

Incompleteness, Reflection, and Implicit Commitment

Volker Halbach

AXDEF project workshop · IHPST
28th April 2026



This is partially joint work with Riccardo Bruni and Andrea Cantini.

The plan

1. Implicit definitions and axioms
2. Reflection and implicit commitment
3. The Implicit Commitment Thesis
4. Preliminary Answers
5. Examples
6. Conclusion

Implicit definitions and axioms

I am going to consider deductive systems S . They are all closed under classical first-order logic with the rules of natural deduction in the respective language. They differ in their (non-logical) axioms.

Given the occasion, I consider the possibility that there are axiomatic definitions, euphemistically called *implicit definitions*.

I am going to consider deductive systems S . They are all closed under classical first-order logic with the rules of natural deduction in the respective language. They differ in their (non-logical) axioms.

Given the occasion, I consider the possibility that there are axiomatic definitions, euphemistically called *implicit definitions*.

Do all systems S implicitly define something?

No! If S is inconsistent, S doesn't define anything.

Example

The language \mathcal{L}_T is the language of arithmetic plus a unary predicate T .
Example PA plus *all* sentences $T\overline{\varphi} \leftrightarrow \varphi$ does not define truth.

If you accept an inconsistent system S , you (explicitly) commit to accepting φ and $\neg\varphi$ for some sentences. No structure or model satisfies φ and $\neg\varphi$; hence you have not defined anything.

Do all systems S implicitly define something?

No! If S is inconsistent, S doesn't define anything.

Example

The language \mathcal{L}_T is the language of arithmetic plus a unary predicate T .
Example PA plus *all* sentences $T\overline{\varphi} \leftrightarrow \varphi$ does not define truth.

If you accept an inconsistent system S , you (explicitly) commit to accepting φ and $\neg\varphi$ for some sentences. No structure or model satisfies φ and $\neg\varphi$; hence you have not defined anything.

Do all systems S implicitly define something?

No! If S is inconsistent, S doesn't define anything.

Example

The language \mathcal{L}_T is the language of arithmetic plus a unary predicate T .
Example PA plus *all* sentences $T\overline{\varphi} \leftrightarrow \varphi$ does not define truth.

If you accept an inconsistent system S , you (explicitly) commit to accepting φ and $\neg\varphi$ for some sentences. No structure or model satisfies φ and $\neg\varphi$; hence you have not defined anything.

If you accept S , you are also *implicitly* committed to accepting the consistency statement $\neg \text{Bew}_S(\overline{\perp})$ for S – or so they say.

Consider the system $S_{\perp} := \text{PA} + \text{Bew}_{\text{PA}}(\overline{\perp})$, which is consistent by Gödel's second incompleteness theorem. If you accept S_{\perp} , you are implicitly committed to accepting also $\neg \text{Bew}_{S_{\perp}}(\overline{\perp})$, which implies $\neg \text{Bew}_{\text{PA}}(\overline{\perp})$.

Thus, if you accept S_{\perp} you are committed to accepting $\text{Bew}_{\text{PA}}(\overline{\perp})$ and $\neg \text{Bew}_{\text{PA}}(\overline{\perp})$. Nothing satisfies both sentences. Thus you cannot define anything by accepting S_{\perp} .

Consequently, there are consistent systems S , such that, if you accept S , you have not defined anything.

If you accept S , you are also *implicitly* committed to accepting the consistency statement $\neg \text{Bew}_S(\overline{r_{\perp}})$ for S – or so they say.

Consider the system $S_{\perp} := \text{PA} + \text{Bew}_{\text{PA}}(\overline{r_{\perp}})$, which is consistent by Gödel's second incompleteness theorem. If you accept S_{\perp} , you are implicitly committed to accepting also $\neg \text{Bew}_{S_{\perp}}(\overline{r_{\perp}})$, which implies $\neg \text{Bew}_{\text{PA}}(\overline{r_{\perp}})$.

Thus, if you accept S_{\perp} you are committed to accepting $\text{Bew}_{\text{PA}}(\overline{r_{\perp}})$ and $\neg \text{Bew}_{\text{PA}}(\overline{r_{\perp}})$. Nothing satisfies both sentences. Thus you cannot define anything by accepting S_{\perp} .

Consequently, there are consistent systems S , such that, if you accept S , you have not defined anything.

If you accept S , you are also *implicitly* committed to accepting the consistency statement $\neg\text{Bew}_S(\overline{r_{\perp}})$ for S – or so they say.

Consider the system $S_{\perp} := \text{PA} + \text{Bew}_{\text{PA}}(\overline{r_{\perp}})$, which is consistent by Gödel's second incompleteness theorem. If you accept S_{\perp} , you are implicitly committed to accepting also $\neg\text{Bew}_{S_{\perp}}(\overline{r_{\perp}})$, which implies $\neg\text{Bew}_{\text{PA}}(\overline{r_{\perp}})$.

Thus, if you accept S_{\perp} you are committed to accepting $\text{Bew}_{\text{PA}}(\overline{r_{\perp}})$ and $\neg\text{Bew}_{\text{PA}}(\overline{r_{\perp}})$. Nothing satisfies both sentences. Thus you cannot define anything by accepting S_{\perp} .

Consequently, there are consistent systems S , such that, if you accept S , you have not defined anything.

S_{\perp} is arithmetically unsound and unsound as a syntax theory.

I will present examples of systems S that are arithmetically sound, but that become inconsistent under reflection just like S_{\perp} .

Reflection does not only implicitly commit us to consistency, but further statements expressing soundness.

S_{\perp} is arithmetically unsound and unsound as a syntax theory.

I will present examples of systems S that are arithmetically sound, but that become inconsistent under reflection just like S_{\perp} .

Reflection does not only implicitly commit us to consistency, but further statements expressing soundness.

We must be more cautious. Showing only that S is consistent, does not guarantee that you can define anything by accepting S .

Hilbert and his students should not have only tried to show that PA is consistent, but also that PA with all accompanying implicit commitments such as $\neg \text{Be}_{WPA}(\overline{\ulcorner \perp \urcorner})$ is consistent.

*All this depends on whether you believe in the *Implicit Commitment Thesis* (Dean 2015) and how you understand it.*

Reflection and implicit commitment

Let S be a recursively axiomatized system extending Q and, for all φ , $\text{Bew}_S(\overline{\ulcorner \varphi \urcorner})$ be provable in S whenever $S \vdash \varphi$. Assume also that S proves its own soundness:

$$(\text{Rfn}(S)) \quad S \vdash \text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi \quad \text{for all } \varphi \in \mathcal{L}_S$$

$$S \vdash \gamma \leftrightarrow \neg \text{Bew}_S(\overline{\ulcorner \gamma \urcorner}) \quad \text{diagonalization}$$

$$S \vdash \text{Bew}_S(\overline{\ulcorner \gamma \urcorner}) \rightarrow \neg \gamma$$

$$S \vdash \text{Bew}_S(\overline{\ulcorner \gamma \urcorner}) \rightarrow \gamma \quad \text{Rfn}(S)$$

$$S \vdash \neg \text{Bew}_S(\overline{\ulcorner \gamma \urcorner})$$

$$S \vdash \gamma \quad \text{first line}$$

$$S \vdash \text{Bew}_S(\overline{\ulcorner \gamma \urcorner}) \quad \text{first derivability condition}$$

Thus, S is inconsistent.

Theorem

With minimal assumptions, no consistent system S proves all instances of $\text{Rfn}(S)$.

Only necessitation for $\text{Bew}_S(x)$ is used.

By Löb's theorem, a consistent S cannot prove *any* instance of $\text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi$, except for the trivial instances with $S \vdash \varphi$. This requires stronger assumptions on S and $\text{Bew}_S(x)$.

Theorem

With minimal assumptions, no consistent system S proves all instances of $\text{Rfn}(S)$.

Only necessitation for $\text{Bew}_S(x)$ is used.

By Löb's theorem, a consistent S cannot prove *any* instance of $\text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi$, except for the trivial instances with $S \vdash \varphi$. This requires stronger assumptions on S and $\text{Bew}_S(x)$.

THE MYSTERY OF REFLECTION

If you accept S , you should better believe in its soundness and thus accept all instances of $\text{Rfn}(S)$.

This assumes that $\text{Bew}_S(x)$ expresses provability in S and $\text{Rfn}(S)$ expresses the soundness of S (at least partially).

$$S_1 := S \cup \{ \text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi : \varphi \in \mathcal{L}_S \}$$

S_1 satisfies the assumptions above. By the above principle, you should also believe in the soundness of S_1 , which leads to S_2 and so on.

THE MYSTERY OF REFLECTION

If you accept S , you should better believe in its soundness and thus accept all instances of $\text{Rfn}(S)$.

This assumes that $\text{Bew}_S(x)$ expresses provability in S and $\text{Rfn}(S)$ expresses the soundness of S (at least partially).

$$S_1 := S \cup \{\text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi : \varphi \in \mathcal{L}_S\}$$

S_1 satisfies the assumptions above. By the above principle, you should also believe in the soundness of S_1 , which leads to S_2 and so on.

Turing 1939, p. 161

The well-known theorem of Gödel [...] shows that every system of logic is in a certain sense incomplete, but at the same time it indicates means whereby from a system L of logic a more complete system L' may be obtained. By repeating the process we get a sequence $L, L_1 = L', L_2 = L'_1, \dots$ each more complete than the preceding. A logic L_ω may then be constructed in which the provable theorems are the totality of theorems provable with the help of the logics L, L_1, L_2, \dots

Turing makes some qualifications. Turing (1939) used also $\text{Rfn}(S)$.

Feferman 1991, p. 1

To what extent can mathematical thought be analyzed in formal terms? Gödel's theorems show the inadequacy of single formal systems for this purpose, except in relatively restricted parts of mathematics. However at the same time they point to the possibility of systematically generating larger and larger systems whose acceptability is implicit in acceptance of the starting theory. The engines for that purpose are what have come to be called reflection principles. These may be iterated into the constructive transfinite, leading to what are called recursive progressions of theories.

The Implicit Commitment Thesis

ICT

If a subject P accepts a system S , P is committed to accepting also statements expressing the soundness of S . This can be iterated into the transfinite.

The commitment to the soundness statements is implicit in the sense that they are typically not deductive consequences of S .

Thus, there are two kinds of consequences of a system S .

ICT

If a subject P accepts a system S , P is committed to accepting also statements expressing the soundness of S . This can be iterated into the transfinite.

The commitment to the soundness statements is implicit in the sense that they are typically not deductive consequences of S .

Thus, there are two kinds of consequences of a system S .

Overthinking objection against ICT

By accepting a reasonable system S , by ICT, you may implicitly commit yourself to accepting all sentences. Thus, ICT precludes you from accepting certain reasonable systems. Thus, ICT should be rejected.

If ICT is correct, we should not only check S for its consistency but also for the consistency of the autonomous progression of reflection based on S .

I try to sharpen ICT by applying it to different systems S , following the recent trend in looking at general versions of ICT.

Overthinking objection against ICT

By accepting a reasonable system S , by ICT, you may implicitly commit yourself to accepting all sentences. Thus, ICT precludes you from accepting certain reasonable systems. Thus, ICT should be rejected.

If ICT is correct, we should not only check S for its consistency but also for the consistency of the autonomous progression of reflection based on S .

I try to sharpen ICT by applying it to different systems S , following the recent trend in looking at general versions of ICT.

Overthinking objection against ICT via definitions

By accepting a reasonable system S , by ICT, you may implicitly commit yourself to accepting all sentences. Thus, ICT precludes you from defining something by accepting certain reasonable systems. But you are free to accept any reasonable system S to define something. As long as S is consistent, you have succeeded in defining something. Thus, ICT should be rejected.

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

Dean's formulation of ICT

Dean's version of ICT · Dean 2015, p. 32

Anyone who accepts the axioms of a mathematical theory S is thereby also committed to accepting various additional statements Δ which are expressible in the language of S but which are formally independent of its axioms.

Dean mentions proof-theoretic reflection principles. These principles need not be independent of S . There are decidable theories such as Presburger's arithmetic for which ICT fails because there are no independent sentences in the language whatsoever. Moreover, there are theories that are not decidable, but incompatible with their own consistency statement. Dean thus imposes the following restriction on S before stating ICT:

Dean's conditions on S · Dean 2015, p. 32

Let S be a recursively axiomatizable ω -consistent theory extending Elementary Arithmetic (EA) [...].

Dean's version of ICT · Dean 2015, p. 32

Anyone who accepts the axioms of a mathematical theory S is thereby also committed to accepting various additional statements Δ which are expressible in the language of S but which are formally independent of its axioms.

Dean mentions proof-theoretic reflection principles. These principles need not be independent of S . There are decidable theories such as Presburger's arithmetic for which ICT fails because there are no independent sentences in the language whatsoever. Moreover, there are theories that are not decidable, but incompatible with their own consistency statement. Dean thus imposes the following restriction on S before stating ICT:

Dean's conditions on S · Dean 2015, p. 32

Let S be a recursively axiomatizable ω -consistent theory extending Elementary Arithmetic (EA) [...].

Głowacki and Zicchetti 2026 · abstract

The Implicit Commitments Thesis (ICT) states that in accepting a formal theory S , one is implicitly committed to additional statements, such as S 's consistency.

Łełyk and Nicolai 2022

Anyone who is justified in believing a sufficiently powerful, consistent mathematical formal system S is also implicitly committed to various additional statements which are expressible in the language of S but which are formally independent of its axioms.

Głowacki and Zicchetti 2026 · abstract

The Implicit Commitments Thesis (ICT) states that in accepting a formal theory S , one is implicitly committed to additional statements, such as S 's consistency.

Łęłyk and Nicolai 2022

Anyone who is justified in believing a sufficiently powerful, consistent mathematical formal system S is also implicitly committed to various additional statements which are expressible in the language of S but which are formally independent of its axioms.

Horsten's version of ICT

In the course of the debate about implicit commitment generated by acceptance of a theory, it has been claimed that in certain circumstances where you know a mathematical theory S and nothing more, you can come to know proof theoretic reflection principles for S without justifying them. Let us call this the Implicit Commitment Thesis (ICT). This thesis is a bold claim because if S is sufficiently strong (and consistent), then proof theoretic reflection principles for S are logically independent of S .

Horsten's version of ICT

In the course of the debate about implicit commitment generated by acceptance of a theory, it has been claimed that **in certain circumstances** where you **know** a mathematical theory S and nothing more, you can come to know proof theoretic reflection principles for S without justifying them. Let us call this the Implicit Commitment Thesis (ICT). **This thesis is a bold claim because if S is sufficiently strong (and consistent), then proof theoretic reflection principles for S are logically independent of S .**

The claim is *too* bold: $PA + Bew_{PA}(\overline{\perp})$ is consistent, but inconsistent with its own consistency statement. Cf. also $S + \{\exists x \neg Px, P\bar{0}, P\bar{1}, P\bar{2}, \dots\}$ plus $RFN(S)$.

Preliminary Answers

1. To which systems S (if any) does ICT apply?
2. What does it mean to accept a system in the relevant sense?
3. What are the statements of soundness?
4. How exactly can we come to accept them?
5. Where does the reflection process lead?

S must contain a theory of syntax. S contains $I\Sigma_1$ or PA and may be formulated in an expansion of the language \mathcal{L}_S of arithmetic.

Any conditions on S must be *transparent* in the sense that P must be able to determine whether S satisfies the conditions. The implicit commitment can be made explicit.

Therefore, I do not restrict ICT to systems that are consistent, ω -consistent, or sound in some sense. Example: $PA + Bew_{PA}(\overline{\ulcorner 1 \urcorner})$

S must contain a theory of syntax. S contains $I\Sigma_1$ or PA and may be formulated in an expansion of the language \mathcal{L}_S of arithmetic.

Any conditions on S must be *transparent* in the sense that P must be able to determine whether S satisfies the conditions. The implicit commitment can be made explicit.

Therefore, I do not restrict ICT to systems that are consistent, ω -consistent, or sound in some sense. Example: $PA + Bew_{PA}(\overline{\perp})$

What does it mean to accept a system?

We must accept S as *able of expressing syntax theory*. I exclude acceptances of PA as the theory of *all* or *nonstandard* models of PA.

Maybe one can correctly accept $\text{PA} + \text{Bew}_{\text{PA}}(\overline{\ulcorner \perp \urcorner})$ as a theory of some nonstandard model.

I use $|\Sigma_1$ as a theory of syntax, namely the theory of strings of strokes. I don't believe in a separation of syntax theory from arithmetic. There are also no differences in determinacy or indeterminacy (Picollo and Waxman 2025).

What does it mean to accept a system?

We must accept S as *able of expressing syntax theory*. I exclude acceptances of PA as the theory of *all* or *nonstandard* models of PA.

Maybe one can correctly accept $\text{PA} + \text{Bew}_{\text{PA}}(\overline{\ulcorner \perp \urcorner})$ as a theory of some nonstandard model.

I use $|\Sigma_1$ as a theory of syntax, namely the theory of strings of strokes. I don't believe in a separation of syntax theory from arithmetic. There are also no differences in determinacy or indeterminacy (Picollo and Waxman 2025).

The following are not precise definitions:

$$\text{(Con(S))} \quad \neg \text{Bew}_S(\overline{\ulcorner \perp \urcorner})$$

$$\text{(Rfn(S))} \quad \text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi \text{ for all formulæ } \varphi$$

$$\text{(RFN(S))} \quad \forall x (\text{Bew}_S(\overline{\ulcorner \varphi(\dot{x}) \urcorner}) \rightarrow \varphi(x)) \text{ for all formulæ } \varphi(x)$$

$$\text{(GRfn(S))} \quad \forall x (\text{Bew}_S(x) \wedge \text{Sent}_T(x) \rightarrow Tx)$$

Con(S) is the *consistency statement*, Rfn(S) the *local reflection principle*, RFN(S) the *uniform reflection principle*, and GRfn(S) the *global reflection principle* for the given (representation of) the system S.

In the presence of the derivability conditions, the Gödel sentence γ is equivalent to Con(S).

The following are not precise definitions:

$$\text{(Con(S))} \quad \neg \text{Bew}_S(\overline{\ulcorner \perp \urcorner})$$

$$\text{(Rfn(S))} \quad \text{Bew}_S(\overline{\ulcorner \varphi \urcorner}) \rightarrow \varphi \text{ for all formulæ } \varphi$$

$$\text{(RFN(S))} \quad \forall x (\text{Bew}_S(\overline{\ulcorner \varphi(\dot{x}) \urcorner}) \rightarrow \varphi(x)) \text{ for all formulæ } \varphi(x)$$

$$\text{(GRfn(S))} \quad \forall x (\text{Bew}_S(x) \wedge \text{Sent}_T(x) \rightarrow Tx)$$

Con(S) is the *consistency statement*, Rfn(S) the *local reflection principle*, RFN(S) the *uniform reflection principle*, and GRfn(S) the *global reflection principle* for the given (representation of) the system S.

In the presence of the derivability conditions, the Gödel sentence γ is equivalent to Con(S).

Con(S) is weak and a minimal condition of soundness.

Rfn(S) expresses soundness as a ‘schema’.

RFN(S) is not obviously a statement of soundness and makes use of ‘*de re* provability’. It adds a formalization of the ω -rule. Over PA, RFN(S) is equivalent to $\forall x \text{ Bew}_S(\overline{\ulcorner \varphi(x) \urcorner}) \rightarrow \forall x \varphi(x)$ (Feferman 1962). There are rule versions and versions with partial truth predicates.

Con(S) is weak and a minimal condition of soundness.

Rfn(S) expresses soundness as a 'schema'.

RFN(S) is not obviously a statement of soundness and makes use of 'de re provability'. It adds a formalization of the ω -rule. Over PA, RFN(S) is equivalent to $\forall x \text{ Bew}_S(\overline{\ulcorner \varphi(x) \urcorner}) \rightarrow \forall x \varphi(x)$ (Feferman 1962). There are rule versions and versions with partial truth predicates.

Con(S) is weak and a minimal condition of soundness.

Rfn(S) expresses soundness as a 'schema'.

RFN(S) is not obviously a statement of soundness and makes use of '*de re* provability'. It adds a formalization of the ω -rule. Over PA, RFN(S) is equivalent to $\forall x \text{ Bew}_S(\overline{\ulcorner \varphi(\dot{x}) \urcorner}) \rightarrow \forall x \varphi(x)$ (Feferman 1962). There are rule versions and versions with partial truth predicates.

The consistency statement implies via the formalized completeness theorem the existence of a model.

See (Kreisel and Lévy 1968, p. 101) for proof-theoretic vs. set-theoretic reflection and, for incompleteness with relative interpretability and formalized completeness, (Friedman 2023).

Kreisel and Lévy (1968, p. 98)

By a “reflection principle” for a formal system S we mean, roughly, the formal assertion stating the soundness of S :

If a statement φ (in the formalism S) is provable in S then φ is valid.

Kreisel and Lévy (1968, p. 98):

Literally speaking, the intended reflection principle cannot be formulated in S itself by means of a single statement. This would require a truth definition T_S , with a variable a over (GÖDEL' numbers [*sic*] of, or, simply, over) formulas of S , and a definition of the proof relation $Prov_S(p, a)$ (read: p is (the GÖDEL number of) a proof of a in S). The reflection principle for S would be

$$(*) \quad \forall p \forall a [Prov_S(p, a) \rightarrow T_S(a)].$$

Such a truth definition T_S does not exist (TARSKI [...]). What is, trivially, possible is to express the reflection principle by means of the schema

$$(R_S) \quad \forall p [Prov_S(p, \ulcorner \varphi \urcorner) \rightarrow \varphi], \text{ for every sentence } \varphi \text{ of } S,$$

where $\ulcorner \varphi \urcorner$ denotes GÖDEL number of φ . (Since we use “denotes” we have in mind an interpretation of S ; in this case $\ulcorner \varphi \urcorner$ denotes the formal object φ of S , namely, the GÖDEL number of the intuitive object φ — and φ itself denotes, of course, its interpretation.) We shall refer to (R_S) as the *local reflection principle* for S . We use the word “local” here to distinguish the principle (R_S) from the principle $(*)$ above, which we shall call the *global reflection principle* for S .

Global reflection principle

$$(GRfn(S)) \quad \forall x (Bew_S(x) \wedge Sent_T(x) \rightarrow Tx)$$

$Sent_T(x)$ expresses that x is a sentence.

In contrast to $Rfn(S)$ and $RFN(S)$, $GRfn(S)$ is a single sentence.

$GRfn(S)$ brings in a new concept, which can be axiomatized or defined.

Cf. (Dean 2015) on stability.

Global reflection principle

$$(GRfn(S)) \quad \forall x (Bew_S(x) \wedge Sent_T(x) \rightarrow Tx)$$

$Sent_T(x)$ expresses that x is a sentence.

In contrast to $Rfn(S)$ and $RFN(S)$, $GRfn(S)$ is a single sentence.

$GRfn(S)$ brings in a new concept, which can be axiomatized or defined.

Cf. (Dean 2015) on stability.

Many versions of ICT contain a restriction that the additional statements must be formulated in the original language of S . I don't know what exactly that means. The language of S can already contain many idling symbols.

Only $GRfn(S)$ fully expresses the soundness of S . $Con(S)$, $Rfn(S)$, and $RFN(S)$ are consequences of $GRfn(S)$ without T .

If only $\text{GRfn}(S)$ fully expresses the soundness of S , we should use it as the main reflection principle.

Its strength depends very much on the axioms for T . Rule of thumb: Typed disquotational axioms for T will give us only the strength of $\text{Rfn}(S)$ or $\text{RFN}(S)$.

If we move from PA to $\text{CT}(\text{PA})$, we can prove $\text{GRfn}(\text{PA})$ from the truth-theoretic axioms. This observations do not so easily generalize to other systems. For some S , $\text{GRfn}(S)$ (with uniform disquotation) is stronger than $\text{CT}(S)$.

If only $\text{GRfn}(S)$ fully expresses the soundness of S , we should use it as the main reflection principle.

Its strength depends very much on the axioms for T . Rule of thumb: Typed disquotational axioms for T will give us only the strength of $\text{Rfn}(S)$ or $\text{RFN}(S)$.

If we move from PA to $\text{CT}(\text{PA})$, we can prove $\text{GRfn}(\text{PA})$ from the truth-theoretic axioms. This observations do not so easily generalize to other systems. For some S , $\text{GRfn}(S)$ (with uniform disquotation) is stronger than $\text{CT}(S)$.

Why confine ourselves to approximations to soundness if we can have the full global reflection principle?

Over PA, global reflection is much stronger than uniform reflection:

Theorem (Feferman??)

The autonomous progression of $\text{RFN}(S)$ based on PA is arithmetically equivalent to $\text{RT}_{<\omega}$ (or ω -times iterated ACA), while the autonomous progression based on truth yields $\text{RT}_{<\Gamma_0}$.

Horsten (2021, p. 741)

A proof theoretic reflection principle for a mathematical theory S says that, or approximates saying that everything that S asserts, i.e., everything that is provable in the theory, is true (Kreisel and Levy 1968, p. 98). By Gödel's incompleteness theorems, proof theoretic reflection principles for S are under very general circumstances logically independent of S . A proof theoretic reflection principle for S can be said to be a presupposition of the cognitive project S .

According to Horsten (2021), one can come to be epistemically entitled to believe $\neg \text{Bew}_S(\overline{\perp})$ without a justification in the sense of a soundness proof (although quite a few cognitive steps are involved such as change of notation and self-awareness).

Consistency is easier than other reflection principles.

Soundness proofs are not trivial. People have made bad mistakes already in the formulation of logical calculi (Pelletier 1999).

I will not accept Leon's trick instead of an inductive soundness proof in any logic exam.

The commitment to the soundness statements in ICT arises only if we can become committed to a notion of soundness and truth (or other notions). Otherwise S may be stable (Dean 2015). Cf. Isaacson's thesis.

Just expressing the full soundness claim requires a truth predicate. See (Feferman 1964, 1987, 1991). See (Kreisel 1967) above. The initial reluctance to expand the language of arithmetic has some technical motivations (re finite axiomatizability etc).

Soundness proofs are not trivial. People have made bad mistakes already in the formulation of logical calculi (Pelletier 1999).

I will not accept Leon's trick instead of an inductive soundness proof in any logic exam.

The commitment to the soundness statements in ICT arises only if we can become committed to a notion of soundness and truth (or other notions). Otherwise S may be stable (Dean 2015). Cf. Isaacson's thesis.

Just expressing the full soundness claim requires a truth predicate. See (Feferman 1964, 1987, 1991). See (Kreisel 1967) above. The initial reluctance to expand the language of arithmetic has some technical motivations (re finite axiomatizability etc).

Soundness proofs are not trivial. People have made bad mistakes already in the formulation of logical calculi (Pelletier 1999).

I will not accept Leon's trick instead of an inductive soundness proof in any logic exam.

The commitment to the soundness statements in ICT arises only if we can become committed to a notion of soundness and truth (or other notions). Otherwise S may be stable (Dean 2015). Cf. Isaacson's thesis.

Just expressing the full soundness claim requires a truth predicate. See (Feferman 1964, 1987, 1991). See (Kreisel 1967) above. The initial reluctance to expand the language of arithmetic has some technical motivations (re finite axiomatizability etc).

Definition (CT(S))

The system CT(S) is given by all the axioms of S with schemata expanded to \mathcal{L}_T and the following axioms:

$$\text{CT1 } \forall s \forall t (T(s=t) \leftrightarrow s^\circ = t^\circ)$$

and analogously for other predicate symbols other than T.

$$\text{CT3 } \forall x (\text{Sent}(x) \rightarrow (T(\neg x) \leftrightarrow \neg Tx))$$

$$\text{CT4 } \forall x \forall y (\text{Sent}(x \wedge y) \rightarrow (T(x \wedge y) \leftrightarrow T(x) \wedge T(y)))$$

$$\text{CT5 } \forall v \forall x (\text{Sent}(\forall v x) \rightarrow (T(\forall v x) \leftrightarrow \forall t T(x(t/v))))$$

Sent(x) expresses that x is a sentence without T.

If the signature of the language is finite and S axiomatized with finitely many schemata, CT(S) will prove all proof-theoretic reflection principles.

Instead of adding a primitive T , we can add second-order quantifiers (or do something else). For PA arithmetical comprehension ACA is a natural theory. Iterations give Ramified Analysis. Ramified truth theories avoid awkward infinitary rules (limit generalization rule). I prefer to add only what is needed.

At any rate I don't know how to obtain reflection principles without invoking additional resources.

Traditionally, the reflection process is continued into the transfinite. The process is restricted by an autonomy restriction. When we start with PA, we can iterate at least up to all levels lower than ϵ_0 .

Instead of adding a primitive T , we can add second-order quantifiers (or do something else). For PA arithmetical comprehension ACA is a natural theory. Iterations give Ramified Analysis. Ramified truth theories avoid awkward infinitary rules (limit generalization rule). I prefer to add only what is needed.

At any rate I don't know how to obtain reflection principles without invoking additional resources.

Traditionally, the reflection process is continued into the transfinite. The process is restricted by an autonomy restriction. When we start with PA, we can iterate at least up to all levels lower than ϵ_0 .

Instead of adding a primitive T , we can add second-order quantifiers (or do something else). For PA arithmetical comprehension ACA is a natural theory. Iterations give Ramified Analysis. Ramified truth theories avoid awkward infinitary rules (limit generalization rule). I prefer to add only what is needed.

At any rate I don't know how to obtain reflection principles without invoking additional resources.

Traditionally, the reflection process is continued into the transfinite. The process is restricted by an autonomy restriction. When we start with PA, we can iterate at least up to all levels lower than ϵ_0 .

Examples

It is not that difficult to find consistent systems that become inconsistent when reflection principles, truth theories etc. are added. But are there any such systems that are also *reasonable*?

We would accept only reasonable systems as sound for syntax theory. Unreasonable, even if consistent, may just fail to define anything.

Overthinking objection against ICT via definitions

By accepting a reasonable system S , by ICT, you may implicitly commit yourself to accepting all sentences. Thus, ICT precludes you from defining something by accepting certain reasonable systems. But you are free to accept any reasonable system S to define something. As long as S is consistent, you have succeeded in defining something. Thus, ICT should be rejected.

If ICT is correct, we cannot define anything by accepting the following theories.

The system $S_{\perp} := PA + Bew_{PA}(\overline{\perp})$ is inconsistent with its own consistency statement $\neg Bew_{S_{\perp}}(\overline{\perp})$.

Assuming ICT, accepting unreasonable theories carries the risk of implicit commitment to inconsistency. Cf. (Cieśliński 2017).

S_{\perp} is arithmetically unsound. $CT(PA) \vdash \neg Bew_{PA}(\overline{\perp})$. S_{\perp} is thus unreasonable and, therefore, not an example for the Overthinking Objection.

By accepting S_{\perp} we cannot define anything. We may be able to accept S_{\perp} in a weak sense as not sound for syntax.

The system $S_{\perp} := PA + Bew_{PA}(\overline{\perp})$ is inconsistent with its own consistency statement $\neg Bew_{S_{\perp}}(\overline{\perp})$.

Assuming ICT, accepting unreasonable theories carries the risk of implicit commitment to inconsistency. Cf. (Cieśliński 2017).

S_{\perp} is arithmetically unsound. $CT(PA) \vdash \neg Bew_{PA}(\overline{\perp})$. S_{\perp} is thus unreasonable and, therefore, not an example for the Overthinking Objection.

By accepting S_{\perp} we cannot define anything. We may be able to accept S_{\perp} in a weak sense as not sound for syntax.

Example II: Full compositionality of truth

The following system $FS_{-1}(S)$ is the obvious theory for proving the soundness of logic, if a little induction in \mathcal{L}_T is available.

Definition (FS_{-1})

The system $FS_{-1}(S)$ is given by all the axioms of S with full induction in \mathcal{L}_T and the following axioms:

$$FS2 \quad \forall x \left(\text{Sent}_T(x) \rightarrow (T(\neg x) \leftrightarrow \neg T x) \right)$$

$$FS3 \quad \forall x \forall y \left(\text{Sent}_T(x \wedge y) \rightarrow (T(x \wedge y) \leftrightarrow T(x) \wedge T(y)) \right)$$

$$FS4 \quad \forall v \forall x \left(\text{Sent}_T(\forall v x) \rightarrow (T(\forall v x) \leftrightarrow \forall t T(x(t/v))) \right)$$

$\text{Sent}_T(x)$ expresses that x is a sentence possibly with T .

I will use $FS_{-1}(PA)$ as an example of a base theory S about which we will reflect.

$FS_{-1}(PA)$ proves $GRfn(\emptyset)$, that is, global reflection (and thus soundness) for logic (without identity) in the full language with T .

Lemma

$$FS_{-1}(PA) \vdash \forall x (\text{Bew}_{\emptyset}^-(x) \wedge \text{Sent}_T(x) \rightarrow Tx)$$

It also proves that deducibility in Natural Deduction preserves truth under all substitutional interpretations (Halbach 2025).

This lemma is required for stating the *universality of logic*, e.g., $\forall x (\text{Sent}_T(x) \rightarrow T(x \forall \neg x))$.

$FS_{-1}(PA)$ proves $GRfn(\emptyset)$, that is, global reflection (and thus soundness) for logic (without identity) in the full language with T .

Lemma

$$FS_{-1}(PA) \vdash \forall x (\text{Bew}_{\emptyset}^-(x) \wedge \text{Sent}_T(x) \rightarrow Tx)$$

It also proves that deducibility in Natural Deduction preserves truth under all substitutional interpretations (Halbach 2025).

This lemma is required for stating the *universality of logic*, e.g., $\forall x (\text{Sent}_T(x) \rightarrow T(x \forall \neg x))$.

Example II: The harmless theory $FS_1(PA)$

$FS_{-1}(PA)$ has ω -models and is thus ω -consistent.

$FS_{-1}(PA)$ is conservative over PA and thus arithmetically correct.

To me $FS_{-1}(PA)$ looks reasonable.

$$\text{RL}(0) := \text{FS}_{-1}(\text{PA})$$

$$\text{RL}(n+1) := \text{RL}(n) + \forall x (\text{Bew}_{\text{RL}(n)}(x) \wedge \text{Sent}_{\text{T}}(x) \rightarrow \text{T}x)$$

$$\text{RL}(\omega) := \bigcup_{n \in \omega} \text{RL}(n)$$

This doesn't really define $\text{RL}(\omega + 1)$, but you know what I mean.

Example II: Reflecting too much on logic

Lemma

RL(1) proves the uniform T-sentences, i.e.,

$\text{RL}(1) \vdash \forall x (\text{T}(\overline{\varphi(\dot{x})}) \leftrightarrow \varphi(x))$ for $\varphi(x) \in \mathcal{L}_{\text{PA}}$.

Proposition

For all $n \in \omega$, RL($n+1$) proves uniform reflection for RL(n) for arithmetical instances (but not necessarily for instances with T).

Proposition

For all $n \in \omega$, RL(n) proves only true arithmetical statements. Thus RL(ω) is arithmetically sound.

Example II: Reflecting too much on logic

Lemma

RL(1) proves the uniform T-sentences, i.e,
$$\text{RL}(1) \vdash \forall x (\text{T}(\overline{\varphi(\dot{x})}) \leftrightarrow \varphi(x)) \text{ for } \varphi(x) \in \mathcal{L}_{\text{PA}}.$$

Proposition

For all $n \in \omega$, RL($n + 1$) proves uniform reflection for RL(n) for arithmetical instances (but not necessarily for instances with T).

Proposition

For all $n \in \omega$, RL(n) proves only true arithmetical statements. Thus RL(ω) is arithmetically sound.

Proposition

$RL(\omega)$ is ω -inconsistent; $RL(\omega + 1)$ is inconsistent.

Cf. (McGee 1985, Halbach and Horsten 2006, Halbach 2014) $RL(\omega)$ is FS without co-necessitation (or, possibly, also with it).

Proposition

$RL(\omega)$ is inconsistent with uniform reflection $RFN(RL(\omega))$.

One could also give a version with proofs for global reflection.

Proposition

$RL(\omega)$ is ω -inconsistent; $RL(\omega + 1)$ is inconsistent.

Cf. (McGee 1985, Halbach and Horsten 2006, Halbach 2014) $RL(\omega)$ is FS without co-necessitation (or, possibly, also with it).

Proposition

$RL(\omega)$ is inconsistent with uniform reflection $RFN(RL(\omega))$.

One could also give a version with proofs for global reflection.

Reflecting too much on logic leads to doom!

Overthinking objection against ICT

By accepting a reasonable system S , you may implicitly commit yourself to accepting all sentences. Thus, ICT precludes you from accepting certain reasonable systems. Thus, ICT should be rejected.

Is $FS_{-1}(PA)$ reasonable? It is ω -consistent, arithmetically sound, and the natural theory for proving the soundness of logic, and thus sound for syntax theory.

By accepting $FS_{-1}(PA)$ you cannot define a compositional truth predicate. If you accept $FS_{-1}(PA)$ in a weak sense (not as sound for syntax), you get a predicate that does not apply to sentences (but rather elements of a weird structure).

Suspicion: The source of the inconsistency result is the use of a type-free truth predicate.

The truth predicate used in reflection should be ‘fresh’.

If we can prove only the truth of sentences without T , we are safe.

Example III: IRT

Definition

The language $\mathcal{L}_{>n}$ is the language \mathcal{L}_{PA} expanded with all truth predicates T_i for finite ordinals $i > n$.

Definition

The theory $\text{IRT}_{\leq n}$ is given by all axioms of Peano arithmetic with full induction in \mathcal{L}_{\top} and the following axioms:

$$\text{IRT1 } \forall s \forall t (\text{T}_n(s \doteq t) \leftrightarrow s^\circ = t^\circ)$$

$$\text{IRT2 } \forall x (\text{Sent}_{>n}(x) \rightarrow (\text{T}_n(\neg x) \leftrightarrow \neg \text{T}_n x))$$

$$\text{IRT3 } \forall x \forall y (\text{Sent}_{>n}(x \wedge y) \rightarrow (\text{T}_n(x \wedge y) \leftrightarrow \text{T}_n(x) \wedge \text{T}_n(y)))$$

$$\text{IRT4 } \forall v \forall x (\text{Sent}_{>n}(\forall v x) \rightarrow (\text{T}_n(\forall v x) \leftrightarrow \forall t \text{T}_n(x(t/v))))$$

$$\text{IRT5 } \forall t (\text{Sent}_{>i}(t^\circ) \rightarrow (\text{T}_n(\text{T}_i t) \leftrightarrow \text{T}_i t^\circ)) \text{ for } i > n$$

Each system $\text{IRT}_{\leq n}$ axiomatizes T_n as a compositional or ‘Tarskian’ truth predicate for $\mathcal{L}_{>n}$.

Proposition

$\text{IRT}_{\leq n} \vdash T_n \overline{\varphi} \leftrightarrow \varphi$ for all sentences $\varphi \in \mathcal{L}_{>n}$.

Proposition (cf. McCarthy 1988)

Each system $\text{IRT}_{\leq n}$ has an ω -model. All $\text{IRT}_{\leq n}$ and thus $\text{IRT} := \bigcup_{n \in \omega} \text{IRT}_{\leq n}$ are arithmetically sound.

Each system $\text{IRT}_{\leq n}$ axiomatizes T_n as a compositional or ‘Tarskian’ truth predicate for $\mathcal{L}_{>n}$.

Proposition

$\text{IRT}_{\leq n} \vdash T_n \overline{\varphi} \leftrightarrow \varphi$ for all sentences $\varphi \in \mathcal{L}_{>n}$.

Proposition (cf. McCarthy 1988)

Each system $\text{IRT}_{\leq n}$ has an ω -model. All $\text{IRT}_{\leq n}$ and thus $\text{IRT} := \bigcup_{n \in \omega} \text{IRT}_{\leq n}$ are arithmetically sound.

$IRT_{>0}$ is IRT only without the topmost predicate T_0 . It contains all T_i with $i > 0$ and the axioms for these T_i .

IRT is $CT(IRT_{>0})$.

Proposition

$IRT \not\vdash \forall x (BeW_{IRT_{>0}}(x) \wedge Sent_{>0}(x) \rightarrow T_0x)$

Under weak assumptions Tarskian truth $CT(S)$ proves $GRfn(S)$, as long as S is axiomatized with finitely many schemata only (but here we have infinitely many). The Tarski-schema is not a schema!

If S is not axiomatized with finitely many schemata, we cannot always arrive at $\text{GRfn}(S)$ via $\text{CT}(S)$.

Reflection may require more in such cases. Of course $\text{IRT}_{>0}$ is still recursively axiomatized.

Since we cannot prove global reflection

$\forall x (\text{Bew}_{\text{IRT}_{>0}}(x) \wedge \text{Sent}_{>0}(x) \rightarrow \text{T}_0x)$ for $\text{IRT}_{>0}$ in IRT , we add it as an axiom:

Proposition

IRT is inconsistent with global reflection

$\forall x (\text{Bew}_{\text{IRT}_{>0}}(x) \wedge \text{Sent}_{>0}(x) \rightarrow \text{T}_0x)$. IRT is also inconsistent with uniform reflection $\text{RFN}(\text{IRT})$.

This follows from (Visser 1989).

We have another example for the Overthinking Objection. Compositional truth can be added to $IRT_{>0}$, but not global (with uniform disquotation) or uniform reflection .

Typed truth predicates do not protect us from the Overthinking Objection. Cf. also $CT(PA + \neg Bew_S(\overline{\ulcorner \perp \urcorner}))$.

We have an example of an arithmetically sound theory IRT that is a classical Tarskian theory for $IRT_{>0}$ that is inconsistent with global reflection for $IRT_{>0}$.

Thus, in some cases, global reflection is stronger than compositional truth. In such cases compositional truth is not an adequate way to establish soundness.

By accepting $IRT_{>0}$ you cannot define anything. In particular, we cannot define all the truth predicates of $IRT_{>0}$.

We have another example for the Overthinking Objection. Compositional truth can be added to $IRT_{>0}$, but not global (with uniform disquotation) or uniform reflection .

Typed truth predicates do not protect us from the Overthinking Objection. Cf. also $CT(PA + \neg Bew_S(\overline{\ulcorner \perp \urcorner}))$.

We have an example of an arithmetically sound theory IRT that is a classical Tarskian theory for $IRT_{>0}$ that is inconsistent with global reflection for $IRT_{>0}$.

Thus, in some cases, global reflection is stronger than compositional truth. In such cases compositional truth is not an adequate way to establish soundness.

By accepting $IRT_{>0}$ you cannot define anything. In particular, we cannot define all the truth predicates of $IRT_{>0}$.

We have another example for the Overthinking Objection. Compositional truth can be added to $IRT_{>0}$, but not global (with uniform disquotation) or uniform reflection .

Typed truth predicates do not protect us from the Overthinking Objection. Cf. also $CT(PA + \neg Bew_S(\overline{\perp}))$.

We have an example of an arithmetically sound theory IRT that is a classical Tarskian theory for $IRT_{>0}$ that is inconsistent with global reflection for $IRT_{>0}$.

Thus, in some cases, global reflection is stronger than compositional truth. In such cases compositional truth is not an adequate way to establish soundness.

By accepting $IRT_{>0}$ you cannot define anything. In particular, we cannot define all the truth predicates of $IRT_{>0}$.

We have another example for the Overthinking Objection. Compositional truth can be added to $IRT_{>0}$, but not global (with uniform disquotation) or uniform reflection .

Typed truth predicates do not protect us from the Overthinking Objection. Cf. also $CT(PA + \neg Bew_S(\overline{\perp}))$.

We have an example of an arithmetically sound theory IRT that is a classical Tarskian theory for $IRT_{>0}$ that is inconsistent with global reflection for $IRT_{>0}$.

Thus, in some cases, global reflection is stronger than compositional truth. In such cases compositional truth is not an adequate way to establish soundness.

By accepting $IRT_{>0}$ you cannot define anything. In particular, we cannot define all the truth predicates of $IRT_{>0}$.

Conclusion

If ICT is correct, nothing is defined implicitly by accepting certain arithmetically sound systems. They are just as bad as inconsistent systems.

Options:

- (i) Put up with the fact that accepting certain arithmetically sound systems does not define anything, even though the systems have models. Only systems whose reflective closure is consistent can be used to define something.
- (ii) Reject ICT.

If ICT is correct, nothing is defined implicitly by accepting certain arithmetically sound systems. They are just as bad as inconsistent systems.

Options:

- (i) Put up with the fact that accepting certain arithmetically sound systems does not define anything, even though the systems have models. Only systems whose reflective closure is consistent can be used to define something.
- (ii) Reject ICT.

If ICT is correct, nothing is defined implicitly by accepting certain arithmetically sound systems. They are just as bad as inconsistent systems.

Options:

- (i) Put up with the fact that accepting certain arithmetically sound systems does not define anything, even though the systems have models. Only systems whose reflective closure is consistent can be used to define something.
- (ii) Reject ICT.

References

- Castaldo, Luca and Maciej Głowacki (2024), 'Implicit commitments of instrumental acceptance: A case study', *The Philosophical Quarterly* pp. ??-??
URL: <https://doi.org/10.1093/pq/pqae108>
- Cieśliński, Cezary (2010), 'Truth, conservativeness, and provability', *Mind* 119, 409–422.
- Cieśliński, Cezary (2017), *The Epistemic Lightness of Truth: Deflationism and its Logic*, Cambridge University Press.
- Dean, Walter (2015), 'Arithmetical reflection and the provability of soundness', *Philosophia Mathematica* 23, 31–64.
- Feferman, Solomon (1959), 'Some completeness results for recursive progressions of theories (ordinal logics)', *Journal of Symbolic Logic* 24, 312–313.
- Feferman, Solomon (1960), 'Arithmetization of metamathematics in a general setting', *Fundamenta Mathematicae* 49, 35–91.
- Feferman, Solomon (1962), 'Transfinite recursive progressions of axiomatic theories', *Journal of Symbolic Logic* 27, 259–316.
- Feferman, Solomon (1964), 'Systems of predicative analysis', *Journal of Symbolic Logic* 29, 1–30.
- Feferman, Solomon (1987), Reflecting on incompleteness (outline draft). handwritten notes.

- Feferman, Solomon (1991), 'Reflecting on incompleteness', *Journal of Symbolic Logic* 56, 1–49.
- Friedman, Harvey (2023), Adventures in Gödel incompleteness.
URL: <https://u.osu.edu/friedman.8/foundational-adventures/downloadable-manuscripts/>
- Friedman, Harvey and Michael Sheard (1987), 'An axiomatic approach to self-referential truth', *Annals of Pure and Applied Logic* 33, 1–21.
- Fujimoto, Kentaro (2010), 'Relative truth definability', *Bulletin of Symbolic Logic* 16, 305–344.
- Fujimoto, Kentaro and Volker Halbach (2024), 'Classical determinate truth I', *Journal of Symbolic Logic* 89, 218–261.
- Głowacki, Maciej and Matteo Zicchetti (2026), 'Implicit commitments, epistemic stability, and the acceptability of consistency', *Erkenntnis* .
URL: <https://doi.org/10.1007/s10670-025-00985-x>
- Halbach, Volker (1994), 'A system of complete and consistent truth', *Notre Dame Journal of Formal Logic* 35, 311–327.
- Halbach, Volker (2014), *Axiomatic Theories of Truth*, revised edn, Cambridge University Press, Cambridge. (first edition 2011).
- Halbach, Volker (2025), *The Definition of Logical Validity*, Oxford University Press, Oxford.
- Halbach, Volker and Graham Leigh (2024), *The Road to Paradox: A Guide to Syntax, Truth, and Modality*, Cambridge University Press.

- Halbach, Volker and Leon Horsten (2006), 'Axiomatizing Kripke's theory of truth', *Journal of Symbolic Logic* 71, 677–712.
- Heck, Richard Kimberly (2025), 'Truth, reflection, and implicit commitment', ?? ??, ??–??
- Horsten, Leon (2021), 'On Reflection', *The Philosophical Quarterly* 71, 738–757.
- Kreisel, Georg (1960), Ordinal logics and the characterization of informal notions of proof, in J.Todd, ed., 'Proceedings of the International Congress of Mathematicians, Edinburgh 1958', Cambridge University Press, Cambridge, pp. 289–299.
- Kreisel, Georg (1967), Informal rigour and completeness proofs, in I.Lakatos, ed., 'The Philosophy of Mathematics', North Holland, Amsterdam, pp. 138–171.
- Kreisel, Georg and Azriel Lévy (1968), 'Reflection principles and their use for establishing the complexity of axiomatic systems', *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 14, 97–142.
- Łełyk, Mateusz (2023), 'Model theory and proof theory of the global reflection principle', *Journal of Symbolic Logic* 88, 738–779.
- Łełyk, Mateusz and Carlo Nicolai (2022), 'A theory of implicit commitment', *Synthese* 200.
- McCarthy, Timothy (1988), 'Ungroundedness in classical languages', *Journal of Philosophical Logic* 17, 61–74.
- McGee, Vann (1985), 'How truthlike can a predicate be? A negative result', *Journal of Philosophical Logic* 14, 399–410.

- Pakhomov, Fedor, Michael Rathjen and Dino Rossegger (2025), 'Feferman's completeness theorem,' *Bulletin of Symbolic Logic* 31, 462–487.
- Pelletier, Francis J. (1999), 'A brief history of natural deduction,' *History and Philosophy of Logic* 20, 1–31.
- Piccolo, Lavinia and Daniel Waxman (2025), 'Arithmetical pluralism and the objectivity of syntax,' *Noûs* 59, 372–391.
- Turing, Alan (1939), 'Systems of logic based on ordinals,' *Proceedings of the London Mathematical Society* 45, 161–228.
- Visser, Albert (1989), Semantics and the liar paradox, in D.Gabbay and F.Guenther, eds, 'Handbook of Philosophical Logic,' Vol. 4, Reidel, Dordrecht, pp. 617–706.
- Zicchetti, Matteo (2025), 'Soundness arguments for consistency and their epistemic value: A critical note,' *The Philosophical Quarterly* 75, 1210–1228.