Inferences and Metainferences in ST

Pablo Cobreros[1] ⓘ · Paul Egré[2] · David Ripley[3] · Robert van Rooij[4]

Abstract

In a recent paper, Barrio, Tajer and Rosenblatt establish a correspondence between metainferences holding in the strict-tolerant logic of transparent truth ST^+ and inferences holding in the logic of paradox LP^+. They argue that LP^+ is ST^+'s external logic and they question whether ST^+'s solution to the semantic paradoxes is fundamentally different from LP^+'s. Here we establish that by parity of reasoning, ST^+ can be related to LP^+'s dual logic $K3^+$. We clarify the distinction between internal and external logic and argue that while ST^+'s nonclassicality can be granted, its self-dual character does not tie it to LP^+ more closely than to $K3^+$.

Keywords Strict-tolerant logic · Metainferences · Proof theory ·
Internal vs external logic · Paradoxes

The strict-tolerant logic ST was proposed to deal with paradoxes of vagueness and with the semantic paradoxes [8, 9]. There is something very distinctive about ST:

✉ Pablo Cobreros
pcobreros@unav.es

Paul Egré
paul.egre@ens.fr

David Ripley
davewripley@gmail.com

Robert van Rooij
R.A.M.vanRooij@uva.nl

1 Department of Philosophy, University of Navarra, 31009 Pamplona, Spain

2 Département de Philosophie & Département d'Etudes Cognitives de l'ENS, Institut Jean-Nicod (CNRS-EHESS-ENS), PSL University, 29, rue d'Ulm, 75005, Paris, France

3 Philosophy Department SoPHIS, Monash University, Building 11, Melbourne, VIC 3800, Australia

4 Institute for Logic, Language and Computation, Universiteit van Amsterdam, P.O. Box 94242, 1090 GE, Amsterdam, The Netherlands

namely, it is classical logic for a classical language, but it provides ways of *strengthening* classical logic to deal with paradoxes in enriched languages. For example, the logic ST^+ (ST for a language with a transparent truth predicate T and self-referential sentences) is an *inference-preserving extension* of classical logic. That is, ST^+ is not only non-trivial, but it extends all the valid inferences of classical logic to cover the full (T-involving) language [10, 23].

How is this possible? Well, because ST^+ preserves all classically valid *inferences* but not some classical *metainferences*. The question then arises of exactly which are the metainferences of ST^+. In a recent paper, Eduardo Barrio, Lucas Rosenblatt and Diego Tajer show that ST^+'s metainferences are closely related to inferences in LP^+, the logic LP extended with a transparent truth predicate. In this paper we review their result and put the connection in a broader context. In particular, we show that in much the same way in which ST^+ is related to the paraconsistent logic LP^+, it can be related to the paracomplete logic $\mathsf{K3}^+$ when we look at the logic's *meta-anti-inferences*.

The results in this paper are either results already proved in Barrio et al. [4] or corollaries based on duality considerations. On the technical side, the contribution of this paper is in the simplification of the proofs in Barrio et al. [4]. On the philosophical side, we believe that the dual result presented in this paper sets a richer context for the understanding of the logic ST and toward the assessment of its merits and shortcomings as a solution to paradoxes.

1 Three-Valued Logics Based on Strong Kleene

Let $\mathcal{L}$ be a propositional language with the usual connectives: $\wedge, \vee, \supset, \neg$. Let an interpretation $\mathbb{I}$ be a function from propositional letters to $\{1, \frac{1}{2}, 0\}$. Interpretations extend to formulas according to the following *Strong Kleene scheme*:[1]

- $\mathbb{I}(\neg A) = 1 - \mathbb{I}(A)$
- $\mathbb{I}(A \wedge B) = min(\mathbb{I}(A), \mathbb{I}(B))$
- $\mathbb{I}(A \vee B) = max(\mathbb{I}(A), \mathbb{I}(B))$
- $\mathbb{I}(A \supset B) = max(1 - \mathbb{I}(A), \mathbb{I}(B))$

If we think of logical consequence as necessary preservation of truth, there are two standard ways in which we can define logical consequence over this semantics. First, if *being true* means taking the value 1 (strict truth), then an argument is valid just in case no interpretation gives all premises the value 1 and all conclusions a value less than 1. Second, if *being true* means taking a value greater than 0 (tolerant truth), then an argument is valid just in case no interpretation makes all premises greater than 0 and all conclusions equal to 0.

Each definition leads to a familiar three-valued logic. The first to the Strong Kleene logic K3 and the second to Priest's *Logic of Paradox* LP. K3 is *paracomplete* in the sense that $\nvDash^{\mathsf{K3}} A \vee \neg A$ and LP is *paraconsistent* in the sense that $A \wedge \neg A \nvDash^{\mathsf{LP}}$.

---

[1]Throughout the paper we use the letters $A, B, C, ...$ for metavariables of formulae, and the letters Γ, Δ for metavariables of sets of formulae.

Fig. 1 Four three-valued logics
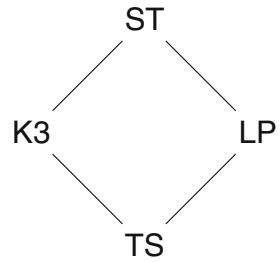
Furthermore, if we assume the language does not contain constant proposition symbols like $\top$ or $\bot$, K3 is "classical on the left" but empty on the right in the sense that for any sets of formulas Γ and Δ:

$$\Gamma \vDash^{\mathsf{K3}} \emptyset \;\; \text{iff} \;\; \Gamma \vDash^{\mathsf{CL}} \emptyset \qquad \text{and} \qquad \emptyset \nvDash^{\mathsf{K3}} \Delta.$$

In words: Γ is K3-unsatisfiable just in case it is classically unsatisfiable and no set Δ is K3-valid. Similarly, LP is empty on the left but classical on the right in the sense that,

$$\Gamma \nvDash^{\mathsf{LP}} \emptyset \qquad \text{and} \qquad \emptyset \vDash^{\mathsf{LP}} \Delta \;\; \text{iff} \;\; \emptyset \vDash^{\mathsf{CL}} \Delta.$$

In words: no set Γ is LP-unsatisfiable but for any Δ, it is LP-valid just in case it is classically valid.

The definitions of logical consequence for K3 and LP are based on the idea that logical consequence should preserve designated values from premises to conclusions. By liberalizing the notion of logical consequence – from the idea of preserving the same set of designated values from premises to conclusions to the idea of going from one designated set to another set from premises to conclusions – we arrive, in the present context, at the following two alternative definitions of logical consequence (see Cobreros et al. [10]):

- An argument is ST-valid if there is no interpretation giving all premises the value 1 and all conclusions the value 0.
- An argument is TS-valid if there is no interpretation giving all premises a value greater than 0 and all conclusions a value less than 1.

The logics are ordered by inclusion (see Fig. 1): in the present language, TS is empty (in the sense that no inference is valid) and ST is classical logic (in the sense that for any inference, it is classically valid iff ST-valid). K3 and LP are both weaker than ST and stronger than TS. K3 and LP are each other's duals while TS and ST are self-dual.[2]

---

[2]Two logics X and Y are duals just in case $\Gamma \vDash^{\mathsf{X}} \Delta$ iff $\neg\Delta \vDash^{\mathsf{Y}} \neg\Gamma$ (where $\neg\Gamma$ is shorthand for $\{\neg A \mid A \in \Gamma\}$), see Cobreros et al. [8]. Duality will play a role in Section 3.

The logics K3, LP, ST and TS have been vindicated as suitable logics to deal with different paradoxes.[3] Although our discussion in this paper will be restricted to propositional logic, we consider for a moment a first-order language in order to illustrate how these logics can be used to deal with semantic paradoxes. Suppose $\mathcal{L}^+$ is a first-order language with identity (for simplicity: no functions) with the constant propositions $\top$ and $\bot$, a distinguished predicate T (our truth predicate) and a distinguished name $\langle A \rangle$ for each formula A. We further assume the possibility of diagonal formulas so that there are Liar-like sentences such as,

$$\lambda \quad = \quad \neg T \langle \lambda \rangle$$

Definition 1 A three-valued model M for the first-order language $\mathcal{L}^+$ is a structure $\langle \mathbb{D}, \mathbb{I} \rangle$ such that:

- $\mathbb{D}$ is a domain such that $\mathcal{L}^+ \subseteq \mathbb{D}$ and
- $\mathbb{I}$ is an interpretation function:

  - For a name or variable a, $\mathbb{I}(a) \in \mathbb{D}$
  - For a name $\langle A \rangle$ of a formula A, $\mathbb{I}(\langle A \rangle) = A$
  - For an n-ary predicate P, $\mathbb{I}(P) \in \{0, \frac{1}{2}, 1\}^{\mathbb{D}^n}$
  - For atomic sentences $P t_1 t_2 ... t_n$,
    $$\mathbb{I}(P t_1 t_2 ... t_n) = \mathbb{I}(P)(\langle \mathbb{I}(t_1), \mathbb{I}(t_2), ..., \mathbb{I}(t_n) \rangle)$$
  - $\mathbb{I}(s = t) = 1$ iff $\mathbb{I}(s) = \mathbb{I}(t)$, and 0 otherwise,
  - $\mathbb{I}(\top) = 1; \mathbb{I}(\bot) = 0$,
  - $\mathbb{I}(\neg A) = 1 - \mathbb{I}(A)$
  - $\mathbb{I}(A \wedge B) = min(\mathbb{I}(A), \mathbb{I}(B))$
  - $\mathbb{I}(A \vee B) = max(\mathbb{I}(A), \mathbb{I}(B))$
  - $\mathbb{I}(A \supset B) = max(1 - \mathbb{I}(A), \mathbb{I}(B))$
  - $\mathbb{I}(\exists x A) = max(\{\mathbb{I}^\dagger(A) \mid \mathbb{I}^\dagger$ is an x-variant of $\mathbb{I}\})$

A Kripke-construction (see Kripke [17]) shows how to find a three-valued model M^+ for $\mathcal{L}^+$ out of a model M for $\mathcal{L}^+$ such that the new model M^+ is *conservative* (agrees with M on the interpretation of all T-free formulas of $\mathcal{L}^+$) and *transparent* (the truth-value of A equals the truth-value of $T \langle A \rangle$ for every formula A of $\mathcal{L}^+$). As a consequence, paradoxical sentences such as λ get value $\frac{1}{2}$. Once we have these models, we can define logical consequence in the style of ST, LP, K3 or TS.[4] The resulting logics can arranged in the diagram of Fig. 2, where logics are ordered by inclusion. Double bars indicate *conservative extensions*, that is, the logic X^+ is the logic X for the T-free fragment. Therefore, ST^+ is a conservative extension of classical logic with a transparent truth-predicate and self-reference. This means that ST^+ does not add or subtract valid inferences for the language without T. Furthermore, ST^+ is an

---

[3]See, for example, [17, 27] for K3, [22] for LP, [8] for ST and [14, 20] for TS.
[4]In this language, TS is no longer empty. For example, $\vDash^{TS} \top$.

Fig. 2 Four three-valued logics and their $\mathcal{L}^+$-extensions
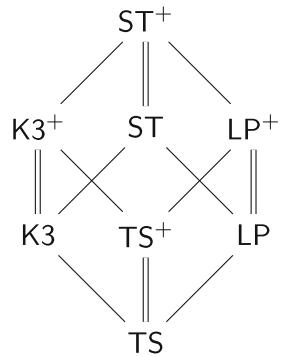
inference-preserving extension of classical logic, in the sense that if an inference is valid in the language without T, it is valid for any uniform substitution on the language containing T.[5] These facts ground our claim that ST^+ preserves classical logic [9, 853].

Kripke's construction is powerful in the sense that the original (first-order) language for the construction can be very rich. The construction can be used, therefore, as a way to find transparent and conservative models for interesting first-order theories. Our discussion below will be restricted, however, to the propositional case. For this reason, from now on, our language $\mathcal{L}^+$ for semantic paradoxes will be very simple: it will be the propositional language $\mathcal{L}$ enriched with the constants $\top$, $\bot$ and λ taken to denote the values 1, 0 and $\frac{1}{2}$ in every interpretation. The constant λ plays the role of the Liar sentence in the propositional setting in the sense that $\mathbb{I}(\lambda) = \mathbb{I}(\neg\lambda)$ in every model.[6]

2 Metainferences

In the same way in which an inference can be presented as a relation between sets of formulae, a metainference can be presented as a relation between sets of inferences. Here we lay out our framework in order to present Barrio et al.'s main result as well as our dualization of it.

2.1 Motivations

We can think of a consequence relation extensionally as a set of pairs of sets of formulas. As with any other relation, we may be interested in different properties the

---

[5]See [9, 23], and [10] for more on ST^+.

[6]Barrio et al. [4] present their main result in the context of a language that includes a truth predicate, like the one we consider here. For discussion of related results in languages without a truth predicate, see for example Přenosil [21].

relation might exhibit. We can wonder whether the relation is, say, Euclidean. That is, we can wonder whether the statement,

$$\forall x \forall y \forall z ((x R y \wedge x R z) \supset y R z)$$

is true for the consequence relation R. As an example, take R to be LP's consequence relation. Then the statement above is certainly false since, for example, $\langle \{p \wedge q\}, \{p\} \rangle \in \models^{\mathsf{LP}}$ and $\langle \{p \wedge q\}, \{q\} \rangle \in \models^{\mathsf{LP}}$ but $\langle \{p\}, \{q\} \rangle \notin \models^{\mathsf{LP}}$.

One way of representing metainferences, is in terms of schemata involving arbitrary formulae. Thus, letting the turnstile symbol "⊢" denote the consequence relation and "$\Longrightarrow$" the metalanguage material conditional, and using ";" to represent conjunctions (in antecedent position) and disjunctions (in consequent position), the Euclidean property above may be written:

$$A \vdash B; \ A \vdash C \Longrightarrow B \vdash C$$

We can think of these metainferences as expressing 'structural properties' in the sense that some of them will correspond to 'structural rules' (rules not involving specific logical vocabulary) in the context of some sequent calculus.[7] For example, the property expressed by the metainference

$$\top \Longrightarrow A \vdash A$$

corresponds to the Identity rule of various sequent calculi.

Likewise, we may represent the property for a consequence relation to be explosive by the following metainference:

$$A \vdash B \wedge \neg B \Longrightarrow A \vdash C$$

If we take again LP 's consequence relation, this metainference won't hold, since, for example, $p \wedge \neg p \models^{\mathsf{LP}} p \wedge \neg p$ but $p \wedge \neg p \not\models^{\mathsf{LP}} q$.

2.2 Definitions

Barrio et al. [4]'s central result establishes a correspondence between ST^+'s consequence exhibiting a metainferential property and LP^+ validating a specific argument. Barrio et al. [4] make use of two languages: one for metainferences (making use of schematic letters $A, B, C \ldots$) and another for inferences (making use of propositional variables $p, q, r \ldots$).

---

[7]The word 'structural' can also be understood in the sense of validity preservation under uniform substitutions of the language (see for example Arieli et al. [1, 34]).

Instead, in what follows we shall represent metainferences in terms of propositional formulae, instead of formula schemata: we will say that a metainference holds if every uniform substitution of its propositional letters is true. Our definitions below of a metainference, a metainference instance, and a metainference holding, allow us to present the discussion in the single language $\mathcal{L}^+$.[8]

As we mentioned above, we will restrict our attention to the propositional fragment of $\mathcal{L}^+$ with propositional letters: p, q, r, ...; the usual constants $\wedge$, $\vee$, $\supset$, $\neg$; plus the constant propositions $\top$, $\bot$ and λ. We leave out propositions involving truth and quantifiers (compare [4, 558], Definition 6).

Definition 2 (Metainference, ST^+-instance, holding for ST^+)

- A **metainference** is a conditional statement of the following form:

  (MI) $\Gamma_1 \vdash \Delta_1; \ \ldots \ ; \Gamma_n \vdash \Delta_n \implies \Gamma'_1 \vdash \Delta'_1; \ \ldots \ ; \Gamma'_k \vdash \Delta'_k$

  where the Γ's and Δ's are sets of formulae of $\mathcal{L}^+$.
- An ST^+**-instance** of a metainference is the result of replacing each propositional variable with some formula of $\mathcal{L}^+$ and all turnstile symbols by the double turnstile symbol with ST^+ superscripted ($\models^{\mathsf{ST}^+}$).
- We will say that **a metainference holds for** ST^+ when all its ST^+-instances are true.

For example,

$$p \vdash q \wedge \neg q \implies p \vdash r$$

is a metainference, and

$$\lambda \models^{\mathsf{ST}^+} p \wedge \neg p \implies \lambda \models^{\mathsf{ST}^+} \bot,$$

is an ST^+-instance of it. This particular instance is true since it is true that $\lambda \models^{\mathsf{ST}^+} \bot$. The metainference, however, does not hold for ST^+ since other ST^+-instances fail to hold (as shown in Example 1 below).

---

[8]Importantly, the notion of a metainference is susceptible of alternative notions of validity. In this paper, following Barrio et al. [4], we use a *global-substitutional* reading of the validity of a metainference in the sense that a metainference holds when, for every substitution of its propositional letters, if its premises are true, its conclusion is true. In Cobreros et al. [9] we assumed a *global* (non-substitutional) reading of the validity of a metainference, saying that a metainference holds when, if its premises are valid, so is its conclusion. Finally, the *local* reading states that a metainference holds when every valuation satisfying its premises, also satisfies its conclusion. See Dicher and Paoli [13], Barrio et al. [5], Scambler [25], Cobreros et al. [11], and Teijeiro [26]. The issue of the relation between the global, global-substitutional, and local readings – and of which should be used to capture the notion of a metainference – is of particular interest but lies beyond the scope of this paper.

3 ST$^+$-Metainferences and LP$^+$-Inferences

Barrio et al.'s first *collapse result* establishes the following connection between ST$^+$-metainferences and LP$^+$-inferences (cf. Barrio et al. [4, 557]):

$$\Gamma_1 \vdash \Delta_1; \; \dots \; ; \Gamma_n \vdash \Delta_n \implies \Gamma'_1 \vdash \Delta'_1; \; \dots \; ; \Gamma'_k \vdash \Delta'_k \quad \text{holds for } \mathsf{ST}^+$$

just in case

$$\bigwedge \Gamma_1 \supset \bigvee \Delta_1, \dots, \bigwedge \Gamma_n \supset \bigvee \Delta_n \vDash^{\mathsf{LP}^+} \bigwedge \Gamma'_1 \supset \bigvee \Delta'_1, \dots, \bigwedge \Gamma'_k \supset \bigvee \Delta'_k$$

That is, a metainference holds for ST$^+$ exactly when the result of "lowering" arguments to conditionals and substituting the metalinguistic conditional by a consequence relation symbol yields an argument that is LP$^+$-valid.

For readability we will give the proof for the simple case with a single inference on either side of the conditional (see Theorem 1). Our results below do not bring any novelty over Barrio et al. though they are presented in a simpler and more straightforward way. We begin by introducing a definition and a lemma.

Definition [Fixation of a formula relative to an interpretation] Let A be a $\mathcal{L}^+$ formula and $\mathbb{I}$ a three-valued interpretation. The **fixation** of A relative to $\mathbb{I}$, written $f_{\mathbb{I}}(A)$, is defined by induction as follows:

- $f_{\mathbb{I}}(\bot) = \bot$, $f_{\mathbb{I}}(\lambda) = \lambda$, $f_{\mathbb{I}}(\top) = \top$
- if $\mathbb{I}(p) = 1$ then $f_{\mathbb{I}}(p) = \top$
  if $\mathbb{I}(p) = \frac{1}{2}$ then $f_{\mathbb{I}}(p) = \lambda$
  if $\mathbb{I}(p) = 0$ then $f_{\mathbb{I}}(p) = \bot$
- $f_{\mathbb{I}}(\neg A) = \neg f_{\mathbb{I}}(A)$
- $f_{\mathbb{I}}(A \circ B) = f_{\mathbb{I}}(A) \circ f_{\mathbb{I}}(B)$ for every binary connective $\circ$.

Since a fixation is made up of constant symbols, its evaluation is invariant across interpretations, that is $\mathbb{I}'(f_{\mathbb{I}}(A)) = \mathbb{I}(A)$ for every $\mathbb{I}'$, as can be shown by induction. We may therefore write $[\![f_{\mathbb{I}}(A)]\!]$ to talk about the truth-value $\mathbb{I}(A)$ of fixation $f_{\mathbb{I}}(A)$.

Lemma. $\mathbb{I}(A \supset B) > 0$ iff $f_{\mathbb{I}}(A) \vDash^{\mathsf{ST}^+} f_{\mathbb{I}}(B)$

Proof

$$\mathbb{I}(A \supset B) = 0$$
$$\text{iff}$$

$$[\![f_{\mathbb{I}}(A \supset B)]\!] = 0$$
$$\text{iff}$$

$$[\![f_{\mathbb{I}}(A)]\!] = 1 \text{ and } [\![f_{\mathbb{I}}(B)]\!] = 0$$
$$\text{iff}$$

$$f_{\mathbb{I}}(A) \nvDash^{\mathsf{ST}^+} f_{\mathbb{I}}(B) \qquad \qquad \square$$

In words, a conditional is LP^+-true if and only if the fixation of its antecedent ST^+-entails the fixation of its consequent (see Cobreros et al. [10]).[9]

Theorem 1 *The metainference $\Gamma \vdash \Delta \implies \Gamma' \vdash \Delta'$ holds for ST^+*

if and only if

$$\bigwedge \Gamma \supset \bigvee \Delta \models^{\mathsf{LP}^+} \bigwedge \Gamma' \supset \bigvee \Delta'$$

Proof (Right to left)

Suppose $\Gamma \vdash \Delta \implies \Gamma' \vdash \Delta'$ does not hold for ST^+. Then there is some ST^+-instance such that $\Gamma_i \models^{\mathsf{ST}^+} \Delta_i$ and $\Gamma'_k \not\models^{\mathsf{ST}^+} \Delta'_k$.

There is, therefore, some interpretation $\mathbb{I}$ for which $\mathbb{I}(A) = 1$ for all $A \in \Gamma'_k$ and $\mathbb{I}(B) = 0$ for all $B \in \Delta'_k$. Thus, $\mathbb{I}(\bigwedge \Gamma'_k) = 1$ and $\mathbb{I}(\bigvee \Delta'_k) = 0$ and so $\mathbb{I}(\bigwedge \Gamma'_k \supset \bigvee \Delta'_k) = 0$. On the other hand, since $\Gamma_i \models^{\mathsf{ST}^+} \Delta_i$, either $\mathbb{I}(A) < 1$ for some $A \in \Gamma_i$ or $\mathbb{I}(B) = 0$ for some $B \in \Delta_i$; in either case $\mathbb{I}(\bigwedge \Gamma_i \supset \bigvee \Delta_i) > 0$. The interpretation $\mathbb{I}$ then shows that $\bigwedge \Gamma_i \supset \bigvee \Delta_i \not\models^{\mathsf{LP}^+} \bigwedge \Gamma'_k \supset \bigvee \Delta'_k$. Since $\Gamma_i, \Delta_i, \Gamma'_k$ and Δ'_k are substitution instances of Γ, Δ, Γ' and Δ' respectively, we have also that $\bigwedge \Gamma \supset \bigvee \Delta \not\models^{\mathsf{LP}^+} \bigwedge \Gamma' \supset \bigvee \Delta'$ □

Proof (Left to right)

Suppose $\bigwedge \Gamma \supset \bigvee \Delta \not\models^{\mathsf{LP}^+} \bigwedge \Gamma' \supset \bigvee \Delta'$. Then there is an interpretation $\mathbb{I}$ such that $\mathbb{I}(\bigwedge \Gamma \supset \bigvee \Delta) > 0$ and $\mathbb{I}(\bigwedge \Gamma' \supset \bigvee \Delta') = 0$. By our Lemma above, the fixation $f_{\mathbb{I}}$ is such that, $f_{\mathbb{I}}(\bigwedge \Gamma) \models^{\mathsf{ST}^+} f_{\mathbb{I}}(\bigvee \Delta)$ and $f_{\mathbb{I}}(\bigwedge \Gamma') \not\models^{\mathsf{ST}^+} f_{\mathbb{I}}(\bigvee \Delta')$. Note finally that the fixation is a uniform substitution of propositional letters in Γ, Δ, Γ' and Δ' so that the latter is an ST^+-instance showing that the metainference $\Gamma \vdash \Delta \implies \Gamma' \vdash \Delta'$ does not hold for ST^+. □
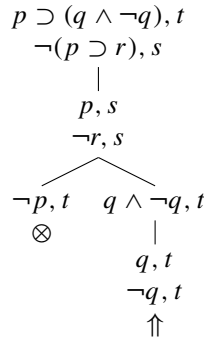
The connection gives us a decision procedure to check whether a metainference is ST^+ valid out of a decision procedure to test whether an inference is LP^+-valid. We use here the trees in Cobreros et al. [8]. Basically, a branch closes when a formula and its negation both appear on it with the labels s and t (or s and s) respectively (crucially, a branch does not close when a formula and its negation appear with labels t and t).

Example 1 $p \vdash q \land \neg q \implies p \vdash r$ does not hold in ST^+

---

[9]A similar reasoning establishes that $\mathbb{I}(A \supset B) = 1$ iff $f_{\mathbb{I}}(A) \models^{\mathsf{TS}^+} f_{\mathbb{I}}(B)$, that is, a conditional is $\mathsf{K3}^+$-true if and only if the fixation of its antecedent TS^+-entails the fixation of its consequent. This way, the relations between ST^+, LP^+ and $\mathsf{K3}^+$ below can be straightforwardly extended to TS^+. This allows us to verify an observation in French [14, 120] that although the logic TS^+ is empty at the level of inferences, it is not empty at the level of metainferences.

Proof

$$p \supset (q \wedge \neg q), t$$
$$\neg(p \supset r), s$$
$$|$$
$$p, s$$
$$\neg r, s$$

$$\neg p, t \qquad q \wedge \neg q, t$$
$$\otimes \qquad\qquad |$$
$$q, t$$
$$\neg q, t$$
$$\Uparrow$$

$\square$

The countermodel for the LP^+-inference is $\mathbb{I}(p) = 1$, $\mathbb{I}(q) = \frac{1}{2}$ and $\mathbb{I}(r) = 0$. Thus, the fixation showing that the metainference does not hold is:

$$\top \models^{\mathsf{ST}^+} \lambda \wedge \neg \lambda \quad \text{BUT} \quad \top \not\models^{\mathsf{ST}^+} \bot$$

4 ST^+-Meta-Anti-Inferences and $\mathsf{K3}^+$-Inferences

The key ingredient in the connection above between LP^+- inferences and ST^+-metainferences is the observation in the Lemma that a conditional is LP^+-true in an interpretation just in case the fixation of its antecedent attending to that interpretation ST^+-entails the fixation of its consequent attending to the interpretation:

$$\mathbb{I}(A \supset B) > 0 \quad \text{iff} \quad f_{\mathbb{I}}(A) \models^{\mathsf{ST}^+} f_{\mathbb{I}}(B).$$

This statement is equivalent to the statement that a conditional is LP^+-untrue in an interpretation just in case the fixation of its antecedent attending to that interpretation does not ST^+-entail the fixation of its consequent attending to the interpretation:

$$\mathbb{I}(A \supset B) = 0 \quad \text{iff} \quad f_{\mathbb{I}}(A) \not\models^{\mathsf{ST}^+} f_{\mathbb{I}}(B).$$

which in turn, informally reads: a conditional is $\mathsf{K3}^+$-*false* in an interpretation (i.e., its negation is $\mathsf{K3}^+$-true) just in case the fixation of its antecedent attending to that interpretation does not ST^+-entail the fixation of its consequent attending to the interpretation. So there is a connection between a conditional being $\mathsf{K3}^+$-false and an argument being ST^+-invalid.

A meta-anti-inference is a conditional statement relating kinds of anti-inferences to kinds of anti-inferences, that is, a meta-anti-inference is a statement of the following form (simple case):

$$\Gamma \not\vdash \Delta \implies \Gamma' \not\vdash \Delta'$$

Where the Γ's and Δ's are sets of formulae of $\mathcal{L}^+$.

The following corollary follows from Theorem 1 with the observation that $\mathsf{K3}$ and LP are duals.

Corollary. The meta-anti-inference $\Gamma \nvDash \Delta \implies \Gamma' \nvDash \Delta'$ holds for ST^+
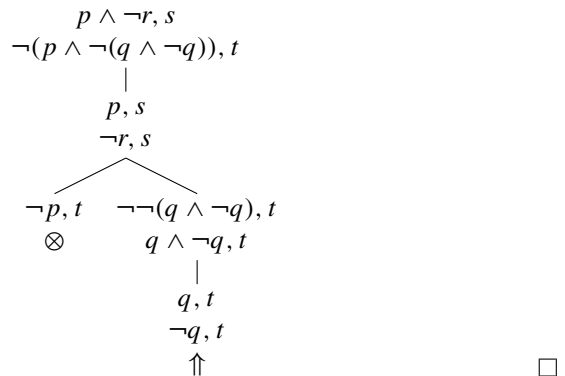
$$\text{if and only}$$

$$\bigwedge \Gamma \wedge \neg \bigvee \Delta \vDash^{\mathsf{K3}^+} \bigwedge \Gamma' \wedge \neg \bigvee \Delta'.$$

Proof Suppose $\Gamma \nvDash \Delta \implies \Gamma' \nvDash \Delta'$ holds for ST^+. Then, contrapositively, $\Gamma' \vdash \Delta' \implies \Gamma \vdash \Delta$ holds for ST^+. By Theorem 1, $\bigwedge \Gamma' \supset \bigvee \Delta' \vDash^{\mathsf{LP}^+} \bigwedge \Gamma \supset \bigvee \Delta$. By duality, $\neg(\bigwedge \Gamma \supset \bigvee \Delta) \vDash^{\mathsf{K3}^+} \neg(\bigwedge \Gamma' \supset \bigvee \Delta')$, which is $\bigwedge \Gamma \wedge \neg \bigvee \Delta \vDash^{\mathsf{K3}^+} \bigwedge \Gamma' \wedge \neg \bigvee \Delta'$. □

As in the previous case, a procedure to decide for $\mathsf{K3}^+$-validity can be used as a procedure to decide whether a given meta-anti-inference holds for ST^+.

Example 2 $p \nvDash r \implies p \nvDash q \wedge \neg q$ does not hold in ST^+

Proof

$$p \wedge \neg r, s$$
$$\neg(p \wedge \neg(q \wedge \neg q)), t$$
$$|$$
$$p, s$$
$$\neg r, s$$

$$\neg p, t \qquad \neg\neg(q \wedge \neg q), t$$
$$\otimes \qquad q \wedge \neg q, t$$
$$|$$
$$q, t$$
$$\neg q, t$$
$$\Uparrow$$ □

The countermodel for the $\mathsf{K3}^+$-inference is $\mathbb{I}(p) = 1$, $\mathbb{I}(q) = \frac{1}{2}$ and $\mathbb{I}(r) = 0$. Thus, the fixation showing that the meta-anti-inference does not hold is:

$$\top \nvDash^{\mathsf{ST}^+} \bot \quad \text{BUT} \quad \top \vDash^{\mathsf{ST}^+} \lambda \wedge \neg \lambda.$$

Our corollary from Theorem 1 shows that ST^+ is connected to $\mathsf{K3}^+$ in much the same way in which it is connected to LP^+; at least under the assumption that metainferences and meta-anti-inferences are important alike. This will be discussed in Section 7.

5 Hybrid Metainferences

We saw that there is a connection between an inference being ST^+-valid and a conditional being LP^+-true, and an analogous connection between an inference being ST^+-invalid and a conditional being $\mathsf{K3}^+$-false. From these connections we can now decide more articulated questions about ST^+'s metainferences. A *hybrid*

metainference is a conditional whose antecedent and consequent might contain inferences and anti-inferences alike, for example:

$$\text{(HMI) } \Gamma \vdash \Delta; \; \Gamma' \nvdash \Delta' \implies \Gamma^\dagger \nvdash \Delta^\dagger; \; \Gamma^* \vdash \Delta^*.$$

We want to know whether there is a uniform substitution of the propositional letters appearing in the hybrid metainference such that: the inference in the premise is ST^+-valid, the anti-inference in the premise is ST^+-invalid, the anti-inference in the conclusion is ST^+-valid, and the inference in the conclusion ST^+- invalid (if any such substitution exists, the metainference does not hold for ST^+). The previous connections between ST^+-inferences and the logics $K3^+$ and LP^+ allow us to extract the corresponding initial list for the tree:[10]
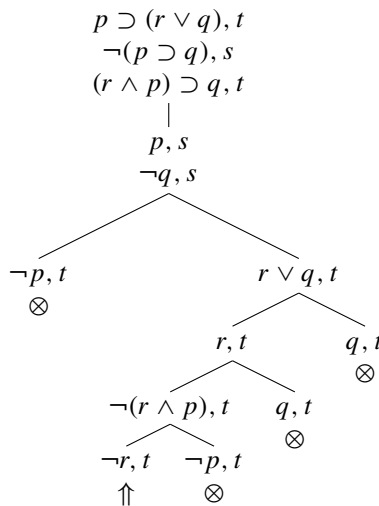
Recall that the trees provide a systematic search for a counterexample, and our initial list embodies the assumption of the existence of such a counterexample. Observe also that if a conditional in the initial list appears negated it goes with an s and if unnegated it goes with a t. This gives us a general pattern to extract the initial list for the tree to test for metainferences in ST^+ (see Fig. 3).

In the particular case of metainferences involving single-premise single-conclusion arguments, we have:

- if $A \vdash B$ appears in premises write: $A \supset B, t$
- if $A \nvdash B$ appears in premises write $\neg(A \supset B), s,$
- if $A \vdash B$ appears in conclusions write $\neg(A \supset B), s,$
- if $A \nvdash B$ appears in conclusions write $A \supset B, t.$

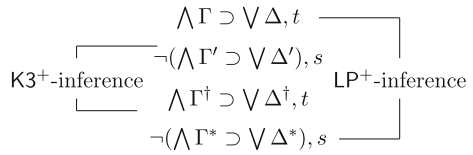*Example 3* $p \vdash r, q \; ; \; p \nvdash q \implies r, p \nvdash q$ does not hold for ST^+.

Proof

$$
\begin{array}{c}
p \supset (r \vee q), t \\
\neg(p \supset q), s \\
(r \wedge p) \supset q, t \\
| \\
p, s \\
\neg q, s
\end{array}
$$

□

---

[10] The same procedure can be used to decide about TS^+-metainferences by relabelling the nodes of the tree.

Fig. 3 Initial list for a hybrid metainference

$$\bigwedge \Gamma \supset \bigvee \Delta, t$$
$$\neg(\bigwedge \Gamma' \supset \bigvee \Delta'), s$$
$$\text{K3}^+\text{-inference} \qquad \bigwedge \Gamma^\dagger \supset \bigvee \Delta^\dagger, t \qquad \text{LP}^+\text{-inference}$$
$$\neg(\bigwedge \Gamma^* \supset \bigvee \Delta^*), s$$

The falsifying instance to show that this metainference does not hold for ST^+ is $p = \top$, $q = \bot$ and $r = \lambda$, which renders the well-known counterexample to transitivity in ST^+.

This raises the following interesting question:[11] what is the logic corresponding to hybrid metainferences of ST^+? More precisely: we have that a meta-inference holds for ST^+ just in case the "lowering" renders a valid LP^+ inference. In particular:

The metainference $A \vdash B \Longrightarrow C \vdash D$ holds for ST^+
iff
The inference $A \supset B \vDash C \supset D$ is LP^+-valid.

We also have that a meta-(anti)-inference holds for ST^+ just in case the "lowering" renders a valid K3^+-inference:

The metainference $A \nvdash B \Longrightarrow C \nvdash D$ holds for ST^+
iff
The inference $\neg(A \supset B) \vDash \neg(C \supset D)$ is K3^+-valid.

So, the question goes, what happens when the metainference is hybrid:

The metainference $A \vdash B, C \nvdash D \Longrightarrow E \nvdash F, G \vdash H$ holds for ST^+
iff
The inference $A \supset B, \neg(C \supset D) \vDash \neg(E \supset F), G \supset H$ is X-valid.

We don't know the answer to this question. The important point here is that this logic X cannot be simply either LP^+ or K3^+. Since contraposition holds for the metainference conditional, we can always translate a hybrid metainference into a non-hybrid metainference. But in that case, we have some freedom on how to do this: we can translate into metainferences or into meta(anti)inferences. Thus, metainferences in ST^+ relate both to LP^+ and K3^+. We return to this discussion in Section 7. In the next section, we first consider another issue raised by Barrio et al. [4] about the connection between ST^+ and LP^+.

6 Internal vs External Logic

According to Barrio, Tajer and Rosenblatt, the logic LP^+ corresponds to ST^+'s *external* logic. As a result, ST^+'s external logic would fail to coincide with its internal

---

[11] Thanks to an anonymous reviewer of the JPL for the suggestion.

logic, a situation they consider problematic. This claim is contentious and calls for clarification. In what follows, we critically examine the notions of internal and external logic, and argue that although Barrio et al.'s claim can be vindicated to some extent, the situation is not as straightforward as they claim.

6.1 Three-Sided Sequents

In order to present Barrio et al.'s argument, we need to say more about the proof theory of ST^+ and LP^+. The logics ST^+ and LP^+ can be presented in proof-theoretic terms via three-sided sequents of type $\Gamma|\Sigma|\Delta$. Semantically, a sequent $\Gamma|\Sigma|\Delta$ is satisfied by an interpretation $\mathbb{I}$ just in case either some element of Γ takes value 0, or some in Σ takes value $\frac{1}{2}$, or some in Δ value 1, relative to $\mathbb{I}$. Barrio et al. describe a proof system $\mathcal{S}$ for these three-sided sequents [4, 562]. We will not present $\mathcal{S}$ here, since the details of its presentation don't matter for what follows. In order to foster intuitions about $\mathcal{S}$, suffice it to note that we can understand a range of two-sided consequence relations via these three-sided sequents. In particular:

Definition 3 (ST^+ proof-theoretic validity) An argument with premises Γ and conclusions Δ is proof-theoretically valid in ST^+, written $\Gamma \vdash^{ST^+} \Delta$, just in case the sequent $\Gamma|\Gamma, \Delta|\Delta$ can be proved in $\mathcal{S}$.

Definition 4 (LP^+ proof-theoretic validity) An argument with premises Γ and conclusions Δ is proof-theoretically valid in LP^+, written $\Gamma \vdash^{LP^+} \Delta$, just in case the sequent $\Gamma|\Delta|\Delta$ can be proved in $\mathcal{S}$.

Thinking these through in terms of interpretations: the sequent $\Gamma|\Gamma, \Delta|\Delta$ is satisfied by an interpretation $\mathbb{I}$ means that either some element in Γ takes value less than 1 or some element in Δ takes value more than 0 on $\mathbb{I}$, in agreement with the definition of ST-validity. Similarly, that the sequent $\Gamma|\Delta|\Delta$ is satisfied by $\mathbb{I}$ means that either some element in Γ takes value less than $\frac{1}{2}$ or some element in Δ takes value more than 0 on $\mathbb{I}$, in agreement with the definition of LP-validity.

The system $\mathcal{S}$ is sound and complete for three-sided sequents, meaning that a sequent is provable in $\mathcal{S}$ iff it is satisfied by every model. From that it follows that $\vdash^{ST^+}$ and $\vdash^{LP^+}$ are sound and complete for ST^+ and LP^+, respectively.

6.2 Internal vs External

It is sometimes useful to distinguish the *internal logic* of a sequent system from its *external logic*. The distinction between internal and external appears in these terms in Avron [2], where they are referred to as 'two methods that are traditionally used for associating a consequence relation with a Gentzen-type formalism'. For two-sided systems, the distinction is clear and captured by the following definitions [4, 564]:

Definition 5 (Internal logic) An argument from premises Γ to conclusions Δ is valid in the internal logic of a sequent system iff the sequent $\Gamma \Rightarrow \Delta$ is derivable in that system.

Definition 6 (External logic) An argument from premises Γ to conclusions Δ is valid in the external logic of a sequent system iff the sequent $\Rightarrow \Delta$ is derivable in the *new* sequent system arrived at by adding initial sequents $\Rightarrow \gamma$ for every $\gamma \in \Gamma$ to the original system.

Although Barrio, Tajer and Rosenblatt make reference to Avron [2] and Mares and Paoli [19] in order to motivate the notion of external logic, their definition of external logic differs in an important way. For Avron and for Mares and Paoli, the definition of external logic involves adding not just new initial sequents, but also adding (if it is not already present) the rule of Cut as a primitive (cf. Avron [2, 163] and Mares and Paoli [19, 451]). Barrio, Tajer and Rosenblatt cannot take on this extra assumption on pain of begging the question against the defender of ST^+, and so they do not do so. In what follows we stick to Barrio, Tajer and Rosenblatt's definition.

The internal logic of a system considers the premises of an argument as living on the *left* side of a sequent. By contrast, the external logic of a system considers premises and conclusions alike as living on the *right* side, with the difference between premises and conclusions being that the premises live at the *top* end of a sequent derivation. Roughly, internal logic is about *following from*, whereas external logic is about *preservation of theoremhood*. Because of this, the internal logic of a system is often useful for connecting the system to natural deduction proof systems, and the external logic for connecting to axiomatic proof systems.[12]

For three-sided systems like $\mathcal{S}$, it is not immediately obvious how to extend these notions, as there are multiple ways to understand 'left' and 'right'. For example, the two consequence relations $\vdash^{ST^+}$ and $\vdash^{LP^+}$ defined above from $\mathcal{S}$ are both recognizably *internal*: they consider only sequents derivable in $\mathcal{S}$ itself, not some expansion of $\mathcal{S}$; and they understand premises and conclusions as occupying different locations in the sequent, with premises to the 'left' (for two different understandings of 'left') and conclusions to the 'right' (for a particular understanding of 'right').

So how should we understand *external* consequence for $\mathcal{S}$? We should consider expansions of $\mathcal{S}$ with new initial sequents, including the premises of our argument on the 'right', and ask after the derivability in this expanded system of a sequent with our conclusions on the 'right'. But we have two workable notions of 'right' available: either the rightmost position alone, or spread across the middle and rightmost positions. (Compare the two understandings of 'left' implicit in the definitions of $\vdash^{ST^+}$ and $\vdash^{LP^+}$.)

So just as the notion of internal consequence breaks apart into multiple nonequivalent notions for three-sided systems like $\mathcal{S}$, so too does the notion of external consequence. And just as the multiple internal consequences allow us to determine both ST^+ and LP^+ via the single sequent system $\mathcal{S}$, the multiple external consequences also allow us to determine multiple logics.

---

[12]This is how they are used, for example, in Avron [2], as well as in Avron [3], where they are called the "truth" consequence relation and "validity" consequence relation. See Mares and Paoli [19] for an application of this distinction to the case of paradoxes, and Dicher [12] for an application of the distinction to intra-theoretical logical pluralism.

Only two versions of external consequence for $\mathcal{S}$ will concern us here:

Definition 7 (*tt*-external validity) An argument with premises Γ and conclusions Δ is *tt*-externally valid, written $\Gamma \vdash_e^{tt} \Delta$, just in case the sequent $\emptyset|\Delta|\Delta$ can be proved in the system that adds initial sequents $\emptyset|\gamma|\gamma$ for each γ in Γ to $\mathcal{S}$.

Definition 8 (*st*-external validity) An argument with premises Γ and conclusions Δ is *st*-externally valid, written $\Gamma \vdash_e^{st} \Delta$, just in case the sequent $\emptyset|\Delta|\Delta$ can be proved in the system that adds initial sequents $\emptyset|\emptyset|\gamma$ for each γ in Γ to $\mathcal{S}$.

As Barrio et al. [4] show, $\vdash_e^{tt}$ is LP^+ and $\vdash_e^{st}$ is ST^+. Barrio et al., however, also speak of 'ST^+'s external logic', casting $\vdash_e^{tt}$ in this role. They then argue that it should be a desideratum for a theory of truth that its internal logic be contained in its external logic.[13] If $\vdash_e^{tt}$ is really ST^+ 's external logic, then ST^+ fails this desideratum, since $\vdash_e^{tt}$ is not ST^+ but LP^+.

6.3 Against the Meaningfulness of "ST^+ 's External Logic"

We think an important point can be recovered from this argument, but to do this, it is important first to note that the notion of "ST^+'s external logic" makes no sense on its face. It is only for particular proof systems that the notion of an external logic is defined. There is no such thing, for example, as "the external logic of classical logic"; there are only external logics of particular sequent systems for classical logic, and these can easily differ from each other, even between systems all of whose internal logic is classical logic.

For concreteness, take Gentzen's calculus LK [15], and let LK$^-$ be the corresponding system without the rule of Cut. As Gentzen showed, LK and LK$^-$ have the same internal logic: they derive the same sequents, exactly the first-order classically-valid arguments. So for example, the sequent $p \wedge q \Rightarrow p$ is valid in the internal logic of both LK and LK$^-$. It is also valid in the external logic of LK. But it is *not* valid in the external logic of LK$^-$, for the simple reason that LK$^-$ has the subformula property: in any LK$^-$ derivation, every formula that appears at all is a subformula of some formula in the conclusion-sequent. This remains true of any extension of LK$^-$ by extra initial sequents. But $p \wedge q$ is not a subformula of p, and so when we add $\Rightarrow p \wedge q$ as an initial sequent to LK$^-$, we know that this initial sequent cannot occur in any derivation of $\Rightarrow p$. That is, $\Rightarrow p$ is derivable in this extended system iff it is derivable in LK$^-$ already, and of course it is not. So p does not follow from $p \wedge q$ according to the external consequence relation of LK$^-$. It would not make any sense, then, to speak of 'classical logic's external consequence relation', since both LK and

---

[13]This is reminiscent of desideratum (g) in Leitgeb [18], that '[t]he outer logic and the inner logic should coincide', but the similarity is a coincidence. Leitgeb's 'outer logic' and 'inner logic' have to do with reasoning 'inside' or 'outside' the truth predicate, and have nothing to do with proof systems.

LK^- are paradigm sequent systems for classical logic, but have strikingly different external consequence relations.[14]

In the present paper as in Barrio et al. [4], however, ST^+ is not a particular proof system, but rather a consequence relation determined by three-valued interpretations. So there is no good sense to the claim that $\vdash_e^{tt}$—or any other consequence relation—is "ST^+'s external logic". Consequence relations like ST^+ do not have external logics; only proof systems do.

This should lead us to be suspicious of Barrio et al. [4]'s proposed desideratum: that the internal logic should be contained in the external.[15] It can only be understood as a desideratum for proof systems. But as a desideratum for proof systems, it is not particularly plausible; as we've already seen, such elegant and well-motivated proof systems as LK^- fail this desideratum pretty spectacularly.

6.4 Making Sense of Barrio et al.'s Desideratum

Perhaps we can understand the desideratum as a desideratum on consequence relations, however, by sticking a quantifier in, by desiring of a consequence relation C that there be *some* proof system $\mathcal{P}$ with C as its internal logic that obeys the original desideratum. Classical logic, for example, would meet this new desideratum, since although LK^-'s internal logic is not contained in its external logic, LK's is.

And indeed ST^+ fails this new desideratum, at least if we restrict our attention to usual two-sided systems. The reason is that any proof system $\mathcal{P}$ whose internal logic contains ST^+ will have to prove $\lambda \Rightarrow p$, since this is valid in ST^+. Supposing this argument comes out externally valid for $\mathcal{P}$, there must be a derivation of $\Rightarrow p$ in the system that adds the initial sequent $\Rightarrow \lambda$ to $\mathcal{P}$. But $\Rightarrow \lambda$ must have *already* been derivable in $\mathcal{P}$, since it is valid in ST^+. So there must be a derivation of $\Rightarrow p$ in $\mathcal{P}$ itself, and so the internal logic of $\mathcal{P}$ is not sound for ST^+.

So understood, then, the desideratum makes good sense, and is satisfied by classical logic (and many others), but not by ST^+. We thus offer this as a way of making sense of Barrio et al. [4]'s comments about ST^+'s external logic: that *there is no* proof system for ST^+ whose internal logic is contained in its external logic.

So suppose that indeed ST^+ does not meet (our interpretation of) Barrio et al.'s desideratum, that there is no proof system of any sort for ST^+ whose internal logic is contained in its external logic. (And so suppose, in particular, that $\mathcal{S}$ doesn't count as such a system, for whatever reason.)

---

[14]The difficulties that attend Cut elimination in the presence of axioms (for which see e.g. Buss [7, §2.4.4–2.4.7]) are manifestations of this phenomenon: the external consequence relation of a system with Cut is often much stronger than the external consequence relation of the corresponding system without Cut.

[15]"It seems prima facie desirable that anything claimed in the internal logic, should also be claimed in the "external" logic". They intend this claim double-barrelledly: as about both metainferences and proof-theoretic external consequence. Here, we respond only to the proof-theoretic barrel, having already discussed the metainferential barrel. They also talk of the stronger desideratum that internal and external logics coincide. However, the only direction that matters for their discussion or ours is the one we focus on.

Why would this be a problem? Barrio et al. [4] suggest that it requires us to endorse logical pluralism. We take no stand here on the debate around logical pluralism (see eg [6, 16, 24]). But whether or not logical pluralism is correct, it would certainly be awkward if we were forced, merely by our advocacy of ST^+, also to advocate logical pluralism. It should at least be *possible* to be a logical monist about ST^+. So if Barrio et al. were correct that ST^+'s failing their desideratum blocked logical monism about ST^+, that would indeed be a problem for us. Here is what they say [4, 567]:

> [S]upporting a logic in which the external and the internal points of view do not coincide seems to be at least inconvenient, because the question 'which logic do you support?' has now an ambiguous answer. Moreover, it brings up the issue of the universality of logic: the question 'which logic does it right?' cannot have only one right response anymore.

We don't think this would follow at all. Whatever endorsing a logic amounts to, endorsing a logic does not require at the same time endorsing the external logic of proof systems related to that logic. Those who advocate classical logic do not thereby become pluralists implicitly endorsing the strikingly weak external logic of LK^-; the connection to this external logic is too indirect for endorsement to spread like this. And similarly, those who advocate ST^+ do not thereby become pluralists implicitly endorsing $\vdash_e^{tt}$, or any other external logic of a proof system for ST^+.

Of course, someone who endorsed classical logic *could* be a pluralist, also endorsing the external logic of LK^-. And someone who endorsed ST^+, similarly, *could* be a pluralist, also endorsing $\vdash_e^{tt}$ (which, recall, is LP^+). But there is nothing in endorsing classical logic, or in endorsing ST^+, that would suggest these positions. The existence of external logics is not something that should disturb a logical monist.

Now, classical logic has a proof system whose internal and external logics coincide; we are supposing ST^+ does not. But this doesn't matter at all for what an endorser of ST^+ is committed to. Endorsement simply doesn't spread to external logics. So, if someone were a monist about ST^+, they would have an unambiguous answer to the questions Barrio et al. pose above: 'ST^+ does it right'.

7 LP^+ in Sheep's Clothing?

Barrio, Rosenblatt and Tajer's result provides an insightful response to the question of which metainferences hold for the logic ST^+. Their paper provides a sharp and elegant characterization of metainferential validity and their main result allows us to translate the question of whether some metainference holds in ST^+ into a question about whether some related inference holds in LP^+.[16] In addition to this, Barrio, Rosenblatt and Tajer take the results in their paper to argue for two claims (these claims appear intertwined in the paper):

---

[16]In their forthcoming JPL paper [5] Barrio, Pailos and Szmuc show even more: that the features of a logic need not be written in its valid metainferences of level n ($n \in \omega$). This important result lies beyond the scope of this paper. See Scambler [25] for a discussion.

Claim 1: That, contrary to what it is claimed in Cobreros et al. [9], ST$^+$ is not classical logic.

Claim 2: That ST$^+$ is no more illuminating than LP$^+$ as a solution to paradoxes (put in Quinean slogan: ST$^+$ is LP$^+$ in sheep's clothing).

In order to sustain these claims, Barrio et al. take two routes. The first, from a semantic perspective, showing a link between ST$^+$'s metainferences and LP$^+$'s inferences. The second, from a proof-theoretic perspective, arguing that ST$^+$'s external logic leads to LP$^+$.

Note that Claim 2 is stronger than Claim 1, for LP$^+$ is the paradigm of a non-classical theory. Thus, evidence for the second claim is evidence for the first. It might happen, as we will comment in a moment, that no evidence is able to establish anything like Claim 1. In this discussion we therefore concentrate on a rebuttal of Claim 2.

The expression 'classical logic' seems to express an unambiguous and precise concept when restricted to a language containing only the usual connectives. When we consider a language equipped with a truth predicate, things become murky. Is it right to claim that ST$^+$ *is* classical logic for such a language?

Claim 1 is clearly true if taken very precisely: classical logic says nothing about the truth predicate, while ST$^+$ does, so they cannot be the same. As to the more suggestive (but less precise) question whether ST$^+$ is or isn't a 'classical theory of truth', this turns on exactly which features of classical logic must be exhibited by a theory of truth for that theory to count as 'classical'. If validating every instance of every classically-valid inference is enough, ST$^+$ is classical. If being closed under every instance of every metainference that holds of classical logic is required, though, ST$^+$ is not classical, since instances of Cut are such metainferences. In their paper, Barrio, Rosenblatt and Tajer focus on metainferences to make a case for ST$^+$'s nonclassicality in this way.

This is where Claim 2 ("ST$^+$ is LP$^+$ in sheep's clothing") enters the stage. If ST$^+$'s metainferences can be characterised as inferences of LP$^+$, then ST$^+$'s metainferences have indeed a non-classical flavour. While we don't want to deny that there is something non-classical about ST$^+$ (this is, indeed, much of its attractiveness) we believe Barrio et al. go too far in claiming that ST$^+$ is no more illuminating than LP$^+$ itself. For ST$^+$ strikes a very specific balance between a paracomplete and a paraconsistent approach, one that is missing in LP$^+$.

In connection to ST$^+$'s external logic (second route towards Claim 2) our response is that external consequence is a feature of proof systems, not consequence relations. We cannot talk about "the" external logic of a consequence relation C. So we see no good reason why the commitment to a consequence relation C should extend to some (perhaps various) external consequences associated to proof systems having C as its internal consequence.

In connection to ST$^+$'s metainferences (first route towards Claim 2) our response draws on the connection between ST$^+$'s meta(anti)inferences and K3$^+$.

A conditional $A \supset B$ can be read in two directions: as expressing that A is a sufficient condition for B (forwards) or as expressing that B is a necessary condition for

A (backwards). The backwards reading can be explicitly represented by contraposing the conditional: $\neg B \supset \neg A$. As with any other conditional, metainferences can also be read in two directions. In the case of metainferences the forwards reading tells us: if such and such inferences are valid, then so are some of these inferences. The backwards reading tells us: if such and such inferences are not valid, then some of these are not valid either.

Although both readings are equivalent, they do not put the same emphasis. Consider the following metainference which is a close relative to the Cut rule:

$$p, q \vdash r \; ; \; p, \neg q \vdash r \implies p \vdash r$$

In its forwards reading it expresses the idea that if we can prove *r* from *p* and *q* and we can prove *r* from *p* and $\neg q$ then we can as well prove *r* from *p* alone. In its backwards reading it says:

$$p \nvdash r \implies p, q \nvdash r \; ; \; p, \neg q \nvdash r$$

which points to the idea that if *p* is consistent with $\neg r$, then either *p* and *q* is consistent with $\neg r$, or *p* and $\neg q$ is. The property expressed by either reading is the same by ST^+'s lights (since ST^+ is self-dual and the metalinguistic conditional is contraposable) but the emphasis is different. Similarly in the case of ST^+, metainferences relate to LP^+, whereas meta-anti-inferences to $\mathsf{K3}^+$. The point is that ST^+ is no more tightly related to LP^+ than to $\mathsf{K3}^+$. Whatever logical creature ST^+ might be in sheep's clothing, that creature is Janus-faced.

References

1. Arieli, O., Avron, A., Zamansky, A. (2011). Maximal and premaximal paraconsistency in the framework of three-valued semantics. *Studia Logica*, *97*(1), 31–60.
2. Avron, A. (1988). The semantics and proof theory of linear logic. *Theoretical Computer Science*, *57*(2–3), 161–184.
3. Avron, A. (1991). Simple consequence relations. *Information and Computation*, *92*(1), 105–139.
4. Barrio, E., Rosenblatt, L., Tajer, D. (2015). The logics of strict-tolerant logic. *Journal of Philosophical Logic*, *44*(5), 551–571.
5. Barrio, E.A., Pailos, F., Szmuc, D. (2020). A hierarchy of classical and paraconsistent logics. *Journal of Philosophical Logic*, *49*, 93–120.
6. Beall, J., & Restall, G. (2006). *Logical Pluralism*. Oxford: Oxford University Press.
7. Buss, S.R. (1998). Introduction to proof theory. In Buss, S.R. (Ed.) *Handbook of proof theory* (pp. 1–78). Amsterdam: Elsevier.
8. Cobreros, P., Egré, P., Ripley, D., van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic*, *41*(2), 347–385.

9. Cobreros, P., Egré, P., Ripley, D., van Rooij, R. (2013). Reaching transparent truth. *Mind*, *122*(488), 841–866.

10. Cobreros, P., Egré, P., Ripley, D., van Rooij, R. (2015). Vagueness, truth and permissive consequence. In Achouriotti, D., Galinon, H., Martinez, J. (Eds.) *Unifying the philosophy of truth* (pp. 409–430): Springer.

11. Cobreros, P., Tranchini, L., La Rosa, E. (2019). (I can't get no) antisatisfaction. *Synthese*. https://doi.org/10.1007/s11229-020-02570-x.

12. Dicher, B. (2018). Variations on intra-theoretical logical pluralism: internal versus external consequence. *Philosophical Studies*, 1–20.

13. Dicher, B., & Paoli, F. (2019). ST, LP, and tolerant metainferences. In Baskent, C., & Ferguson, T.M. (Eds.) *Graham Priest on dialetheism and paraconsistency*. Dordrecht: Springer. Forthcoming.

14. French, R. (2016). Structural reflexivity and the paradoxes of self-reference. *Ergo, an Open Access Journal of Philosophy*, 3.

15. Gentzen, G. (1969). Investigations into logical deduction. In Szabo, M.E. (Ed.) *The collected papers of Gerhard Gentzen* (pp. 68–131). Amsterdam: North-Holland Publishing Company.

16. Kouri Kissel, T. (2018). Logical pluralism from a pragmatic perspective. *Australasian Journal of Philosophy*, *96*(3), 578–591.

17. Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, *72*(19), 690–716.

18. Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophy Compass*, *2*(2), 276–290.

19. Mares, E., & Paoli, F. (2014). Logical consequence and the paradoxes. *Journal of Philosophical Logic*, *43*(2-3), 439–469.

20. Nicolai, C., & Rossi, L. (2016). Principles for object-linguistic consequence: from logical to irreflexive. *Journal of Philosophical Logic*, 1–29.

21. Přenosil, A. (2017). Cut elimination, identity elimination, and interpolation in super-Belnap logics. *Studia Logica*, *105*(6), 1255–1289.

22. Priest, G. (1979). Logic of paradox. *Journal of Philosophical Logic*, *8*, 219–241.

23. Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *The Review of Symbolic Logic*, *5*(02), 354–378.

24. Russell, G. (2019). Logical pluralism. In Zalta, E.N. (Ed.) *The Stanford Encyclopedia of Philosophy*. *Metaphysics research lab, Stanford University, summer 2019 edition*.

25. Scambler, C. (2019). Classical logic and the strict tolerant hierarchy. *Journal of Philosophical Logic*, https://doi.org/10.1007/s10992-019-09520-0.

26. Teijeiro, P. (2019). Strength and stability. *Análisis Filosófico*. To appear.

27. Tye, M. (1994). Sorites paradoxes and the semantics of vagueness. *Philosophical Perspectives*, 189–206.