Evidence at the Wrong Grain

Truth, Comprehension, and Restricted Reductio

Abstract

Inferential deflationists distinguish inferences that preserve commitment from those that preserve evidence. They use this distinction to retain unrestricted truth rules while preventing those rules from generating the Liar paradox: truth inferences remain valid, but may not occur within the evidence-preserving reductio that yields strong rejection. The restriction is motivated partly by cases in which a speaker can justifiably endorse a sentence as true without understanding, or even identifying, the sentence endorsed. I argue that these cases establish less than the solution requires. They show that the unrestricted truth rules are not uniformly evidence-preserving, but not that every transparent application must be excluded from reductio. The difference matters because the solution is not invariant under conservative refinements of its rules. One can retain the unrestricted commitment rule while separately certifying transparent applications as evidence-preserving, adding no new primitive commitment transition but changing what may occur under reductio. Once transparent disquotation is admitted, the Liar returns; if it is excluded because the Liar returns, the paradox rather than the independent epistemic cases is doing the explanatory work. The same problem recurs in a recent inferentialist treatment of unrestricted comprehension and Russell's paradox. The resulting dilemma does not refute inferential expressivism, but it does undermine its advertised evidential diagnosis of the paradoxes.

Keywords: truth; inferentialism; bilateralism; evidence; Liar paradox; Russell's paradox

1. A blind inference and a transparent one

Truth lets us endorse what we cannot conveniently repeat. I can say that everything in a long report is true without reciting it, or that what a speaker just said is true without remembering her exact words. More strikingly, I can endorse what someone says without understanding the language in which it was said. A devoted but monolingual disciple might insist that everything her multilingual guru says is true, including pronouncements in a language she cannot understand. Truth is useful partly because it permits endorsement at this distance.

But truth is also used at no distance at all. Someone who understands *Snow is white* can move from *'Snow is white' is true* to *Snow is white*. Here the sentence is displayed, its quotation is canonical, and nothing prevents the speaker from drawing the conclusion. This is the humdrum use most naturally represented by the disquotational inference from $(T\ulcorner A\urcorner)$ to (A). The disciple's endorsement and the humdrum inference instantiate the same general relationship between truth predication and assertion. Epistemically, however, they look very different.

This contrast lies at the centre of Luca Incurvati and Julian J. Schlöder's inferential deflationism. Their theory distinguishes inferences that preserve *commitment* from those that preserve *evidence*. A speaker who endorses her guru is committed to the consequences of the guru's claims even when she does not understand them. But she cannot use that endorsement to obtain evidence for a conclusion she is unable to understand or formulate. Truth rules therefore preserve commitment without preserving evidence. That distinction, Incurvati and Schlöder argue, provides a new diagnosis of the Liar paradox. The truth rules may remain unrestricted and meaning-conferring, but they must be excluded from a form of reductio reserved for evidence-preserving reasoning. The Liar and its negation can then both be rejected without contradiction (Incurvati and Schlöder 2023a, 2023b, ch. 7).

There is an immediate question about the movement from the disciple to the Liar. The disciple case shows that *some* applications of the truth rules do not transmit evidence. The solution to the Liar excludes *every* application of those rules from evidence-preserving reductio, including applications in which the relevant sentence is displayed and fully understood. Incurvati and Schlöder see no gap here. Evidence preservation, they insist, is a property of rules rather than particular inferences. One counterexample suffices to establish that a rule is not evidence-preserving; that some of its applications are epistemically impeccable is "neither here nor there" (2023a, 551).

That reply is correct as far as it goes. A rule is uniformly evidence-preserving only if all of its applications preserve evidence. But it leaves open a prior question: what makes a collection of applications one rule for the purpose of restricting reductio? The question matters because the proof theory is sensitive to how rules are individuated. We can retain an unrestricted truth rule as a commitment-preserving, meaning-conferring rule while also recognising a restricted, transparent subrule as evidence-preserving. Doing so adds no new primitive transition at the level of commitment: every application of the subrule was already licensed by the unrestricted rule. It nevertheless changes which derivations qualify for evidence-preserving reductio. In the case of truth, the change is catastrophic. The transparent subrule is enough to restore the Liar.

My claim is not that unrestricted transparency is consistent, nor that the inferentialist calculi are inconsistent as formulated. It is that the proposed evidential restriction is imposed at the wrong grain. The opacity cases establish that the unrestricted truth rules are not *uniformly* evidence-preserving. They do not establish that no transparent restriction of those rules preserves evidence. If transparent applications are nevertheless excluded because admitting them restores paradox, then the paradox, rather than the independent opacity cases, is determining the restriction. The account may still provide a coherent non-transparent theory of truth. But its claim to offer an independently motivated evidential diagnosis is undermined.

I first reconstruct the small part of the multilateral framework at which the issue arises. I then isolate what the opacity cases establish and formulate a general requirement on rule-restricted proof theories: their evidential consequences should be invariant under conservative

refinements that leave the primitive commitment transitions and the underlying epistemic differences unchanged. The truth-theoretic restriction violates this requirement. I apply the result first to the Liar and then to grounded truth reasoning that the coarse restriction also blocks. Finally, I show that the same fault line appears in Schlöder's recent treatment of unrestricted comprehension and Russell's paradox. The result is not a general objection to multilateralism. On the contrary, the multilateral distinction between weak and strong rejection is what makes the problem visible.

2. Where the paradox is stopped

The basic machinery can be introduced without reconstructing the whole of multilateral logic. Incurvati and Schlöder represent several attitudes toward a sentence (A). Write $(+A)$ for strongly asserting (A), $(-A)$ for strongly rejecting it, and $(\ominus A)$ for rejecting it in a weaker sense: declining or refraining from committing oneself to (A). Finally, $(\bot)$ marks the act of registering an incoherence. Strong assertion and strong rejection express opposed beliefs. Weak rejection need express no belief in the negation of the sentence rejected.

The distinction between the two rejections responds to a problem pressed by Dickie (2010). A rejection can be justified in different ways. One might reject *Homer wrote the Iliad* because Homer did not write it, but one might instead reject it because Homer never existed. Only the first evidence also supports the corresponding negative assertion. Rejection *tout court* is therefore evidentially unspecific: from the fact that rejecting (A) is justified, it does not follow that asserting $(\neg A)$ is justified. Marking strong rejection separately allows the bilateral negation rules to connect $(-A)$ with $(+\neg A)$ while leaving $(\ominus A)$ weaker (Incurvati and Schlöder, 2017).

The same problem arises for reductio. Suppose that an assumption $(+A)$ leads to $(\bot)$. That may show only that one ought not commit to (A), not that one has evidence for $(\neg A)$. Ordinary Smileian reductio accordingly yields weak rejection:

> If an assumption $(+A)$ leads to $(\bot)$, discharge it and infer $(\ominus A)$.

To obtain strong rejection, Incurvati and Schlöder introduce a stronger rule:

> If an assumption $(+A)$ leads to $(\bot)$ by an evidence-preserving derivation, discharge it and infer $(-A)$.

Call the latter rule strong reductio. Given the rules connecting strong rejection and negation, it plays the role of classical reductio in the assertoric fragment. The restriction to evidence-preserving subderivations ensures that the evidence hypothetically supplied for (A) could be transformed into evidence of incoherence and hence into evidence supporting $(\neg A)$. A route to

$(\bot)$ that merely records incompatible commitments can support withdrawal without supporting disbelief.

The distinction relies on an agentive conception of inference associated with Prawitz. The point of an inference is to justify its conclusion: a subject who possesses grounds for its premises can use the inference to acquire grounds for its conclusion (Prawitz 2015, 71–73; 2024). Evidence preservation is therefore not simply truth preservation. Whether a subject can use an inference depends on whether she can grasp and perform it. This remains so when reasoning occurs under assumptions. An open derivation represents a hypothetical transformation: if grounds for its assumptions were supplied, its inferential steps would transform them into grounds for its conclusion. Strong reductio is legitimate only when the derivation under its assumption has this character.

This machinery becomes philosophically interesting when the language is extended. Inferential deflationism takes the meaning of the truth predicate to be given partly by its relation to assertion. Its strongly asserted truth rules are:

$$\frac{+A}{+T^{\ulcorner}A^{\urcorner}}\ (+T\text{-IN}) \qquad \frac{+T^{\ulcorner}A^{\urcorner}}{+A}\ (+T\text{-OUT}).$$

There are corresponding rules for strong rejection. Together they capture truth's roles in indirect and compendious endorsement and opposition. Crucially, the rules are unrestricted at the level of commitment. If a speaker asserts that everything the guru says is true, then she is committed to what the guru says even when she cannot identify or understand all of it. This is precisely why the truth predicate cannot simply be limited to sentences already understood by a speaker. Its ability to bridge such gaps is part of its point (Shapiro 2003; Incurvati and Schlöder 2023b, §7.3).

Unrestricted truth rules produce the Liar. Let (L) be a sentence materially equivalent to $(\neg T^{\ulcorner}L^{\urcorner})$. Using truth introduction, one can derive $(+\neg L)$ from $(+L)$: from $(+L)$, infer $(+T^{\ulcorner}L^{\urcorner})$, and use the Liar equivalence to obtain $(+\neg L)$. In the other direction, the Liar equivalence takes $(+\neg L)$ to $(+T^{\ulcorner}L^{\urcorner})$, from which truth elimination yields $(+L)$. Thus:

$$+L \vdash +\neg L \qquad \text{and} \qquad +\neg L \vdash +L.$$

From the assumption $(+L)$, the first derivation produces both $(+L)$ and $(+\neg L)$, and hence $(\bot)$. If strong reductio applies, one obtains $(-L)$, and therefore $(+\neg L)$. But the second derivation then produces $(+L)$, yielding contradiction.

Incurvati and Schlöder block the final movement. The derivations from (L) to $(\neg L)$ and back remain valid because they preserve commitment. They do not count as evidence-preserving because they use truth rules. Ordinary reductio therefore permits us to infer $(\ominus L)$ and $(\ominus\neg L)$: both the Liar and its negation ought to be weakly rejected. Strong reductio cannot be applied,

so neither weak rejection can be converted into the assertion of the opposite sentence. The two weak rejections are jointly consistent.

The decisive restriction is thus easy to state: a derivation containing any application of a truth rule cannot occur beneath strong reductio. This includes apparently innocent applications such as the move from *'5 + 7 = 12' is true* to *5 + 7 = 12*. Incurvati and Schlöder accept this consequence. Many ordinary arguments can be rearranged so that the truth application occurs outside the restricted subderivation, and full transparency is in any event unavailable to a consistent theory of truth (2023b, §7.4). The present issue is not whether some restriction is required. It is whether the epistemic cases offered for the restriction justify imposing it by rule name, at this degree of generality.

3. What opacity establishes

Consider first Incurvati and Schlöder's whiteboard case. Thomas writes one sentence on a whiteboard each day and reliably writes truths. A speaker who knows Thomas's habit can have excellent evidence for *The sentence written on Thomas's whiteboard is true* without knowing which sentence is presently on the board. Suppose the sentence is *It is raining*. The speaker may lack any information about the weather and may not even know that *It is raining* is the relevant sentence. She can justify the truth predication by citing Thomas's reliability, but she cannot thereby formulate, much less justify, the particular assertion *It is raining*. Truth elimination does not transmit evidence in this situation (Incurvati and Schlöder, 2023a, 548–50).

Shapiro's disciple makes the point more vivid. The disciple endorses every utterance of a guru who speaks both arithmetic and set theory, although the disciple speaks only the language of arithmetic. The guru asserts a set-theoretic sentence (A), together with bridge principles connecting it to an arithmetical sentence (B). A logician informs the disciple that (B) follows. The disciple can recognise the arithmetical conclusion and become committed to it, perhaps even coming to believe it on the logician's testimony. Yet she cannot use her evidence for (A) *is true* to obtain evidence for (A), because she cannot understand (A) or grasp what would count as evidence for it. Truth predication can therefore be meaningful and inferentially consequential even in the absence of competence with the sentence called true (Shapiro, 2003).

These are genuine counterexamples to a uniform evidence-preservation claim. But notice what makes them work. In the whiteboard case, the inference apparently represented as

$$T(d) \ \therefore \ A$$

suppresses several epistemically significant transitions. The subject would need to identify the referent of (d) as $(\ulcorner A \urcorner)$, substitute the canonical quotation into the truth predication, and then disquote. She cannot perform the first two transitions. In the disciple case, the canonical

quotation may be available, but the subject cannot grasp (A) and so cannot perform an inference whose conclusion is (A). In each case, the subject lacks something required to turn a distant endorsement into a grasped conclusion.

Nothing similar is missing when a competent speaker moves from *'Snow is white' is true* to *Snow is white*. The sentence is presented, the quotation is recognised, and the conclusion is understood. This does not by itself prove that pure disquotation always preserves evidence. It does establish that the independent explanation of failure in the opaque cases no longer applies. The epistemic comparison supports a mixed verdict: the general rule has applications that cannot be used to acquire evidence, while at least some transparent applications have a prima facie claim to do exactly that.

Incurvati and Schlöder respond that one counterexample is enough. They define evidence preservation, like truth preservation, over a whole rule: a rule preserves evidence just when every application does. On this definition the unrestricted truth rule plainly fails. Let $(I(R))$ be the class of applications of a rule (R), and let $(EP(i))$ mean that application (i) can transmit grounds in the required way. The whiteboard and disciple cases establish:

$$\neg\forall i \in I(R)\, EP(i).$$

That is sufficient for (R) not to be uniformly evidence-preserving. What it does not establish is:

$$\forall i \in I(R)\, \neg EP(i),$$

or, more importantly, that $(I(R))$ contains no proper and independently specifiable subclass all of whose applications preserve evidence. The distinction is not verbal. Strong reductio is sensitive to which *rule* a proof step instantiates. Whether a transparent subclass is printed as a rule in its own right can therefore change what the calculus permits.

4. Evidence at the wrong grain

The problem can be seen without any special assumptions about truth. Imagine that a proof system contains two generic rules, one evidence-preserving and the other not. Strong reductio is allowed to use the first but not the second. Now replace the two displayed rules with a single rule whose side condition is the disjunction of their respective side conditions. The new rule licenses exactly the same premise-conclusion transitions as the original pair. Since it has counterexamples to evidence preservation, however, strong reductio may use none of its applications. A merely presentational merger has removed previously available strong conclusions. Splitting the rule again restores them. Nothing about the evidence possessed by any subject has changed.

This elementary example identifies a general issue for proof theories that classify derivations by

the generic rules they contain. There are three levels to distinguish. An *application* is a particular transition made in a derivation. A *generic inference* groups applications of a certain form. A *presentation* groups generic inferences into the rules of a calculus. Prawitz (2024) carefully distinguishes generic inferences from their applications, since a generic inference specifies, among other things, which assumptions may be discharged while an application specifies what is actually discharged. But generic inferences themselves can be presented more or less coarsely. The fact that validity is predicated of a generic inference does not settle which collection of applications constitutes the relevant generic inference.

We can make this precise with very little formalism. Treat a rule (R) extensionally as a set $(I(R))$ of licensed applications. A *refinement* of (R) is a family $(R_1, \ldots, R_n)$ such that

$$I(R) = I(R_1) \cup \cdots \cup I(R_n).$$

The refinement changes how the applications are classified but not which local transitions are licensed. In the case at hand, let (R^τ) comprise the transparent applications and (R^ω) the opaque ones. The relevant distinction need not be reduced to syntax alone. (R^τ) may carry epistemic side conditions: the subject understands the conclusion, recognises the relevant quotation or extension term, and is able to perform the transition as a means of acquiring evidence. The target account already appeals to precisely such facts when explaining why the opaque applications fail.

It is important that this refinement need not replace the original rule. Inferential deflationists can retain (R) as the unrestricted, meaning-conferring rule that preserves commitment. They can additionally recognise (R^τ) as an evidentially certified route. This is a conservative addition at the level of primitive commitment transitions. Every transition licensed by (R^τ) was already licensed by (R), so any derivation using (R^τ) can be transformed into a derivation of the same conclusion by relabelling that occurrence as (R). The primitive commitment transitions, and hence the inferential role used to explain the relevant expression's meaning, remain unchanged.

Call this the **Base-Conservativity Lemma**:

> If (R^τ) licenses only applications already licensed by (R), adding (R^τ) as a separately certified rule introduces no new primitive commitment transition.

The proof is immediate: replace every occurrence of (R^τ) with the corresponding occurrence of (R). The importance of the lemma is not its difficulty but the contrast it enables. Although the refinement changes no local transition, it can change which subderivations qualify for strong reductio. An occurrence labelled (R) is disallowed because (R) has opaque counterexamples; the very same transition labelled (R^τ) is allowed if the transparent subrule preserves evidence. Consequently, strong rejection and strong assertion can differ between the two presentations.

This gives us the **Restricted Non-Invariance Lemma**:

> A proof theory that restricts strong reductio by the evidential status of generic rule labels need not be invariant under conservative refinement. Two presentations can agree on every primitive, meaning-conferring commitment transition while disagreeing on strong-rejection consequences.

Again the proof is simple. Take a derivation of $(\bot)$ from an assumption $(+A)$ whose only prohibited occurrence is a transparent application of (R). In the coarse presentation, the occurrence of (R) prevents strong reductio and yields at most $(\ominus A)$. In the refined presentation, label that occurrence (R^{τ}). The commitment derivation is otherwise unchanged, but strong reductio now yields $(-A)$. If the language connects $(-A)$ with $(+\neg A)$, assertoric consequences change as well.

The lemma motivates a constraint on evidentially restricted proof systems. Call it the **Evidential-Grain Requirement**:

> When a proof restriction is justified by whether an inferential transition transmits evidence, its consequences should be invariant under conservative refinements that preserve both the primitive commitment transitions and the epistemic profile of their applications.

Without this requirement, evidential consequences depend on editorial decisions about the rule sheet. We can make an evidence-preserving inference unavailable beneath reductio by merging it with an unsafe rule, or restore it by splitting the rule. A restriction that is supposed to track the transmission of grounds should not be manipulable in this way.

One might object that the same point could be made about truth preservation. If a schematic rule has a counterexample, logicians call it invalid even when some instances happen to preserve truth. That is correct, but it does not answer the present problem. Logicians also freely recognise valid restrictions of invalid schemata. The invalidity of an unrestricted rule does not make every application of each valid subrule unusable in a proof. What is distinctive here is that a negative verdict about the coarse schema is used to prohibit individual applications inside an otherwise legitimate higher-order rule. The prohibition therefore requires not merely a definition of uniform validity but a reason to regard the coarse schema as the governing unit of evidential permission.

The agentive character of evidence strengthens this point. Truth preservation abstracts from what any subject understands. Evidence preservation, as Incurvati and Schlöder conceive it, does not. Their counterexamples succeed because one subject cannot identify a sentence and another cannot understand it. If those differences matter when finding counterexamples, it is

difficult to see why they become irrelevant when classifying the transparent applications. A rule-wide universal is useful for saying whether a schema is safe without qualification. It is not yet a theory of which inferential performances may transmit grounds under a supposition.

Nor can the grain be fixed merely by appeal to meaning. The unrestricted truth rules are supposed to explain both direct and compendious endorsement. Restricting their *meaning-conferring range* to sentences already understood would indeed sacrifice part of truth's function. But the present proposal does not do that. The unrestricted commitment rule remains in place and continues to explain distant endorsement. The transparent rule is an additional evidential certificate, not a substitute meaning rule. Inferential expressivism already distinguishes the rules that determine commitments from the subclass of reasoning that transmits evidence. There is no general reason why the latter classification must reproduce the former's grain.

There are, moreover, precedents for this flexibility within the programme. Incurvati and Schlöder show that many ordinary arguments involving truth can be rearranged so that a truth rule is applied in the global proof context rather than inside a restricted subderivation (2023b, §7.4). Schlöder later introduces an *Absoluteness* metarule to recover a class of evidence-preserving inferences among extension claims, even though the underlying extension rules are globally classified as non-preserving (2025, 7–8). These devices do not establish that transparent disquotation or comprehension preserves evidence. They do show that the framework can separately certify epistemically acceptable routes without abandoning its unrestricted meaning-conferring rules.

The remaining question is therefore substantive. Are the transparent subrules evidence-preserving? The opacity cases do not decide the issue. But the paradoxes exert immediate pressure on the answer.

5. Disquotation under supposition

Consider a transparent form of truth elimination:

$$\frac{+T^{\ulcorner}A^{\urcorner}}{+A}\ (+T\text{-OUT}^{\tau}),$$

where the subject understands (A), recognises its canonical quotation, and is capable of performing the displayed inference. There is a corresponding transparent truth-introduction rule. The side conditions are not chosen to protect a preferred derivation. They negate the very features that make the whiteboard and disciple cases work. No denoting expression must be resolved, no unknown sentence must be recovered, and no conclusion lies outside the subject's competence.

On the Prawitzian conception invoked by inferential deflationism, these rules have a strong prima facie claim to preserve evidence. Suppose a subject has testimonial evidence for *'Snow is white' is true*, understands *Snow is white*, and recognises the quotation. She can cite the testimony in support of *Snow is white*; indeed, it would be peculiar to regard the testimony as supporting only

the metalinguistic predication and not the proposition transparently predicated true. Conversely, someone with evidence that snow is white can use the meaning-conferring introduction rule to justify saying that *'Snow is white' is true*. In neither direction does the subject need to discover a hidden identity or acquire competence she lacks.

This is not an argument for unrestricted transparency. Even among sentences a speaker could in principle understand, evidence may be acquired through an opaque description of a sentence whose referent she does not presently know. The transparent rule is narrower. It certifies the inferential performance in which the content and its canonical truth predication are both before a competent subject. If that performance still fails to transmit grounds, the explanation must be different from the one supplied by the guru and whiteboard cases.

Unfortunately, the Liar uses exactly this kind of disquotation. The sentence (L) is displayed; its canonical quotation is explicitly available; and anyone following the proof must understand the vocabulary required to formulate both (L) and its truth predication. The derivation from $(+L)$ to $(+\neg L)$ employs transparent truth introduction, and the derivation from $(+\neg L)$ to $(+L)$ employs transparent truth elimination. Certify these applications and the restriction on strong reductio no longer blocks the paradox:

$$\begin{array}{c}[+L]\\ \vdots\\ (+T\text{-IN}^{\tau})\\ +\neg L\\ \bot\end{array} \quad \Rightarrow \quad -L \quad \Rightarrow \quad +\neg L \quad \Rightarrow \quad +L \quad \Rightarrow \quad \bot.$$

The point can be stated as a corollary of restricted non-invariance. The coarse presentation contains only the unrestricted truth rules, classified as non-preserving because of opaque applications. The refined presentation retains those rules while adding transparent certificates. Both presentations have the same primitive, meaning-conferring commitment transitions. But the refined presentation permits the transparent Liar steps beneath strong reductio and is inconsistent. The paradox solution is therefore not invariant under a conservative refinement motivated by the target's own account of why evidence transmission sometimes fails.

Incurvati and Schlöder have a direct response. They explicitly consider restricting the truth rules to pure disquotation and reject the proposal. The semantic paradoxes, they say, show that the truth rules fail to preserve evidence even in pure applications (2023b, §7.3). The response is coherent. The Liar proves that the transparent rules cannot combine with all the other evidence-preserving principles of their theory, including strong reductio and the classical laws governing negation. Something in that package must give.

But this response also reveals the explanatory limit of the independent argument. The disciple and whiteboard show that distant endorsement does not uniformly transmit evidence. They do not show what goes wrong in the pure Liar steps. The reason for classifying those steps as

non-preserving is that allowing them beneath reductio produces paradox. At this point the Liar, rather than opacity, fixes the evidential boundary. Inferential deflationists can still say that the paradox supports their diagnosis just as it supports the paracompletist's restrictions on classical reasoning or the paraconsistentist's tolerance of gluts (Field 2008; Priest 1979). What they lose is an independent argument that specifically selects their diagnosis over these alternatives.

This is more than a point about advertising. The independent argument was supposed to explain *why* reductio is misapplied in the Liar. Without it, "the derivation does not preserve evidence" risks becoming a proof-theoretic redescription of the fact that the derivation must be blocked somewhere. The theory needs a non-paradoxical criterion that identifies the pure Liar steps as defective while distinguishing them from ordinary transparent disquotation. Global classification by rule name supplies a formally sufficient boundary, but the opacity cases do not show that this is the epistemically correct boundary.

The price of that boundary appears in grounded reasoning. Let (S) be an ordinary, fully understood sentence. Consider:

$$\neg\big(S \wedge T^{\ulcorner}\neg S^{\urcorner}\big). \tag{1}$$

This is one of the weak truth-theoretic principles highlighted by Heck (2012). Its instances are compelling even in the absence of unrestricted transparency. It cannot be both that snow is white and that *'Snow is not white' is true*. A natural justification assumes the conjunction, extracts (S) and the truth predication of $(\neg S)$, disquotes the latter to $(\neg S)$, and registers the resulting incoherence. Strong reductio then supports the negation of the conjunction.

In the inferential-deflationist calculus, that natural proof contains a truth rule beneath strong reductio. It can therefore justify weakly rejecting the conjunction, but not strongly rejecting it or asserting (1). This is not merely a cost involving an exotic self-referential sentence. Set (S) to *5 + 7 = 12*. A competent reasoner appears able to transform any putative grounds for

$$5 + 7 = 12 \text{ and } \textit{'5 + 7} \neq \textit{12' is true}$$

into grounds for incoherence, using only conjunction elimination and canonical disquotation. The opaque cases supply no reason why this hypothetical transformation fails.

The book's procedure for moving truth rules outside restricted contexts does not recover this case. The procedure applies when the desired output of a truth rule does not itself depend on assumptions that must be discharged. One can then assume that output inside the restricted context and apply the truth rule only after returning to the global context. In the proof of (1), however, the truth predication is obtained from the conjunction assumed for reductio. Its disquoted output $(\neg S)$ therefore depends on the very assumption to be discharged. Moving the truth rule outside would leave no categorical truth premise from which to apply it.

Heck shows that principles such as (1) cannot be retained without limit in a sufficiently expressive self-referential setting. His argument is a reason not to demand their unrestricted validity. It does not show that the arithmetic instance is epistemically defective. On the contrary, it sharpens the present dilemma. If even grounded instances must be withheld because their schema becomes paradoxical under self-reference, the restriction is sensitive to the global paradox-generating potential of a principle rather than to whether this subject can use this inference to acquire evidence. That may be a defensible theory of truth, but it is no longer the straightforward agentive diagnosis suggested by the opacity cases.

The truth case therefore leaves inferential deflationism with three possibilities. It can admit transparent evidential subrules, in which case the Liar returns. It can exclude them because they return the Liar, in which case paradox rather than opacity does the classificatory work. Or it can supply a further, independently motivated account of why even competent canonical disquotation fails to transmit evidence under a supposition. The existing argument establishes none of these further claims.

6. The same fault line in comprehension

The problem is not peculiar to truth or quotation. Schlöder (2025) has recently extended the same strategy to unrestricted comprehension. The characteristic extension rules allow movement between an object's falling under a concept and its belonging to the concept's extension:

$$\frac{+Px}{+x \in \varepsilon P}\ (+\varepsilon\text{-IN}) \qquad \frac{+x \in \varepsilon P}{+Px}\ (+\varepsilon\text{-OUT}).$$

There are corresponding rules for strong rejection. Schlöder argues that these rules preserve commitment but not evidence. Excluding them from strong reductio blocks Russell's paradox while leaving them unrestricted at the commitment level. The resulting calculus is consistent and proves full second-order arithmetic. The ambition is substantial: classical arithmetic is recovered in a logic with unrestricted comprehension, while Russell's paradox receives the same evidential diagnosis as the Liar.

The independent case for non-preservation again exploits opacity. A person may know that a cube belongs to the extension of *Platonic solid* while not understanding the concept expressed by the predicate. Perhaps she knows only that the relevant extension comprises the tetrahedron, cube, octahedron, dodecahedron, and icosahedron. She can justify the extensional claim by consulting the list. She cannot use that evidence to justify the intensional claim that the cube is a Platonic solid, understood as a regular convex polyhedron in three-dimensional space, because she does not know which properties constitute the concept. Nevertheless, the extensional assertion commits her to consequences of the intensional claim once those consequences are explained (Schlöder, 2025, 6–7).

Schlöder carefully notes that natural language obscures a difference here. *The cube is a Platonic solid* may use *Platonic solid* as shorthand for an extension or may express an intension. These uses have different inferential roles: only the latter supports an immediate inference to the cube's being convex. From an inferentialist perspective this qualification is important. The motivating failure occurs when a subject moves from an extensional understanding to an intensional content she does not grasp. It shows that the unrestricted extension rules are not uniform evidence transformers. It does not establish that the rules fail for a subject who understands (P), recognises (εP) as its extension, and has both sides of the conversion transparently before her.

Here, too, a transparent certificate would add no commitment transition. Every licensed inference from (Px) to membership in the extension of (P), or back, is already licensed by the unrestricted extension rule. The certificate would matter only beneath strong reductio. And once it is admitted, Russell's paradox returns. Let

$$R = [\lambda v.\neg v \in v].$$

Then:

$$+\varepsilon R \in \varepsilon R \ \vdash \ +R(\varepsilon R) \ \vdash \ +\neg\varepsilon R \in \varepsilon R,$$

while in the other direction:

$$+\neg\varepsilon R \in \varepsilon R \ \vdash \ +R(\varepsilon R) \ \vdash \ +\varepsilon R \in \varepsilon R.$$

The concept (R) is explicitly displayed and constructed from familiar logical vocabulary. The derivation does not rely on a disciple who has failed to grasp which concept the guru expressed. If the relevant extension conversions are certified as transparent, strong reductio converts the first derivation into rejection of self-membership and the second back into its assertion. The calculus becomes inconsistent for the same reason as in the truth case.

Schlöder's Absoluteness principle makes the grain issue especially clear. Suppose that a relation between concepts can be grasped without understanding the concepts themselves—for example, that a conjunctive concept entails each conjunct. Absoluteness carries an evidence-preserving inference between concept claims over to a corresponding evidence-preserving inference between extension claims. Thus a disciple may infer from membership in the extension of a conjunctive concept to membership in the extension of a conjunct even while remaining ignorant of the concepts' internal contents. Schlöder adds this metarule because the blanket restriction on the extension rules would otherwise withhold too much ordinary reasoning (2025, 7–8).

Absoluteness does not certify the direct movement from extension to intension used in Russell's paradox, so it does not by itself create inconsistency. But it confirms the structural point. The

proof theory can distinguish epistemically acceptable routes involving extension vocabulary from the unrestricted extension rules that determine commitment. The question is why this selectivity should stop before competent, transparent conversion. The extensional-list example shows that conversion can fail when competence is missing. It does not supply an independent answer for Russell's displayed concept. Once again, the paradox itself supplies the pressure to treat every direct conversion as non-preserving.

Truth and comprehension therefore instantiate a more general pattern. A meaning-conferring bridge rule supports both opaque and transparent uses. Opacity establishes that the bridge is not a universally safe evidence transformer. A proof theory then bans every application of the bridge inside a higher-order rule, such as strong reductio. But the ban is not invariant under conservative refinement, and a transparent refinement restores the paradox. Any account that follows this pattern needs a principled reason why the evidential classification should be made at the coarse, paradox-blocking grain.

7. What remains of the diagnosis

The argument leaves much of inferential expressivism untouched. The distinction between commitment and evidence is genuine and philosophically useful. The guru and whiteboard cases show that speakers can undertake inferential commitments that they are not yet in a position to appreciate or use as evidence. Truth rules can explain those commitments without being construed as uniformly evidence-preserving. Similarly, extension talk may carry commitments beyond a speaker's present grasp of a concept. None of this depends on the success of the proposed paradox solution.

What fails is the inference from those observations to the particular restriction on strong reductio. The independent cases justify a negative verdict about unrestricted rules. They do not identify the proper evidential grain. At the level of primitive, meaning-conferring commitment transitions, coarse and refined presentations agree. Once evidence-sensitive discharge is added, they diverge. The inferentialist cure therefore needs a reason why the meaning-conferring rule, rather than some transparent subrule, is the unit whose failures determine every permissible occurrence under reductio.

There are several ways forward. One is to develop genuinely application-sensitive annotations for evidence transmission. This would fit the agentive motivation of the view, but it would immediately confront the task of excluding the transparent Liar and Russell steps without merely invoking their paradoxical consequences. A second is to add a semantic or grounding-sensitive boundary. That may produce a defensible theory resembling familiar supervaluationist or grounded approaches (Kripke, 1975), but it would surrender the claim that ordinary epistemic opacity independently explains the paradoxical failures. A third is to retain the coarse syntactic ban and accept its consequences: the system remains formally coherent, but its evidence-sensitive fragment is presentation-dependent and withholds strong grounded conclusions such as ordinary

instances of (1).

The methodological lesson extends beyond these cases. Kortenbach, Incurvati, and Schlöder (2025) show that multilateral supervaluationist systems can be classical at the levels of theorems and ordinary inferences while remaining weaker at meta- and higher inferential levels. The present problem arises at just such a higher level. Two systems can agree on every local commitment transition while differing over which derivations may be embedded under reductio. Once epistemic properties of rules regulate such embeddings, conservative equivalence at the base level is no longer enough. The evidential classification must itself be stable under harmless changes of presentation.

This is precisely what the multilateral machinery allows us to see. Mainstream epistemology can describe the difference between testimony whose content one understands and testimony one endorses blindly. Mainstream theories of truth can restrict transparency. What they do not by themselves display is the difference between a derivation that warrants only weak rejection and one that warrants strong rejection because of the rule labels occurring beneath reductio. Inferential expressivism exposes that difference. It thereby exposes the burden as well: if evidence is the reason one derivation is barred, the restriction must track evidence rather than the accidental grain of the rulebook.

A blind application shows that an unrestricted rule is too coarse to serve as a universal evidence transformer. It does not show that evidence should be tracked at that same coarse grain. Until inferential deflationism supplies a principled reason to exclude transparent applications independently of the paradoxes they generate, opacity cannot bear the weight of its cure.

References

Imogen Dickie. Negation, anti-realism, and the denial defence. *Philosophical Studies*, 150:161–185, 2010.

Hartry Field. *Saving Truth from Paradox*. Oxford University Press, Oxford, 2008.

Richard Kimberly Heck. A liar paradox. *Thought*, 1:36–40, 2012.

Luca Incurvati and Julian J. Schlöder. Weak rejection. *Australasian Journal of Philosophy*, 95: 741–760, 2017.

Luca Incurvati and Julian J. Schlöder. Epistemic multilateral logic. *Review of Symbolic Logic*, 15: 505–536, 2022.

Luca Incurvati and Julian J. Schlöder. Inferential deflationism. *Philosophical Review*, 132:529–578, 2023a.

Luca Incurvati and Julian J. Schlöder. *Reasoning with Attitude: Foundations and Applications of Inferential Expressivism*. Oxford University Press, New York, 2023b.

Bas Kortenbach, Luca Incurvati, and Julian J. Schlöder. Multilateral supervaluationism and classicality. *Journal of Philosophical Logic*, 54:247–290, 2025.

Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

Dag Prawitz. Explaining deductive inference. In Heinrich Wansing, editor, *Dag Prawitz on Proofs and Meaning*, pages 65–100. Springer, Dordrecht, 2015.

Dag Prawitz. The aim and validity of inference and argument. *Theoria*, 90:515–527, 2024.

Graham Priest. The logic of paradox. *Journal of Philosophical Logic*, 8:219–241, 1979.

Julian J. Schlöder. Basic rules of arithmetic. *Australasian Journal of Philosophy*, 2025. doi: 10.1080/00048402.2025.2521739. URL `https://doi.org/10.1080/00048402.2025.2521739`. Advance online publication.

Stewart Shapiro. The guru, the logician, and the deflationist: Truth and logical consequence. *Noûs*, 37:113–132, 2003.