



Expressing logical disagreement from within

Andreas Fjellstad¹

Received: 13 August 2021 / Accepted: 20 March 2022 / Published online: 10 April 2022
© The Author(s) 2022

Abstract

Against the backdrop of the frequent comparison of theories of truth in the literature on semantic paradoxes with regard to which inferences and metainferences are deemed valid, this paper develops a novel approach to defining a binary predicate for representing the valid inferences and metainferences of a theory within the theory itself under the assumption that the theory is defined with a classical meta-theory. The aim with the approach is to obtain a tool which facilitates the comparison between a theory and its competitors within the theory itself, thereby expressing the disagreement between the theories within the theories. After discussing what we can and should require of an object-linguistic representation of a theory for that purpose, this paper proposes to restrict the representation of valid metainferences to locally valid metainferences, a requirement which turns out to be ω -consistent and conservative over classical first-order arithmetic. This approach is then applied to four theories definable on strong Kleene models using a labelled nested sequent calculus.

Keywords Validity predicate · Validity curry · Curry's paradox · Logical disagreement · Non-classical logics · Metainferences · Local metainferential validity · Global metainferential validity · Deflationism · Substructural logics · ω -consistency · Labelled sequent calculus

1 Logical disagreement and comparing alternatives

It seems typical to portray the various formal theories of truth presented to resolve the liar paradox as competing proposals for which inferential and metainferential schemas are valid for the relevant vocabulary. Moreover, it is not uncommon that an advocate of

The research for this paper was supported by the Research Council of Norway under Grant Number 262837/F10 (co-funded by the European Union's Seventh Framework Programme for research, technological development and demonstration under Marie Curie Grant Agreement No 608695).

✉ Andreas Fjellstad
afjellstad@gmail.com

¹ Department of Philosophy, University of Bergen, Postboks 7805, 5020 Bergen, Norway

one formal theory of truth compares that formal theory of truth with other candidates in terms of whether every instance of an inferential or metainferential schema is deemed valid by it as if validating some particular pattern of inference should be considered a virtue.¹ In that sense, the theories disagree about what is valid and the aim of the comparison is to highlight this disagreement. One pertinent example is the case of modus ponens which may hold or fail either as inferential or metainferential schema.² With regard to modus ponens as inferential schema, a theory T could be such that

- for every formula A and B : $A, A \supset B \vDash_T B$, or
- there are formulas A and B such that: $A, A \supset B \not\vDash_T B$

Two theories T_1 and T_2 can thus be said to disagree over modus ponens as inferential schema just in case the former is true of T_1 and the latter is true of T_2 .³ With regard to modus ponens as metainferential schema, a theory T could be such that:

- for every formula A and B , if $\vDash_T A$ and $\vDash_T A \supset B$ then $\vDash_T B$, or
- there are formulas A and B such that $\vDash_T A$ and $\vDash_T A \supset B$ but $\not\vDash_T B$.

A natural question about such universally and existentially generalised claims is the extent to which a comparison with regard to such claims between T and its competitors can be formalised within T using binary predicates representing what is valid according to T and its competitors. This paper is concerned with a question very much in the vicinity of that question. Suppose that T_1, \dots, T_n are theories based on a common language \mathcal{L} that disagree with regard to some inferential or metainferential schema and which are defined with a classical meta-theory. To which extent can we extend \mathcal{L} with the binary predicates P_1, \dots, P_n and define the theories $T'_1 \supseteq T_1, \dots, T'_n \supseteq T_n$ based on the obtained language \mathcal{L}' such that each T'_i now contains inferences involving the predicates P_1, \dots, P_n that express the relevant universally and existentially quantified claims over which T_1, \dots, T_n disagree in terms of T'_1, \dots, T'_n ? In other words, to which extent can we within T'_1, \dots, T'_n express the disagreement between the original theories T_1, \dots, T_n as a disagreement between T'_1, \dots, T'_n ?

Considering how theories are proposals for what is valid, it is natural to think of the predicate P for a theory T as an *internal* validity predicate for T .⁴ Each obtained

¹ The discussions in Beall et al. (2017), Hjortland (2021) and Ripley (2015) may serve to illustrate this approach to evaluating formal theories of truth. For other ways of evaluating and comparing theories of truth based on various measures of strength, see for example Halbach (2011), Leigh and Rathjen (2010) and Łetyk and Wcisło (2019).

² Sentential schemas such as $(A \wedge (A \supset B)) \supset B$ are treated as zero-premise inferential schemas.

³ But if we agree that modus ponens holds in T_1 and fails in T_2 , where is then the disagreement, a reader might ask. The disagreement comes into the picture when we assume that T_1 and T_2 are competing proposals for what is valid *tout court*. Of course, if one understands T_1 and T_2 as presenting merely different non-competing conceptions of validity along pluralist lines, then there is no disagreement. A reader with such a perspective is free to ignore the disagreement talk and focus on how this paper presents a proposal for how to make theories' validity facts through the introduction of object-linguistic predicates.

⁴ The formulation of the question thus excludes an approach based on *typed* predicates in the sense that the predicates should concern the theories T'_1, \dots, T'_n as opposed to the theories T_1, \dots, T_n . Of course, one could reformulate the question and thus the project to rather pursue a typed approach where the validity predicates only concern the theories T_1, \dots, T_n ; I have no objection to that other than that it would be a different project. The aim here is not to argue against a typed approach, but to develop one particular untyped approach.

theory T should thus be ‘self-referential’ and it is reasonable to expect that there will be some issues arising from Curry- and liar-like formulas as suggested by the recent literature on validity predicates such as Beall and Murzi (2013) and Murzi (2014). As it turns out, there is no issue arising from requiring merely that an inference is represented as valid in T if and only if that inference is valid according to T as long as we have counterexamples for invalid inferences that are not only invalid but unsatisfiable. Instead, the main challenge is to ensure that this also holds for metainferences, i.e. that a metainference is represented as valid in T just in case that metainference is valid for T . Hlobil (2018) has shown that the obvious way of parsing that principle trivialises any theory defined with a classical meta-theory. Inspired by recent research on metainferences such as Barrio et al. (2015), Barrio et al. (2020) and Dicher and Paoli (2019), this paper proposes and develops an approach according to which the requirement for metainferences should be restricted from (globally) valid metainferences to locally valid metainferences; that a metainference is represented as valid in T if and only if the metainference is *locally valid* for T .

The rest of the paper is organised into two sections. Section 2 presents the proposal in more detail and Sect. 3 illustrates how the proposal can be applied on four theories of transparent truth definable on strong Kleene models.

2 An approach to expressing valid inferences and metainferences

To give the presentation of the proposal some structure, this section is divided into four subsections. Section 2.1 elaborates on what it means to represent an inference and a metainference within a theory using a validity predicate and moreover clarifies how the expressive nature of the project justifies conservativeness and preservation as desiderata. The paper proceeds then in Sect. 2.2 to discuss which principles we could want a validity predicate to satisfy. This includes a presentation of some problems caused by ‘self-referential’ inferences for the obvious candidate principles. Section 2.3 presents and discusses the merits of the novel proposal before Sect. 2.4 illustrates how one can expand a sequent calculus for classical logic with a validity predicate satisfying the novel principle.

2.1 Representing inferences and metainferences

An inference will in this paper be defined as a pair of finite sets of formulas and theories will be defined as sets of inferences.⁵ X , Y and variations thereof will be used as variables for sets of formulas throughout the paper. An inference from X to Y is said to be valid according to T just in case $\langle X, Y \rangle \in T$. This, in turn, is written as $X \vDash_T Y$.

In addition to comparing theories with regard to which inferences they contain, theories are also compared with regard to which metainferences they are closed under.

⁵ Theories represented as pairs of multisets of formulas will thus not be considered in this paper. Examples of such theories include Zardini (2011) and Rosenblatt (2019) in which there are formulas A such $A, A \vDash_T$ but $A \not\vDash_T$. This is not a significant restriction since we can in these and other similar cases obtain a theory T' from T based on sets as opposed to multisets with the help of multiplicative connectives.

In this paper we define a metainference as a pair of finite sets of inferences and say a theory T is closed under a metainference $\langle \Theta, \Upsilon \rangle$ just in case if $X \vDash_T Y$ for each $\langle X, Y \rangle \in \Theta$ then $X' \vDash_T Y'$ for some $\langle X', Y' \rangle \in \Upsilon$. A metainference is valid according to T just in case T is closed under that metainference.⁶

The definition of a valid metainference makes it clear that whether a metainference is valid depends on the meaning of "if..then" in the meta-theory. As mentioned in the introduction and following contemporary practice with regard to comparing formal theories of truth, the meta-theory will remain classical throughout this paper.⁷

With theories being sets of pairs of sets of formulas, it is natural to let the predicates representing what is valid according to a theory be binary, thus following the recent literature on validity predicates such as Barrio et al. (2016), Beall and Murzi (2013), Field (2017), Hlobil (2019), Murzi (2014), Murzi and Rossi (2021), Nicolai and Rossi (2018), Rosenblatt (2017), Zardini (2013) and Zardini (2014). However, as discussed by Rosenblatt (2017), the same literature also typically treats the validity predicate as applying to pairs of formulas in the same way as the truth predicate applies to formulas even if theories are sets of pairs of sets of formulas. One could for example be interested in distinguishing between the inferences $\langle \emptyset, \{A \vee B\} \rangle$ and $\langle \emptyset, \{A, B\} \rangle$. We will therefore take such predicates to apply to pairs of sets of formulas rather than pairs of formulas. To that purpose it will throughout the paper be assumed that the language contains numerals for the natural numbers and that a Gödel-coding is at disposal which associates not only each formula with a unique natural number, but through which each finite subset of the set of formulas is assigned a unique natural number, including the empty set \emptyset . $\ulcorner A \urcorner$ will be used as metalinguistic expression for the natural number associated with the formula A and $\ulcorner X \urcorner$ will be used as metalinguistic expression for the natural number associated with the set of formulas X . The same expression will be used for the corresponding numeral as long as there is no ambiguity. With Val_T as a validity predicate for a theory T , the statement that the inference from X to Y is valid according to T can now be represented with the formula $Val_T(s, t)$ where s and t are closed terms such that $s = \ulcorner X \urcorner$ and $t = \ulcorner Y \urcorner$ are true arithmetical equations.

In order to also express universal and existential generalisations, it will be assumed that the language contains \wedge , \vee , \neg and \forall as primitives and with \supset and \exists defined in the usual way. In addition, the language is assumed to contain function-symbols for the primitive recursive functions which are defined accordingly. The language will thus have function-symbols representing the following primitive recursive functions:

- $f_{\{\}}(\ulcorner A \urcorner)$ returns the code of the singleton containing the formula A .
- $f_{\supset}(\ulcorner A \urcorner, \ulcorner B \urcorner)$ returns the code of the formula $\neg A \vee B$.
- $f_{\cup}(\ulcorner X \urcorner, \ulcorner Y \urcorner)$ returns the code of the union of X and Y .

⁶ Metainferences are thus restricted to finitely many premises and conclusions. Metainferences with infinitely many premises will be discussed briefly towards the end of Sect. 2.3.

⁷ Exceptions to this practice include Bacon (2013), Girard and Weber (2019), McAllister (forthcoming), Zardini (2014) and Rosenblatt (2021). Importantly, the fact that this paper assumes throughout that the meta-theory is classical does not imply that the author of this paper believes that we should use a classical meta-theory. Instead, classical meta-theory is assumed because it is the contemporary standard. In fact, I do not have any reasons to believe that we need to use classical meta-theory for the current project and I think it would be quite interesting to pursue it from within a non-classical meta-theory.

\dot{f} will be used as function-symbol for the function f . The language is also assumed to contain a predicate *Sent* that is true of terms that are equal to the code of a sentence and false of terms that aren't.

To illustrate the function-symbols in action, consider the formula

$$Val_T(\ulcorner A, \neg A \vee B \urcorner, \ulcorner B \urcorner)$$

representing the claim that an instance of modus ponens as inference for some arbitrary formulas A and B is valid according to T . The function-symbols are then used to "access" the formulas in order to quantify over them:

$$Val_T(\dot{f}_\cup(\dot{f}_\cap(\ulcorner A \urcorner), \dot{f}_\supset(\dot{f}_\supset(\ulcorner A \urcorner, \ulcorner B \urcorner))), \dot{f}_\cap(\ulcorner B \urcorner))$$

This is not very easy to read, and it will thus be simplified in the following way:⁸

$$Val_T(\{\ulcorner A \urcorner\} \cup \{\ulcorner A \urcorner \supset \ulcorner B \urcorner\}, \{\ulcorner B \urcorner\})$$

Replacing the closed terms with variables, one can now express the following:

$$\begin{aligned} \forall xy((Sent(x) \wedge Sent(y)) \supset Val_T(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \\ \exists xy((Sent(x) \wedge Sent(y)) \wedge \neg Val_T(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \end{aligned}$$

For this formula to adequately represent the meta-theoretic statement, \supset and \neg should behave classically with regard to the validity predicate, that is, classical logic (inferences and metainferences) should hold for the fragment based on \supset , \neg and the validity predicate. Otherwise that formula could hardly be said to represent the statement made in the classical meta-theory. This is particularly important in the case of metainferences. Consider for example the following formula intended to represent the universally generalised statement about modus ponens as metainference:

$$\begin{aligned} \forall xy((Sent(x) \wedge Sent(y)) \supset ((Val_T(\{\emptyset\}, \{x\}) \wedge Val_T(\{\emptyset\}, \{x \dot{\supset} y\})) \\ \supset Val_T(\{\emptyset\}, \{y\}))) \end{aligned}$$

Again, \supset is used to express the classical "if..then" of the meta-theory, and should therefore also behave classically within T with regard to the validity predicate.

With the goal being to expand a theory with a predicate in order to express universally and existentially generalised claims about the expanded theory, it is reasonable to think of this predicate as playing an *expressive* role inspired by deflationism about truth as articulated by for example Beall (2009). Nonetheless, one must here distinguish between assigning the validity predicate for a theory T an expressive role and something like 'deflationism about validity' as a philosophical thesis along the lines of Shapiro (2011). The focus in this paper is on expanding theories understood as

⁸ $\{\ulcorner A \urcorner\}$ is thus used as a metalinguistic expression for the term $\dot{f}_\cap(\ulcorner A \urcorner)$. That $\{$ and $\}$ are outside $\ulcorner A \urcorner$ as opposed to inside should not cause any confusion for the observant reader. The notation will anyway only be used on a few occasions.

a proposals for what is valid with predicates that express generalisations about the resulting theories. Even if the validity predicate is thus assigned an expressive role, a deflationist thesis about validity in analogy to deflationism about truth doesn't follow.

Moreover, and in that regard, it is also important to stress that the aim is therefore not to capture what is 'really' valid with a predicate, but rather what the theory as a proposal claims is valid. In other words, even if every instance of modus ponens as inference is valid in some analytic, metaphysical or epistemic sense, maybe because classical logic is 'the true logic', a validity predicate for a theory in which some instances of modus ponens fail should not express that every instance of modus ponens is valid according to that theory. Similarly, a validity predicate for first-order Peano Arithmetic shouldn't express that the inference $\langle \emptyset, \{\neg Prov_{PA}(\ulcorner 0 = 1 \urcorner)\} \rangle$ is valid where $Prov_{PA}$ is the standard provability predicate for PA even if $\neg Prov_{PA}(\ulcorner 0 = 1 \urcorner)$ is true in the standard model of arithmetic, and it is surely true that Peano Arithmetic doesn't prove $0 = 1$. A theory is a set of inferences, and the only thing that matters is what inferences that theory contains and what metainferences it is closed under, not what one can say about the intended model for that theory or what one might think is valid in 'the true logic'.

Such considerations suggests that the resulting theory T' should be conservative over the original theory T where T' is conservative over T if and only if, if an inference of \mathcal{L} is valid according to T' then that inference is also valid according T , where \mathcal{L} is the original language over which T is formulated. Consider the case where modus ponens as inferential schema fails to hold for T but holds for T' . Then we cannot use what is valid in T' to say something about what is valid in T . Similarly if T is Peano Arithmetic and the inference $\langle \emptyset, \{\neg Prov_{PA}(\ulcorner 0 = 1 \urcorner)\} \rangle$ is valid and thus expressed with a validity predicate in T' . In fact, it is natural to extend the notion of conservativeness from inferences to metainferences in the obvious way.

Moreover, it shouldn't be enough that T' is conservative over T . It should also be *inferentially and metainferentially preservative* over T , where that is the case if and only if, if an inferential or a metainferential schema holds for T with regard to \mathcal{L} then it also holds for T' with regard to \mathcal{L}' . Consider for example the theory of truth STT explored by Ripley (2013a) and Ripley (2012) and Cobreros et al. (2013). The resulting theory of truth conservatively extends Peano Arithmetic but this comes at the cost of modus ponens as metainferential schema. Even if the same approach can be used to conservatively extend Peano Arithmetic with a validity predicate, the resulting theory is not metainferentially preservative over Peano Arithmetic since modus ponens as metainferential schema is valid in Peano Arithmetic but not in the resulting theory. An advocate of 'going non-transitive' could argue that the resulting validity predicate still expresses that modus ponens as metainferential schema holds, and that one can thus still proceed with the comparison. However, this requires that the base theory is itself transitive: the result of expanding the non-transitive theory of truth STT with a validity predicate defined along the same lines will still tell us that the resulting theory is transitive even if STT is not transitive according to the classical meta-theory. With an inferentially and metainferentially preservative theory we avoid this issue.⁹

⁹ I should stress that this is not an argument against STT as such, but rather an argument against the use of such an approach for the current project. There are certainly cases where we are better served with a theory

2.2 Metainferential problems with validity

To ensure that a validity predicate is fit for purpose, it should satisfy some inferential or metainferential conditions. One perhaps obvious condition considering the objective, is that it shouldn't be the case that $\models_{T_i} \text{Val}_{T_j}(\ulcorner X \urcorner, \ulcorner Y \urcorner)$ when $X \not\vdash_{T_j} Y$ and vice versa. How could it facilitate a comparison if each predicate $\text{Val}_{T'_i}$ was not true to the theory T'_i with regard to which inferences are valid? The same goes with regard to invalid inferences; if an inference schema has an instance which is not valid according to the theory, then this should be expressible in the theory along the above lines using an existential quantifier. The guiding idea is thus that the principles for the predicates should facilitate a comparison with regard to the differences between the theories that are relevant within the debate where the theories are proposals. We take such a comparison to be facilitated if the theories contain theorems expressing the relevant differences between the theories. Considering that the theories may also disagree over metainferences, it is natural to extend the above requirement from inferences to metainferences.

Moreover, there could also be further principles that one might find “intuitively compelling” for a validity predicate. However, if that principle is not needed to achieve the objective, then it can be discarded. In fact, as long as paradoxes can be formulated using the available vocabulary, there will be principles that are “intuitively compelling” but that we should nonetheless discard. To obtain paradoxes of the relevant kind we shall in this paper rely on the availability of function-symbols for the primitive recursive functions. Thanks to the strong diagonal lemma there is for every formula $A(y)$ with exactly y free a closed term τ such that $\models_T \tau = \ulcorner A(\tau) \urcorner$.¹⁰ This generalises straightforwardly to the singleton containing $A(y)$ and it is thus the case that $\models_T \tau = \{\ulcorner A(\tau) \urcorner\}$. It follows that there are certain important limitations on which principles a validity predicate for a theory T can satisfy within T as long as substitution of identicals holds.

For example, it follows from the observations made by Beall and Murzi (2013) that a transitive theory T defined with a classical meta-theory is trivial if it satisfies the following two principles:¹¹

Validity proof (VP): if $A \models_T B$ then $\models_T \text{Val}_T(\{\ulcorner A \urcorner\}, \{\ulcorner B \urcorner\})$

Validity detachment (VD): $\text{Val}_T(\{\ulcorner A \urcorner\}, \{\ulcorner B \urcorner\}), A \models_T B$

Footnote 9 continued

that does not preserve the inferences and metainferences of the theory it expands. For example, STT and PosFS are both conservative extensions of Peano Arithmetic, but only PosFS preserves its inferences and metainferences. This comes at a cost for PosFS since STT can prove more things about the liar sentence, things that one might be interested in saying.

¹⁰ See e.g. Jeroslow (1973) and Burgess (1986).

¹¹ As discussed above in footnote 5, this paper ignores the option to reject “structural contraction” by using sets rather than multisets to define theories. That option can be simulated by formulating (VD) using conjunction and then add idempotence of conjunction as a third principle. On the other hand, it should be clear from the discussion in this section that the option to accommodate a validity predicate by rejecting the idempotence of conjunction is not really on the table for us since the connectives should behave classically with regard to validity predications as discussed in Sect. 2.1.

There are variations on this pair obtained by replacing (VD) with alternatives that are, under the assumption of transitivity and reflexivity, equivalent.¹² However, if the principles required for a validity predicate are incompatible with either a reflexive or a transitive theory, then the approach itself would be quite limited with regard to the current project. Indeed, even if a proponent of the pair (VD) and (VP) such as Zardini (2014) takes their “intuitive plausibility” together with the triviality result to show that one should revise our conception of validity, such results should not be considered as arguments for revising a “naive” conception of validity from within the current perspective. Instead, it follows that (VD) and (VP) taken together are unsuitable for the task at hand.

The principle (VD) has been criticised from a variety of perspectives, examples being the points made by Cook (2014), Field (2017), Hlobil (2019), and Zardini (2013). As I do not intend to express dissatisfaction with their reasons against (VD), I shall not repeat them here. The focus will instead be directed at how (VD) is not required for the task at hand and that (VD) can thus be discarded.

In elaborating on the principles (VP) and (VD), Beall and Murzi (2013, p.150) introduce a further principle

V-Schema (VS): $\models_T \text{Val}_T(\{\ulcorner A \urcorner\}, \{\ulcorner B \urcorner\})$ if and only if $A \models_T B$

and explain that

“putting VP and VD together yields what, by analogy with truth and exemplification, may be called the *V-schema*. What we now note is that VP and VD -or, simply, the V-schema- along with the standard structural rules, are the ingredients for v-Curry paradox.”

As suggested above, it seems reasonable to maintain that it is sufficient for a validity predicate to satisfy (VS) in order to ensure that an inference is expressed as valid according to T within T just in case it is actually valid according to T . However, as pointed out by Field (2017, p.9), (VS) is a weaker requirement than (VP) and (VD) taken together, and is in itself insufficient for triviality. It is not mentioned by Field (2017, p.9), but it is actually the case that if T is a recursively enumerable theory extending Robinson Arithmetic, then $\text{Val}_T(\ulcorner A \urcorner, \ulcorner B \urcorner)$ defined as $\text{Prov}_T(\ulcorner A \supset B \urcorner)$ where Prov_T is its standard provability predicate will satisfy (VS). The consistency of (VS) with classical logic follows from this observation.

On the other hand, (VD) is arguably useful for representing invalid inferences as invalid. A natural candidate for a principle to represent invalid inferences would be the following:

$$\text{if } X \not\models_T Y \text{ then } \text{Val}_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \models_T$$

As shown by Zardini (2014), this principle and (VP) jointly suffice for triviality if T is transitive and monotonic as long as the meta-theory is classical. A better candidate

¹² Consider for example the principle that if $C \models_T \text{Val}_T(\{\ulcorner A \urcorner\}, \{\ulcorner B \urcorner\})$ and $C \models_T A$ then $C \models_T B$. With this principle rather than (VD) then the assumption that T is reflexive suffices for triviality as shown by Murzi (2014).

for a principle to this purpose is the following restricted version implied by (VD) if T is transitive:

$$(VI) \text{ If } \vDash_T A \text{ and } B \vDash_T \text{ then } Val_T(\{\ulcorner A \urcorner\}, \{\ulcorner B \urcorner\}) \vDash_T$$

With $\vDash_T A$ and $B \vDash_T$ implying that $A \not\vdash_T B$ when T is not only transitive but also monotonic, one can with (VI) represent that this particular inference is invalid. Under the assumption that the connectives behave classical with regard to $Sent$ and Val_T , it can also be used to represent that the inferential schema of which that inference is an instance has some instance which is invalid as follows:

$$\vDash_T \exists xy(Sent(x) \wedge Sent(y) \wedge \neg Val_T(\{x\}, \{y\}))$$

From the perspective of this particular usage of (VD) then, (VD) can be replaced by (VI) as the partner of (VS).

Moreover, the pair (VS) and (VI) is consistent with classical logic. Let T be a recursively enumerable theory extending Robinson Arithmetic. Then $Val_T(\ulcorner A \urcorner, \ulcorner B \urcorner)$ defined as $Prov_T^R(\ulcorner A \supset B \urcorner)$ where $Prov_T^R$ is its standard Rosser provability predicate will satisfy (VS) and (VI) (Cf. Arai (1990)).

So far so good. There are nonetheless two reasons for being dissatisfied with (VS) and (VI). Firstly, (VI) will overgenerate in the case where T is such that $A \vDash_T$ and $\vDash_T A$ for some formula A but nontransitive and thus not trivial. Secondly, neither (VS) nor (VI) can be used to represent metainferences. Let us focus on the second issue first since our solution to the second issue will also solve the first issue.

A first stab at a requirement for representing metainferences is presented in the following form by Hlobil (2018) but the idea of a requirement along these lines is also suggested by Barrio et al. (2016) and Rosenblatt (2017):¹³

$$\begin{aligned} &\text{if } X_0 \vDash_T Y_0 \text{ and } \dots \text{ and } X_{n-1} \vDash_T Y_{n-1} \text{ then } X_n \vDash_T Y_n \\ &\qquad\qquad\qquad \text{if and only if} \\ &Val_T(\ulcorner X_0 \urcorner, \ulcorner Y_0 \urcorner), \dots, Val_T(\ulcorner X_{n-1} \urcorner, \ulcorner Y_{n-1} \urcorner) \vDash Val_T(\ulcorner X_n \urcorner, \ulcorner Y_n \urcorner) \end{aligned} \tag{GVS}$$

This requirement, the *generalised validity schema*, would ensure the representation of a valid metainference just in case it is valid. Since one direction descends to the object-theory from the meta-theory and the other direction ascends to the meta-theory from the object-theory, we will refer to each direction as the *ascending* and *descending* direction respectively.

However, as anticipated by Hlobil (2018), it turns out that every theory satisfying this requirement is trivial if the meta-theory is classical. In particular, Hlobil (2018)

¹³ Whereas Hlobil (2018) presents this principle the same way as here but refer to it as "faithfulness", Barrio et al. (2016) and Rosenblatt (2017) present the left side as a "metarule" and say that a theory internalises a metarule if it proves the corresponding sequent with validity predicates. A metarule is here basically a sequent calculus rule. The focus of Barrio et al. (2016) is on how ST internalises the cut rule, and Rosenblatt (2017) looks in addition at other substructural logics and observe for example that affine logic fails to internalise some admissible metarule. That Rosenblatt (2017) includes not only primitive and derivable but also admissible rules in the discussion, suggest that Hlobil (2018)'s change in notation from a metarule to the "if...then..." expression involving validity claims is in line with Barrio et al. (2016) and Rosenblatt (2017).

shows that a theory T is trivial if it is based on a language containing a binary predicate Adm and closed terms functioning as names for inferences, and T satisfies the equivalence

$$\text{If } X \vDash_T Y \text{ then } X' \vDash_T Y' \text{ if and only if } \vDash_T Adm(\ulcorner X, Y \urcorner, \ulcorner X', Y' \urcorner)$$

Hlobil (2018) notes that the corresponding result “holds for every logic that satisfies” (GVS) if $Adm(\ulcorner X, Y \urcorner, \ulcorner X', Y' \urcorner)$ is replaced with $Val_T(\ulcorner X \urcorner, \ulcorner Y' \urcorner) \supset Val_T(\ulcorner X' \urcorner, \ulcorner Y \urcorner)$.

An alternative version of Hlobil (2018)’s result for a validity predicate is obtained directly with (GVS) by considering a closed term κ such that $\kappa = \{\ulcorner Val_T(\kappa, \tau) \urcorner\}$ holds in virtue of the strong diagonal lemma where τ is an abbreviation for $\{\ulcorner Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \urcorner\}$. With κ at hand the following is an instance of the ascending direction of (GVS):

$$(1) \text{ if } Val_T(\kappa, \tau) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \text{ then, if } Val_T(\kappa, \tau) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \text{ then } X \vDash_T Y$$

It follows immediately by contraction for the meta-theoretic conditional used to state (GVS) that

$$(2) \text{ If } Val_T(\kappa, \tau) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \text{ then } X \vDash_T Y$$

Applying the descending direction of (GVS) on (2), it follows that

$$(3) Val_T(\kappa, \tau) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$$

Assuming furthermore that the meta-theoretic conditional satisfies the rule of modus ponens as applied to theorems of the meta-theory, (2) and (3) imply $X \vDash_T Y$. Considering the commitment to a classical meta-theory, (GVS) is not a good candidate for a principle to ensure representations of valid metainferences of T within T .

2.3 Expressing locally valid metainferences

The aim of this subsection is to present a novel proposal for how to restrict (GVS) and elaborate on some of its implications. To that purpose, it is useful to first highlight the extent to which any suitable restriction on (GVS) will have significant limitations with regard to the task at hand.

Consider the following four principles where the last three are labelled according to their counterparts in modal logic:

- (VS) $X \vDash_T Y$ if and only if $\vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$
- (V4) $Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \vDash_T Val_T(\ulcorner \emptyset \urcorner, \ulcorner Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \urcorner)$
- (VC4) $Val_T(\ulcorner \emptyset \urcorner, \ulcorner Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner) \urcorner) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$
- (VK) $Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner), \{Val_T(\ulcorner \emptyset \urcorner, \ulcorner A \urcorner) \mid A \in X\} \vDash_T Val_T(\ulcorner \emptyset \urcorner, \ulcorner Y \urcorner)$

Their relationship to (GVS) is as follows. Each instance of (VS) is a premise-free instance of (GVS). (VK) is obtained by applying the descending direction of (GVS) on the relevant instance of the metainference that T is transitive. (V4) and (VC4) are obtained by applying the descending direction of (GVS) on (VS). With inspiration

from the last contradiction for theories of truth presented by Friedman and Sheard (1987) one can show that the four principles imply triviality under the assumption that the meta-theory is classical and T is transitive. With κ as above we have the following instance of (VK):

$$Val_T(\kappa, \tau), Val_T(\ulcorner \emptyset \urcorner, \kappa) \vDash_T Val_T(\ulcorner \emptyset \urcorner, \tau)$$

By transitivity of T , (V4) and (VC4) it follows that

$$Val_T(\kappa, \tau) \vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$$

With (VS) it follows that

$$\vDash_T Val_T(\kappa, \tau)$$

Transitivity applied on the latter two claims yields $\vDash_T Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$ which by (VS) implies $X \vDash_T Y$.

It follows that a transitive theory cannot contain a validity predicate which satisfies (VS), inferences representing (VS) with the validity predicate in the form of (V4) and (VC4), and inferences representing that it is transitive in the form of (VK). Since transitivity is a property that is up for discussion in the literature on theories of truth, surely a validity predicate should express that the theory is transitive just in case it is transitive. Moreover, (VS), (V4) and (VC4) come as a package in the sense that if (VS) does not hold for T , then neither should (V4) nor (VC4). Under the assumption that the converse also holds one seems left with the option to reject all three.

However, there is a pragmatic alternative available. Considering the overall aim of this project, namely to facilitate comparison between theories, and under the assumption that every theory is supposed to satisfy (VS), then a comparison with regard to whether they satisfy (VS) is not particularly enlightening; it is significantly more interesting to compare them with regard to some metainference over which they disagree, for example transitivity. From such a perspective it is not much of a loss to give up on (V4) or (VC4). Uniformity even suggests that one might as well give up on both. A proposal along these lines is obtained by restricting (GVS) to *locally* valid metainferences as opposed to valid metainferences.¹⁴

¹⁴ An alternative way to restrict (GVS) by Hlobil (2019) is presented as consisting in restricting (GVS) to derivable metainferences. Before discussing some details with Hlobil (2019)'s proposal, it is worth pointing out that restricting (GVS) to derivable metainferences only makes sense if one accepts that a theory may be closed under certain metainferences that must remain undervivable come what may in any proof system for that theory. Instead, the proofs of their admissibility may only be obtained by other means such as for example proof analysis or completeness with suitable models. With regard to the reasoning presented in Sect. 2.2, this would mean that the rule $Val(\kappa, \tau) \Rightarrow Val(\ulcorner A \urcorner, \ulcorner B \urcorner) / A \Rightarrow B$ cannot be derivable in any proof system for the theory but will be admissible about the theory and thus about the proof system. By using derivability as criterion to distinguish between metainferences that are safe to make explicit and those that are not, the proposal is in effect to let proof theory be structurally incomplete, that there should not be an alignment between proof theory and models beyond inferences. For a definition of structural completeness, see Iemhoff (2015). For those who see proof theory and model theory as complementing tools, such a necessary and enforced structural incompleteness is certainly undesirable. The same holds for those who put proof theory first. The proposal presented in this paper, on the other hand, will be

The notion of a locally valid metainference has surfaced in the recent literature on metainferences such as Dicher and Paoli (2019) and Barrio et al. (2020). It is fruitfully articulated from a model-theoretic perspective. Suppose that we have defined a theory T in such a way that we first define the notion of an inference $\langle X, Y \rangle$ being satisfied by an interpretation I , here represented with the notation $I \Vdash [X \Rightarrow_T Y]$, and then define $X \vDash_T Y$ as that $I \Vdash [X \Rightarrow_T Y]$ for every $I \in \mathbb{I}$ where \mathbb{I} is the set of permissible interpretations. In the case of first-order models for classical logic, $I \Vdash [X \Rightarrow_T Y]$ would mean that I is such that either some formula in X is false or some formula in Y is true. We can read the notation as saying that I T -satisfies the inference $\langle X, Y \rangle$, and furthermore say that the inference is T -valid or valid in T if and only if every interpretation T -satisfies it.

A (globally) valid and a locally valid metainference can now be defined as follows.¹⁵ Suppose that T is a theory based on \mathcal{L}' , is defined with I in the above manner and that $\langle \Theta, \Upsilon \rangle$ is a \mathcal{L}' -metainference. Then $\langle \Theta, \Upsilon \rangle$ is

- (i) *globally valid of T* just in case, if for each $\langle X, Y \rangle \in \Theta$, $\forall I \in \mathbb{I}$, $I \Vdash [X \Rightarrow_T Y]$, then for some $\langle X, Y \rangle \in \Upsilon$, $\forall I \in \mathbb{I}$, $I \Vdash [X \Rightarrow_T Y]$.
- (ii) *locally valid of T* just in case, $\forall I \in \mathbb{I}$, if for each $\langle X, Y \rangle \in \Theta$, $I \Vdash [X \Rightarrow_T Y]$, then for some $\langle X, Y \rangle \in \Upsilon$, $I \Vdash [X \Rightarrow_T Y]$.

A quick comparison of the definitions yields the following observations:

- (a) For every T , every single-conclusion locally valid metainference is globally valid.
- (b) For some T , some locally valid multiple-conclusion metainferences are not globally valid.
- (c) For some T , some globally valid single-conclusion metainferences are not locally valid.

An example of (b) is the metainference $\langle \emptyset, \{ \langle \emptyset, \{A\} \rangle, \langle \emptyset, \{\neg A\} \rangle \} \rangle$, as this is locally but not globally valid of classical logic. In the local case it is equivalent to excluded middle. An example of (c) is the rule for uniform substitution of formulas for propositional variables in theorems in classical propositional logic. Consider for example

Foonote 14 continued

compatible with the above rule being not only admissible but also derivable because neither will imply $Val(\kappa, \tau) \vDash Val(\ulcorner A \urcorner, \ulcorner B \urcorner)$. Instead, the rule comes out as globally and but not locally valid, and we cannot thus conclude $Val(\kappa, \tau) \vDash Val(\ulcorner A \urcorner, \ulcorner B \urcorner)$.

That being said, there are also some features with Hlobil (2019)'s proposal in particular that makes it unsuitable for the current project. Hlobil (2019)'s proposal is illustrated with a "naturalised" sequent calculus, that is, a sequent calculus with a rule of assumption for sequents together with rules for a validity predicate that take advantage of the rule of assumption. The rules for a validity predicate (and the rule of assumption) extend a sequent calculus for the non-transitive theory of truth STT that follows the presentation by Ripley (2013b), and the resulting theory is also non-transitive. The resulting calculus does not contain a cut rule. In fact, the calculus cannot be expanded with a cut rule restricted to validity predicates as that would make the rule $Val(\kappa, \tau) \Rightarrow Val(\ulcorner A \urcorner, \ulcorner B \urcorner) / A \Rightarrow B$ derivable and thus the theory trivial. To see why this is the case, a curious reader may adapt the derivation in the final paragraph of Sect. 2.4 below to Hlobil (2019)'s setting. Such a tree would with Hlobil (2019)'s rules for the validity predicate suffice to obtain a derivation of $Val(\kappa, \tau) \Rightarrow Val(\ulcorner A \urcorner, \ulcorner B \urcorner)$ from no assumptions. In other words, the formal details of Hlobil (2019)'s proposal is not suitable for our purposes because we are also interested in transitive logics, and the idea of preservation of metainferences is incompatible with the expanded theory becoming non-transitive.

¹⁵ We thus adhere to the terminology in the recent literature on metainferences such as Dicher and Paoli (2019) and Barrio et al. (2020) in which metainferential validity is referred to as 'global' in order to contrast it from local metainferential validity. See also Humberstone (1996).

how classical models are such that $\langle\{\emptyset, \{p\}\}, \{\emptyset, \{A\}\}\rangle$ for arbitrary A is globally valid since there is a model where p is false, but not locally valid because there will be a model where p is true but A false for some formula A .¹⁶

The new proposal then, is to require that a validity predicate for a theory T should be such that

$$\begin{aligned} \langle\langle X_0, Y_0 \rangle, \dots, \langle X_{n-1}, Y_{n-1} \rangle\rangle, \langle X_n, Y_n \rangle \text{ is locally valid for } T \\ \text{if and only if} \tag{GVS*} \\ Val_T(\ulcorner X_0 \urcorner, \ulcorner Y_0 \urcorner), \dots, Val_T(\ulcorner X_{n-1} \urcorner, \ulcorner Y_{n-1} \urcorner) \vDash_T Val_T(\ulcorner X_n \urcorner, \ulcorner Y_n \urcorner) \end{aligned}$$

Whereas (GVS) is presented for single-conclusion metainferences simply because that is how it was presented by Hlobil (2018), Barrio et al. (2016) and Rosenblatt (2017), (GVS*) is restricted to single-conclusion metainferences as a consequence of (b) above. Moreover, it follows from (a) that whenever a single-conclusion metainference is expressed in the theory with a predicate satisfying (GVS*), that single-conclusion metainference is going to be globally valid. As expected, however, the approach will undergenerate since some single-conclusion metainferences will be globally valid but not locally valid and thus not represented within the theory. In particular, it follows from (GVS*) that (VS) is globally valid, but this does not make (VS) locally valid, and (V4) and (VC4) are thus not implied by (GVS*).

In addition to (VS), one can also show that (GVS*) delivers (VI) for a theory T if transitivity is locally valid for T . Assume that $\vDash_T A$ and $B \vDash_T$. Then every I is such that $I \Vdash [\Rightarrow_T A]$ and $I \Vdash [B \Rightarrow_T]$. It follows by transitivity that $I \not\Vdash [A \Rightarrow_T B]$ for every I . Any metainference with $\langle A, B \rangle$ as premise will thus be locally valid. It follows that $Val_T(\ulcorner A \urcorner, \ulcorner B \urcorner) \vDash_T$.

In fact, the approach is consistent with classical logic. A validity predicate which satisfies (GVS*) can be defined in the (classical) theory of truth PosFS explored by Horsten et al. (2012) and Fjellstad (2020) by defining $Val_T(\ulcorner X \urcorner, \ulcorner Y \urcorner)$ as $Tr(\ulcorner \bigwedge X \supset \bigvee Y \urcorner)$. As opposed to (GVS) then, (GVS*) is consistent not only with a classical meta-theory, but also with an object-theory closed under classical logic. In addition to consistency, it also follows from the definability of such a validity predicate within PosFS that satisfying (GVS*) is compatible with the two features that was considered

¹⁶ Dicher and Paoli (2019) take this to illustrate a problem with the global notion: such metainferences are vacuously valid. I take this feature to be necessary in order for globally valid metainferences to align with the notion of admissible rules in sequent calculi. Consider for example how the cut rule is admissible for propositional variables in the sequent calculus for STT presented by Ripley (2013a), and how the same instances are globally but not locally valid with strict-tolerant entailment on Strong Kleene models. It actually follows from this example that global metainferential validity is not closed under uniform substitution of arbitrary formulas for propositional variables since the cut rule is globally valid for propositional variables but is not globally valid for the liar sentence. The reader preferring an example that does not involve the truth predicate may consider the cut-free version of the sequent calculus for the modal logic S5 presented by Ohnishi and Kazuo (1959) for which the cut rule is not admissible. In that calculus the cut rule will be admissible for propositional variables but not for the modal formulas. In any sound and complete semantics for that calculus, the cut rule will be globally valid for propositional variables but not for modal formulas. Another example can be found in Golan (2021) who also proves that local metainferential validity is closed under uniform substitution. We will take advantage of this feature with locally valid metainferences in Sect. 3.4. Now, Golan (2021) treats the failure of uniform substitution for global metainferential validity as a reason to reject the notion. I do not share this worry for the same reasons as I do not worry about the vacuously true globally valid metainferences.

desirable in Sect. 2.1, namely preservation and conservativeness. PosFS is metainferentially and inferentially preservative over Peano Arithmetic; if an inferential or a metainferential schema holds for Peano Arithmetic, then it also holds for PosFS. This follows immediately from its axiomatic presentation. Moreover, PosFS is a subtheory of a theory of truth which is to proven to be conservative over Peano Arithmetic by Friedman and Sheard (1987); adding the principles to Peano Arithmetic doesn't prove any new theorems involving only the arithmetical vocabulary.

Although the above observations are nice, things are not perfect and I will here mention two significant issues.

The first issue is that a theory T satisfying (GVS*) is such that any two formulas represented as invalid are equivalent if T is monotonic and transitive. Assume that $Val_T(\ulcorner \emptyset \urcorner, \ulcorner A \urcorner) \vDash_T$ and $Val_T(\ulcorner \emptyset \urcorner, \ulcorner B \urcorner) \vDash_T$. If T is transitive, then we have $A \vDash_T$ and $B \vDash_T$ from which $A \vDash_T B$ and $B \vDash_T A$ follows by monotonicity. This is certainly not a desirable feature but it is consistent with the requirements set forth in the introduction. After all, this does not hinder us from proving existential generalisations expressing that there is a counterexample to the validity of an inferential or a metainferential schema. However, this does illustrate an awkward consequence of the fact that the approach will only guarantee $\vDash_T \neg Val(\ulcorner A \urcorner, \ulcorner B \urcorner)$ if for every $I, I \not\llcorner_T [A \Rightarrow B]$, i.e. if the inference $(\{A\}, \{B\})$ is T -unsatisfiable. The counterexample to an inferential or a metainferential schema must therefore be an instance which is T -unsatisfiable.

The second issue concerns infinitary metainferences. A validity predicate satisfying (GVS*) is also definable with the theory of truth shown to be ω -inconsistent by McGee (1985). However, (GVS*) itself does not imply all the conditions corresponding to those presented by McGee (1985) for a theory of truth to be ω -inconsistent. In particular, (GVS*) does not imply the Barcan formula as formulated for the validity predicate:

$$\forall x Val_T(\ulcorner \emptyset \urcorner, \{\ulcorner A \dot{x} \urcorner\}) \vDash_T Val_T(\ulcorner \emptyset \urcorner, \{\ulcorner \forall x Ax \urcorner\}) \tag{VBF}$$

This follows from the fact that PosFS is conservative over Peano Arithmetic and results by Leigh (2015) imply that the theory obtained by expanding PosFS with the Barcan formula for a truth predicate is not conservative over Peano Arithmetic.

Now, this is also the reason why (GVS*) is restricted to finite inferences and metainferences. Suppose that (GVS*) holds for infinitary metainferences, the ω -rule holds as a locally valid (infinitary) metainference and infinitary inferences are permitted. Then the following inference would be valid:

$$Val_T(\ulcorner \emptyset \urcorner, \{\ulcorner A(\bar{0}) \urcorner\}), Val_T(\ulcorner \emptyset \urcorner, \{\ulcorner A(\bar{1}) \urcorner\}) \dots \vDash_T Val_T(\ulcorner \emptyset \urcorner, \{\ulcorner \forall x Ax \urcorner\})$$

With \forall behaving classically with regard to Val_T , (VBF) would be an immediate consequence. One could thus make a case for the claim that (VBF) expresses that the ω -rule is a valid metainference. It follows that certain theories cannot both satisfy the ω -rule and (VBF) without thereby being inconsistent, and thus moreover that we cannot generalise the approach from finitary to infinitary metainferences. This is perhaps

the most significant drawback with the current proposal, as it implies that we cannot, in general, compare theories with regard to the ω -rule.¹⁷

In conclusion, we are left with the following observations about the proposal to restrict (GVS) from globally to locally valid metainferences through the adoption of (GVS*). A validity predicate satisfying (GVS*) will not overgenerate with regard to which metainferences are represented as valid, but it will undergenerate in three ways. Firstly, there will be valid metainferences that cannot be represented, namely those that are not locally valid. This includes in particular (VS). Secondly, it will undergenerate with regard to invalid inferences in the sense that only inferences that are unsatisfiable will be represented as invalid. Finally, it cannot be extended to include metainferences with infinitely many premises.¹⁸

2.4 Expressing validity in infinitary Peano arithmetic

The aim of this subsection is three-fold. Firstly, it is useful to illustrate how we can expand a standard two-sided sequent calculus for classical logic with a validity predicate satisfying (GVS*) and moreover in a way that preserves the admissibility of cut. Secondly, by doing this for infinitary Peano Arithmetic, we can show that the resulting theory is ω -consistent. Finally, we can also use the sequent calculus to explain how the approach blocks the paradoxical result of Sect. 2.2.

The starting point will be a standard sequent calculus for infinitary Peano Arithmetic along the lines of Rathjen and Sieg (2018) that we will expand with certain rules for the truth predicate typically associated with deontic logic.

Let \mathcal{L} be a first-order language based on the connectives \neg , \wedge , \vee and \forall with countably many variables, a constant 0, a binary predicate $=$ and function-symbols for every primitive recursive function. Let \mathcal{L}^{Tr} be obtained from \mathcal{L} by adding a unary predicate Tr . As usual, $A \supset B$ will be treated as a metalinguistic abbreviation for $(\neg A) \vee B$.

A standard bilateral sequent is an expression of the form $A_0, \dots, A_n \Rightarrow B_0, \dots, B_m$ that represents a pair of finite sets or multisets of formulas. Sequent(s) above the line in a rule is referred to as premise-sequent(s) and the sequent below the line as the conclusion-sequent. Each rule consists of zero or more active formulas and zero or more principal formulas. The active formulas of a rule are the displayed formulas in the premise-sequent(s) and the principal formulas are the displayed for-

¹⁷ But why not avoid this problem by simply accepting that the expanded theory is ω -inconsistent, one reader might wonder. The literature offers various arguments against ω -inconsistency such as those presented by Leitgeb (2007), Barrio and Picollo (2013) and Barrio and Ré (2018), but the issue with ω -inconsistency from the perspective of this approach is simply that it is incompatible with preservation. As far as this approach goes, it only makes sense to include (VBF) in the case where the ω -rule is a locally valid metainference for a theory T . In fact, (VBF) follows by closing such a T under infinitary (GVS*). Assume now that T is classical logic expanded with Peano Arithmetic and the ω -rule. If we close it under infinitary (GVS*) to define T' , T' will be inconsistent. We could accept that consequence by allowing for example transitivity or explosion (as locally valid metainference) to fail along the lines of ST or LP. However, T' would in any such case no longer preserve the metainferences of T .

¹⁸ For the reader in search of expressive limitations with the current proposal that could be understood as so-called "revenge paradoxes", perhaps along the lines of Scharp (2014), my guess is that the first and third issue could count as such, depending on how one wishes to understand the idea of a revenge paradox.

mulas in the conclusion-sequent. Γ and Δ are referred to as the context of the rule: they represent arbitrary permissible expressions. Derivations are trees constructed by using initial sequents and zero-premise rules as leaves with the root being the sequent for which the tree is a derivation.

We shall work with the following sequent calculus for infinitary Peano arithmetic with a deontic truth predicate where sequents are pairs of finite sets of formulas and $s^{\mathbb{N}} = t^{\mathbb{N}}$ means that $s = t$ is true in the standard model of arithmetic:

$$\begin{array}{c}
 \frac{Tr(s), \Gamma \Rightarrow \Delta, Tr(t) \quad \text{if } s^{\mathbb{N}} = t^{\mathbb{N}}}{s = t, \Gamma \Rightarrow \Delta \quad \text{if } s^{\mathbb{N}} \neq t^{\mathbb{N}}} =L \quad \text{and} \quad \frac{}{\Gamma \Rightarrow \Delta, s = t \quad \text{if } s^{\mathbb{N}} = t^{\mathbb{N}}} =R \\
 \\
 \frac{\Gamma \Rightarrow \Delta, A}{\neg A, \Gamma \Rightarrow \Delta} \neg L \quad \frac{A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A} \neg R \\
 \\
 \frac{A, B, \Gamma \Rightarrow \Delta}{A \wedge B, \Gamma \Rightarrow \Delta} \wedge L \quad \frac{\Gamma \Rightarrow \Delta, A \quad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B} \wedge R \\
 \\
 \frac{A, \Gamma \Rightarrow \Delta \quad B, \Gamma \Rightarrow \Delta}{A \vee B, \Gamma \Rightarrow \Delta} \vee L \quad \frac{\Gamma \Rightarrow \Delta, A, B}{\Gamma \Rightarrow \Delta, A \vee B} \vee R \\
 \\
 \frac{A(t/x), \Gamma \Rightarrow \Delta}{\forall x Ax, \Gamma \Rightarrow \Delta} \forall L \quad \frac{\Gamma \Rightarrow \Delta, A(\bar{n}/x) \quad \text{for each } n < \omega}{\Gamma \Rightarrow \Delta, \forall x Ax} \forall R \\
 \\
 \frac{B_0, \dots, B_n \Rightarrow A}{\Gamma, Tr(s_0), \dots, Tr(s_n) \Rightarrow Tr(t), \Delta \quad \text{whenever } s_i^{\mathbb{N}} = \ulcorner B_i \urcorner^{\mathbb{N}} \text{ and } t^{\mathbb{N}} = \ulcorner A \urcorner^{\mathbb{N}}} T_K \\
 \\
 \frac{B_0, \dots, B_n \Rightarrow}{\Gamma, Tr(s_0), \dots, Tr(s_n) \Rightarrow \Delta \quad \text{whenever } s_i^{\mathbb{N}} = \ulcorner B_i \urcorner^{\mathbb{N}}} T_D
 \end{array}$$

The set of formulas represented by B_0, \dots, B_n in T_K and T_D is possibly empty. The contexts Γ and Δ in T_K and T_D are added to the conclusion-sequent of each rule to ensure admissibility of weakening. The rules are inspired by the rules for modal logic presented by for example Negri (2011).

The cut-rule,

$$\frac{\Gamma \Rightarrow \Delta, A \quad A, \Gamma' \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'} \text{ cut}$$

is admissible in this sequent calculus.

Now, the admissibility of cut cannot be shown through the standard argument going back to Gentzen (1934) and along the lines of Negri and von Plato (2001), namely through an induction on the complexity of a formula with a subsidiary induction on the height of a derivation. This is due to the presence of the rules T_K and T_D and is a well-known issue with sequent calculi for semantic predicates; see for example Kremer (1988). Instead, one can follow Cantini (1990)’s strategy by relying on a triple induction where the second and third measure are based on the complexity of a formula and the height of a derivation respectively whereas the first measure tracks the supremum of the number of applications of truth-rules in a derivation. In fact, it suffices with only minor modifications to the cut-elimination proof presented by

Cantini (1990) in order to show that cut is admissible. We thus spare the reader for those details, but nonetheless provide an explanation for why the strategy works.

The main reason this approach works is that cuts are straightforwardly eliminated from a derivation in the case where the formula A is not principal in $\Gamma \Rightarrow \Delta, A$ (or $A, \Gamma' \Rightarrow \Delta'$) and the last applied rule is T_D or T_K . In this case the desired conclusion $\Gamma, \Gamma' \Rightarrow \Delta, \Delta'$ is already obtainable directly with an application of T_K or T_D from the same sequent as $\Gamma \Rightarrow \Delta, A$ (or $A, \Gamma' \Rightarrow \Delta'$) is obtained from. Consider the following case where X_{T_r} contains truth predications obtained from X and A is thus not principal in the application of T_D :

$$\frac{X \Rightarrow}{X_{T_r} \Rightarrow A} T_D \quad A, \Gamma' \Rightarrow \Delta'$$

The desired conclusion $X_{T_r}, \Gamma' \Rightarrow \Delta'$ is now obtained directly from $X \Rightarrow$ using T_D itself without thereby increasing the number of times the rule has been applied.¹⁹

We concluded in the previous subsection that satisfying (GVS*) is not sufficient for McGee (1985)’s theorem as formulated for a validity predicate to hold because (GVS*) doesn’t imply (VBF). This alone does not settle the question whether the sequent calculus just defined is ω -consistent. As it turns out, ω -consistency is in our case guaranteed by the admissibility of cut together with the shape of the rules. To see that, we reason as follows. If the theory defined with the sequent calculus is ω -inconsistent, then we have derivations $\Rightarrow A(n)$ for each $n \in \omega$ but also $\forall x A \Rightarrow$. One application of the ω -rule yields $\Rightarrow \forall x A$, which by admissibility of cut implies the empty sequent \Rightarrow . But there is no derivation of the empty sequent because the calculus does not contain any elimination rules, and the theory can thus not be ω -inconsistent.

Let us now turn our attention to showing that (GVS*) holds and moreover use the explanation for why that is the case to illustrate how the approach blocks the paradoxical reasoning involving (GVS) from Sect. 2.2.

Consider a metainference $\langle \Theta, \langle X, Y \rangle \rangle$ and the sequent

$$\{\bigwedge X' \supset \bigvee Y' \mid \langle X', Y' \rangle \in \Theta\}, X \Rightarrow Y \tag{d}$$

representing that metainference using the connectives. In the case of interpretations for classical logic, the sequent (d) is valid just in case the corresponding metainference is locally valid.²⁰

¹⁹ This is the crucial difference between this application of Cantini (1990)’s strategy on the one hand, and the use of it by Fischer and Gratzl (2018) to show that cut is admissible for a sequent calculus defining a formal theory of transparent truth. The application of the strategy by Fischer and Gratzl (2018) is erroneous precisely because it allows contexts in the premise-sequents of the relevant rules for the truth predicate. When contexts Γ and Δ are included in the premise-sequent on either side of \Rightarrow (as opposed to weakening them into the conclusion-sequent), then the above procedure doesn’t work because the application of cut must be pushed up through the rule with the implication that the number of times a truth-rule has been applied could increase.

²⁰ The same also holds for infinitary Peano Arithmetic with deontic truth. To show this, it suffices to define suitable models with regard to which the sequent calculus presented in this subsection is sound and complete, and then prove that the desired equivalence is the case in those models. Suitable models are obtained with relatively standard neighbourhood models where the modal operator is replaced with a unary predicate

Since the derivability of the sequent (d) is equivalent to the local validity of the thereby represented metainference, it should be clear that a validity-predicate defined as $Tr(\ulcorner \bigwedge X \supset \bigvee Y \urcorner)$ will result in a validity-predicate that, in virtue of the rules T_K and T_D , immediately satisfies the descending direction of (GVS*).

The ascending direction of (GVS*) is established by showing that the context-free inverses of T_K and T_D are admissible, i.e. the following rules:

$$\frac{Tr(s_0), \dots, Tr(s_n) \Rightarrow Tr(t)}{B_0, \dots, B_n \Rightarrow A \text{ whenever } s_i^{\mathbb{N}} = \ulcorner B_i \urcorner^{\mathbb{N}} \text{ and } t^{\mathbb{N}} = \ulcorner A \urcorner^{\mathbb{N}}} T_{KI}$$

$$\frac{Tr(s_0), \dots, Tr(s_n) \Rightarrow}{B_0, \dots, B_n \Rightarrow \text{ whenever } s_i^{\mathbb{N}} = \ulcorner B_i \urcorner^{\mathbb{N}}} T_{DI}$$

This is established by (transfinite) induction on the height of a derivation defined as the supremum of the height of its subderivations. The inductive argument must include the limit case because applications of the ω -rule lead to derivations of a transfinite height. We illustrate the reasoning to establish the admissibility of T_{KI} . For the base case, it suffices to observe that the premise-sequent is an initial sequent in which one of s_i is such that $s_i = t$, and the desired result therefore follows from the derivability of $A, \Gamma \Rightarrow \Delta, A$. For the successor stage, the premise-sequent can only have been obtained with T_K or T_D in which case there is a derivation of a sequent $X \Rightarrow Y$ where every formula in X is one of B_i and Y contains at most one formula, A . This follows by weakening on the premise-sequent. With regard to the limit case, we note that there is no derivation of the premise-sequent where the height is a limit ordinal (since the last applied rule will in that case be the ω -rule which introduces a quantified formula), so the desired conclusion follows trivially.

Now, the derivability of the sequent (d) is also equivalent to the following rule being admissible:

$$\frac{X', \Gamma \Rightarrow \Delta, Y' \text{ for } \langle X', Y' \rangle \in \Theta}{X, \Gamma \Rightarrow \Delta, Y}$$

To show the admissibility of the rule assuming the derivability of the sequent, we reason as follows using the cut-rule:

$$\frac{\frac{X', \Gamma \Rightarrow \Delta, Y' \text{ for } \langle X', Y' \rangle \in \Theta}{\Gamma \Rightarrow \Delta, \bigwedge X' \supset \bigvee Y' \text{ for } \langle X', Y' \rangle \in \Theta} \quad \{\bigwedge X' \supset \bigvee Y' \mid \langle X', Y' \rangle \in \Theta\}, X \Rightarrow Y}{X, \Gamma \Rightarrow \Delta, Y}$$

Footnote 20 continued

and where the neighbourhoods of each world contain the unit, is proper, and is closed under supersets and binary intersections (cf. Pacuit (2017) for an introduction to this terminology). Presenting the models and the relevant proofs would take us beyond the scope of this paper.

To derive the sequent using the rule, we apply the rule on sequents that are basically instances of modus ponens as inference and thus derivable:

$$\frac{\{\wedge X' \supset \vee Y' \mid \langle X', Y' \rangle \in \Theta\}, X' \Rightarrow Y' \text{ for } \langle X', Y' \rangle \in \Theta}{\{\wedge X' \supset \vee Y' \mid \langle X', Y' \rangle \in \Theta\}, X \Rightarrow Y}$$

Notice that this derivation requires that the rule assumed to be admissible permits contexts. We must thus distinguish between the following two rules:

$$\frac{X', \Gamma \Rightarrow \Delta, Y' \text{ for } \langle X', Y' \rangle \in \Theta}{X, \Gamma \Rightarrow \Delta, Y} \quad \frac{X' \Rightarrow Y' \text{ for } \langle X', Y' \rangle \in \Theta}{X \Rightarrow Y}$$

The left rule implies the right rule, but the right rule does not imply the left rule. Now, as shown by Humberstone (1996), the notion of an admissible rule corresponds to that of a globally valid metainference. However, since the sequent (d) represents a locally valid metainference and the sequent (d) is equivalent to the left rule, it follows that the left rule also represents a locally valid metainference. This corresponds to proposition (a) and (c) in Sect. 2.3: every (single-conclusion) locally valid metainference is globally valid, but some globally valid (single-conclusion) metainferences are not locally valid. The rules T_K and T_D are examples of the right rule that do not imply the corresponding left rule since they do not permit contexts. In our setting, the admissibility of the left rule means that the metainference is not only globally but also locally valid, whereas the admissibility of only the right rule means that the metainference is only globally valid.²¹

It follows that the rules T_K and T_D introduce globally valid metainferences involving truth-predications. This means that the corresponding metainferences for a validity predicate Val defined as $Tr(\ulcorner \wedge X \supset \vee Y \urcorner)$ are also globally valid, which in turn does not guarantee that they are locally valid.

Let us use this wisdom to illustrate how the paradoxical reasoning involving (GVS) in Sect. 2.2 is blocked in this setting where (GVS) is restricted to locally valid metainferences. First, we note that the following rules are admissible:

$$\frac{X \Rightarrow Y}{\Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)} \text{ VP} \quad \frac{\Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)}{X \Rightarrow Y} \text{ VPC}$$

Let as before κ be a closed term such that $\kappa = \{\ulcorner Val(\kappa, \tau) \urcorner\}$ holds in virtue of the strong diagonal lemma where τ is an abbreviation for $\{\ulcorner Val(\ulcorner X \urcorner, \ulcorner Y \urcorner) \urcorner\}$. We now

²¹ Note that the proof of the equivalence between (d) and the left rule with contexts uses cut. In fact, the equivalence may fail for calculi where cut is not admissible, and we can even in that case use cut as an example. Consider for example how cut is admissible for propositional variables (but not for every formula) in the sequent calculus for STT presented by Ripley (2013a). This rule, cut for propositional variables, is globally but not locally valid on models for STT. Note also that the cut rule corresponding to that rule in the labelled sequent calculus presented in Sect. 3.2, a cut on the sequents $\Gamma \Rightarrow \Delta, \tau : p$ and $S : p, \Gamma' \Rightarrow \Delta'$ to obtain $\Gamma, \Gamma' \Rightarrow \Delta, \Delta'$, is not admissible in the labelled sequent calculus.

observe that the following metarule is admissible:

$$\frac{Val(\kappa, \tau) \Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)}{X \Rightarrow Y}$$

This is obtained as follows:

$$\frac{\frac{Val(\kappa, \tau) \Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)}{\Rightarrow Val(\kappa, \tau)} \text{ VP} \quad Val(\kappa, \tau) \Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)}{\Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)} \text{ cut} \\ \frac{\quad}{X \Rightarrow Y} \text{ VPC}$$

Luckily for us, this does not imply the derivability of the sequent

$$Val(\kappa, \tau) \Rightarrow Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)$$

through (GVS*) because the rule is not admissible with contexts. After all, behind the scene we will find applications of T_K . Without that sequent we cannot use the admissible rule to show that arbitrary inferences are valid.

3 Validity predicates for theories definable on strong Kleene models

3.1 Four theories definable on strong Kleene models

The aim of this section is to apply the approach developed in the previous section on four theories definable on strong Kleene models, four theories that are typically seen as competing proposals for what is valid.²² This subsection presents the theories in question.

Using the set $\{0, \frac{1}{2}, 1\}$ with its natural ordering as values, and assuming that every object in the domain is denoted by a term, strong Kleene models for a first-order language based on the connectives \vee, \wedge, \neg and \forall can be represented as a function I from the language to $\{1, \frac{1}{2}, 0\}$ such that,

- $I(\neg A) = 1 - I(A)$
- $I(A \vee B) = \max(I(A), I(B))$
- $I(A \wedge B) = \min(I(A), I(B))$
- $I(\forall x A) = \inf\{I(A(t/x)) : t \text{ is a term}\}$

Strong Kleene models are popular in the literature on semantic paradoxes because they are, as shown by Kripke (1975), compatible with a truth predicate Tr satisfying the condition that $I(A) = I(Tr(\ulcorner A \urcorner))$.

The models can be used to define a variety of theories by varying the conditions for being a premise and for being a conclusion in a sound inference. Consider the following four ways of varying the conditions for a multiple conclusion inference to be satisfied in a model:

²² An exception to a presentation of them as rivals is the pluralist proposal by Hjortland (2013).

- (i) If every premise is assigned 1 then some conclusion is assigned 1
- (ii) If every premise is not assigned 0 then some conclusion is not assigned 0
- (iii) If every premise is assigned 1 then some conclusion is not assigned 0
- (iv) If every premise is not assigned 0 then some conclusion is assigned 1

The conditions give rise to distinct sets of multiple-conclusion inferences. Theories based on these proposals have been presented and explored in more historical, recent and contemporary literature on formal theories of truth such as Priest (1979), Field (2008), Ripley (2012) and Cobreros et al. (2013). In addition, there is also the variant according to which an inference is sound in a model if and only if both (i) and (ii) holds. This can be found for example in the work of Halbach and Horsten (2006).

This paper will focus on the four theories obtained with the conditions (i)-(iv) respectively, thus ignoring the fifth option. The four theories in question are the para-complete theory K_3 , the paraconsistent theory LP, the nonreflexive theory TS and the nontransitive theory ST.²³ We shall in this paper refer to them using the strict-tolerant schema from Cobreros et al. (2013) as ST, TS, SS (=K₃) and TT (=LP).

The four theories disagree for example over whether modus ponens holds as inference or metainference:

- Modus ponens as inferential schema is valid SS and ST but invalid in TT and TS.
- Modus ponens as metainferential schema is valid in SS and TS but invalid in TT and ST.

Our objective then, is to express these disagreements within the theories using the approach developed in the previous section. The rest of the section is therefore organised as follows. Section 3.2 presents a labelled sequent calculus for the four theories expanded with transparent truth. The calculus is designed for the purpose of expressing locally valid metainferences through labelled internal sequents. Section 3.3 shows that the material conditional together with transparent truth is not suitable to define validity predicates for those theories of truth. Section 3.4 employs the labelled internal sequents to define validity predicates satisfying (GVS*) for the four theories of truth and finally illustrates that we can express the above disagreements within each theory.

3.2 A labelled sequent calculus for strict and tolerant satisfaction

The strict-tolerant schema is based on the distinction between strict and tolerant satisfaction, where a formula is strictly satisfied if assigned 1 and tolerantly satisfied if assigned 1 or $\frac{1}{2}$. With $I \Vdash_S A$ and $I \Vdash_T A$ representing respectively that A is strictly and tolerantly satisfied at I , it follows that

²³ The main reason for ignoring the fifth option sometimes referred to as PKF is that the relevant sequent calculus rules wouldn't be as elegant as for the other options in the current setting. Also, it will keep the discussion more streamlined. In any case, it should be clear from the discussion in the following subsections that a validity predicate defined as $Tr(\ulcorner A \supset B \urcorner)$ in PKF will have the same issues with regard to satisfying (GVS*) as a validity predicate defined using the material conditional in the four theories under consideration. In addition, the approach to define validity predicates for the four theories that this paper proposes is also straightforwardly generalised to PKF.

if $I \Vdash_S A$ then $I \Vdash_T A$

Moreover, the above clauses imply the following where χ and γ are uniformly substituted with either S or T:

- $I \Vdash_\chi A \wedge B$ iff $I \Vdash_\chi A$ and $I \Vdash_\chi B$
- $I \Vdash_\chi A \vee B$ iff $I \Vdash_\chi A$ or $I \Vdash_\chi B$
- $I \Vdash_\chi \neg A$ iff $I \not\Vdash_\gamma A$ where $\chi \neq \gamma$
- $I \Vdash_\chi \forall x A$ iff $I \Vdash_\chi A(t/x)$ for every term t

With those principles at hand it is easy to see that the following labelled sequent calculus with {S, T} as labels so that $S : A$ and $T : A$ are labelled formulas based on the calculus presented by Fjellstad (forthcoming) is sound and complete with regard to such models:

$$\begin{array}{c}
 \chi : P, \Gamma \Rightarrow \Delta, \chi : P \quad S : P, \Gamma \Rightarrow \Delta, T : P \\
 \frac{\chi : A, \Gamma \Rightarrow \Delta \quad \chi : B, \Gamma \Rightarrow \Delta}{\chi : A \vee B, \Gamma \Rightarrow \Delta} \vee_L \quad \frac{\Gamma \Rightarrow \Delta, \chi : A, \chi : B}{\Gamma \Rightarrow \Delta, \chi : A \vee B} \vee_R \\
 \frac{\chi : A, \chi : B, \Gamma \Rightarrow \Delta}{\chi : A \wedge B, \Gamma \Rightarrow \Delta} \wedge_L \quad \frac{\Gamma \Rightarrow \Delta, \chi : A \quad \Gamma \Rightarrow \Delta, \chi : B}{\Gamma \Rightarrow \Delta, \chi : A \wedge B} \wedge_R \\
 \frac{\Gamma \Rightarrow \Delta, \chi : A}{\gamma : \neg A, \Gamma \Rightarrow \Delta} \neg_L (\chi \neq \gamma) \quad \frac{\chi : A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \gamma : \neg A} \neg_R (\chi \neq \gamma) \\
 \frac{\chi : A(x/t), \Gamma \Rightarrow \Delta}{\chi : \forall x A, \Gamma \Rightarrow \Delta} \forall_L \quad \frac{\Gamma \Rightarrow \Delta, \chi : A(x/y)}{\Gamma \Rightarrow \Delta, \chi : \forall x A} \forall_R (y \text{ free})
 \end{array}$$

In Fjellstad (forthcoming) only the propositional fragment is presented and explored, but extending the completeness proof to the first-order case is straightforward. It follows from soundness and completeness that the four logics above are represented as follows:

$$\begin{array}{c}
 A_0, \dots, A_n \Vdash_{\chi\gamma} B_0, \dots, B_m \\
 \text{if and only if} \\
 \chi : A_0, \dots, \chi : A_n \Rightarrow \gamma : B_0, \dots, \gamma : B_m \text{ is derivable}
 \end{array}$$

To obtain the non-classical theories of truth, we assume that the sequent calculus is defined for the language \mathcal{L}^{Tr} , add the rules

$$\frac{\chi : s = t, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta \text{ if } s^{\mathbb{N}} = t^{\mathbb{N}}} = \text{ and } \frac{}{\chi : s = t, \Gamma \Rightarrow \Delta \text{ if } s^{\mathbb{N}} \neq t^{\mathbb{N}}} \neq$$

for the arithmetic to obtain self-reference, and the following rules for the predicate Tr :

$$\frac{\chi : A, \Gamma \Rightarrow \Delta}{\gamma : u = \ulcorner A \urcorner, \chi : Tr(u), \Gamma \Rightarrow \Delta} TrL \quad \frac{\Gamma \Rightarrow \Delta, \chi : A}{\gamma : u = \ulcorner A \urcorner, \Gamma \Rightarrow \Delta, \chi : Tr(u)} TrR$$

We also include initial sequents of the following forms:

$$\begin{aligned} \gamma : s = t, \chi : Tr(s), \Gamma \Rightarrow \chi : Tr(t) \\ \gamma : s = t, s : Tr(s), \Gamma \Rightarrow \top : Tr(t) \end{aligned}$$

The following derivations may serve to illustrate the calculus in action. They involve the liar sentence obtained with the equality $l = \ulcorner \neg Tr(l) \urcorner$ through the strong diagonal lemma:

$$\frac{\frac{\frac{s : Tr(l) \Rightarrow \top : Tr(l)}{s : \neg Tr(l), s : Tr(l) \Rightarrow} \neg L}{s : l = \ulcorner \neg Tr(l) \urcorner, s : Tr(l) \Rightarrow} TrL}{s : Tr(l) \Rightarrow} = \frac{\frac{\frac{s : Tr(l) \Rightarrow \top : Tr(l)}{\Rightarrow \top : \neg Tr(l), \top : Tr(l)} \neg R}{\top : l = \ulcorner \neg Tr(l) \urcorner \Rightarrow \top : Tr(l)} TrR}{\Rightarrow \top : Tr(l)} =$$

Notice the difference in label. The sequent to the left says that the liar is never strictly satisfied and the sequent to the right says that the liar is always tolerantly satisfied. This is precisely how it ought to be according to the strict-tolerant schema.

Now, adding the arithmetic and the rules for truth means that we cannot (in a straightforward manner that I am aware of at the time of writing) show that the following rule is admissible:

$$\frac{\Gamma \Rightarrow \Delta, \chi : A \quad \chi : A, \Gamma' \Rightarrow \Delta'}{\Gamma, \Gamma' \Rightarrow \Delta, \Delta'} \text{ cut}$$

Now, this is not because the calculus becomes inconsistent as in the case of an unlabelled sequent calculus such as that for STT presented by Ripley (2013a). Instead, the issue is merely that a semantic argument along the lines of Ripley (2012) and Fjellstad (2017) is not available because of the arithmetical content and Gödel’s incompleteness theorems. On the other hand, soundness with the models presented in Cobreros et al. (2013) is sufficient to guarantee that the sequent calculus is consistent with regard to assignment of labels. We therefore add this cut rule to our system as a primitive rule.

The labelled sequent calculus is quite practical from the perspective of this paper. In addition to obtaining the valid inferences of each of the four theories of truth, we can along the lines of Fjellstad (forthcoming) expand the calculus with expressions to represent that an inference is satisfied according to some standard.

Let an expression of the form $\chi\gamma : [X \Rightarrow Y]$ be a labelled internal sequent where X and Y are sets of formulas and $\chi\gamma$ is composed from formula labels. They are governed by the following rules:

$$\frac{\Gamma \Rightarrow \Delta, \chi : A \text{ for every } A \in X \quad \gamma : B, \Gamma \Rightarrow \Delta \text{ for every } B \in Y}{\chi\gamma : [X \Rightarrow Y], \Gamma \Rightarrow \Delta} [\Rightarrow]L$$

$$\frac{\{\chi : A \mid A \in X\}, \Gamma \Rightarrow \Delta, \{\gamma : B \mid B \in Y\}}{\Gamma \Rightarrow \Delta, \chi\gamma : [X \Rightarrow Y]} [\Rightarrow]R$$

A quick check should convince the reader that these rules are adequate to represent that an inference is $\chi\gamma$ -satisfied where $\chi\gamma$ is TT, TS, ST or SS. Moreover, Fjellstad (forthcoming) establishes that we can use internal sequent expressions to represent locally valid metainferences of ST, TS, SS and TT. The following derivation illustrates that the inverse of the right intro rule for \neg is locally TS-valid:

$$\frac{\frac{\frac{\Gamma \Rightarrow \Delta, \neg A}{\Gamma \Rightarrow \Delta, \neg A} \quad \text{TS} : [\Rightarrow \neg A], \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta, \neg A} \quad \text{TS} : [A \Rightarrow]}{\Gamma \Rightarrow \Delta, \neg A} \Rightarrow \text{TS} : [A \Rightarrow]$$

At this point the reader might at first think that contexts are missing from the internal sequents of the sequent $\text{TS} : [\Rightarrow \neg A] \Rightarrow \text{TS} : [A \Rightarrow]$. After all, in an (unlabelled) sequent calculus, it is the following rule that is admissible:

$$\frac{\Gamma \Rightarrow \Delta, \neg A}{A, \Gamma \Rightarrow \Delta}$$

That would however conflate the role of the contexts in such a rule as elaborated on in Sect. 2.4: A rule presented with contexts corresponds to a metainference which is not only globally, but also locally valid.

Using the cut rule we can show that to each locally valid metainference there is a corresponding admissible rule, here illustrated with the same rule:

$$\frac{\frac{\Gamma \Rightarrow \Delta, \neg A}{\Gamma \Rightarrow \Delta, \neg A} \quad \text{TS} : [\Rightarrow \neg A] \Rightarrow \text{TS} : [A \Rightarrow]}{\Gamma \Rightarrow \Delta, \neg A} \quad \frac{\Gamma \Rightarrow \Delta, \neg A \quad \text{TS} : [A \Rightarrow]}{\Gamma \Rightarrow \Delta, \neg A} \quad \frac{\Gamma \Rightarrow \Delta, \neg A \quad \text{TS} : [A \Rightarrow]}{\Gamma \Rightarrow \Delta, \neg A} \Rightarrow \text{TS} : [A \Rightarrow]$$

The converse direction for any metainference $\langle \Theta, \langle X, Y \rangle \rangle$ is established with the following derivation:

$$\frac{\chi\gamma : [X' \Rightarrow Y'], \chi : X' \Rightarrow \gamma : Y' \quad \langle X', Y' \rangle \in \Theta}{\{\chi\gamma : [X' \Rightarrow Y'] \mid \langle X', Y' \rangle \in \Theta\}, \chi : X \Rightarrow \gamma : Y} \Rightarrow \chi\gamma : [X \Rightarrow Y]$$

Notice how the rule assumed to be admissible must permit contexts. After all, we apply the rule on a set of sequents containing more expressions (i.e. $\chi\gamma : [X' \Rightarrow Y']$) than the active expressions of the rule (i.e. $\chi : X'$ and $\gamma : Y'$). The following equivalence thus holds:

$$\frac{\Gamma, \chi : X' \Rightarrow \gamma : Y', \Delta \quad \langle X', Y' \rangle \in \Theta}{\Gamma, \chi : X \Rightarrow \gamma : Y, \Delta} \text{ if and only if } \{\chi\gamma : [X' \Rightarrow Y'] \mid \langle X', Y' \rangle \in \Theta\} \Rightarrow \chi\gamma : [X \Rightarrow Y]$$

This shows that we can use both internal sequent expressions and admissible rules permitting contexts to represent that a metainference is locally valid.

Finally, we can also see with both the derivations as illustrations and their rules that an internal sequent expression $[X \Rightarrow Y]$ behaves in this sequent calculus as the formula $\bigwedge X \supset \bigvee Y$ in the sequent calculus for classical logic from Sect. 2.4 with regard to representing that an inference is satisfied. Importantly, this does not imply that we could manage without internal sequents also in this case by using the formula $\bigwedge X \supset \bigvee Y$ together with the transparent truth-predicate to define a suitable validity predicate for any of the four theories. The next subsection explains why.

3.3 Issues with validity predicates as strong Kleene material conditional

We find in the literature on semantic paradoxes and validity predicates proposals that amount to defining a validity predicate along the lines of a (possibly modalised) material conditional in strong Kleene models or proof-theoretic variants thereof. Examples include Ripley (2013a), Nicolai and Rossi (2018), Murzi and Rossi (2021) and Golan (forthcoming). Such an approach is not suitable for our purposes. This subsection explains why that is the case.

Let $Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)$ be defined as $\bigwedge_{A \in X} Tr(\ulcorner A \urcorner) \supset \bigvee_{B \in Y} Tr(\ulcorner B \urcorner)$, and thus that the special cases of $Val(\ulcorner \emptyset \urcorner, \ulcorner Y \urcorner)$ and $Val(\ulcorner X \urcorner, \ulcorner \emptyset \urcorner)$ are defined as $\bigvee_{B \in Y} Tr(\ulcorner B \urcorner)$ and $\neg \bigwedge_{A \in X} Tr(\ulcorner A \urcorner)$ respectively.

We can now derive the sequent $s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner) \Rightarrow s : Val(\ulcorner \emptyset \urcorner, \ulcorner A \supset B \urcorner)$ as follows:

$$\frac{\frac{\frac{T : A, s : \neg A \Rightarrow \quad s : B \Rightarrow s : B}{s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner), T : A \Rightarrow s : B}}{s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner) \Rightarrow s : \neg A \vee B}}{s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner) \Rightarrow s : Val(\ulcorner \emptyset \urcorner, \ulcorner A \supset B \urcorner)}$$

Let us assume that this sequent tells us that that inference representing a metainference is valid according to SS and that the validity predicate represents SS-validity to the extent that it satisfies (GVS*). Then the following rule should be admissible:

$$\frac{\Gamma, s : A \Rightarrow s : B, \Delta}{\Gamma \Rightarrow s : A \supset B, \Delta}$$

However, assuming this will lead to inconsistent labelling:

$$\frac{s : Tr(l) \Rightarrow s : 0 = 1}{\Rightarrow s : \neg Tr(l) \vee 0 = 1} \quad \frac{s : \neg Tr(l) \Rightarrow \quad s : 0 = 1 \Rightarrow}{s : \neg Tr(l) \vee 0 = 1 \Rightarrow}$$

We obtain with this and similar pieces of reasoning the following four observations, one for each theory:

- (i) The sequent $s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner) \Rightarrow s : Val(\ulcorner \emptyset \urcorner, \ulcorner A \supset B \urcorner)$ is derivable for every formula A and B but the expressed metainference is not locally valid in SS for some formulas A and B .

- (ii) The sequent $\top : Val(\ulcorner A \urcorner, \ulcorner B \urcorner), \top : Val(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow \top : Val(\ulcorner A \urcorner, \ulcorner C \urcorner)$ is not derivable for some formulas A, B and C but the expressed metainference is locally valid in TT for every formula A, B and C .
- (iii) The sequent $\top : Val(\ulcorner A \urcorner, \ulcorner B \urcorner), \top : Val(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow s : Val(\ulcorner A \urcorner, \ulcorner C \urcorner)$ is not derivable for some formulas A, B and C but the expressed metainference is locally valid in TS for every formula A, B and C .
- (iv) The sequent $s : Val(\ulcorner A \urcorner, \ulcorner B \urcorner), s : Val(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow \top : Val(\ulcorner A \urcorner, \ulcorner C \urcorner)$ is derivable for every formula A, B and C , but the expressed metainference is not locally valid for ST for some formulas A, B and C .

In addition, the following observations illustrate the extent to which the validity predicate occurring in one theory is actually about what is valid according to another theory. In the previous subsection we concluded that $\langle \Theta, \langle X, Y \rangle \rangle$ is locally valid according to $\chi\gamma$ if and only if rule

$$\frac{\Gamma, \chi : X' \Rightarrow \gamma : Y', \Delta \quad \langle X', Y' \rangle \in \Theta}{\Gamma, \chi : X \Rightarrow \gamma : Y, \Delta}$$

is admissible. Let now $\chi : V_\Theta$ be $\{\chi : Val(\ulcorner X' \urcorner, \ulcorner Y' \urcorner) \mid \langle X', Y' \rangle \in \Theta\}$ and $V_{\langle XY \rangle}$ be $Val(\ulcorner X \urcorner, \ulcorner Y \urcorner)$. It follows that:

- (i) $\langle \Theta, \langle X, Y \rangle \rangle$ is locally ST-valid if and only if $\top : V_\Theta \Rightarrow \top : V_{\langle XY \rangle}$ is derivable.
- (ii) $\langle \Theta, \langle X, Y \rangle \rangle$ is locally TS-valid if and only if $s : V_\Theta \Rightarrow s : V_{\langle XY \rangle}$ is derivable.
- (iii) If $\langle \Theta, \langle X, Y \rangle \rangle$ is locally TT-valid or locally SS-valid then $s : V_\Theta \Rightarrow \top : V_{\langle XY \rangle}$ is derivable.

For proofs of a proposition similar to (i) but regarding the material conditional, see Dicher and Paoli (2019) and Barrio et al. (2015).

Even if such observations certainly support the pluralist proposal of Hjortland (2013) about theories definable on strong Kleene models, it also shows that such a validity predicate doesn't have the desired features with regard to the aim here.²⁴ Let's move on to another and better proposal.

3.4 Defining validity predicates

We have now reached the finale of this section where the aim is to show how to expand the labelled sequent calculus with rules for validity predicates that satisfy (GVS*) and illustrate how we can prove generalisations which express that the theories disagree over certain inferences and metainferences.

To that purpose we shall introduce four new binary validity predicates into the language: $Val_{SS}, Val_{TT}, Val_{ST}, Val_{TS}$. The basic idea is to pair each validity predicate with its respective labelled internal sequent. In analogy with the rules T_K and T_D from Sect. 2.4, we introduce the following rules where X_\square is a possibly empty set containing only labelled internal sequents, and X_V and X_\equiv are obtained from X_\square by adding $\zeta : Val_{\chi'\gamma'}(u_i, v_i)$ to X_V and the equalities $\zeta' : \ulcorner X_i \urcorner = u_i$ and $\zeta'' : \ulcorner Y_i \urcorner = v_i$

²⁴ See for example Barrio et al. (2016), Dicher and Paoli (2019) and Rosenblatt (2017) for more on these issues with regard to ST.

to $X_{=}$ for each labelled internal sequent $\chi'\gamma' : [X_i \Rightarrow Y_i]$ in X_{\square} where the ζ 's are arbitrary labels (that may differ for each member of X_{\square}). In the case of V_K one also adds $\zeta' : \ulcorner X \urcorner = u$ and $\zeta'' : \ulcorner Y \urcorner = v$ to $X_{=}$:

$$\frac{X_{\square} \Rightarrow \chi\gamma : [X \Rightarrow Y]}{\Gamma, X_{=}, X_V \Rightarrow \zeta : Val_{\chi\gamma}(u, v), \Delta} V_K \quad \frac{X_{\square} \Rightarrow}{\Gamma, X_{=}, X_V \Rightarrow \Delta} V_D$$

ζ in V_K is also an arbitrary label and may differ from every label in $X_{=}, X_V$. Notice the generality of these rules. X_{\square} may contain internal sequents of any kind (and not just of one kind), and $\chi\gamma : [X \Rightarrow Y]$ may be an internal sequent of any kind. The following figure illustrates an application of V_K in which the equalities are removed for readability:

$$\frac{SS : [A \Rightarrow B], TT : [B \Rightarrow C] \Rightarrow TS : [A \Rightarrow C]}{T : Val_{SS}(\ulcorner A \urcorner, \ulcorner B \urcorner), S : Val_{TT}(\ulcorner B \urcorner, \ulcorner C \urcorner) \Rightarrow T : Val_{TS}(\ulcorner A \urcorner, \ulcorner C \urcorner)}$$

Even if these rules are straightforward generalisations of the rules T_D and T_K , we cannot establish the ascending direction of (GVS*) in the same way as in Sect. 2.4. The rules as they stand only guarantee the descending direction of (GVS*) because the labelled calculus has elimination rules, both for removing true equations with the rule labelled = and the cut rule. The argument for the admissibility of the inverses of T_K and T_D (from which the ascending direction of (GVS*) followed) relied on that the calculus did not have any elimination rules.

We therefore add the following rules where X_{\square}, X_V and $X_{=}$ are defined as above:

$$\frac{X_V \Rightarrow \zeta : Val_{\chi\gamma}(u, v)}{\Gamma, X_{=}, X_{\square} \Rightarrow \chi\gamma : [X \Rightarrow Y], \Delta} V_{KI} \quad \frac{X_V \Rightarrow}{\Gamma, X_{=}, X_{\square} \Rightarrow \Delta} V_{DI}$$

Considering the relationship between internal sequents and locally valid metainferences, the four rules V_K, V_{KI}, V_D and V_{DI} are together clearly sufficient for (GVS*) to be satisfied by each theory.

Now, in addition to satisfying (GVS*), we would like to prove certain universal and existential generalisations that we use to express the disagreement between the theories. While existential generalisations are easily sorted out, complications arise in the case of universal generalisations because $\forall R$ requires free variables. If the ω -rule had been admissible for the labelled sequent calculus that we are working with, then this would not have been a problem. After all, we could in that case simply provide a derivation of each instance and then apply the ω -rule to obtain the desired result. However, in order to avoid the complaint that the approach relies on the ω -rule, we will in this section not rely on the ω -rule. Moreover, we will also not follow the standard approach for theories of truth as exemplified in Halbach (2011) which would consist in introducing compositional principles for the validity predicates. Instead, we will here pursue a slightly non-standard approach to obtain the desired universal generalisations.

First, we expand the language with countably many propositional variables, that is, zero-place predicates. Now, let's say that a formula A is *accessible* in a formula B if

and only if B contains the term $\ulcorner A \urcorner$. As applied to the truth-predicate for simplicity, A is accessible in both $Tr(\ulcorner A \urcorner)$ and $Tr(f_{Tr}(\ulcorner A \urcorner))$ where $f_{Tr}(u) = \ulcorner Tr(u) \urcorner$ and $Tr(\ulcorner A \urcorner \dot{\vee} \ulcorner B \urcorner)$ but not in $Tr(\ulcorner Tr(\ulcorner A \urcorner) \urcorner)$ and $Tr(\ulcorner A \vee B \urcorner)$. This generalises in the obvious way to numerals denoting sets of formulas.

Consider now the following two restrictions on applications of TrL and TrR :

- (i) If a labelled formula $\chi : A$ contains a propositional variable p as subformula and $\chi : B$ is the labelled formula replacing $\chi : A$ after an application of TrL/TrR , then p must be accessible in B .
- (ii) If a propositional variable p is accessible in a formula A and $\chi : B$ is the labelled formula replacing $\chi : A$ after an application of TrL/TrR , then p must be accessible in B .

And the following two restrictions on applications of V_K and V_D :

- (i) If a formula A in a labelled internal sequent $\chi \gamma : [X \Rightarrow Y]$ contains a propositional variable p as subformula and $z : B$ is the labelled formula replacing $\chi \gamma : [X \Rightarrow Y]$ after an application of V_K/V_D , then p must be accessible in B .
- (ii) If a propositional variable p is accessible in a formula A occurring in a labelled internal sequent $\chi \gamma : [X \Rightarrow Y]$ and $z : B$ is the labelled formula replacing $\chi \gamma : [X \Rightarrow Y]$ after an application of V_K/V_D , then p must be accessible in B .

To illustrate the restrictions as applied to truth-predications, the result of applying a truth-rule on the labelled formula $\chi : A \vee p$ should not be $\chi : Tr(\ulcorner A \vee p \urcorner)$ but rather $\chi : Tr(\ulcorner A \urcorner \dot{\vee} \ulcorner p \urcorner)$. Correspondingly, applying the rule on $\chi : Tr(\ulcorner A \urcorner \dot{\vee} \ulcorner p \urcorner)$ should not result in $\chi : Tr(\ulcorner Tr(\ulcorner A \urcorner \dot{\vee} \ulcorner p \urcorner) \urcorner)$ but $\chi : Tr(f_{Tr}(\ulcorner A \urcorner \dot{\vee} \ulcorner p \urcorner))$.

The point of these requirements is to make it safe to uniformly replace Gödel-codes of propositional variables with variables in order to permit quantification over the variables. The substitution proceeds through the following rule:

$$\frac{\Gamma(x/\ulcorner p \urcorner) \Rightarrow \Delta(x/\ulcorner p \urcorner)}{\chi : Sent(x), \Gamma \Rightarrow \Delta} \text{PS}$$

Finally, we must also expand the sequent calculus with initial sequents of the following form where the Gödel-codes for X and Y must be constructed so as to ensure the accessibility of propositional variables:

$$\zeta : n = n', \zeta : m = m', \zeta : n = \ulcorner X \urcorner, \zeta : m = \ulcorner Y \urcorner, \zeta : Val_{\chi \gamma}(n, m), \Gamma \Rightarrow \Delta, \zeta' : Val_{\chi \gamma}(n', m')$$

ζ and ζ' are arbitrary formula labels.

The additional initial sequents are required to ensure that connectives behave classically with regard to validity predications. That connectives should behave classically follows from the reliance on a classical meta-theory. Indeed, one virtue with strong Kleene models with regard to the current project is that the connectives behave classically on formulas that are assigned 1 or 0. This is captured proof-theoretically by allowing the label of the validity predication in the initial sequent to differ across \Rightarrow . This trick makes the labels intersubstitutable for validity predications.

The overall aim of this section has been to expand the four theories definable on strong Kleene models with suitable validity predicates in order to express comparisons

between the theories within the theories themselves, focusing on differences that are familiar from the literature on semantic paradoxes. Let’s illustrate what has been achieved.

The four theories disagree over whether modus ponens as inference is valid; modus ponens as inference is valid in SS and ST but invalid in TT and TS. Consider the following formula which arguably expresses this fact:

$$\begin{aligned} &\forall xy((Sent(x) \wedge Sent(y)) \supset (Val_{SS}(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \wedge \\ &\forall xy((Sent(x) \wedge Sent(y)) \supset (Val_{ST}(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \wedge \\ &\exists xy((Sent(x) \wedge Sent(y)) \wedge (\neg Val_{TT}(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \wedge \\ &\exists xy((Sent(x) \wedge Sent(y)) \wedge (\neg Val_{TS}(\{x\} \cup \{x \dot{\supset} y\}, \{y\})) \end{aligned}$$

Let us call that formula *MPi*. The sequents $\Rightarrow S : MPi$ and $\Rightarrow T : MPi$ are derivable. In the case of SS and ST, the following sequents are derivable for some propositional variables *p* and *q*:

$$S : p, S : \neg p \vee q \Rightarrow S : q \quad S : p, S : \neg p \vee q \Rightarrow T : q$$

It follows that the sequents

$$\Rightarrow SS : [p, \neg p \vee q \Rightarrow q] \quad \Rightarrow ST : [p, \neg p \vee q \Rightarrow q]$$

are derivable and thus also that the following sequents are derivable:

$$\begin{aligned} &\Rightarrow \chi : Val_{SS}(\{\ulcorner p \urcorner\} \cup \{\ulcorner p \urcorner \dot{\supset} \ulcorner q \urcorner\}, \{\ulcorner q \urcorner\}) \\ &\Rightarrow \chi : Val_{ST}(\{\ulcorner p \urcorner\} \cup \{\ulcorner p \urcorner \dot{\supset} \ulcorner q \urcorner\}, \{\ulcorner q \urcorner\}) \end{aligned}$$

The universal generalisations can now be obtained through the PS rule since the propositional variables are accessible. Correspondingly, we can use the liar sentence to obtain counterexamples for the existential generalisations. It follows that each of the four theories can express the fact represented by *MPi*.

The four theories also disagree over whether modus ponens as metainference is valid; modus ponens as metainference is valid in SS and TS but invalid in TT and ST. This is arguably expressed with the

$$\begin{aligned} &\forall xy((Sent(x) \wedge Sent(y)) \supset ((Val_{SS}(\{\emptyset\}, \{x\}) \wedge Val_{SS}(\{\emptyset\}, \{x \dot{\supset} y\})) \supset Val_{SS}(\{\emptyset\}, \{y\})) \wedge \\ &\forall xy((Sent(x) \wedge Sent(y)) \supset ((Val_{TS}(\{\emptyset\}, \{x\}) \wedge Val_{TS}(\{\emptyset\}, \{x \dot{\supset} y\})) \supset Val_{TS}(\{\emptyset\}, \{y\})) \wedge \\ &\exists xy((Sent(x) \wedge Sent(y)) \wedge ((Val_{TT}(\{\emptyset\}, \{x\}) \wedge Val_{TT}(\{\emptyset\}, \{x \dot{\supset} y\})) \wedge \neg Val_{TT}(\{\emptyset\}, \{y\})) \wedge \\ &\exists xy((Sent(x) \wedge Sent(y)) \wedge ((Val_{ST}(\{\emptyset\}, \{x\}) \wedge Val_{ST}(\{\emptyset\}, \{x \dot{\supset} y\})) \wedge \neg Val_{ST}(\{\emptyset\}, \{y\})) \end{aligned}$$

Let us name that formula *MPmi*. The sequents $\Rightarrow S : MPmi$ and $\Rightarrow T : MPmi$ are derivable. The proofs for these sequents are reconstructed very much in the same way as above, the twist being that we derive e.g. $SS : [\Rightarrow p]$, $SS : [\Rightarrow \neg p \vee q] \Rightarrow SS : [\Rightarrow q]$ from $S : p, S : \neg p \vee q \Rightarrow S : q$. Each of the four theories can thus express the fact represented by *MPmi*.

4 Concluding remarks

The aim of this paper has been to present an approach which uses validity predicates to express the kind of comparison between theories that can be found in the literature on semantic paradoxes within the theories themselves. The approach allows us to prove formulas that arguably represent the relevant differences between them within each theory under the assumption that the theories under comparison are defined with a classical meta-theory. In line with how the literature on semantic paradoxes has turned its attention to metainferences, the approach takes into consideration not only inferences but also metainferences. The main novelty with the approach is that the representation of metainferences is restricted to finitary locally valid metainferences in order to avoid the triviality result by Hlobil (2018) regarding the representation of valid metainferences. Expanding first-order classical arithmetic with a validity predicate satisfying the new requirement results in a theory which is ω -consistent and conservative over first-order arithmetic.

To further illustrate the approach, Sect. 3 applied the approach to four theories definable on strong Kleene models. Even if there is a sense in which we have through that exercise not learned anything new about the four original theories, it should be stressed that learning something new about them was not the aim; they were used as a case study in order to illustrate the approach.

From a metaphysical perspective, it is worth underlining the extent to which the resulting validity predicates play an expressive role in a clear analogy to claims made in the debate on deflationism about truth with examples including Horsten (1995), Beall (2009) and more recently Picollo and Schindler (2019). One could even use the approach to validity predicates presented in this paper as a springboard towards an account of a deflationary truth predicate for classical logic. Consider the truth-predicate as defined with the multiple-conclusion variant of T_K and T_D . In this case we can obtain a validity predicate equivalent to that defined in Sect. 2.4 by expanding the theory with a KD-modality represented by a unary operator \Box , thus defining the validity predicate as $\Box Tr(\Gamma(A_1 \wedge \dots \wedge A_n) \supset (B_1 \vee \dots \vee B_m) \neg)$. If we replace the D axiom with the T axiom, then we can prove that validity is truth-preserving. This truth predicate arguably plays an expressive role with regard to the representation of valid inferences and locally valid metainferences. This role, in turn, could then be used to justify the requirement that the truth predicate should be conservative over the base theory.²⁵

Acknowledgements It took me a while to write this paper, and different bits and pieces of the material has been presented at various workshops. I would thus like to thank the audiences for their comments and questions at SADAF Workshop on Metainferences and Substructural Logics in Buenos Aires (2019), the Bergen Workshop on Logical Disagreements (2019), the SADAF Workshop on Validity and Metainferences in Buenos Aires (2018), the Workshop on New Perspectives on Truth and Deflationism at University of Salzburg (2018), the Workshop on Current Topics in Disagreement at University of Bologna (2017) and

²⁵ Importantly, it is the kind of expressive role in question, i.e. what is being expressed, that justifies conservativeness as requirement. Consider for example the proposal presented by Picollo and Schindler (2019) according to which the truth predicate's role as expressive device relates to the imitation of certain second-order machinery. With such an expressive role, conservativeness is no longer a desired property.

finally the SADAF W6 Buenos Aires Workshop on Philosophy of Logic (2017). I would also like to thank all the reviewers for their careful reading and very helpful comments.

Funding Open access funding provided by University of Bergen (incl Haukeland University Hospital).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Arai, T. (1990). Derivability conditions on Rosser's provability predicates. *Notre Dame Journal of Formal Logic*, 31(4), 487–497. <https://doi.org/10.1305/ndjfl/1093635585>.
- Bacon, A. (2013). Non-classical metatheory for non-classical logics. *Journal of Philosophical Logic*, 42(2), 335–355. <https://doi.org/10.1007/s10992-012-9223-9>.
- Barrio, E., & Da Ré, B. (2018). Truth without standard models: Some conceptual problems reloaded. *Journal of Applied Non-Classical Logics*, 28(1), 122–139. <https://doi.org/10.1080/11663081.2017.1397326>.
- Barrio, E., Pailos, F., & Szmuc, D. (2020). A hierarchy of classical and paraconsistent logics. *Journal of Philosophical Logic*, 49(1), 93–120. <https://doi.org/10.1007/s10992-019-09513-z>.
- Barrio, E., & Picollo, L. (2013). Notes on Ω -inconsistent theories of truth in second-order languages. *Review of Symbolic Logic*, 6(4), 733–741. <https://doi.org/10.1017/s1755020313000269>.
- Barrio, E., Rosenblatt, L., & Tajer, D. (2015). The logics of strict-tolerant logic. *Journal of Philosophical Logic*, 44(5), 551–571. <https://doi.org/10.1007/s10992-014-9342-6>.
- Barrio, E., Rosenblatt, L., & Tajer, D. (2016). Capturing naive validity in the cut-free approach. *Synthese*. <https://doi.org/10.1007/s11229-016-1199-5>.
- Beall, J. C. (2009). *Spandrels of truth*. Oxford University Press.
- Beall, J. C., Glanzberg, M., & Ripley, D. (2017). Liar paradox. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (fall 2017 ed.). Metaphysics Research Lab, Stanford University.
- Beall, J. C., & Murzi, J. (2013). Two flavors of curry paradox. *Journal of Philosophy*, 110, 143–165. <https://doi.org/10.5840/jphil2013110336>.
- Burgess, J. P. (1986). The truth is never simple. *Journal of Symbolic Logic*, 51(3), 663–681. <https://doi.org/10.2307/2274021>.
- Cantini, A. (1990). A theory of formal truth arithmetically equivalent to ID₁. *Journal of Symbolic Logic*, 55(1), 244–259. <https://doi.org/10.2307/2274965>.
- Cobrerros, P., Égré, P., Ripley, D., & van Rooij, R. (2013). Reaching transparent truth. *Mind*, 122(488), 841–866. <https://doi.org/10.1093/mind/fzt110>.
- Cook, R. T. (2014). There is no paradox of logical validity. *Logica Universalis*, 8(3–4), 447–467. <https://doi.org/10.1007/s11787-014-0094-4>.
- Dicher, B., & Paoli, F. (2019). St, Ip and tolerant metainferences. In C. Bakent, & T. Ferguson (Eds.), *Graham Priest on dialetheism and paraconsistency*. Springer.
- Field, H. (2008). *Saving truth from paradox*. Oxford University Press.
- Field, H. (2017). Disarming a paradox of validity. *Notre Dame Journal of Formal Logic*, 58(1), 1–19. <https://doi.org/10.1215/00294527-3699865>.
- Fischer, M., & Gatzl, N. (2018). Truth, partial logic and infinitary proof systems. *Studia Logica*, 106(3), 515–540. <https://doi.org/10.1007/s11225-017-9751-y>.
- Fjellstad, A. (2017). Non-classical elegance for sequent calculus enthusiasts. *Studia Logica*, 105(1), 93–119. <https://doi.org/10.1007/s11225-016-9683-y>.
- Fjellstad, A. (2020). Herzberger's limit rule with labelled sequent calculus. *Studia Logica*, 108(4), 815–855. <https://doi.org/10.1007/s11225-019-09878-x>.

- Fjellstad, A. Metainferential reasoning on strong kleene models. *Journal of Philosophical Logic*, pp. 1–18, forthcoming. <https://doi.org/10.1007/s10992-021-09627-3>.
- Friedman, H., & Sheard, M. (1987). An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33(1), 1–21. [https://doi.org/10.1016/0168-0072\(87\)90073-X](https://doi.org/10.1016/0168-0072(87)90073-X).
- Gentzen, G. (1934). Untersuchungen über das logische Schliessen i, ii. *Mathematische Zeitschrift*, 39(176–210), 405–431.
- Girard, P., & Weber, Z. (2019). Modal logic without contraction in a metatheory without contraction. *Review of Symbolic Logic*, 12(4), 685–701. <https://doi.org/10.1017/s1755020319000030>.
- Golan, R. (2021). There is no tenable notion of global metainferential validity. *Analysis*, 81(3), 411–420. <https://doi.org/10.1093/analys/anab032>.
- Golan, R. Metainferences from a proof-theoretic perspective, and a hierarchy of validity predicates. *Journal of Philosophical Logic*, 1, 1–31, forthcoming. <https://doi.org/10.1007/s10992-021-09616-6>.
- Halbach, V. (2011). *Axiomatic theories of truth*. Cambridge University Press.
- Halbach, V., & Horsten, L. (2006). Axiomatizing Kripke's theory of truth. *Journal of Symbolic Logic*, 71(2), 677–712. <https://doi.org/10.2178/jsl/1146620166>.
- Hjortland, O. T. (2013). Logical pluralism, meaning-variance, and verbaldisputes. *Australasian Journal of Philosophy*, 91(2), 355–373. <https://doi.org/10.1080/00048402.2011.648945>.
- Hjortland, O. T. (2021). Theories of truth and the maxim of minimal mutilation. *Synthese*, 199(Suppl 3), 787–818. <https://doi.org/10.1007/s11229-017-1612-8>.
- Hlobil, U. (2018). The cut-free approach and the admissibility-curry. *Thought: A Journal of Philosophy*, 7, 40–48. <https://doi.org/10.1002/tht3.267>.
- Hlobil, U. (2019). Faithfulness for naive validity. *Synthese*, 196(11), 4759–4774. <https://doi.org/10.1007/s11229-018-1687-x>.
- Horsten, L. (1995). The semantical paradoxes, the neutrality of truth and the neutrality of the minimalist theory of truth. In P. Cartois (ed.), *The many problems of realism (Studies in the General Philosophy of Science: Volume 3)*. Tilberg University Press.
- Horsten, L., Leigh, G. E., Leitgeb, H., & Welch, P. (2012). Revision revisited. *Review of Symbolic Logic*, 5(4), 642–664. <https://doi.org/10.1017/s175502031100030x>.
- Humberstone, L. (1996). Valuational semantics of rule derivability. *Journal of Philosophical Logic*, 25(5), 451–461. <https://doi.org/10.1007/BF00257380>.
- Iemhoff, R. (2015). On rules. *Journal of Philosophical Logic*, 44(6), 697–711. <https://doi.org/10.1007/s10992-015-9354-x>.
- Jeroslow, R. G. (1973). Redundancies in the Hilbert-Bernays derivability conditions for gödel's second incompleteness theorem. *Journal of Symbolic Logic*, 38(3), 359–367. <https://doi.org/10.2307/2273028>.
- Kremer, M. (1988). Logic and meaning: The philosophical significance of the sequent calculus. *Mind*, 97(385), 50–72. <https://doi.org/10.1093/mind/XCVII.385.50>.
- Kripke, S. A. (1975). Outline of a theory of truth. *Journal of Philosophy*, 72(19), 690–716. <https://doi.org/10.2307/2024634>.
- Leigh, G. E. (2015). Some weak theories of truth. In K. Fujimoto, J. M. Fernández, H. Galinon, & T. Achourioti (Eds.), *Unifying the philosophy of truth*. Springer.
- Leigh, G. E., & Rathjen, M. (2010). An ordinal analysis for theories of self-referential truth. *Archive for Mathematical Logic*, 49(2), 213–247. <https://doi.org/10.1007/s00153-009-0170-2>.
- Leitgeb, H. (2007). What theories of truth should be like (but cannot be). *Philosophy Compass*, 2(2), 276–290. <https://doi.org/10.1111/j.1747-9991.2007.00070.x>.
- Łełyk, M., & Wcisło, B. (2019). Models of positive truth. *Review of Symbolic Logic*, 12(1), 144–172. <https://doi.org/10.1017/s1755020318000400>.
- McAllister, I. Classical logic is not uniquely characterizable. *Journal of Philosophical Logic*, pp. 1–21, forthcoming. <https://doi.org/10.1007/s10992-021-09633-5>.
- McGee, V. (1985). How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14(4), 399–410. <https://doi.org/10.1007/BF00649483>.
- Murzi, J. (2014). The inexpressibility of validity. *Analysis*, 74(1), 65–81. <https://doi.org/10.1093/analys/ant096>.
- Murzi, J., & Rossi, L. (2021). Naïve validity. *Synthese*, 199(Suppl 3), 819–841. <https://doi.org/10.1007/s11229-017-1541-6>.
- Negri, S. (2011). Proof theory for modal logic. *Philosophy Compass*, 6(8), 523–538. <https://doi.org/10.1111/j.1747-9991.2011.00418.x>.

- Negri, S., & von Plato, J. (2001). *Structural proof theory*. Cambridge University Press.
- Nicolai, C., & Rossi, L. (2018). Principles for object-linguistic consequence: From logical to irreflexive. *Journal of Philosophical Logic*, 47, 549–577. <https://doi.org/10.1007/s10992-017-9438-x>.
- Ohnishi, M., & Kazuo, M. (1959). Gentzen method in modal calculi II. *Osaka Mathematical Journal*, 11, 115–120.
- Pacuit, E. (2017). *Neighborhood semantics for modal logic*. Mit Press.
- Picollo, L., Schindler, T. (2019). Deflationism and the function of truth. *Philosophical Perspectives*. <https://doi.org/10.1111/phpe.12113>.
- Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, 8(1), 219–241. <https://doi.org/10.1007/BF00258428>.
- Rathjen, M., & Sieg, W. (2018). Proof theory. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (fall 2018 ed.). Metaphysics Research Lab, Stanford University.
- Ripley, D. (2012). Conservatively extending classical logic with transparent truth. *The Review of Symbolic Logic*, 5, 354–378. <https://doi.org/10.1017/s1755020312000056>.
- Ripley, D. (2013). Paradoxes and failures of cut. *Australasian Journal of Philosophy*, 91(1), 139–164. <https://doi.org/10.1080/00048402.2011.630010>.
- Ripley, D. (2013b). Revising up: Strengthening classical logic in the face of paradox. *Philosophers' Imprint*, 13.
- Ripley, D. (2015). Comparing substructural theories of truth. *Ergo: An Open Access Journal of Philosophy*. <https://doi.org/10.3998/ergo.12405314.0002.013>.
- Rosenblatt, L. (2017). Naive validity, internalization, and substructural approaches to paradox. *Ergo: An Open Access Journal of Philosophy*. <https://doi.org/10.3998/ergo.12405314.0004.004>.
- Rosenblatt, L. (2019). Noncontractive classical logic. *Notre Dame Journal of Formal Logic*, 60(4), 559–585. <https://doi.org/10.1215/00294527-2019-0020>.
- Rosenblatt, L. (2021). Towards a non-classical meta-theory for substructural approaches to paradox. *Journal of Philosophical Logic*, 50(5), 1007–1055. <https://doi.org/10.1007/s10992-020-09589-y>.
- Scharp, K. (2014). Truth, revenge, and internalizability. *Erkenntnis*, 79(S3), 597–645. <https://doi.org/10.1007/s10670-013-9562-0>.
- Shapiro, L. (2011). Deflating logical consequence. *Philosophical Quarterly*, 61(243), 320–342. <https://doi.org/10.1111/j.1467-9213.2010.678.x>.
- Zardini, E. (2011). Truth without contra(di)ction. *Review of Symbolic Logic*, 4(4), 498–535. <https://doi.org/10.1017/s1755020311000177>.
- Zardini, E. (2013). Naive logical properties and structural properties. *Journal of Philosophy*, 110(11), 633–644. <https://doi.org/10.5840/jphil2013110118>.
- Zardini, E. (2014). Naive truth and naive logical properties. *Review of Symbolic Logic*, 7, 351–384. <https://doi.org/10.1017/s1755020314000045>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.