mathrm{T} \vdash \delta \leftrightarrow \neg P(\ulcorner \delta \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner) \rightarrow \delta$	Rfn
iii)	$\mathrm{T} \vdash \neg P(\ulcorner \delta \urcorner)$	i), ii)
iv)	$\mathrm{T} \vdash \delta$	i), iii)
v)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner)$	Nec, iv)
vi)	$\mathrm{T} \vdash \bot$	iii), v)

Montague's Theorem

T ($=$ closure of $\mathrm{Z} + \mathrm{Rfn}$ under Nec) is **inconsistent**:

i)	$\mathrm{T} \vdash \delta \leftrightarrow \neg P(\ulcorner \delta \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner) \rightarrow \delta$	Rfn
iii)	$\mathrm{T} \vdash \neg P(\ulcorner \delta \urcorner)$	i), ii)
iv)	$\mathrm{T} \vdash \delta$	i), iii)
v)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner)$	Nec, iv)
vi)	$\mathrm{T} \vdash \bot$	iii), v)

- Nec can replace $\varphi \rightarrow Tr(\ulcorner \varphi \urcorner)$ in the Liar.

Montague's Theorem

T (= closure of $\mathrm{Z} + \mathrm{Rfn}$ under Nec) is **inconsistent**:

i)	$\mathrm{T} \vdash \delta \leftrightarrow \neg P(\ulcorner \delta \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner) \rightarrow \delta$	Rfn
iii)	$\mathrm{T} \vdash \neg P(\ulcorner \delta \urcorner)$	i), ii)
iv)	$\mathrm{T} \vdash \delta$	i), iii)
v)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner)$	Nec, iv)
vi)	$\mathrm{T} \vdash \bot$	iii), v)

- Nec can replace $\varphi \rightarrow Tr(\ulcorner \varphi \urcorner)$ in the Liar.
- History: Gödel (1933), Myhill (1960), Montague (1963), Kreisel (1962), Goodman (1970)

Montague's Theorem

T ($=$ closure of $\mathrm{Z} + \mathrm{Rfn}$ under Nec) is **inconsistent**:

i)	$\mathrm{T} \vdash \delta \leftrightarrow \neg P(\ulcorner \delta \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner) \rightarrow \delta$	Rfn
iii)	$\mathrm{T} \vdash \neg P(\ulcorner \delta \urcorner)$	i), ii)
iv)	$\mathrm{T} \vdash \delta$	i), iii)
v)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner)$	Nec, iv)
vi)	$\mathrm{T} \vdash \bot$	iii), v)

- Nec can replace $\varphi \rightarrow Tr(\ulcorner \varphi \urcorner)$ in the Liar.
- History: Gödel (1933), Myhill (1960), Montague (1963), Kreisel (1962), Goodman (1970)
- Does this represent an antimony for constructive truth?

Montague's Theorem

T ($=$ closure of $\mathrm{Z} + \mathrm{Rfn}$ under Nec) is **inconsistent**:

i)	$\mathrm{T} \vdash \delta \leftrightarrow \neg P(\ulcorner \delta \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner) \to \delta$	Rfn
iii)	$\mathrm{T} \vdash \neg P(\ulcorner \delta \urcorner)$	i), ii)
iv)	$\mathrm{T} \vdash \delta$	i), iii)
v)	$\mathrm{T} \vdash P(\ulcorner \delta \urcorner)$	Nec, iv)
vi)	$\mathrm{T} \vdash \bot$	iii), v)

- Nec can replace $\varphi \to Tr(\ulcorner \varphi \urcorner)$ in the Liar.
- History: Gödel (1933), Myhill (1960), Montague (1963), Kreisel (1962), Goodman (1970)
- Does this represent an antimony for constructive truth?
  - Weaver (2012): "yes"
  - Our view: "no". **The problem lies with $\mathbf{Nec}$.** (But this derivation doesn't get at the core of the problem.)

The Theory of Constructions ($\mathcal{C}$)

- Semantics for intuitionistic logic: Kripke/Beth models, realizability, Curry-Howard isomorphism, ...

The Theory of Constructions ($\mathcal{C}$)

- Semantics for intuitionistic logic: Kripke/Beth models, realizability, Curry-Howard isomorphism, ...
- Kreisel: try to directly formalize BHK by taking proofs and operations on proofs as primitives.

The Theory of Constructions ($\mathcal{C}$)

- Semantics for intuitionistic logic: Kripke/Beth models, realizability, Curry-Howard isomorphism, ...
- Kreisel: try to directly formalize BHK by taking proofs and operations on proofs as primitives.
- Goodman: to be faithful to BHK, such an approach should be "type- and logic-free".

The Theory of Constructions ($\mathcal{C}$)

- Semantics for intuitionistic logic: Kripke/Beth models, realizability, Curry-Howard isomorphism, ...
- Kreisel: try to directly formalize BHK by taking proofs and operations on proofs as primitives.
- Goodman: to be faithful to BHK, such an approach should be "type- and logic-free".
- Basic ingredients:
  - untyped lambda calculus (or combinatorial logic)
  - combinatorial operations on proofs
  - primitive proof operator
  - **NB: no arithmetic**

The Theory of Constructions ($\mathcal{C}$)

- Semantics for intuitionistic logic: Kripke/Beth models, realizability, Curry-Howard isomorphism, ...
- Kreisel: try to directly formalize BHK by taking proofs and operations on proofs as primitives.
- Goodman: to be faithful to BHK, such an approach should be "type- and logic-free".
- Basic ingredients:
  - untyped lambda calculus (or combinatorial logic)
  - combinatorial operations on proofs
  - primitive proof operator
  - **NB: no arithmetic**
- Question for later: why is this foundational bedrock?

The syntax and proof theory of $\mathcal{C}$

- Terms:
  - variables $x, y, z, \ldots$ over constructive proofs
  - λ (*abstraction*)
  - xy (*application*)
  - Dxy (*pairing*)
  - $D_i(D(x_1, x_2)) = x_i$, $i \in \{1, 2\}$ (*projection*)
  - $\top$ (*truth* = $\lambda x.\lambda y.x$), $\bot$ (*falsity* = $\lambda x.\lambda y.y$)
  - πxy (*proof operator*)
  - $\equiv$ (*intensional identity*)
  - $\ldots$

The syntax and proof theory of $\mathcal{C}$

- Terms:
  - variables $x, y, z, \ldots$ over constructive proofs
  - λ (*abstraction*)
  - xy (*application*)
  - Dxy (*pairing*)
  - $D_i(D(x_1, x_2)) = x_i$, $i \in \{1, 2\}$ (*projection*)
  - $\top$ (*truth* = $\lambda x.\lambda y.x$), $\bot$ (*falsity* = $\lambda x.\lambda y.y$)
  - πxy (*proof operator*)
  - $\equiv$ (*intensional identity*)
  - $\ldots$
- $\mathcal{C}$ is an *equational calculus* – i.e. all statements are of the form $s \equiv t$.
  - e.g. $\lambda x.D_2x((\lambda y.y)u) \equiv u$, $\pi xy \equiv \top$
  - terms may be **undefined** – e.g. $(\lambda x.xx)(\lambda x.xx)$

The syntax and proof theory of $\mathcal{C}$

- Terms:
  - variables $x, y, z, \dots$ over constructive proofs
  - λ (*abstraction*)
  - xy (*application*)
  - Dxy (*pairing*)
  - $D_i(D(x_1, x_2)) = x_i$, $i \in \{1, 2\}$ (*projection*)
  - $\top$ (*truth* $= \lambda x.\lambda y.x$), $\bot$ (*falsity* $= \lambda x.\lambda y.y$)
  - πxy (*proof operator*)
  - $\equiv$ (*intensional identity*)
  - $\dots$
- $\mathcal{C}$ is an *equational calculus* – i.e. all statements are of the form $s \equiv t$.
  - e.g. $\lambda x.D_2 x((\lambda y.y)u) \equiv u$, $\pi xy \equiv \top$
  - terms may be **undefined** – e.g. $(\lambda x.xx)(\lambda x.xx)$
- $\Gamma \vdash s \equiv t$ iff

$s \equiv t$ is provable from assumptions Γ in the untyped lambda calculus with special axioms for D, D_i and π.

The proof operation π

- Intended interpretation:

$$\pi uv \equiv \top \text{ iff } v \text{ is a proof that } u \equiv \top$$

- So $\pi uv \equiv \top$ *approximately* expresses $\mathsf{Proof}(\overline{v}, \ulcorner \varphi_u \urcorner)$.
- Some desirable properties:

(Decidability) $\mathrm{Z} \vdash \mathsf{Proof}(\overline{n}, \ulcorner \varphi \urcorner)$ or $\mathrm{Z} \vdash \neg\mathsf{Proof}(\overline{n}, \ulcorner \varphi \urcorner)$
So $\pi uv \equiv \top$ should be decidable.

(Explicit Reflection) $\mathcal{N} \models \mathsf{Proof}(\overline{n}, \ulcorner \varphi \urcorner) \rightarrow \varphi$
So $\pi uv \equiv \top$ should entail $u \equiv \top$.

(Internalization) $\mathrm{Z} \vdash \varphi \Rightarrow \mathrm{Z} \vdash \mathsf{Proof}(n, \ulcorner \varphi \urcorner)$ for some n
So if $\vdash u \equiv \top$, we should be able to construct v such that $\vdash \pi uv \equiv \top$.

Formalizing the desirable properties

$$(\mathrm{Dec})\ \frac{\pi uv \equiv \top \vdash t \equiv s \qquad \pi uv \equiv \bot \vdash t \equiv s}{\vdash t \equiv s}$$

$$(\mathrm{Dec}')\ \frac{\pi uv \equiv \top \vdash \top \equiv \bot}{\vdash \pi uv \equiv \bot}$$

$$(\mathrm{ExpRef})\ \pi uv \equiv \top \vdash u \equiv \top$$

(Int) If $\vdash u \equiv \top$, then $\vdash \pi uv \equiv \top$ for some v.

- Dec' is derivable from Dec by identity axioms and Cut.
- Int is a **metatheorem** (like Σ^0_1-completeness for Z).

Interpreting the intuitionisitic connectives

- We ultimately want to construct a term $t(x)$ which expresses

  "x is **not** a proof of $t(x)$"

Interpreting the intuitionisitic connectives

- We ultimately want to construct a term $t(x)$ which expresses

  "x is **not** a proof of $t(x)$"

- So we need a way of expressing intuitionistic negation in $\mathcal{C}$.

Interpreting the intuitionisitic connectives

- We ultimately want to construct a term $t(x)$ which expresses

  "x is **not** a proof of $t(x)$"

- So we need a way of expressing intuitionistic negation in $\mathcal{C}$.
- $\neg\varphi =_{df} (\varphi \rightarrow \bot)$
- $(\varphi \rightarrow \psi)^* = \lambda y.\lambda x.(\pi\varphi^* x \supset_c \pi\psi^*(yx))$

Interpreting the intuitionisitic connectives

- We ultimately want to construct a term $t(x)$ which expresses

  "x is **not** a proof of $t(x)$"

- So we need a way of expressing intuitionistic negation in $\mathcal{C}$.
- $\neg\varphi =_{df} (\varphi \to \bot)$
- $(\varphi \to \psi)^* = \lambda y.\lambda x.(\pi\varphi^* x \supset_c \pi\psi^*(yx))$
- This approximates Kriesel's analysis of $\mathrm{BHK}_{\to}$:
  - Recall: $\pi uv \equiv \top$ is assumed be *decidable*.
  - So we can apply *classical* implication ($\supset_c$).

Interpreting the intuitionisitic connectives

- We ultimately want to construct a term $t(x)$ which expresses

  "x is **not** a proof of $t(x)$"

- So we need a way of expressing intuitionistic negation in $\mathcal{C}$.
- $\neg\varphi =_{df} (\varphi \to \bot)$
- $(\varphi \to \psi)^* = \lambda y.\lambda x.(\pi\varphi^* x \supset_c \pi\psi^*(yx))$
- This approximates Kriesel's analysis of $\mathrm{BHK}_{\to}$:
  - Recall: $\pi uv \equiv \top$ is assumed be *decidable*.
  - So we can apply *classical* implication ($\supset_c$).
- Classical (i.e. truth functional) implication is definable in $\mathcal{C}$ as

$$x \supset_c y =_{df} \lambda x.\lambda y.(xy)\top$$

where $\top =_{df} \lambda u.\lambda v.u$ and $\bot =_{df} \lambda u.\lambda v.v$.

Fixed point (aka “paradoxical”) combinators

- Curry (and others) independently discovered a “purely combinatorial” form of self-reference for the lambda calculus.

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A *fixed point combinator* is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A <u>*fixed point combinator*</u> is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner \delta \urcorner)$.

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A <u>*fixed point combinator*</u> is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner \delta \urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A *fixed point combinator* is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner\delta\urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.
- Example: $Y =_{df} \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$.

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A *fixed point combinator* is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner \delta \urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.
- Example: $Y =_{df} \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$.
- $YF \equiv$

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A *fixed point combinator* is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner\delta\urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.
- Example: $Y =_{df} \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$.
- $YF \equiv (\lambda x.F(xx))(\lambda x.F(xx)) \quad \equiv$

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A _fixed point combinator_ is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner\delta\urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.
- Example: $Y =_{df} \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$.
- $YF \equiv (\lambda x.F(xx))(\lambda x.F(xx)) \quad \equiv$
  $F((\lambda x.F(xx))(\lambda x.F(xx))) \equiv$

Fixed point (aka "paradoxical") combinators

- Curry (and others) independently discovered a "purely combinatorial" form of self-reference for the lambda calculus.
- A *fixed point combinator* is a lambda-term Y s.t. for any F

$$\vdash YF \equiv F(YF)$$

- Roughly: think of F as $\varphi(x)$ and YF as δ s.t. $\vdash \delta \leftrightarrow \varphi(\ulcorner\delta\urcorner)$.
- Better: F as $\phi_e(x)$, YF as $fix(e)$ s.t. $\phi_{fix(e)}(x) = \phi_{\phi_e(fix(e))}(x)$.
- Example: $Y =_{df} \lambda f.(\lambda x.f(xx))(\lambda x.f(xx))$.
- $YF \equiv (\lambda x.F(xx))(\lambda x.F(xx)) \equiv$
  $F((\lambda x.F(xx))(\lambda x.F(xx))) \equiv F(YF)$

Formalizing non-provability and self-reference

- "x does not prove y" is formalized as $\pi y x \equiv \bot$.

Formalizing non-provability and self-reference

- "x does not prove y" is formalized as $\pi y x \equiv \bot$.
- We can reformulate this as a term:

$$H(y, x) =_{df} \lambda y.\lambda x.(\pi y x \supset_c \bot)$$

Formalizing non-provability and self-reference

- "x does not prove y" is formalized as $\pi y x \equiv \bot$.
- We can reformulate this as a term:

$$H(y, x) =_{df} \lambda y.\lambda x.(\pi y x \supset_c \bot)$$

- Applying fixed point combinator Y to H we get YH s.t.

$$(\star) \qquad \vdash YH(y, x) \equiv H(YH, x)$$

Formalizing non-provability and self-reference

- "x does not prove y" is formalized as $\pi yx \equiv \bot$.
- We can reformulate this as a term:

$$H(y, x) =_{df} \lambda y.\lambda x.(\pi yx \supset_c \bot)$$

- Applying fixed point combinator Y to H we get YH s.t.

$$(\star) \qquad \vdash YH(y, x) \equiv H(YH, x)$$

- $YH(y, x) \approx$ "I am not proved by x."

Formalizing non-provability and self-reference

- "x does not prove y" is formalized as $\pi y x \equiv \bot$.
- We can reformulate this as a term:

$$H(y, x) =_{df} \lambda y.\lambda x.(\pi y x \supset_c \bot)$$

- Applying fixed point combinator Y to H we get YH s.t.

$$(\star) \qquad \vdash YH(y, x) \equiv H(YH, x)$$

- $YH(y, x) \approx$ "I am not proved by x."
- A simplified form of the Kreisel-Goodman paradox can be constructed by mimicking the reasoning of a "free variable" form of Montague's Theorem in $\mathcal{C}$.

The simplified Kreisel-Goodman Paradox

i)		$\vdash YH \equiv H(YH, x)$	$(\star)$
ii)	$\pi(YH)x \equiv \top$	$\vdash (YH) \equiv \top$	ExpRef
iii)	$\pi(YH)x \equiv \top$	$\vdash H(YH, x) \equiv \top$	i)
iv)	$\pi(YH)x \equiv \top$	$\vdash (\pi(YH)x \supset_c \bot) \equiv \top$	defn of $H(x)$
v)	$\pi(YH)x \equiv \top$	$\vdash \bot \equiv \top$	defn $\supset_c$
vi)		$\vdash \pi(YH)x \equiv \bot$	Dec′
vii)		$\vdash (\pi(YH)x \supset_c \bot) \equiv \top$	defn $\supset_c$
viii)		$\vdash H(YH, x) \equiv \top$	defn $H(x)$
ix)		$\vdash YH \equiv \top$	i)
x)		$\vdash \pi(YH)a \equiv \top$	Int (for some a)
xi)		$\vdash \pi(YH)a \equiv \bot$	substitution vi)
xii)		$\vdash \top \equiv \bot$	x), xi)

Ingredients in the paradox

1) "combinatory completeness" (e.g. unrestricted λ-abstraction)
2) decidability of the proof predicate – i.e. $\pi yx \equiv \top$ *or* $\pi yx \equiv \bot$
3) "explicit" reflection – i.e. $\pi ux \vdash u \equiv \top$ (with x free)
4) internalization – i.e. $\vdash u \equiv \top \Rightarrow \exists v$ s.t. $\vdash \pi uv \equiv \top$
5) free proof variables/proof quantifiers
6) lacking of typing/stratification of proofs
7) ...

Ingredients in the paradox

1) “combinatory completeness”: Kreisel (1962?)
2) decidability of π: Beeson (1985), Weinstein (1983?)
3) “explicit” reflection: McCarty (1983)
4) internalization: **D & K**
5) free proof variables/proof quantifiers: Goodman (1970), **D & K**
6) lacking of typing/stratification of proofs: Goodman (1970)
7) . . .

Re-arithmetizing the simplified K-G paradox (1)

- Introduce a predicate $R(x, y)$ to $\mathcal{L}_a$ intended to express

$$R(\overline{n}, \ulcorner\varphi\urcorner) \approx n \text{ codes a constructive proof of } \varphi$$

Re-arithmetizing the simplified K-G paradox (1)

- Introduce a predicate $R(x, y)$ to $\mathcal{L}_a$ intended to express

$$R(\overline{n}, \ulcorner\varphi\urcorner) \approx n \text{ codes a constructive proof of } \varphi$$

- Plausible properties about $R(x, y)$:

$$\begin{aligned}
&(\mathrm{P}) && \forall x \forall y[\mathsf{Proof}_{\mathrm{Z}}(x, y) \to R(x, y)] \\
&(\text{ExpRef-N}) && R(t(\overline{x}), \ulcorner\varphi\urcorner) \to \varphi \text{ for } t \in Term_{\mathcal{L}_a} \\
&(\mathrm{Int}^-) && \text{If } \mathrm{Z} + \mathrm{P} \vdash \varphi \text{, then } \vdash R(f(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \\
&(\mathrm{Int}^+) && \text{If } \mathrm{Z} + \mathrm{P} + \text{ExpRef-N} \vdash \varphi \text{, then} \\
& && \vdash R(g(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner).
\end{aligned}$$

Re-arithmetizing the simplified K-G paradox (1)

- Introduce a predicate $R(x, y)$ to $\mathcal{L}_a$ intended to express

$$R(\overline{n}, \ulcorner\varphi\urcorner) \approx n \text{ codes a constructive proof of } \varphi$$

- Plausible properties about $R(x, y)$:

$$\begin{aligned}
&(\mathrm{P}) && \forall x \forall y[\mathsf{Proof}_{\mathrm{Z}}(x, y) \rightarrow R(x, y)] \\
&(\text{ExpRef-N}) && R(t(\overline{x}), \ulcorner\varphi\urcorner) \rightarrow \varphi \text{ for } t \in Term_{\mathcal{L}_a} \\
&(\mathrm{Int}^-) && \text{If } \mathrm{Z} + \mathrm{P} \vdash \varphi\text{, then } \vdash R(f(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \\
&(\mathrm{Int}^+) && \text{If } \mathrm{Z} + \mathrm{P} + \text{ExpRef-N} \vdash \varphi\text{, then} \\
& && \vdash R(g(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner).
\end{aligned}$$

- Int^- is derivable as a metatheorem about $\mathrm{Z} + \mathrm{P}$ (we can construct a prim. rec. $f(n)$ from Σ^0_1-completeness).

Re-arithmetizing the simplified K-G paradox (1)

- Introduce a predicate $R(x, y)$ to $\mathcal{L}_a$ intended to express

$$R(\overline{n}, \ulcorner\varphi\urcorner) \approx n \text{ codes a constructive proof of } \varphi$$

- Plausible properties about $R(x, y)$:

$$\begin{array}{rl} \text{(P)} & \forall x \forall y [\mathsf{Proof}_{\mathrm{Z}}(x, y) \rightarrow R(x, y)] \\ \text{(ExpRef-N)} & R(t(\overline{x}), \ulcorner\varphi\urcorner) \rightarrow \varphi \text{ for } t \in Term_{\mathcal{L}_a} \\ (\mathrm{Int}^-) & \text{If } \mathrm{Z} + \mathrm{P} \vdash \varphi\text{, then } \vdash R(f(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \\ (\mathrm{Int}^+) & \text{If } \mathrm{Z} + \mathrm{P} + \mathrm{ExpRef\text{-}N} \vdash \varphi\text{, then} \\ & \vdash R(g(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \end{array}$$

- Int^- is derivable as a metatheorem about $\mathrm{Z} + \mathrm{P}$ (we can construct a prim. rec. $f(n)$ from Σ^0_1-completeness).
- So $\mathrm{T}_0 =$ the closure of $\mathrm{Z} + \mathrm{P} + \mathrm{ExpRef\text{-}N}$ under Int^- is consistent.

Re-arithmetizing the simplified K-G paradox (1)

- Introduce a predicate $R(x, y)$ to $\mathcal{L}_a$ intended to express

$$R(\overline{n}, \ulcorner\varphi\urcorner) \approx n \text{ codes a constructive proof of } \varphi$$

- Plausible properties about $R(x, y)$:

$$\begin{array}{rl} \text{(P)} & \forall x \forall y[\mathsf{Proof}_{\mathrm{Z}}(x, y) \rightarrow R(x, y)] \\ \text{(ExpRef-N)} & R(t(\overline{x}), \ulcorner\varphi\urcorner) \rightarrow \varphi \text{ for } t \in Term_{\mathcal{L}_a} \\ (\mathrm{Int}^-) & \text{If } \mathrm{Z} + \mathrm{P} \vdash \varphi, \text{ then } \vdash R(f(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \\ (\mathrm{Int}^+) & \text{If } \mathrm{Z} + \mathrm{P} + \mathrm{ExpRef\text{-}N} \vdash \varphi, \text{ then} \\ & \vdash R(g(\ulcorner\varphi\urcorner), \ulcorner\varphi\urcorner). \end{array}$$

- Int^- is derivable as a metatheorem about $\mathrm{Z} + \mathrm{P}$ (we can construct a prim. rec. $f(n)$ from Σ^0_1-completeness).
- So T_0 = the closure of $\mathrm{Z} + \mathrm{P} + \mathrm{ExpRef\text{-}N}$ under Int^- is consistent.
- But T_1 = the closure of $\mathrm{Z} + \mathrm{P} + \mathrm{ExpRef\text{-}N}$ under Int^+ is **inconsistent** for any total $g : \mathbb{N} \rightarrow \mathbb{N}$.

Re-arithmetizing the simplified K-G paradox (2)

i)	$\mathrm{T}_1 \vdash D(x) \leftrightarrow \neg R(x, \ulcorner D(x) \urcorner)$	Diagonal Lemma
ii)	$\mathrm{T}_1 \vdash R(x, \ulcorner D(x) \urcorner) \rightarrow D(x)$	ExpRef-N
iii)	$\mathrm{T}_1 \vdash \neg R(x, \ulcorner D(x) \urcorner)$	i), ii)
iv)	$\mathrm{T}_1 \vdash D(x)$	i), iii)
v)	$\mathrm{T}_1 \vdash R(g(\ulcorner D(x) \urcorner), \ulcorner D(x) \urcorner)$	Int^+, iv)
vi)	$\mathrm{T}_1 \vdash \forall x \neg R(x, \ulcorner D(x) \urcorner)$	UG iii)
vii)	$\mathrm{T}_1 \vdash \neg R(g(\ulcorner D(x) \urcorner), \ulcorner D(x) \urcorner)$	UI vi)
vi)	$\mathrm{T}_1 \vdash \bot$	v), vii)

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner \varphi \urcorner) \to \varphi$.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \rightarrow \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \rightarrow \varphi$.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \to \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi$.
- By Int^+, $\mathrm{T}_1 \vdash \exists y(y, \ulcorner\exists x(R(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x,y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \rightarrow \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \rightarrow \varphi$.
- By Int^+, $\mathrm{T}_1 \vdash \exists y(y, \ulcorner\exists x(R(x, \ulcorner\varphi\urcorner) \rightarrow \varphi)\urcorner)$.
- But
  - $\mathcal{N} \not\models \exists y \mathsf{Proof}_{\mathrm{Z}}(y, \ulcorner\exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \rightarrow \varphi)\urcorner)$ – i.e.
  - $\mathcal{N} \not\models \mathsf{Prov}_{\mathrm{Z}}(\ulcorner\mathsf{Prov}_{\mathrm{Z}}(\ulcorner\varphi\urcorner) \rightarrow \varphi\urcorner)$

  unless $\mathrm{Z} \vdash \varphi$.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \to \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi$.
- By Int^+, $\mathrm{T}_1 \vdash \exists y(y, \ulcorner\exists x(R(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$.
- But
  - $\mathcal{N} \not\models \exists y \mathsf{Proof}_{\mathrm{Z}}(y, \ulcorner\exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$ – i.e.
  - $\mathcal{N} \not\models \mathsf{Prov}_{\mathrm{Z}}(\ulcorner\mathsf{Prov}_{\mathrm{Z}}(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$

  unless $\mathrm{Z} \vdash \varphi$.
- So the problem can be localized in internalizing instances of $\mathrm{ExpRef\text{-}N}$ with free variables.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \to \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi$.
- By Int^+, $\mathrm{T}_1 \vdash \exists y(y, \ulcorner\exists x(R(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$.
- But
  - $\mathcal{N} \not\models \exists y \mathsf{Proof}_{\mathrm{Z}}(y, \ulcorner\exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$ – i.e.
  - $\mathcal{N} \not\models \mathsf{Prov}_{\mathrm{Z}}(\ulcorner\mathsf{Prov}_{\mathrm{Z}}(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$

  unless $\mathrm{Z} \vdash \varphi$.
- So the problem can be localized in internalizing instances of $\mathrm{ExpRef\text{-}N}$ with free variables.
- The same applies (*mutatis muntandis*) to the derivation in $\mathcal{C}$.

Technical diagnosis

- Moral: in-order to get a contradiction we have to internalize reasoning about $R(x, y)$ which
  i) contains x free
  ii) is mediated by $\mathrm{ExpRef\text{-}N}$
- Note $\mathrm{T}_0 \vdash \exists x R(x, \ulcorner\varphi\urcorner) \to \varphi$.
- Okay since $\mathcal{N} \models \exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi$.
- By Int^+, $\mathrm{T}_1 \vdash \exists y(y, \ulcorner\exists x(R(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$.
- But
  - $\mathcal{N} \not\models \exists y \mathsf{Proof}_{\mathrm{Z}}(y, \ulcorner\exists x \mathsf{Proof}_{\mathrm{Z}}(x, \ulcorner\varphi\urcorner) \to \varphi)\urcorner)$ – i.e.
  - $\mathcal{N} \not\models \mathsf{Prov}_{\mathrm{Z}}(\ulcorner\mathsf{Prov}_{\mathrm{Z}}(\ulcorner\varphi\urcorner) \to \varphi\urcorner)$

  unless $\mathrm{Z} \vdash \varphi$.
- So the problem can be localized in internalizing instances of $\mathrm{ExpRef\text{-}N}$ with free variables.
- The same applies (*mutatis muntandis*) to the derivation in $\mathcal{C}$.
- (To see this, it's easiest to embed $\mathcal{C}$ into Fitting's Quantified Logic of Proofs.)

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: “we know a proof when we see one”

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: "we know a proof when we see one"
  - Instrumentally: If $\mathfrak{p} : \varphi$ weren't decidable, then intuitionists (like formalists) would be left with a justification for asserting $\forall x \in \mathbb{N} \ldots$

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: "we know a proof when we see one"
  - Instrumentally: If $\mathfrak{p} : \varphi$ weren't decidable, then intuitionists (like formalists) would be left with a justification for asserting $\forall x \in \mathbb{N} \ldots$
- "Concrete" reflection (i.e. $\mathfrak{p} : \varphi \Rightarrow \varphi$ where $\mathfrak{p}$ is a concrete proof object) *seems* to be analytic of constructive proof.

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: "we know a proof when we see one"
  - Instrumentally: If $\mathfrak{p} : \varphi$ weren't decidable, then intuitionists (like formalists) would be left with a justification for asserting $\forall x \in \mathbb{N} \ldots$
- "Concrete" reflection (i.e. $\mathfrak{p} : \varphi \Rightarrow \varphi$ where $\mathfrak{p}$ is a concrete proof object) *seems* to be analytic of constructive proof.
- But once we start to reason about proofs abstractly, what justification do we possess for $x : \varphi \Rightarrow \varphi$?

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: "we know a proof when we see one"
  - Instrumentally: If $\mathfrak{p} : \varphi$ weren't decidable, then intuitionists (like formalists) would be left with a justification for asserting $\forall x \in \mathbb{N} \ldots$
- "Concrete" reflection (i.e. $\mathfrak{p} : \varphi \Rightarrow \varphi$ where $\mathfrak{p}$ is a concrete proof object) *seems* to be analytic of constructive proof.
- But once we start to reason about proofs abstractly, what justification do we possess for $x : \varphi \Rightarrow \varphi$?
- Before we can accept (let alone *prove*) this it seems we need a characterization of the range of x.

Philosophical diagnosis

- Decidability seems to be an essential feature of constructive proof:
  - Epistemologically: "we know a proof when we see one"
  - Instrumentally: If $\mathfrak{p} : \varphi$ weren't decidable, then intuitionists (like formalists) would be left with a justification for asserting $\forall x \in \mathbb{N} \ldots$
- "Concrete" reflection (i.e. $\mathfrak{p} : \varphi \Rightarrow \varphi$ where $\mathfrak{p}$ is a concrete proof object) *seems* to be analytic of constructive proof.
- But once we start to reason about proofs abstractly, what justification do we possess for $x : \varphi \Rightarrow \varphi$?
- Before we can accept (let alone *prove*) this it seems we need a characterization of the range of x.
- But this is exactly what the BHK intepretation **doesn't** provide . . .

Interpreting Heyting Arithmetic (HA) in $\mathcal{C}^+$

- "Intuisionistic mathematics is … a languageless activity of the mind."
- Corollary: Natural numbers are mental constructions.
- Idea: i) axiomatize the fact that constructions are built up inductively; ii) interpret $n \in \mathbb{N}$ as a special kind of term n^*; iii) prove that induction holds for items satisfying this definition.
- Pairing and BHK:
  - A proof of $\varphi \wedge \psi$ is a pair $\langle \mathfrak{p}, \mathfrak{q} \rangle$ s.t. $\mathfrak{p} : \varphi$ and $\mathfrak{q} : \psi$.
  - $(\varphi \wedge \psi)^* = \lambda x.(\pi\varphi^*(D_1 x) \ \cap_c \ \pi\psi^*(D_2 x))$
- Pairing in $\mathcal{C}^+$:
  - Dxy is intended to denote the pair $\langle x, y \rangle$.
  - $\vdash D_i(Dx_1x_2) \equiv x_i$
  - "x is a pair" is a *decidable* predicate δ.
  - $\delta x \equiv \bot \vdash D_i(x) \equiv x$

Induction on pairs, the natural numbers, primitive recursion

$$\text{(Ind-P)} \quad \frac{\Gamma, \delta x \equiv \bot \vdash ax \equiv \top \qquad \Gamma, a(D_1 x) \equiv \top, a(D_2 x) \equiv \top \vdash ax \equiv \top}{\Gamma \vdash ax \equiv \top}$$

- Idea: everything is a not a pair or built up by applying D.
- Natural numbers as terms:
  - $\mathrm{K} =_{df} \lambda x.\lambda y.x$
  - $0^* = \mathrm{K}$
  - $(n+1)^* = D\mathrm{K}n$
  - $\mathrm{K}, D\mathrm{KK}, D\mathrm{K}(D\mathrm{KK}), \ldots$
- Representation of primitive recursive functions
  - Successor: $s(n) =_{df} D\mathrm{K}n^*$
  - Suppose that $G(h, n) = k$, $G(h, n+1) = h(G(h, n))$.
  - There is a fixed-point functional $\mathcal{F}$ such that $\mathcal{F}(G) = G(\mathcal{F}(G))$.
  - $\mathcal{F}$ may be represented as a $\mathcal{C}^+$-term Φ such that

$$\vdash \Phi xy \equiv xy(\Phi xy)$$

Defining the natural numbers

- Qxy expresses intentional equality.
- Using definable apparatus we can define a term

$$\nu(x,y) = Qx0^* \cup_c (D(x) \cap_c QK(D_1x) \cap_c y(D_2x))$$

– i.e. either x is 0^* or x is a pair and y holds of D_2x.

- We now use the Φ combinator to define a predicate

$$N(x) = \Phi(\nu(x,y)) \equiv \nu(x, \Phi(\nu(x,y)))$$

- Goodman proves that $N(x)$ is decidable and the following induction rule is derivable:

$$\text{(Ind-N)} \quad \frac{\Gamma \vdash a0^* \equiv \top \qquad \Gamma, N(x) \equiv \top, ax \equiv \top \vdash ax \equiv \top}{\Gamma \vdash ax \equiv \top}$$

The embedding $(\cdot)^*$

- We want to define a mapping $(\cdot)^* : \mathcal{L}_a \to \mathcal{L}_{\mathcal{C}^+}$ such that

$$\mathrm{HA} \vdash \varphi \iff \mathcal{C}^+ \vdash \varphi^* a \equiv \top \text{ for some } a$$

- So intuitively, $\varphi^* \approx \mathsf{Proof}(x, \ulcorner\varphi\urcorner)$.
- $\mathcal{C}^+ =$ the *stratified theory of constructions*.
- $\Pi xyz \approx$ "x is a grasped domain containing y, and z is a proof that $y \equiv \top$"
- Stratified levels of constructions: $L_0 \subset L_1 \subset L_2 \subset \ldots$
- $(\varphi \to \psi)^* = \lambda x.\lambda y.\Pi(L_p)[\varphi^* x \supset_c \psi^* D_2 y)](D_1 y)$
  where $p = \max(\mathsf{rank}(\varphi), \mathsf{rank}(\psi))$
- I.e. the "level" of a proof of $\varphi \to \psi$ is bounded by the complexity of φ, ψ.