Synthese (2021) 198 (Suppl 20):S4831–S4858
https://doi.org/10.1007/s11229-018-02078-5

S.I.: PLURALISTIC PERSPECTIVES ON LOGIC

# Is there a neutral metalanguage?

**Rea Golan**[1]

Received: 25 May 2018 / Accepted: 29 December 2018 / Published online: 12 January 2019


**Abstract**
Logical pluralists are committed to the idea of a neutral metalanguage, which serves as a framework for debates in logic. Two versions of this neutrality can be found in the literature: an agreed upon collection of inferences, and a metalanguage that is neutral as such. I discuss both versions and show that they are not immune to Quinean criticism, which builds on the notion of meaning. In particular, I show that (i) the first version of neutrality is sub-optimal, and hard to reconcile with the theories of meaning for logical constants, and (ii) the second version collapses mathematically, if rival logics, as object languages, are treated with charity in the metalanguage. I substantiate (ii) by proving a collapse theorem that generalizes familiar results. Thus, the existence of a neutral metalanguage cannot be taken for granted, and meaning-invariant logical pluralism might turn out to be dubious.

**Keywords** Logical pluralism · Meaning · Quine · Collapse

## 1 Introduction

Many logicians and philosophers of logic are pluralists (e.g., Beall and Restall 2006; Priest 2003; Shapiro 2011, 2014). Namely, they are prone to adopt, at least in rough terms, sayings such as: "There are different yet equally legitimate logics," "There are many contrasting intuitions about logic that lead to the development of different systems," etc. Unlike Quine, even philosophers and logicians who are monists (e.g.,

This work was supported by a Rothschild Postdoctoral Fellowship at the University of Pittsburgh, and an Edelstein Fellowship at The Hebrew University of Jerusalem. I wish to thank Carl Posy and Robert Brandom for their influence on this work. The paper also benefited from fruitful discussions with David Kashtan, Aviv Keren, and Gil Sagi. My thanks also to the participants of the LOGICA 2016 conference, where I presented the main ideas that are discussed below. Finally, I want to thank to anonymous referees for this journal, for their helpful comments.

✉ Rea Golan
reagolan@gmail.com

1 Sidney M. Edelstein Center for the History and Philosophy of Science, Technology and Medicine, The Hebrew University of Jerusalem, Edmund J. Safra Campus, 91904 Jerusalem, Israel

Read 2006) are nevertheless willing to engage in debates over the correctness question in logic, i.e., the question of which logic is the true one. That is, they are willing to consider (hypothetically) the proliferation of logics. They simply claim that at the end of the day their logic wins the debate.

The present paper purports to examine the very possibility of meaningful debates between rival logics. By a "logic," I mean here a logical theory (based on a deductive system), which is concerned mainly with the validity of deductive arguments, and whose canonical application is to reasoning (Priest 2006, ch. 10). Hence, a logical theory, when applied to reasoning, can be correct or incorrect, or at least better or worse (Priest 2006, ch. 12). By "meaningful debates," I mean debates on the correctness of logical theories, where the disputants are not speaking past each other, i.e., debates which express genuine disagreements rather than mere verbal disputes that result from variations in the meanings of the discussed items. Therefore, the kind of logical pluralism that will be at stake here, via examining the possibility of meaningful debates in logic, is meaning-invariant pluralism, according to which the difference between logics is not an effect of the logical constants having different meanings in different logics (Beall and Restall 2006; Hjortland 2013; Dicher 2016b).

Undoubtedly, all those philosophers and logicians mentioned in the first paragraph acknowledge that logic is a serious, rational, and indeed scientific business that deserves genuine discussions. Therefore, they acknowledge that all such debates, as well as the general business of doing logic, must be carried out rationally. "Rationally" might carry here various meanings, but at the very least, it means that such business must be committed to certain logical norms, underwritten by some logic. There should be, in other words, some (several?) logical framework(s) underlying such debates. Within such a framework, we should be able to express and make sense of statements like: "Intuitionistic logic and classical logic agree on decidable domains," , "Paraconsistent and relevance logics don't admit explosion whereas classical and intuitionistic logics do," etc. We should also be able to make sense, within such a framework, of statements about particular logical constants such as: "Intuitionistic and classical negations agree on double-negation introduction, but differ with double-negation elimination," and so on and so forth.

A suspicion arises at this point that such a meta-framework may not really be neutral: it may well be that the metalanguage[1] forces us to adopt the norms and meanings it confers on all logical constants, thereby making the debate merely verbal. All those philosophers and logicians mentioned would agree, it seems to me, that this sneaking suspicion is disturbing. As a response, they all invoke, in various ways, the idea of a *neutral metalanguage* (Priest 2003; Beall and Restall 2006, pp. 99–100; Field 2009; Shapiro 2014, pp. 213–236; Williamson 2014). Whether it is a unique system, certain specific systems, or all possible systems, a neutral metalanguage must exist, so they reply. Hartry Field, for instance, advocates the widest approach. For him, "There is no good argument that in using a logic L to evaluate itself and other logics, L will always come out best in the evaluation" (Field 2009, p. 256).

[1] I use the terms "metalanguage" and "metalanguage logic" interchangeably in this paper for the sake of simplicity, unless stated otherwise. Hence, by "metalanguage" I also mean the logic that comes with it.

The purpose of the present paper is to provide such an argument. Further, it will be shown that in using such a logic L, it is not even clear whether other logics can be expressed or dealt with. This paper proceeds as follows. First, I examine the two versions of the idea of neutrality that are discussed in the existing literature, which I call the agreed upon collection (AUC) version and the metalinguistic version. Second, I discuss Quine's reservations about logical pluralism that stem from his holistic view of meaning. Accordingly, I proceed to consider theories of meaning for logical constants. It turns out that such theories can hardly account for controversies in logic and, therefore, the AUC version is, at the very least, suboptimal. Next, I turn to discuss the metalinguistic version of neutrality, which is akin to theories of meaning that have to do with the notion of translation. I prove that, under some plausible assumptions, this version collapses mathematically. I conclude that neutrality is really hard to achieve, and that this is a genuine challenge for meaning-invariant logical pluralism.

The scope of my discussion is propositional logics, though I think it can be generalized. Accordingly, by the term "logical constants" I mainly refer to connectives rather than quantifiers. The logics to be considered are all deviant (except for classical logic); expansions such as modal logics will not be taken into account. A further restriction is that I consider only logics with reflexive, transitive, and compact consequence relations. Hence, my discussion doesn't cover all possible logics, but it does cover almost all the systems discussed in the literature, such as classical, intuitionistic, relevance, and paraconsistent logics. The scope of the discussion is thus sufficiently broad.[2]

## 2 Two versions of neutrality

One can find two versions of neutrality in the existing literature. In the AUC version, neutrality is achieved by appealing only to inferences approved by the rival logics. To be precise, by this account different logics are predominantly not at odds, as they agree on the status of most inferences. For example, all logicians accept at least some instantiations of *modus ponens* (MP), albeit not for the same reasons. Thus, there exists some agreed upon collection of inferences on which we can rely neutrally while weighing up pros and cons for different logics, because our "framework" (i.e., this collection) does not prioritize some specific logic over others (Priest 2003, pp. 464–465; Beall and Restall 2006, pp. 99–100).

As far as I understand this version, its basic assumption is that most logicians agree as a matter of fact on most inferences, though not for the same reasons. Take for example:

(1)
$$\frac{All\ men\ are\ mortal,\ \ Socrates\ is\ a\ man}{Socrates\ is\ mortal}$$

[2] Some philosophers (e.g., Beall and Restall 2006, p. 91) don't consider non-transitive and non-reflexive systems to be logics at all, on the grounds that their consequence relations have nothing to do with preservation.

Syllogism (1) is endorsed by almost all logicians, even if they differ over the universal quantifier or the status of singular terms. For such disagreements revolve around the *theoretical* reasons as to *why* (1) is valid; they do not question *whether* it is valid as a matter of fact. That said, the path of neutrality is already paved: if we conduct the debate over rival logics by employing only such unquestionable inferences, we don't appear to prioritize one of the logics over the others. Indeed, one might wonder whether this agreed upon collection of inferences is as rich as required for the purpose of conducting such debates, but if it is, we are ostensibly at home.

It is worth noticing that the AUC version doesn't really require a *meta*language, nor does it require a neutral framework within which to conduct debates in logic. Neutrality isn't achieved, by this account, within some *framework*. The second version is different in this regard.[3] As Williamson puts it:

> How can this anarchy of different systems be reconciled with the apparently scientific, unphilosophical nature of logic? The answer lies in the role of metalogic. All these systems are normally studied from within a first-order non-modal metalanguage, using classical reasoning and set theory. Scientific order is restored at the meta-level. Not only are the systems susceptible to normal methods of mathematical inquiry with respect to their syntax and proof theory, their model theory is also carried out within classical first-order set theory. (Williamson 2014, p. 214)

Williamson's favorite metalogic is classical first-order logic. Yet, if another system can do the job, it is welcomed just as well. As I stressed, Field (2009) believes that there is no good argument that a given metalogic will evaluate itself as the best one.

It is also worth noticing that at the meta-level, this approach might turn out to be monistic: we have to have a fixed metalogic in order to render the debates over different systems possible at the object-language level. Williamson is well aware of this:

> A sort of tacit Quineanism seems to be operating at the meta-level. Any deviation from classical first-order non-modal logic is permitted, because it can be given a model theory in a classical first-order non-modal metalogic. The maxim is: be as unorthodox as you like in your object language, provided that you are rigidly orthodox in your metalanguage. This attitude may even encourage the impression that differences in logic are merely notational, or at least somehow superficial, because we are all agreed in our metatheory. Since contemporary mathematical logic is largely metalogic, no wonder it uses agreed, scientific methods. (Williamson 2014, p. 217)

Williamson mentions Quineanism, which I shall address without further ado, but before that let me just point out that the two approaches aren't necessarily at odds. Dummett (1991, pp. 54–55) seems to be offering a combination of them: On the one hand, he maintains the distinction between metalanguage and object language; the discussion is conducted at the meta-level. On the other hand, he does not postulate

[3] What's important in this version, is that neutrality is achieved by appealing to a unifying framework, rather than by a mere collection of inferences. It is less important whether such a framework involves metalinguistic resources such as quantifiers over sets. Hence, a "neutral metalanguage" can be, e.g., first order logic, if proven neutral.

that the metalanguage be neutral. It is sufficient (and of course necessary) that the disputants agree on some forms of inference at the meta-level.

## 3 The Quinean challenge

Against both these versions stands Quine, whose meaning-variance thesis is incompatible with the very notion of neutrality.[4] For Quine, meaning-variance in logic results from the obviousness of logical truths. These truths are so obvious, he contends, that we cannot really change them without changing the *meanings* of the logical constants (Quine 1986, pp. 82–83). Hence, debates in logic cannot be but verbal. Take for example the debate over paraconsistent logics. Quine comments that:

> My view of this dialogue is that neither party knows what he is talking about. They think they are talking about negation, '¬,' 'not'; but surely the notation ceased to be recognizable as negation when they took to regarding some conjunctions of the form '$p \wedge \neg p$' as true, and stopped regarding such sentences as implying all others. Here, evidently, is the deviant logician's predicament: when he tries to deny the doctrine he only changes the subject. (Quine 1986, p. 81)

If deviant logics differ on the meanings they ascribe to logical constants, then in any such debate the two sides simply talk past each other, and there is no room for genuine disagreement. The challenge Quine poses to logical pluralism is then the fear of *shifts of meaning*: changing the logic entails changing the meanings of all logical constants.[5]

This challenge pertains to both versions of neutrality, though not to the same effect. First, it threatens to dismiss the very idea behind the AUC version: if difference in logic entails difference in meaning, then ostensible agreements on some inferences don't guarantee genuine agreements. It might be the case that rival logicians who say that they agree on this or that rule simply misunderstand one another's words. To prevent this from happening, it seems that we have to fix the meanings of all logical constants. But the moment we do so, the discussion is arguably over. As for the metalinguistic version, things are more complicated. If we fix the metalanguage, then (by Quine's account) we fix the meanings of all logical constants according to that logic. As a result, though we may discuss the mathematical differences between various systems at the object-language level (weighing up the pros and cons of rival systems, etc.), we cannot consider them to be "logics" in a substantive sense, i.e., in a sense that avoids shifts of meaning and makes room for disputes that I called "meaningful," i.e., not merely verbal.

[4] It is commonly accepted that Quine's meaning-variance thesis objects to any kind of logical pluralism, as it implies that debates in logic cannot even be expressed. Yet, an anonymous reviewer suggested that, presumably, one could also see Quine as a kind of meaning variant pluralist. I will not make a stand on this issue.

[5] It is worth pointing out that Quine does *not* want to make logic impossible to revise, nor does he want to argue that there cannot be any rivalry in logic. Indeed, different logics are put to use in the exact sciences as a matter of fact. It's only that there is no way to make sure that no shifts of meaning happen when one logician talks to her opponents in such a debate (see Priest 2006, ch. 10 for a detailed discussion). I wish to thank an anonymous reviewer for clarifying this point.

As a result, each version should deal with Quine's challenge differently. In the AUC version, we should somehow fix the meanings of logical constants in such a way that there is room for controversy. Namely, we should find a theory of meaning for the connectives that prevents shifts of meaning and allows for meaningful disputes. In the metalinguistic version, we have to provide an explanation as to why fixing meanings at the meta-level doesn't make the metalanguage intolerant towards the object language logics. The rest of this paper deals with such attempts.

## 4 Fixing the meanings of logical constants

As I said, in the AUC version we have to somehow fix the meanings of logical constants such that there is room for meaningful disputes. To do so, we first have to have a theory of meaning for the logical constants. Before discussing such theories, I wish to make a preliminary remark about what constraints need to be imposed on the definitions of the connectives in order to face the Quinean challenge in the AUC version. As is well known, Quine himself proposed the idea of *meaning holism*, according to which the meaning of a single proposition depends on one's entire language. Meaning holism entails incommensurability, since any difference of opinion yields totally different views regarding the meaning of everything, and so there is no common ground on which to carry out a comparison between rival views.

Analogously, the Quinean challenge regarding logical pluralism cannot be overridden if meaning is conferred on logical constants *holistically*, i.e., if the meaning of each constant somehow depends on the meanings of them all, in which case even the slightest difference in one constant entails the most colossal difference in all other constants. In other words, if meaning is conferred holistically, then any disagreement about some constant results in totally different views regarding all constants, so that any note of dissent in logic entails incommensurability, thereby undermining the existence of an agreed upon collection of inferences and with it the AUC position.

Let me elaborate on this point. In the AUC version, we should be able to make sense of statements such as:

(2) Intuitionistic logic and classical logic agree on the double-negation introduction rule and differ on the double-negation elimination rule.

For (2) to make sense, the meaning of "negation" (or at least of "double negation") should be fixed in a way that leaves room for expressing the dispute. If, for example, the meaning of "double negation" is fixed solely by the introduction rule, we're out of jeopardy, for in that case classical logic and intuitionistic logic share the meaning of "double negation" and have a difference of opinion in regard to the *same* logical constant, the meaning of which is already fixed. By contrast, if the meaning of "double negation" depends on the meanings of all other logical constants, and given that these logics differ on the meanings of many of them, the meaning of "double negation" cannot be agreed upon. For even the slightest difference of opinion in regard to, say, implication leads to classical logic and intuitionistic logic attributing different meanings to "double negation", thereby rendering any agreement impossible.

To prevent this from happening, I suggest that the meanings of logical constants be fixed in ways that I shall call *molecular* and *modular*, where:

(i) The meaning of a logical constant is fixed *molecularly*, or *non−holistically*, if it is fixed independently of all other constants.
(ii) The meaning of a logical constant is fixed *modularly* if it has several "components" to it, such that it is divisible.

Let me explain this. First, if we want to avoid meaning holism, we need to make sure that meaning is conferred *non−holistically*, i.e., conferred on each constant independently of the others. In this way, we can focus our debates while making sure that they remain meaningful. Suppose that we want to have a meaningful discussion, say, about negation. It would be nothing but natural to make sure that we need not agree about all other constants in order to make the discussion possible; the requirement of molecularity guarantees exactly this.[6]

As for modularity, suppose that we've managed to make sure that our discussion is to revolve solely around negation. What form would the discussion then take? Well, even though meaning is conferred molecularly, we still have to make sure that we are talking about the same thing, and so we still have to agree on some substantial characteristics, while leaving room for the dispute. We may, for example, agree on the introduction rule and differ on the elimination rule, or agree on some truth conditions and differ on others. For such a thing to be possible, $\neg A$ (or any other constant) has to have more than one (substantial) characteristic. This is what modularity guarantees: that meaning is conferred on a given constant in such a way that it is possible to agree on some of its "components" (thereby preventing shifts of meaning), and yet maintain a substantial disagreement.

Fixing the meanings of logical constants molecularly and modularly is thus a good point of departure from which to respond to the Quinean challenge to the AUC version. I therefore turn to discuss theories of meaning for logical constants, the connectives in particular, that arguably meet these requirements. One can find in the literature two theories of meaning for logical constants:

– The *representational* theory accounts for the meanings of logical constants in terms of their *truth conditions*, i.e., their contributions to the truth values of propositions.
– The *inferential* theory accounts for the meaninga of logical constants in terms of their *use*, i.e., the roles they play in inferences.

Let us therefore explore whether these theories meet requirements (i) and (ii).

[6] To be more precise, there are cases where meaning is conferred molecularly, but not *atomically*, strictly speaking. For example, in a system where $\neg A$ is *defined* as $A \rightarrow \bot$, any disagreement about negation would stem from, and depend on, a corresponding disagreement, either about the conditional or about absurdity. Hence, there is no way of having a meaningful dispute only about negation. Needless to say, this doesn't rule out a scenario where the conditional is itself disputable, and meaningfully so. (By the way, negation isn't unique in this regard: there are other examples like, e.g., the biconditional and exclusive disjunction.) There might also be cases where two constants are interdefinable, such that meaning is conferred on both somewhat non-molecularly. I take it that *if* there are such cases, then meaningful disputes are still possible *insofar as* the meaning of a given constant doesn't depend on the entire system as a whole. In any case, unlike strict atomism, molecularism is flexible enough to accommodate this kind of meaning interdependence.

### 4.1 Fixing meaning representationally

The basic principle of the representational theory is that the meaning of a logical constant, a connective in particular, is given in terms of truth conditions, i.e., in terms of its contribution to the truth values of [the] propositions formed with its help (Tarski 1946; Sher 1991).[7] This is easily demonstrated with truth functional connectives: each cell in a truth table can be looked upon as a different "component" of the connective's truth conditions, or meaning, which may be either agreed upon, or in dispute. On this view, meaning is arguably conferred molecularly and modularly as required: molecularly, since each connective has its own truth conditions, independently of others; modularly, because rival logicians may agree about several cells in a given table while disagreeing about others.

Let me demonstrate this idea with the truth conditions of disjunction in Kleene's strong system (Kleene 1938), given in the form of the following table (where "*N*" stands for a third truth value, other than truth and falsity):

| $\vee$ | $T$ | $N$ | $F$ |
|---|---|---|---|
| $T$ | $T$ | $T$ | $T$ |
| $N$ | $T$ | $N$ | $N$ |
| $F$ | $T$ | $N$ | $F$ |

This table seems to be "in agreement" with the table for classical disjunction. Kleene's logic simply has one more truth value, but his disjunction, as a function reduced to the set of "classical" truth values, is identical to classical disjunction. We may then arrive at the conclusion that:

(3) Kleene's strong disjunction and classical disjunction agree on a great deal of the meaning of disjunction. The disagreement between them simply revolves around whether there are more than two truth values and, if so, how disjunction is to be defined over them.

Statement (3) reflects not only the molecular but also the modular character of the meaning of disjunction: an agreement on the "classical" truth values is sufficient for classical logic and Kleene's system to fix the meaning of disjunction in the same way. If we can achieve similar results with all the logical constants, the Quinean challenge will be met.

Compelling as this notion might be, the following considerations show that it is not so simple:

(i) Ostensible agreement doesn't entail genuine agreement: the truth table for disjunction in Priest's logic of paradox (Priest 1979) accords with Kleene's truth table for disjunction. However, Priest understands the "third" value to be a combination of truth and falsity. For him, "classical logic errs in

[7] This idea isn't restricted to connectives or constants that are truth functional; possible-worlds semantics can also be looked upon as specifying the meanings of logical constants in terms of their contributions to making propositions true (or false, or something else), even if one has to take into account the distribution of truth values in all possible worlds (or states) in order to specify such contributions.

assuming that no sentence can be both true and false" (Priest 1979, p. 226). Under this interpretation, there is no agreement on the "classical" truth *values*, and, as a result, there is no agreement on truth *conditions*. For example, unlike classical disjunction, the LP disjunction may be true even if both its disjuncts are false, as one of them may also be true. Hence, one may contend, similar truth conditions cannot guarantee sameness of meaning unless what one means by the "truth values" themselves is *already* agreed upon.

(ii) Even what may appear as perfectly similar truth conditions doesn't guarantee genuine agreement: *prima facie*, Kleene's three-valued logic and LP have the same truth-values and truth conditions for each of the connectives. The difference lies rather in the consequence relation, i.e., in what's required for truth preservation: in Kleene's strong system, only $T$ is a designated value, whereas in LP, both $T, N$ are designated. That is, an argument from $\Gamma$ to $A$ is valid according to Kleene if for every valuation $v$ such that $v(B) = T$ for all $B \in \Gamma$, it must be that $v(A)$ is also $T$; such an argument is valid according to Priest if for every valuation $v$ such that $v(B) \in \{T, N\}$ for all $B \in \Gamma$, it must be that $v(A)$ is either $N$ or $T$. In other words, it seems that both logics assign the same meanings to the connectives, and it's just that their consequence relations are different. But a moment's reflection will remind us that, according to the representational theory, a consequence relation is defined in terms of $truth$-preservation; the dispute over which values are designated (and need to be preserved in order to guarantee that an argument is valid) thus turns out to be a dispute over which truth-values count as $truth$. And, since the meanings of the connectives are given in terms of their truth-conditions, there can be no agreement on them unless there is already an agreement on what counts as truth.

(iii) As mentioned earlier, not all connectives can be given truth tables; intuitionistic negation serves as a well-known counterexample. Therefore, there arises the question of how to account, on this view, for meaning common to, say, classical negation and intuitionistic negation, given that their truth conditions are radically different.[8]

Considerations like (i)–(iii) reflect a deep problem with the representational theory in regard to the AUC version: the very concept of truth is $theory-laden$, depending on the logic held to be true.[9] By contrast, fixing common meanings via truth conditions naively presupposes that there is a somewhat universal, agreed upon account of truth. In a nutshell, the representational theory (in regard to the AUC version) cannot guarantee neutrality by itself, unless there is a universal, agreed upon notion of truth, regardless of the logic used. As there is no such thing, the representational theory cannot account for the AUC version.

---

8 For detailed discussion of this point, see Priest (2006, pp. 204–206), Hjortland (2013, pp. 363–365).

9 This point is made in more detail in Weber et al. (2016). Dummett (1978, pp. 238–239) makes a similar point: that one cannot convey the meaning of the intuitionistic apparatus to classicists, based on its disputable notion of truth. Thus, he goes on to explain the intuitionistic notions of truth and meaning in terms of the agreed upon notion of mathematical proof. To use our terminology, Dummett moves from the representational conception of meaning to the inferential one, because he views truth as theory-laden.

### 4.2 Fixing meaning inferentially

By the inferentialist account, the meanings of logical constants are fixed by the ways we *use* them, namely, by the roles these constants play in inferences. This account seems molecular too; inferentialists invoke natural deduction systems and sequent calculi to specify the inferential roles of the logical constants, and apparently, those settings account for these roles molecularly: the introduction rules, and perhaps the elimination rules as well (or the left and right rules in a sequent calculus), seem to confer meaning on each constant independently, assuming that the meanings of other constants have already been given (Dummett 1991, pp. 215–216).

Further, these rules seem to confer meaning modularly, since each constant comes with at least two rules: the introduction and elimination rules in natural deduction systems, or the left and right rules in sequent calculi. One may thus look upon the rules for a given constant as different "components" of its meaning, as it were, and so having a debate on some of the rules wouldn't necessarily result in changing the constant's entire meaning.

It thus seems that by the inferentialist account we might be able to make sense of meaningful disputes in logic. For, by such an account, an agreement on some "core" of rules may suffice to block meaning-variance. Therefore, rival logicians may have a meaningful dispute, if they agree to rely only on those inferences approved by both logics. That is, according to the inferential theory, an agreement on such a "core" of meanings/rules may suffice to secure a neutral AUC of inferences. Moreover, unlike the representational theory, the inferential theory makes no use of theory-laden concepts such as "truth." Hence, it supposedly manages to avoid the obstacles that the representational theory faces.

The inferential theory does, however, face a central challenge, which is reflected in Prior's well-known "tonk" connective (Prior 1960). In brief, Prior claims that if rules of inference suffice to confer meaning on linguistic expressions, particularly on logical constants, then we can introduce a new connective, "tonk," the meaning of which is given by the following rules:

(4) Itonk:

$$\frac{A}{A tonk B}$$

Etonk:

$$\frac{A tonk B}{B}$$

With tonk, we can derive anything from any proposition, and so it must be illegal or meaningless for some reason. However, by the inferentialist account there is no such reason: rules suffice to confer meaning by themselves, and that's it. Prior thus concludes that rules cannot confer meaning by themselves.

How does this problem arise? Belnap came up with the following ingenious diagnosis:

> [W]e are not defining our connectives ab initio, but rather in terms of an antecedently given context of deducibility, concerning which we have some definite notions. By that I mean that before arriving at the problem of characterizing connectives, we have already made some assumptions about the nature of deducibility. That this is so can be seen immediately by observing Prior's use of the transitivity of deducibility in order to secure his ingenious result. But if we note that we already have some assumptions about the context of deducibility within which we are operating, it becomes apparent that by a too careless use of definitions, it is possible to create a situation in which we are forced to say things inconsistent with those assumptions. (Belnap 1962, p. 131)

That is, the definition of tonk is problematic because it clashes with certain assumptions about the context of deducibility, i.e., about the structural properties of the consequence relation. Therefore, to prevent tonk-like problems, one has to make sure that the definition of each constant is compatible (or "consistent" in Belnap's terms) with those assumptions about the context of deducibility. More specifically, to solve the problem, some conditions need to be imposed on the definition of each logical constant, to guarantee that it is defined in a way that is compatible with the assumptions about the context of deducibility. In this way, we can remain inferentialists: meaning is indeed conferred by rules, yet rules can *successfully* confer meaning only if they meet such conditions.

Now, if meaning is conferred by rules only if they meet certain conditions regarding the context of deducibility, then the meanings of logical constants are context-sensitive: there is no meaning outside of context, so to speak. To be precise, one should distinguish between two kinds of criteria of meaning: global and local. By a global criterion, the meaning of a given constant is context-dependent. That is, rules can confer meaning only in a pre-fixed context of deducibility, and to the extent that the rules somehow "align" with the nature of this context. By a local criterion, by contrast, one doesn't have to fix the entire context in advance; rather, a local test of meaning speaks only of the rules for the particular connective at hand, regardless of the other connectives or the entire context of deducibility. This is *not* to say that local criteria do not require any structural properties; yet, by such a criterion, a constant can be introduced in any context of deducibility, to the extent that it has some desired properties. There is thus no need to fix the context "in advance," as it were.[10]

Now, I shall argue as follows:

(i) A global criterion reduces the AUC version to the metalinguistic version. Moreover, it may require that we fix the context of deducibility in such a way that the demand of molecularity cannot be met.

(ii) Local criteria are more promising for the AUC version, but at the expense of some theoretical constraints, and only for a rather limited range of cases.

Hence, the AUC version is suboptimal, and hard to reconcile with the inferential theories of meaning. In particular, I shall demonstrate claim (i) on Belnap's own criterion (Belnap 1962), and claim (ii) on Dicher's discussion of meaning-invariant pluralism in Dicher (2016b), which presupposes a local criterion; it will become clear

[10] See Dicher (2016b, pp. 738–739) for a detailed discussion of this point, as well as my discussion below.

that the same considerations apply to other local criteria (e.g., Prawitz 1979; Dummett 1991; Read 2010) as well.

Belnap introduces two conditions on the definition of a constant: conservativeness and uniqueness. Let us first focus on the former. A constant $*$ is governed by *conservative* rules if no new sequents in the $*$-free language are derivable after its addition. Consequently, if a new constant forms a conservative extension, it cannot trivialize the context of deducibility at hand, provided that this context isn't trivial to begin with. By contrast, tonk is meaningless because its rules aren't conservative, as they form a trivial extension.

Conservativeness is a global criterion. That's what Belnap literally means by saying that we define the connectives "in terms of an antecedently given context of deducibility," and that:

> It is good to keep in mind that the question of the existence of a connective having such and such properties is relative to our characterization of deducibility. If we had initially allowed $A \vdash B$(!), there would have been no objection to tonk, since the extension would then have been conservative. Also, there would have been no inconsistency had we omitted from our characterization of deducibility the rule of transitivity. (Belnap 1962, p. 133)

But if meanings are context-dependent, then there is no comparing meanings across contexts, because they are not held fixed across contexts![11] That is, the contextual character of conservativeness reduces the AUC version to the metalinguistic version because merely appealing to a contextless collection of inferences common to some rival logics is no longer meaningful; we first have to fix a common context of deducibility—a proof system—rather than count on a contextless collection of inferences.

More precisely, as many authors point out (Read 2000; Dicher 2016b, to name just two), conservativeness actually depends on the entire system under consideration. That is, a test for conservativeness actually tells us how a connective's rules interact with *all* the other rules in the system, not just the structural ones. This is well demonstrated by examples of two connectives with associated sets of rules, where each forms a conservative extension if added to a given proof system, whereas extending the system with both at once results in a non-conservative extension (Humberstone 2011, p. 658). Hence, conservativeness, if taken as a criterion of meaning, makes the meanings of all constants dependent not only on the structural properties of a given system, but also on all other constants. As a result, if conservativeness is our criterion of meaning, there is no way of fixing a constant's meaning unless the entire *system of logic* in which it lives is also fixed. Needless to say, if this is the case, then the requirement of *molecularity* can never be met, and there is no future for the AUC version.[12]

[11] Restall (2014) disagrees with this point. A more detailed criticism of his position can be found in Dicher (2016b).

[12] An anonymous reviewer objected to this claim, arguing that context-sensitivity doesn't necessarily imply such holism: it may be the case that the operational rules of a given connective determine its meaning on their own, and it's just that whether they successfully do this is measured in relation to other (operational as well as structural) rules. Suppose that a connective # is conservative (= definable) in two distinct contexts and undefinable (= non conservative) in a third, also distinct from each of the first two. This would not

Furthermore, sometimes we cannot even introduce the rival constants in one context. Suppose, for example, that we want to account, inferentially, for meaning common to classical and intuitionistic negations. To meet the requirement of conservativeness, we first have to introduce a *context of deducibility* within which these two constants can somehow share meaning by way of agreement on some "core" of rules. However, many times, particularly in the case of our negations, one of the connectives turns out to be *unique*. A logical constant $*$ is unique if any other constant conforming to the $*$-rules turns out to be identical to $*$. This phenomenon is known as "collapse": we introduce two different constants, one of which is designed to be "stronger," yet it turns out that the context forces to two to be equal. Specifically, if we introduce classical negation and intuitionistic negation in the same context of deducibility, the latter collapses into the former.[13] So the two "rival" constants cannot live together within the same context.[14]

How about the local criteria? By such a criterion, an adequate constraint on a meaningful logical constant is a *local* constraint on *its* inference rules, rather than a *global* constraint on the entire *system*. As I implied, the context of deducibility might be important also for proper formulations of local constraints.[15] That is, by such an account, the constraints on the rules of a given constant would require certain structural properties, yet the constant could be introduced in all of those contexts where the requirements are met. As a result, there is no need to fix a context in advance, since constants can share meaning across contexts, to the extent that they meet certain requirements. Hence, the prospects of the AUC version are better with a local criterion.

However, there is an overarching problem here: there is no such agreed upon criterion, let alone perfect one (see, e.g., Prawitz 1979; Dummett 1991; Read 2010; Dicher 2016a). In fact, many of the debates on those criteria are intertwined with debates between rival logics. A notable example is Dummett's well-known criticism of classical negation. He argues, based on his criterion, that "intuitionistic negation. . . has come out of our inquiry very well. . . [but] classical negation. . . is not amenable to any proof theoretic justification procedure based on laws which may reasonably be regarded as self-justifying" (Dummett 1991, p. 299). Whether or not this conclusion is correct, it is clear that the AUC version cannot meet the Quinean challenge when applied to those debates where, essentially, the underlying account of meaning is also

Footnote 12 continued
mean that # has one meaning in the first and another in the second. It would just mean that # has no meaning in the third.
However, as Dicher (2016b, pp. 734–737) argues, the operational rules themselves are not quite the same in different contexts. For instance, the additive and multiplicative rules for disjunction (which are discussed below) turn out to be identical in some contexts and different in others. Thus, then the rules do confer different meanings in different contexts. Indeed, Dicher introduces later on in his paper some criteria for identifying rules across contexts, but his suggestion (which is discussed below) relies on a local version of conservativeness, which is not at stake at this moment.

13 To be exact, given our assumption that the consequence relation is transitive, the two negations cannot be combined in a conservative way. See Cerro et al. (2013) for more detail.

14 I will show later, in regard to the metalinguistic version, that a similar problem arises even if no connective is unique.

15 See Hjortlang (2012) for a detailed discussion of this point.

at issue. Moreover, not every such criterion is applicable to every debate. That is, the AUC strategy doesn't work in certain cases unless we adopt a *specific* criterion, that accounts for the connective at issue as meaning-sharing. Thus, I'm about to claim, the AUC version is suboptimal; it can guarantee neutrality, but only at the expense of some theoretical constraints, and for a rather limited range of cases. To be specific, the AUC strategy is applicable only if: (i) the rival parties first reach agreement on a (local) criterion of meaning, and (ii) the connectives at issue are meaning-sharing according to this criterion.

Let me explain this in more detail, using Dicher's discussion in Dicher (2016b). Dicher's criterion is a local version of conservativeness. By his criterion, a connective is meaningful if it is unique, and forms a conservative extension with respect to the system containing, besides its rules, *only* the basic structural rules of Cut and Identity, which are supposed to guarantee (respectively) that the consequence relation at hand is transitive and reflexive (Dicher 2016b, p. 749).[16] Based on this criterion, Dicher introduces the meaning-individuation thesis that: "the meaning of a connective is given by the rules which define it conservatively and uniquely while at the same time inducing no more structural properties than are required for the connective's definability" (Dicher 2016b, p. 745).[17]

This thesis can secure the AUC strategy in some cases. Consider the following sequent rules for negation:

$$L\neg : \frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta},\quad R\neg : \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta}$$

[16] There are several versions of these rules, but they are all along the lines of:

$$Cut : \frac{\Gamma, A \Rightarrow, \Delta \quad \Gamma \Rightarrow A, \Delta}{\Gamma \Rightarrow \Delta} \quad Id : \ A \Rightarrow A$$

Needless to say, one has to make up one's mind on the version that fits best with the context of deducibility at hand. In any case, I must say I am somewhat reluctant to waive global conservativeness: if we give up this requirement, then a connective can be meaningful even where it forms a non-conservative (let alone inconsistent) extension with respect to a given system. Dicher is well-aware of this, acknowledging that a connective can be meaningful even if it cannot be used in a given system: "[T]here is. . . a gap between a connective being endowed with a coherent meaning and it being usable in a determinate logical system." Yet, he adds: "This is not something that should worry us. It is just another illustration of the truism that whether a connective is usable *hic et nunc* depends on more than just its defining rules" (Dicher 2016b, p. 749).

*But*, recall the inferential theory in the background of this story. The basic insight of the inferential theory is that the meaning of an expression is to be given in terms of its *use*. Hence, a distinction between a connective being endowed with a *coherent meaning*, on the one hand, and it being *usable* in a given system, on the other, is problematic by the very standards of inferentialism. By these standards, in cases where some connective is usable in one system and unusable in another, it should be conceived of as meaningful in the first, and meaningless in the second. Ultimately, the inferential theory is a theory of *use*, and so, according to this theory, an expression unusable in a given context should not be considered meaningful in it. Yet, I cannot establish this point within the scope of the present discussion, and so I leave it to the judgment of the reader. See also Dicher (2016a).

[17] To be precise, Dicher distinguishes between *intrinsic* structural properties, which are induced by the connective rules, and *extrinsic* properties, which are in charge of the connective's interaction with other connectives. What's important for our purposes is only the former, i.e., structural properties that "belong" to the rules themselves, so to speak.

In a standard sequent calculus for classical logic, $L\neg$ and $R\neg$ are the operational rules for negation. However, by restricting the right-hand side of the sequents to sets with at most one formula, we get intuitionistic negation.[18] In addition, both sets of rules (the restricted and the unrestricted) define meaningful connectives, as they are each unique and each forms a conservative extension with respect to Cut and Identity. Moreover, Dicher argues that the defining rules of classical and intuitionistic negations are actually *identical*, as the rules of classical negation do *not* really require a structure where more than one formula are allowed on the right.[19]. If this claim goes through, there is no shift of meaning, since the two connectives share a "core" of meaning/rules even though they are introduced in different contexts. Hence, if the classical-intuitionistic debate is carried out only by drawing on such shared resources, as an AUC of inferences, then it is carried out neutrally, and the Quinean challenge of meaning-variance is met.

On the other hand, consider the right-rules for additive and multiplicative disjunctions:

$$\vee_{add} : \frac{\Gamma \Rightarrow A\ (or\ B), \Delta}{\Gamma \Rightarrow A \vee B, \Delta} \quad \vee_{mul} : \frac{\Gamma \Rightarrow A, B, \Delta}{\Gamma \Rightarrow A \vee B, \Delta}$$

Unlike the additive rule, the multiplicative one *requires* more than one formula on the right: multiplicative disjunction *cannot* be defined in contexts where the succedent (right-hand side) of the sequents is restricted to sets with at most one formula. That is, the rules require different structural properties, and so they cannot be accounted for as meaning-sharing according to the above criterion (Dicher 2016b, pp. 749–752).[20]

The latter example shows that reaching agreement on a local criterion of meaning may not be enough to guarantee a neutral AUC of inferences in all cases. Moreover, it demonstrates how, in the AUC version, the business of deciding between rival logical constants is intertwined with choices of criteria of meaning. Hence my conclusion: the AUC version is, at any rate, suboptimal and hard to reconcile with the inferential theories of meanings. Let us explore whether we can do better with the metalinguistic version.

[18] In addition, by restricting the left-hand side (analogously), we get dual-intuitionistic negation, but I shall not discuss this here.

[19] To be precise, his conclusion is that "negation, via its operational rules, determines the structural properties which make swaps possible—that is, the structural property of having an empty slot and the structural property of having (at least) two formula occurrences on the same side" (Dicher 2016b, pp. 743–744)

[20] This claim may strike one as counterintuitive, as the two are really similar. In fact, if Weakening and Contraction hold:

$$WL : \frac{\Gamma \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta,}, \quad WR : \frac{\Gamma \Rightarrow \Delta}{\Gamma, \Rightarrow A, \Delta}, \quad CL : \frac{\Gamma, A, A \Rightarrow \Delta}{\Gamma, A \Rightarrow \Delta,}, \quad CR : \frac{\Gamma \Rightarrow A, A, \Delta}{\Gamma, \Rightarrow A, \Delta}$$

and the context at hand permits more than one formula on each side of the turnstile, the two connectives can be proven to be identical, and the same goes for multiplicative and additive conjunctions. (In addition, it is rather obvious that the issue is much broader than these examples.) In a more recent paper (Dicher 2018), Dicher approaches this particular issue in more detail, using a different criterion of definability, according to which the two connectives do share meaning. But that is exactly my point: not every criterion can be used in every case, and so the business of deciding between logics is intertwined with choices of criteria of meaning.

## 5 On the very idea of a neutral metalanguage

Recall that the Quinean challenge confronts the metalinguistic version in this way: if we fix the meanings of all logical constants following the metalanguage, we can no longer consider the object languages other "logics." Supposedly, the metalanguage takes over, and at most allows us to discuss mathematical differences between systems,rather than consider them to be "logics" in a sense that avoids shifts of meaning (Williamson 2014, p. 217). Namely, the only true "logic" is the logic of metalanguage, and the meanings of logical constants are fixed according to that logic; other systems discussed within this framework are of mere mathematical interest.

How can we be sure that the metalanguage doesn't take over in such a determinative way? This is a hard problem. In what follows, I describe a failed attempt of mine to meet this challenge, for I believe that there is a lesson to learn from this failure. My idea was to meet this version of the Quinean challenge by treating other logics with charity *from within* the metalanguage. That is, my idea was that the Quinean challenge may be answered by metalanguages that are rich enough to *represent*, within themselves, different logics at the object language level. In this way, fixing meanings and norms at the meta-level, even if determinative for the object language level as well, may still leave room for expressing different logics by means of that same meta-level.[21]

This kind of a solution manifests Davidson's translational theory of meaning (Davidson 1984). So far, in considering both the representational and the inferential theories, we have assumed that agreements in logic posit agreements on the meanings of the logical constants. It turns out that we cannot account for such agreements plausibly. The translational theory of meaning treats controversies in a radically different way. By the translational account, instead of fixing meanings mutually, we never try to get out of our own language and the ways we understand things. Rather, we *translate* other people's viewpoints in charitable ways and treat them from within our own language. Having the means to do so, i.e., the means to *represent* other languages and logics within our own language, we may be able to conduct a comprehensive comparison between two such logics without forcing our norms uncharitably, thereby circumventing the Quinean fear of shifts of meaning.

Let us assume, for the sake of simplicity, that there are only two rival logics at stake. To substantiate the Davidsonian strategy, we have to show that a metalanguage can

[21] In this regard, the metalinguistic version is similar to Hjortalnd's intra-theoretic pluralism (Hjortland 2013). The latter is a kind of pluralism that aims to combine more than one consequence relation in a *single* logical theory. Recall, for example, the operational rules for negation that are mentioned above, and that restricting contexts in various ways yields different consequence relations. The result is a single proof system with different definitions for "derivation" . Since all of these consequence relations live in one system, one may claim, there are no shifts of meaning; in a way, the meanings of logical constants are fixed via their use within the agreed upon (minimal) collection of inferences—those inferences that meet all mentioned restrictions.
However, the issue I wish to explore here is quite different: whether such a combining system can be proven neutral for debates in logic. For the latter purpose, it is not enough to show that certain systems combine more than one consequence relation. What is at stake is rather whether one can *neutrally* discuss different logics with the resources provided by such a system. I address this issue at length in the following pages.

really represent these logics. Specifically, such a metalanguage should have the means to represent both the vocabulary of the rival logics and their different consequence relations. As for vocabulary, the task is rather simple: the metalanguage has to include the vocabulary of both of the object languages. Every logical constant in the object languages at issue should be a constant also in the metalanguage. This is crucial: the metalanguage cannot simply refer to the logical constants by naming them, for mere names don't prevent shifts of meaning. If we are to treat other logics with charity, we have to consider their constants as real logical constants, as our constants: what they take to be negation, for instance, is what we take to be negation, and so on and so forth.

This statement also has implications for representing the consequence relations of the rival logics. On the one hand, the metalanguage should have syntactic resources with which to represent other consequence relations, for otherwise it could not compare other logics, and the Quinean threat would remain alive. In particular, it should have the means to express derivation-statements such as "In the logic $L_1$, $A$ follows from $\Gamma$, but not so in the logic $L_2$." On the other hand, as we saw, the meanings of $A$ and the members of $\Gamma$ have to be fixed, if this statement is to express a real controversy, rather than a mere verbal dispute. That is why the metalanguage cannot express such derivation-statements with some sort of a predicate, like Gödel's *Bew* or the *Val* predicate, used in some discussions of Curry's paradox (e.g., Beall and Murzi 2013). For, suppose that we want to express the above statement with two corresponding validity predicates $Val_1$, $Val_2$ for each rival logic; the statement would then take (in the metalanguage) the form: $Val_1(\ulcorner\Gamma\urcorner, \ulcorner A\urcorner) \wedge \neg Val_2(\ulcorner\Gamma\urcorner, \ulcorner A\urcorner)$, where $\ulcorner\Gamma\urcorner$, $\ulcorner A\urcorner$ stand for the names, or Gödel numbers, of $\Gamma$, $A$, respectively. Yet, in this way $A$ and $\Gamma$ are only *mentioned*, and so their meanings cannot be fixed in a way that prevents meaning-variance; namely, the problem from the last paragraph reproduces itself.

Consequently, if we want to avoid any shift of meaning, not only does the metalanguage have to include the same constants of the object languages, but also the resources it uses in expressing derivation-statements must be integral parts of it. That is to say, to prevent shifts of meaning, such derivation-statements cannot just mention the premises and the conclusion at hand by way of naming them. For this reason, I shall use *implications* to represent, or rather express consequence relations, following the tradition which views conditionals as expressing derivations (e.g., Ryle 2009, pp. 249–250; Dummett 1991, pp. 272–274; Read 1994; Brandom 2008, pp. 44–48). Indeed, in a conditional of the form $A \rightarrow B$ the ingredients $A, B$ figure as propositional components, rather than mere names.

Now, conditionals can be said to express derivations only if each inference licenses a corresponding conditional and each conditional indeed mentions a valid inference, i.e., only if those implications admit corresponding deduction theorems:[22]

$$A \vdash B \text{ iff } \vdash A \rightarrow B$$

[22] The following is a somewhat restricted version of the deduction theorem, but it is all that's required, as we assume compactness.

Luckily, this stipulation isn't really restrictive: most if not all logics can and do represent their consequence relations syntactically. This is surely true for all the systems discussed in the literature on logical pluralism.[23]

Let me put more technical flesh on the theoretical bones. Let $L_1, L_2$ be two rival logics, defined over the same language $\mathcal{L}$.[24] I shall now construct a unifying metalanguage $\mathcal{L}'$, along with its logic $L'$, that can hopefully represent these logics, treat them with charity, and make comparisons between them. As I implied, the vocabulary of $\mathcal{L}'$ should include the vocabulary of $\mathcal{L}$. However, as implications will be used to express derivation-statements, we should bifurcate the implication symbol in $\mathcal{L}$ into two implication symbols, $\rightarrow_1, \rightarrow_2$, in $\mathcal{L}'$, which are supposed to represent the different consequence relations of $L_1, L_2$, respectively. The metalanguage $\mathcal{L}'$ may have additional vocabulary, but I shall assume that it doesn't, for the sake of simplicity.[25] Hence, $\mathcal{L}'$ can be defined as the minimal language generated by adding those two implications to $\mathcal{L}$, after subtracting its own implication symbol.

The unifying metalanguage thus has syntactic means—symbols—to represent both $L_1, L_2$. Put formally, it is supposed to do so in the following way. First, we define two translation functions (for $i = 1, 2$), $\tau_i : \mathcal{L} \rightarrow \mathcal{L}'$, by induction:

(5)

$$\tau_i(\alpha) = \begin{cases} \alpha & if\ \alpha = p,\ where\ p\ is\ atomic \\ \tau_i(\beta) * \tau_i(\gamma) & if\ \alpha = \beta * \gamma,\ where\ * \neq \rightarrow \\ \neg\tau_i(\beta) & if\ \alpha = \neg\beta \\ \tau_i(\beta) \rightarrow_i \tau_i(\gamma) & if\ \alpha = \beta \rightarrow \gamma \end{cases}$$

These functions translate expressions of $\mathcal{L}$ into expressions of $\mathcal{L}'$, making as little changes as possible. Then, representing both consequence relations amounts to the stipulation that (for $i = 1, 2$) for all $\alpha, \beta \in \mathcal{L}$ (i.e., all expressions in the original language $\mathcal{L}$),

(6)

$$\alpha \vdash_{L_i} \beta \quad iff \ \vdash \tau_i(\alpha) \rightarrow_i \tau_i(\beta)$$

where $\vdash_{L_i}, \vdash$ stand for the consequence relations of $L_i, L'$ respectively.

If (6) holds, then each implication, $\rightarrow_i$, represents derivations in the corresponding object-language, by marking out regions where the unifying metalanguage behaves like $L_i$, so to speak. Indeed, there is only one turnstile, the metalinguistic turnstile, as the Davidsonean methodology insists that we cannot get out of our own language in trying to understand other languages/logics. But we can nevertheless specify regions

[23] This is clearly the case with classical, intuitionistic, and relevance logics, but such implications can also be introduced into paraconsistent logics (Hewitt 2008). For the general question of having deduction theorems in propositional logics, see Pigozzi (2001). As the scope of this work is propositional logics, I disregard the problem with free logics.

[24] By "rival logics" I mean that there exist $\alpha \in \mathcal{L}, \Gamma \subseteq \mathcal{L}$ such that $\Gamma \vdash_{L_1} \alpha, \Gamma \nvdash_{L_2} \alpha$.

[25] This assumption isn't necessary for the results that follow.

where our language/logic behaves like the languages/logics we are trying to understand. In this way, the metalanguage will hopefully allow us to make comparisons between the rival logics $L_1$, $L_2$, weigh up the pros and cons of each system, etc. For instance, assume that $L_1$ is classical logic, $L_2$ is intuitionistic logic, and $\alpha, \beta \in \mathcal{L}$ do not involve implications (and so: $\tau_i(\alpha) = \alpha, \tau_i(\beta) = \beta$ for $i = 1, 2$). We want to have: $\alpha \rightarrow_2 \beta \vdash \alpha \rightarrow_1 \beta$, which is supposed to mean in the metalanguage: "I can prove that if $\beta$ is derivable from $\alpha$ in intuitionistic logic, this is also the case in classical logic." On the other hand, we don't want the converse to be true, because classical logic is "stronger" than intuitionistic logic.[26]

From a philosophical point of view, if this strategy goes through, we can say that the metalanguage treats both logics with charity in the sense that:

(i) It can translate the expressions of both logics, as its vocabulary is rich enough.
(ii) These translations/interpretations are as conservative as possible: by the metalinguistic account, the different logics at the object-language level confer the same meaning as we do on the different logical constants except for the implications. The difference thus lies in the consequence relations.
(iii) Thus, the difference between the rival logics and ours is reduced to a minimum by this method: different logics simply argue over inferences, and not over meanings.

Now recall Davidson's famous words:

> [I]f all we know is what sentences a speaker holds true, and we cannot assume that his language is our own, then we cannot take even a first step towards interpretation without knowing or assuming a great deal about the speaker's beliefs. Since knowledge of beliefs comes only with the ability to interpret words, the only possibility at the start is to assume general agreement on beliefs. [...] The guiding policy is to do this as far as possible. [...] The method is not designed to eliminate disagreement, nor can it; its purpose is to make meaningful disagreement possible, and this depends entirely on a foundation—some foundation—in agreement. The agreement may take the form of widespread sharing of sentences held true by speakers of 'the same language,' or agreement in the large mediated by a theory of truth contrived by an interpreter for speakers of another language. (Davidson 1984, pp.196–197)

If (6) holds, conditions (i)–(iii) are met. Hence, Davidson's method is well followed by the speakers of the metalanguage since:

- Condition (i) assures that we have at our disposal syntactic resources to translate rival logicians.
- Condition (ii) guarantees maximum agreement on meanings: we interpret different logics as if there were no shifts of meaning with respect to the logical constants, with the necessary exception of implications.

26 If $\alpha, \beta$ do involve implications, the translation is more complex, and we want to have: $\tau_2(\alpha) \rightarrow_2 \tau_2(\beta) \vdash \tau_1(\alpha) \rightarrow_1 \tau_1(\beta)$ and not the converse. Following the Davidsonean methodology, this is the best we can do to make a comprehensive comparison between the two logics: it allows us to compare straightforwardly the introduction and elimination rules for the connectives (regarding both logics), etc.

– Condition (iii) reduces the disagreement to where it intuitively belongs: different logics simply argue over the validity of some inferences; no shifts of meanings are required (see Beall and Restall 2006, pp. 25–31, 97–98).

To sum up, if this strategy goes through, the meanings ascribed to the logical constants are those conferred by the metalanguage, and it is within this language that controversies are spelled out.

I now turn to discuss in detail how such a unifying metalanguage is constructed. (Readers unwilling to work through the technical details are advised to skip to the last paragraph in the next page which begins with "the rest of the technical details," and from there, skim through the rest of the section.) The Davidsonean methodology has led us to the technique of *combining systems* (Gabbay 1996): supposedly, if we can combine two rival logics in a unifying system such that different implications will mark out regions governed by each logic (thereby representing corresponding consequence relations), then condition (6) is met and the Davidsonean methodology is well-followed. This vision behooves us to construct the logic of the metalanguage (namely, $L'$) by imposing the following conditions:

a. $L'$ must be consistent, or at least non-trivial. Otherwise, it is clearly useless. This means that the logics at issue shouldn't generate pernicious inconsistency while representing them together.
b. More concretely, $\vdash$, i.e., the consequence relation of $L'$, must allow for the possibility of representing the axioms of both $L_i$'s in a way that lines up with condition (6). To that end, I take axiomatizations of the rival logics that are $implication-saturated$, i.e., axiomatizations such that the principal operator in each axiom is implication. One can easily turn any axiomatization into an equivalent implication-saturated axiomatization by simply replacing any axiom $\beta$ that is not implication-saturated with infinitely many axioms of the form $\alpha \to \beta$ for all $\alpha \in \mathcal{L}$. Now, if $\alpha \to \beta$ is an axiom of $L_i$, then $\tau_i(\alpha) \to_i \tau_i(\beta)$ will be an axiom of $L'$. From the point of view of $L'$, we can look upon this axiom as representing a derivation in the corresponding object language, because $L_i$ admits a deduction theorem, (i.e., $\alpha \vdash_{L_i} \beta$ iff $\vdash_{L_i} \alpha \to \beta$). That is, each such axiom is interpreted in $L'$ as "$\tau_i(\beta)$ is derivable from $\tau_i(\alpha)$ in $L_i$," namely, as representing a valid derivation in the object-language logic $L_i$.
c. The consequence relation $\vdash$ should also represent the rules of inference of each $L_i$ with $\to_i$. Accordingly, any such rule for some $L_i$:

$$\alpha_1,...,\alpha_n/\beta$$

will be represented by the axiom: $\vdash \bigwedge_{k=1}^{n} \tau_i(\alpha_k) \to_i \tau_i(\beta)$ in $L'$.[27]

[27] To that end, the metalanguage should have a conjunction that obeys the conjunction introduction and elimination rules. I haven't mentioned this assumption explicitly, because it is quite natural to assume. This method moreover requires that both $L_i$'s be compact, as I've already said. I could have, alternatively, just stipulated that $L'$ obeys rules of each rival logic (it doesn't matter for the results that follow), but this would obfuscate the roles that the implications play in representing consequence relations.

d. Each implication will obey a corresponding MP, in accordance with the deduction theorem for each object-language logic.[28]

e. To represent the consequence relations of both $L_i$'s with corresponding implications successfully, $L'$ will also take care of their transitivity,[29] i.e., $L'$ will have the rules (for $i = 1, 2$):

$$\alpha \rightarrow_i \beta, \beta \rightarrow_i \gamma / \alpha \rightarrow_i \gamma$$

Here is what we finally get: $L'$ represents the axioms and rules of inference of both $L_i$'s. Under our assumptions, it is easy to verify that $\alpha \vdash_{L_i} \beta$ entails straightforwardly that $\vdash \tau_i(\alpha) \rightarrow_i \tau_i(\beta)$. Yet, for the implications to represent derivations in the corresponding object languages successfully, namely, for (6) to hold, we should also stipulate the converse, i.e., we should stipulate that (for $i = 1, 2$) for all $\alpha, \beta \in \mathcal{L}$,

(7)

$$\vdash \tau_i(\alpha) \rightarrow_i \tau_i(\beta) \textit{ entails } \alpha \vdash_{L_i} \beta$$

Condition (7) guarantees that any provable conditional in $L'$, whose antecedent and consequent are expressions taken from $\mathcal{L}$, represents a valid derivation in the corresponding $L_i$. This is a crucial point: we may enrich $L'$ with structural rules, or even additional vocabulary along with further, corresponding axioms and rules of inference, but we want its implications to hold just in case they represent valid derivations in the corresponding $L_i$. Otherwise, we cannot know whether some implication represents a valid derivation or not, and the translational strategy fails.

The rest of the technical details, including definitions of semantics for $L'$ and soundness and completeness proofs, can be found in the "Appendix". Now, the informed reader will surely remember, at this point, that there are well-known cases where (7) doesn't hold: so is the case where one logic is weaker yet uniquely characterized (like the classical-intuitionistic scenario). What I prove in the "Appendix" is that under our (quite general and therefore plausible) assumptions, (7) *never* holds. Technically speaking, Theorem 2 in the "Appendix" states that:

**Theorem (collapse)** Let $L_1$, $L_2$ be two rival logics, and L' the unifying system. Assume without loss of generality that for some $\alpha, \beta \in \mathcal{L}$: $\alpha \vdash_{L_1} \beta, \alpha \nvdash_{L_2} \beta$. The representational roles of the implications collapse: $\vdash \tau_1(\alpha) \rightarrow_1 \tau_1(\beta)$ entails $\vdash \tau_2(\alpha) \rightarrow_2 \tau_2(\beta)$.

It thus seems that the Davidsonean strategy fails, and colossally so. Let me give here a hand-waving proof for this theorem (the proof itself can be found in the "Appendix"),

[28] It is worth noticing that: (i) those implications obey MP in the corresponding object languages, since they admit corresponding deduction theorems in these languages, and (ii) these implications *do not* admit the conditional introduction rules (in the metalanguage), which would bring about a collapse immediately. It is also worth noticing that in the case of implication-saturated axiomatizations with deduction theorems, MP should suffice as a unique rule of inference.

[29] I leave open the question whether $L'$ should have as an axiom $\vdash \alpha \rightarrow_i \alpha$, or whether this stipulation should somehow be derivable. In any case, recall our assumption that the consequence relations of both logics are reflexive and transitive.

for there is a philosophical moral to draw from this failure. It turns out that the unifying metalanguage is sound and complete with respect to lattice-based semantics.[30] Moreover, the class of lattices that can serve as models for $L'$ - $C$—consists exactly of those lattices that can serve as models for both the object-language logics. That is, if we designate by $C_1, C_2$ the classes of lattices that serve as models for $L_1, L_2$ (respectively), we get: $C = C_1 \cap C_2$. Assume, without loss of generality, that for some $\alpha, \beta \in \mathcal{L}$: $\alpha \vdash_{L_1} \beta, \alpha \nvdash_{L_2} \beta$. By the completeness of $L_2$, there is some $c \in C_2$ such that $c \models \alpha, c \not\models \beta$. Now, clearly $c \notin C_1$ and so $c \notin C$. Therefore, $C$ lacks counterexample models in which $\alpha$ holds and $\beta$ doesn't. However, since $L'$ is sound and complete with respect to $C$, there is a corresponding derivation, namely: $\alpha \vdash_{L'} \beta$. Ultimately, this will imply that for all $c \in C$: $c \models \alpha \rightarrow_2 \beta$ (see the "Appendix" for the exact details), and so, by completeness: $\vdash_{L'} \alpha \rightarrow_2 \beta$. Hence, $\rightarrow_2$ marks a derivation that only $\rightarrow_1$ is supposed to mark. $\rightarrow_1$ collapses into $\rightarrow_2$ in a similar way.

In words, the idea behind the proof is that, given two rival logics $L_1, L_2$, for at least one such $L_i$, the unifying metalanguage lacks counterexample models for unproven propositions. However, $L'$ is complete, and since it lacks these counterexample models, it can actually prove those propositions that were left unproven in $L_i$. It is worth noticing that this proof pertains to all unifying metalanguages, and not only to cases where one of the logics at issue is weaker, yet uniquely characterized (like the classical-intuitionistic scenario); it holds even in cases where there is no "weaker" logic.[31] Hence, this theorem is a novel result: the collapse isn't caused by uniqueness; rather, it manifests some fundamental intolerance between rival logics, when being put together.

To be precise, this hand-waving proof illustrates the dilemma with which we are confronted: on the one hand, we want $L'$ to be sound and complete. On the other hand, if $L'$ is sound and complete (as it so turns out), it does not represent both logics faithfully. The philosophical moral to draw here, I think, is this: as a result of admitting a deduction theorem, each implication comes along with a somewhat holistic nature that cannot be avoided: it interacts with the other connectives, by way of "representing" all the inferential relations of its own object-language logic. However, such an implication cannot be both used in its connective role and in representing consequence outside of its "home" object language, so to speak, where more expressive resources are available. In other words, the collapse theorem arguably shows that this holistic nature of each implication is sufficient for inducing incommensurability between rival logics: we cannot represent rival consequence relations in the metalanguage, because this unifying system either forces them to agree with each other, or it doesn't represent different consequence relations at all. Consequently, there is no way to conduct a comparison between rival logics with the help of the translational account of meaning, unless (of course) we change the meanings of the logical constants as Quine predicted. Thus, my attempt to account for the metalinguistic version has failed.

[30] Lattice-based semantics (which is a kind algebraic semantics, see Rasiowa 1974) is used here only as a ladder, to use Wittgenstein's metaphor. It may not be the natural semantics for the metalanguage, for several reasons. Nevertheless, the completeness theorem implies that there is a *syntactic* collapse, regardless of what the natural semantics is. Theorem 2 shows, in other words, that you cannot properly represent rival logics (indeed, at the level of syntax) within one meta-framework.

[31] Take for example the case of intuitionistic logic vs. the system R of relevance logic. Intuitionistic logic has explosion, but R has double negation elimination, so there is no "weaker" logic.

## Conclusion

I take it that the above results show that both versions of neutrality rest on somewhat shaky grounds, and that neutrality is really hard to achieve. Yet the question of what conclusion to draw from this remains open. Those who take the existence of neutral grounds to be a precondition for the success of the pluralist project, might consider the above results an argument against logical pluralism, by way of reconstruction of the Quinean thesis that different logics are incommensurable. On the other hand, there are other forms of pluralism which weren't directly criticized here. For one thing, meaning-variant pluralism (such as Carnapian conventionalism) remains a viable option.[32]

Alternatively, one may try to respond to the critical arguments in Sects. 4 and 5. Regarding Sect. 4, the local accounts of meaning and their usefulness for the AUC version haven't been refuted; it was only argued that they work at the expense of some theoretical constraints, and for a rather limited range of cases. Indeed, it might be fruitful to specify the exact limits of this range. As for Sect. 5, perhaps one can try to construct metalinguistic neutrality by providing the metalanguage with some more heavy tools, like doxastic operators that distinguish clearly between the regions governed by the different logics, and a somewhat neutral zone. It is not clear how to get such a construction up and running, nor is it clear that such a construction will block meaning-variance; even if it can be achieved—the resultant metalanguage will be far more complicated than all the familiar logics. One thing is for sure: the existence of a neutral metalanguage cannot be taken for granted.

## Appendix

As the logics at issue are reflexive, transitive, and compact, and admit deduction theorems, we can use lattice-based semantics to give a general account of their semantics. I follow mainly Rasiowa's notion of *implicative algebras* (Rasiowa 1974) with slight differences. Recall that we are concerned here with a language $\mathcal{L}$ whose expressions are generated by some set of connectives over a countable set of atoms to be designated by $At$. Axioms and rules of inference for each logic (over $\mathcal{L}$) are defined in the usual way.

**Definition 1** A *lattice* is a partial order $\langle A, \leq \rangle$ such that for all $a, b \in A$ there exist $a \vee b, a \wedge b \in A$ with:

1. $a \wedge b \leq a, b$ and if $c \leq a, b$ then $a \wedge b \leq c$.
2. $a, b \leq a \vee b$ and if $a, b \leq c$ then $a \vee b \leq c$.

A lattice is *bounded* if there are $0, 1 \in A$ such that $0 \leq a \leq 1$ for all $a \in A$.
For semantics, we have to define *implication lattices*:

**Definition 2** An *implication lattice* is a tuple $M = \langle A, \leq, Tr \rangle$ where:

1. $\langle A, \leq \rangle$ is a bounded lattice.

[32] One could also endorse Hjortland's intra-theoretic pluralism which is mentioned in footnote 22.

2. For each $n$-ary logical constant, $* \in \mathcal{L}$, there is a corresponding $n$-ary operation $*' : A^n \to A$.
3. In particular, $\wedge, \vee, \bot, \top$ stand for conjunction, disjunction, bottom, and top respectively.
4. $Tr \subseteq A$ is a filter with respect to $\leq$, $\top \in Tr$. We call $Tr$ a *truth filter*.

**Definition 3** Let L be some logic over $\mathcal{L}$, and $M = \langle A, \leq, Tr\rangle$ be an implication lattice. A functor $F : \mathcal{L} \to A$ is a *semantic functor* for $L$ if the following conditions are met:

1. For each $n$-ary logical connective, $* \in \mathcal{L}$, and $\alpha_1, \ldots, \alpha_n \in \mathcal{L}$: $F(*(\alpha_1, \ldots, \alpha_n)) = *'(F(\alpha_1), \ldots, F(\alpha_n))$.
2. For each conditional, $\alpha \to \beta$: $F(\alpha \to \beta) \in Tr$ iff $F(\alpha) \leq F(\beta)$. The interpretation of implications is thus constrained by the lattice partial order, though this restriction isn't unique:—we might have various implications meeting this condition. Under this restriction each implication respects the consequence relation, so that deduction theorems can hold, as required.

The points of $A$ are regarded as truth values. We say that $M$ satisfies $\alpha$ ($M \vDash \alpha$) if $F(\alpha) \in Tr$.

**Lemma 1** *Let $F : \mathcal{L} \to A$ be a semantic functor and $\sigma : At \to At$ some permutation on the atomic propositions. We define by induction for any $\alpha \in \mathcal{L}$ $\sigma(\alpha)$—the translation function generated by $\sigma$ in the obvious way. Then $F_\sigma = F \circ \sigma : \mathcal{L} \to A$ is a semantic functor as well.*

The proof of Lemma 1 results trivially from the formality of logic. Hence a lattice is a model for some logic independently of some specific semantic functor. Note that not every lattice is a model for every logic. Soundness and completeness theorems may specify for each logic its class of models in terms of corresponding algebraic properties.

**Definition 4** A logic is said to be *sound* if $\Gamma \vdash \alpha$ implies $\Gamma \models \alpha$, i.e., for each lattice $M$ and semantic functor, $F : L \to M$: $F(\Gamma) \subseteq Tr$ implies $F(\alpha) \in Tr$. A logic is called *complete* if $\Gamma \vdash \alpha$ entails $\Gamma \models \alpha$.

Let $L_1, L_2$ be two *rival logics*, i.e., there are $\Gamma \subseteq \mathcal{L}$, $\alpha \in \mathcal{L}$ s.t. $\Gamma \vdash_{L_1} \alpha$, $\Gamma \nvdash_{L_2} \alpha$. We define the unifying metalanguage $\mathcal{L}'$ and the corresponding logic $L'$ as explained in Sect. 5. We define models for $\mathcal{L}'$ in the usual way.

**Lemma 2** *Let $C_i$ be the class of models for $L_i$ ($i = 1, 2$); then $M$ is a model for $L'$ iff $M \in C_1 \cap C_2$.*

***Proof*** ($\Leftarrow$) Assume $M \in C_1 \cap C_2$. Then, there are two semantic functors for these logics: $F_i : \mathcal{L} \to A$. Define $F : \mathcal{L}' \to A$ as follows. Without loss of generality (following Lemma 1) we assume that $F_1 \restriction At = F_2 \restriction At$. By definition, for each logical constant $*$ except for implication: $F(*(\alpha_1, \ldots, \alpha_n)) = *'(F(\alpha_1), \ldots, F(\alpha_n))$ is well-defined. In the same way, for each implication $F(\alpha \to_i \beta)$ is well defined in terms of the operator $\to_i'$, since $M \in C_i$ (recall Definition 3). By induction, it is trivial to prove that all the rules of inferences and axioms for both logics are represented successfully in this way, for they are preserved by the original $F_i$'s, and each $\rightarrow_i$ represents the corresponding consequence relation. (⇒) Assume $F : \mathcal{L}' \rightarrow A$ is a semantic functor. For each $i$ define: $F_i : \mathcal{L} \rightarrow A$ by $F_i(\alpha) = F(\tau_i(\alpha))$ where $\tau_i$ is the translation function defined in Sect. 5. The rest of the proof is trivial. □

**Corollary 1** *Soundness If $\Gamma \vdash \alpha$ then $\Gamma \models \alpha$.*

***Proof*** By induction on the derivation of $\alpha$ from $\Gamma$. Each step is either a premise (true by definition), or an axiom or rule of inference representing $L_1$, preserved because $M \in C_1$, or an axiom or rule of inference representing $L_2$, preserved because $M \in C_2$, as Lemma 2 shows. □

**Definition 5** For technical reasons, we define a hierarchy of *incomplete* languages (namely, sets of propositions not necessarily closed under the connectives) such that: $\mathcal{L}' = \bigcup_{n \in \mathbb{N}} \mathcal{L}_n$. Let $\mathcal{L}_1 = (\mathcal{L} \backslash \{\rightarrow\}) \cup \{\rightarrow_1\}$, and define $\mathcal{L}_{2n+1}$ $(\mathcal{L}_{2n})$ by induction: $\alpha \in \mathcal{L}_{2n+1}$ $(\mathcal{L}_{2n})$ iff:

$$\alpha \in \mathcal{L}_{2n}\ (\mathcal{L}_{2n-1})\ or\ \alpha = \beta \rightarrow_1 \gamma\ (\alpha = \beta \rightarrow_2 \gamma)\ where\ \beta, \gamma \in \mathcal{L}_{2n+1}\ (\mathcal{L}_{2n})$$

Let $i(n)$ denote the index of the implication symbol (i.e., 1 or 2) added during this enrichment process at the level $\mathcal{L}_n$. That is, $i(n) = \begin{cases} 1 & n\ is\ odd \\ 2 & n\ is\ even \end{cases}$

Before proving completeness, let me just explain what this hierarchy is good for. To prove completeness, we construct a saturated set which induces a lattice, and then show that this lattice is indeed a model for both logics, relying on the completeness theorems of both $L_1, L_2$. In particular, we show by induction that up to the level of $\mathcal{L}_n$ we have a Tarski–Lindenbaum lattice for $L_{i(n)}$. Thus, for the entire language we have a model for $L'$. This will become clear while reading through the proof.

**Theorem 1** *Completeness: If $\Gamma \models \alpha$ then $\Gamma \vdash \alpha$.*

***Proof*** Assume $\Gamma \nvdash \alpha$. We first construct a saturated set $\Gamma^+ \supseteq \Gamma$ s.t. $\Gamma^+ \nvdash \alpha$. Let $\varphi_1$ be an enumeration of $\mathcal{L}_1$ and $\varphi_{n+1} : \mathbb{N} \rightarrow \mathcal{L}_{n+1} \backslash \mathcal{L}_n$, an enumeration of $\mathcal{L}_{n+1} \backslash \mathcal{L}_n$. Then enumerate $\mathcal{L}'$ with $\varphi : \omega \times \omega \rightarrow \mathcal{L}'$ given by $\varphi(n, m) = \varphi_n(m)$. Now, we construct $\Gamma^+$ by induction:

$$\begin{aligned}
\Gamma_{1,0} &= \Gamma \cap \mathcal{L}_1 \\
\Gamma_{n,m+1} &= \begin{cases} \Gamma_{n,m} \cup \varphi(n,m)\} & if\ \Gamma_{n,m} \cup \{\varphi(n,m)\} \nvdash \alpha \\ \Gamma_{n,m} & otherwise \end{cases} \\
\Gamma_n^+ &= \bigcup_{m<\omega} \Gamma_{n,m} \\
\Gamma_{n+1,0} &= \Gamma_n^+ \cup (\Gamma \cap \mathcal{L}_{n+1}) \\
\Gamma^+ &= \bigcup_{n<\omega} \Gamma_{n,0}
\end{aligned}$$

It is easy to verify that $\Gamma^+$ is closed under the consequence relation and that $\Gamma^+ \nvdash \alpha$.

Now define $\delta \leq \gamma$ if $\Gamma^+, \delta \vdash \gamma$, and $\delta \sim \gamma$ if $\delta \leq \gamma$ and $\gamma \leq \delta$. $\sim$ is then an equivalence relation. Define $A_M = \mathcal{L}'/\sim$, so that $\langle A_M, \leq \rangle$ is a bounded lattice, since for all $\delta \in \mathcal{L}'$: $[\bot] \leq [\delta] \leq [\top]$. Define $Tr = \Gamma^+/\sim$. $Tr$ is by definition a truth filter, and so $M = \langle A_M, \leq, Tr \rangle$ is an implication lattice. Define now: $F(\delta) = [\delta]$. All we have to show is that $F : \mathcal{L}' \to A_M$ is a semantic functor for $L'$. If so, we get $Th(M) = \Gamma^+$ by definition, and $M \nvDash \alpha$ as a result. In particular, we have to show that $M$ is a Tarski–Lindenbaum lattice for both logics (see Font et al. 2003, pp. 22–24). If so, we get:

(1) For each $n$-ary operator $*$ : $F(*(\delta_1, \ldots, \delta_n)) = *'(F(\delta_1), \ldots, F(\delta_n))$, where $*$ is the operator defined for semantic functors $F : \mathcal{L}' \to A_M$, where $M \in C_1 \cap C_2$.
(2) For each implication: $F(\delta \to_i \gamma) \in Tr$ iff $F(\delta) \leq F(\gamma)$.
(3) $F$ satisfies all the axioms and the rules of inference.

The proof rests on the completeness of both $L_1, L_2$. We first define for all $n \in \mathbb{N}$ a projection translation function $\eta_n : \mathcal{L}_n \to \mathcal{L} \cup \{p_n^k\}_{k \in \mathbb{N}}$, where $\{p_n^k\}_{k \in \mathbb{N}}$ are countably many new atoms. $\eta_1$ is the identity function. $\eta_{n+1}$ is defined by induction:

$$\eta_{n+1}(\delta) = \begin{cases} \delta & if\ \delta \in At \\ \eta_{n+1}(\beta) * \eta_{n+1}(\gamma) & if\ \delta = \beta * \gamma,\ where * \neq \to_1, \to_2 \\ p_{n+1}^k & if\ \delta = \beta \to_{i(n)} \gamma,\ where\ p_{n+1}^k\ wasn't\ added\ yet \\ \eta_{n+1}(\beta) \to \eta_{n+1}(\gamma) & if\ \delta = \beta \to_{i(n+1)} \gamma \end{cases}$$

All $\eta_n$'s are by definition bijective. I now argue that $\eta_n(\Gamma_n^+)$ is closed under the consequence relation of $L_{i(n)}$. Assume $\eta_n(\Gamma_n^+) \vdash_{L_{i(n)}} \delta$; we are about to show by induction on the length of this proof that $\delta \in \eta_n(\Gamma_n^+)$. For length=1, if $\delta$ is not an axiom then immediately $\delta \in \eta_n(\Gamma_n^+)$. If it is an axiom of $L_{i(n)}$, then $\eta_n^{-1}(\delta)$ is an axiom of $L'$ whose principal operator is not $\to_{i(n+1)}$. Thus, $\eta_n^{-1}(\delta) \in \mathcal{L}_n$. By the construction, $\eta_n^{-1}(\delta) \in \Gamma_n^+$, so $\delta \in \eta_n(\Gamma_n^+)$. For length>1, suppose that $\delta$ is derived from $\delta_1, \ldots, \delta_k$ in $L_{i(n)}$. By the induction hypothesis, $\delta_1, \ldots, \delta_k \in \eta_n(\Gamma_n^+)$. Thus, $\eta_n^{-1}(\delta_1), \ldots, \eta_n^{-1}(\delta_k) \in \Gamma_n^+$. By the conjunction introduction-rule of $L'$: $\Gamma_n^+ \vdash \bigwedge_{1 \leq j \leq k} \eta^{-1}(\delta_j)$. The above mentioned derivation in $L_{i(n)}$ is by the construction represented in $L'$, with $\Gamma_n^+ \vdash \bigwedge 1 \leq j \leq k \eta^{-1}(\delta_j) \to_{i(n)} \eta^{-1}(\delta)$. By MP: $\Gamma_n^+ \vdash \eta^{-1}(\delta)$. Hence, by the construction: $\eta_n^{-1}(\delta) \in \Gamma_n^+$. Therefore, $\delta \in \eta_n^{-1}(\Gamma_n^+)$.

Define now: $F_n : \mathcal{L} \cup \{p_n^k\}_{k \in \mathbb{N}} \to A_M$ by $F_n(\delta) = F(\eta_n^{-1}(\delta))$. It is easy to verify that due to the soundness and completeness of $L_{i(n)}$, together with the representational properties of $L'$, $F_n$ is a well defined semantic functor for $L_{i(n)}$, and $F_n(\mathcal{L} \cup \{p_n^k\}_{k \in \mathbb{N}}) \subseteq A_M$ forms a Tarski–Lindenbaum lattice for $L_{i(n)}$. In conclusion: $M \in C_{i(n)}$, because $A_M \supseteq F_n(\mathcal{L} \cup \{p_n^k\}_{k \in \mathbb{N}})$.

The last result is true for all $n \in \mathbb{N}$. Moreover, $F_n(\mathcal{L} \cup \{p_n^k\}_{k \in \mathbb{N}}) = F(\mathcal{L}_n)$, and hence $F(\mathcal{L}_n) \in C_{i(n)}$. By definition $F(\mathcal{L}_n) \subseteq F(\mathcal{L}_{n+1})$ (because $\mathcal{L}_n \subseteq \mathcal{L}_{n+1}$), and $F(\mathcal{L}_{n+1}) \in C_{i(n+1)}$. To sum up, for all $n \in \mathbb{N} \cup \{0\}$:

$$F(\mathcal{L}_{2n+1}) \in C_1$$
$$F(\mathcal{L}_{2n}) \in C_2$$
$$F(\mathcal{L}_n) \subseteq F(\mathcal{L}_{n+1})$$

These three results give us:

$$\bigcup_{n\in\mathbb{N}} F(\mathcal{L}_{2n}) = F\left(\bigcup_{n\in\mathbb{N}} \mathcal{L}_{2n}\right) = F\left(\bigcup_{n\in\mathbb{N}} \mathcal{L}_n\right) = A_M = F\left(\bigcup_{n\in\mathbb{N}} \mathcal{L}_n\right)$$
$$= F\left(\bigcup_{n\in\mathbb{N}} \mathcal{L}_{2n+1}\right) = \bigcup_{n\in\mathbb{N}} F(\mathcal{L}_{2n+1})$$

But $\bigcup F(\mathcal{L}_{2n+1})_{n\in\mathbb{N}}, \bigcup F(\mathcal{L}_{2n})_{n\in\mathbb{N}}$ form implication lattices that belong to $C_1, C_2$, respectively. So $M \in C_1 \cap C_2$, and $F$ forms a Tarski–Lindenbaum lattice for both logics. Thus, $M$ is a model for $L'$ such that $M \nvDash \alpha$. □

**Theorem 2** *(collapse): Let $L_1, L_2$ be two rival logics, and L' the unifying system. Assume without loss of generality that for some $\alpha, \beta \in \mathcal{L}$: $\alpha \vdash_{L_1} \beta, \alpha \nvdash_{L_2} \beta$. The representational roles of the implications collapse: $\vdash \tau_1(\alpha) \rightarrow_1 \tau_1(\beta)$ entails $\vdash \tau_2(\alpha) \rightarrow_2 \tau_2(\beta)$.*

***Proof*** We first assume that $\alpha, \beta$ do not involve implications, so that: $\tau_i(\alpha) = \alpha, \tau_i(\beta) = \beta$, and show that $\alpha \vdash_{L_1} \beta$ entails $\vdash \alpha \rightarrow_2 \beta$ (this by itself shows that the representational roles of the implications collapse). Assume then that $\alpha \vdash_{L_1} \beta$ where $\alpha, \beta$ do not involve implications. Therefore, $\vdash \alpha \rightarrow_1 \beta$ as $L'$ represents $L_1$ (this direction is straight forward, as stated in Sect. 5). Hence, for all $M = \langle A_M, \leq, Tr\rangle$ such that $M \in C_1 \cap C_2$ and semantic functor: $F : \mathcal{L}' :\rightarrow A_M$: $F(\alpha \rightarrow_1 \beta) \in Tr$. By definition, for all these models: $F(\alpha) \leq F(\beta)$. But then: $F(\alpha \rightarrow_2 \beta) \in Tr$. That is, for all $M \in C_1 \cap C_2$ and $F : \mathcal{L}' \rightarrow A_M$: $M \models \alpha \rightarrow_2 \beta$. By the completeness of $L'$ we get: $\vdash \alpha \rightarrow_2 \beta$ even though $\alpha \nvdash_{L_2} \beta$. The proof for $\alpha, \beta$ that do involve implications is done by induction on the number of implications in $\alpha$ and $\beta$. Since I've already shown that the representational roles collapse, it is left to the reader. □

## References


Beall, J. C., & Murzi, J. (2013). Two flavors of Curry's paradox. *The Journal of Philosophy*, *110*(3), 143–165.

Beall, J. C., & Restall, G. (2006). *Logical pluralism*. Oxford: Oxford University Press.

Belnap, N. D. (1962). Tonk, plonk and plink. *Analysis*, *22*(6), 130–134.

Blok, W. J., & Pigozzi, D. (2001). Abstract algebraic logic and the deduction theorem. Manuscript. http://orion.math.iastate.edu:80/dpigozzi/.

Brandom, R. B. (2008). *Between saying and doing: Towards ananalytic pragmatism*. Oxford, NY: Oxford University Press.

Cerro, D., Farinas, L., & Herzig, A. (2013). Combining classical and intuitionistic logic, or Intuitionistic implication as a conditional. In F. Baader & K. U. Schulz (Eds.), *Frontiers of combining systems: First international workshop, Munich, March 1996* (Vol. 3). Berlin: Springer Science & Business Media.

Davidson, D. (1984). On the very idea of a conceptual scheme. In D. Davidson (Ed.), *Proceedings and addresses of the American philosophical association: 183–198*. Oxford: Oxford University Press.

Dicher, B. (2016a). Weak Disharmony: Some lessons for proof-theoretic semantics. *The Review of Symbolic Logic*, *9*(3), 583–602.

Dicher, B. (2016b). A proof-theoretic defence of meaning-invariant logical pluralism. *Mind*, *125*(499), 727–757.

Dicher, B (2018). Hopeful monsters: A note on multiple conclusions. *Erkenntnis*. https://doi.org/10.1007/s10670-018-0019-3.

Dummett, M. A. E. (1978). *Truth and other enigmas*. Cambridge, MA: Harvard University Press.

Dummett, M. A. E. (1991). *The logical basis of metaphysics*. Cambridge: Harvard University Press.

Field, H. (2009). Pluralism in logic. *The Review of Symbolic Logic*, *2*(02), 342–359.

Font, J. M., Jansana, R., & Pigozzi, D. (2003). A survey of abstract algebraic logic. *Studia Logica*, *74*(1), 13–97.

Gabbay, D. M. (1996). An overview of fibred semantics and the combination of logics. In D. M. Gabbay & F. Guenthner (Eds.), *Frontiers of combining systems* (pp. 1–55). Berlin: Springer.

Hewitt, C. (2008). *Large-scale organizational computing requires unstratified reflection and strong paraconsistency* (pp. 110–124). Berlin: Springer.

Hjortland, O. T. (2012). Harmony and the context of deducibility. In C. D. Novaes & O. Hjortland (Eds.), Insolubles and consequences: Essays in honour of Stephen Read. New York: College Publications.

Hjortland, O. T. (2013). Logical pluralism, meaning-variance, and verbal disputes. *Australasian Journal of Philosophy*, *91*(2), 355–373.

Humberstone, L. (2011). *The Connectives*. Cambridge, MA: MIT Press.

Kleene, S. C. (1938). On notation for ordinal numbers. *The Journal of Symbolic Logic*, *3*(12), 150–155.

Prawitz, D. (1979). *Proofs and the Meaning and completeness of the logical constants* (pp. 25–40). Dordrecht: Springer.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, *8*(1), 219–241.

Priest, G. (2003). On alternative geometries, arithmetics, and logics: A tribute to Lukasiewicz. *Studia Logica*, *74*(3), 441–468.

Priest, G. (2006). *Doubt truth to be a liar*. Oxford: Oxford University Press.

Prior, A. (1960). The Runabout inference ticket. *Analysis*, *21*(2), 38–9.

Quine, W. V. (1986). *Philosophy of logic*. Harvard: Harvard University Press.

Rasiowa, H. (1974). *An algebraic approach to non-classical logics. Studies in logic and the foundations of mathematics* (1st ed., Vol. 78). Amsterdam: North-Holland Publishing Company.

Read, S. (1994). Formal and material consequence. *Journal of Philosophical Logic*, *23*(3), 247–265.

Read, S. (2000). Harmony and autonomy in classical logic. *Journal of Philosophical Logic*, *29*(2), 123–54.

Read, S. (2006). Monism: The one true logic. In D. de Vidi & T. Kenyon (Eds.), *A logical approach to philosophy* (pp. 193–209). Berlin: Springer.

Read, S. (2010). General-elimination harmony and the meaning of the logical constants. *Journal of Philosophical Logic*, *39*(5), 557–576.

Restall, G. (2014). Pluralism and proofs. *Erkenntnis*, *79*, 279–291.

Ryle, G. (2009). *Collected essays 1929–1968: Collected papers* (Vol. 2). Abingdon: Routledge.

Shapiro, S. (2011). *Varieties of pluralism and relativism for logic* (pp. 526–552). London: Wiley.

Shapiro, S. (2014). *Varieties of logic*. Oxford: Oxford University Press.

Sher, G. (1991). *The bounds of logic: A generalized viewpoint*. Cambridge: MIT Press.

Tarski, A. (1946). *Introduction to logic and to the methodology of deductive sciences*. Mineola: Dover Publications.

Weber, Z., Badia, G., & Girard, P. (2016). What is an inconsistent truth table? *Australasian Journal of Philosophy*, *94*(3), 533–548.

Williamson, T. (2014). Logic, metalogic and neutrality. *Erkenntnis*, *79*(2), 211–231.