This is a ``preproof'' accepted article for *The Bulletin of Symbolic Logic*.
This version may be subject to change during the production process.
DOI: 10.1017/bsl.2025.10083

Carnap's (Categoricity) Problem

Sebastian G.W. Speitel

Abstract

Carnap's (Categoricity) Problem concerns the relationship between (rules of) inference and model-theoretic values. In particular, it asks whether proof-theoretic constraints are 'strong enough' to uniquely determine intended semantic values. Carnap (1943) demonstrated that already in the classical bivalent setting this is not the case for the majority of the usual logical constants. To remedy this underdetermination of 'semantics by syntax' a variety of solution strategies has been explored in the literature. This article is a philosophical-logical survey of these attempts, comparing them with respect to scope, motivation and success. Besides the mathematical interest held by *Carnap's Problem*, the underdetermination it uncovers has significant consequences for a variety of philosophical projects and positions, warranting a systematic study of attempts at resolving it.

Contents

1 Introduction: A Full Formalization of Logic

A recurring theme in Carnap's work on the foundations of logic and mathematics is a concern with notions of *formal completeness* and questions pertaining to the determinacy and uniqueness of formal, i.e., logical and mathematical, concepts. The most prominent example of this can be found in his ultimately abandoned *Untersuchungen zur Allgemeinen Axiomatik*, containing the ill-fated *Gabelbarkeitssatz*, in which he aimed to unify several extant conceptions of the completeness of an axiom system. After Carnap's 'semantic turn', following the rise and wide-spread acceptance of Tarskian model-theory, questions concerning the determinacy of formal notions gained a further layer of complexity.[1] It was in this context that Carnap worried that the usual characterizations of *logical* notions, in spite of the soundness and completeness of the systems they were part of, i.e., in spite of a perfect match at the level of consequence, left essential properties underspecified.

Carnap took up the question of a *full formalization of logic* in a "very little known" (Raatikainen, 2008, 283) work (Carnap, 1943). There, he demonstrated that, surprisingly, the standard rules for almost all of the usual logical constants of FOL severely underdetermine their standard model-theoretic semantics. This state of affairs, Carnap thought, was highly undesirable: a unique determination of the standard semantic values of the constants by their rules of inference was to be a desideratum for a logical system on par with its soundness and completeness. Although tradition has not followed him in this assessment of importance, Carnap's discovery had impactful consequences for a wide variety of philosophical views and projects, even outside the foundations of logic. Due to this, *Carnap's (Categoricity) Problem* has, in recent years, been the subject of intense attention, after only sporadic discussions and rediscoveries for the over 60 years directly following the publication of (Carnap, 1943).[2]

Despite the renewed interest there exists, to the best of my knowledge, no fully systematic study of the solution strategies put forward for resolving *Carnap's Problem*.[3] This is the gap the present paper aims to fill. In doing so, it is bound to remain incomplete as the literature is, by now, vast and scattered. While containing several novel observations and results, its main goal remains expository; it tries to bring under common philosophical perspectives different solution strategies that have been proposed and elaborated in the literature.

Carnap's Problem has consequences for debates in the philosophy of language, mind, mathematics, ontology, epistemology, and logic and constitutes a deep and fruitful starting point for evaluating seemingly unrelated philosophical positions and proposals. Even though Carnap's original interest in and treatment of the issue was much more narrow, this paper deviates in this from his original focus and starting point. It is, moreover, less concerned with a faithful reconstruction of the historical Carnap, but rather attempts a systematic survey and discussion of *Carnap's Problem* in a contemporary context. It is hoped that this, by bringing together various avenues of investigation, will shed light on just how deep, basic and widespread the issues raised by *Carnap's Problem* are.

The structure of the paper is as follows: in Section 2 I will, based on existing treatments of the issue, introduce the basic structure of *Carnap's Problem* and provide a partial and incomplete overview of debates it impacts. Section 3 surveys and discusses the first of the major families of strategies addressing the problem, strategies who have in common the idea that the notion of *inference* needs to be refined or strengthened in order to rule out the underdetermination of the logical constants. Sections 4 and 5 take up the, less widespread, semantic strategy for remedying Carnapian underdetermination and investigate its scope and prospects. Finally, Section 6 concludes with a brief outlook. A short appendix contains definitions and results clarifying several points from the main parts of the paper.

[1] Cf. (Leitgeb and Carus, 2020, Section 6.2) and (Awodey and Carus, 2001).

[2] See the references in note 6.

[3] Though see, e.g., Hjortland (2014), Bonnay and Westerståhl (2016), Button and Walsh (2018) and Brîncus (2024a) for comparative assessments of several of the solutions discussed in this paper.

2 Carnap's Categoricity Problem

Let $\mathscr{L}$ be a propositional language consisting of a countably infinite stock of propositional variables $p, q, r \ldots$ and the usual connectives $\neg, \wedge, \vee, \rightarrow$ and $\leftrightarrow$. The set of sentences of $\mathscr{L}$ is designated by $\textsc{Sent}_{\mathscr{L}}$. A (single-conclusion) *consequence relation* $\vdash_{\mathscr{L}}$ over $\mathscr{L}$ is a relation of the form:[4]

$$\vdash_{\mathscr{L}} \subseteq \mathcal{P}(\textsc{Sent}_{\mathscr{L}}) \times \textsc{Sent}_{\mathscr{L}}$$

As usual, we write $\Gamma \vdash_{\mathscr{L}} \varphi$ for $\langle \Gamma, \varphi \rangle \in \vdash_{\mathscr{L}}$.

A (two-valued) *valuation* v is a total function $v : \textsc{Sent}_{\mathscr{L}} \mapsto \{0, 1\}$. A sentence φ is *true under a valuation* v if $v(\varphi) = 1$. We designate the set of all (two-valued) valuations by $\textsc{Val}_{\mathscr{L}}$. Let $\mathcal{V} \subseteq \textsc{Val}_{\mathscr{L}}$. A sentence φ is a $\mathcal{V}$*-consequence* of a set of sentences Γ, $\Gamma \models_{\mathcal{V}} \varphi$, if, for all $v \in \mathcal{V}$, whenever $v(\gamma) = 1$ for all $\gamma \in \Gamma$ then also $v(\varphi) = 1$ (we write $v(\Gamma) = n$ for $v(\gamma) = n$ for all $\gamma \in \Gamma$).

A valuation v is *consistent with a consequence relation* $\vdash_{\mathscr{L}}$ if, whenever $\Gamma \vdash_{\mathscr{L}} \varphi$, it is not the case that $v(\Gamma) = 1$, but $v(\varphi) = 0$. A set of valuations $\mathcal{V}$ is consistent with a consequence relation $\vdash_{\mathscr{L}}$ if $\vdash_{\mathscr{L}} \subseteq \models_{\mathcal{V}}$. $\mathbb{V}(\vdash_{\mathscr{L}}) = \{v \in \textsc{Val}_{\mathscr{L}} \mid v \text{ is consistent with } \vdash_{\mathscr{L}}\}$ yields the set of valuations consistent with $\vdash_{\mathscr{L}}$. The *semantic value* of a connective $c \in \{\neg, \wedge, \vee, \rightarrow, \leftrightarrow\}$, $[\![c]\!]$, is a set of valuations, i.e., $[\![c]\!] \subseteq \textsc{Val}_{\mathscr{L}}$. $[\![c]\!]$ is *consistent with a consequence relation* $\vdash_{\mathscr{L}}$ if $[\![c]\!] \subseteq \mathbb{V}(\vdash_{\mathscr{L}})$. $[\![c]\!]$ is *(uniquely) determined* if $\mathbb{V}(\models_{[\![c]\!]}) = [\![c]\!]$.[5]

Carnap's Problem concerns the question of what semantic value we are able to recover on the basis of inferential information and demonstrates, in particular, the failure of standard inferential characterizations of the logical notions to determine their standard semantics.[6]

2.1 Classical Propositional Logic

To appreciate the scope and import of *Carnap's Problem* it is best to start with a case of successful meaning-determination. Consider, to this end, *conjunction* as inferentially characterized by the usual clauses – ($\wedge$I) $\varphi, \psi \vdash_{\mathscr{L}} \varphi \wedge \psi$; ($\wedge$E$_1$) $\varphi \wedge \psi \vdash_{\mathscr{L}} \varphi$ and ($\wedge$E$_2$) $\varphi \wedge \psi \vdash_{\mathscr{L}} \psi$ – and endowed with its usual semantic value (as provided by the standard boolean satisfaction-clause):

$$[\![\wedge]\!] = \{v \in \textsc{Val}_{\mathscr{L}} \mid \text{for all } \varphi, \psi \in \textsc{Sent}_{\mathscr{L}}\text{: } v(\varphi \wedge \psi) = 1 \text{ iff } v(\varphi) = 1 \text{ and } v(\psi) = 1\}$$

From the fact that $\mathbb{V}(\cdot)$ and $\models_{\_}$ form an *antitone Galois-connection* it immediately follows that $[\![\wedge]\!] \subseteq \mathbb{V}(\models_{[\![\wedge]\!]})$.[7] Suppose, then, that there exists a $v^* \in \textsc{Val}_{\mathscr{L}}$, s.t. $v^* \in \mathbb{V}(\models_{[\![\wedge]\!]})$ but $v^* \notin [\![\wedge]\!]$. This means

[4]Note that we don't impose any further requirements on $\vdash_{\mathscr{L}}$ to constitute a consequence relation, though all consequence relations considered in this paper are fully *Tarskian*, see Avron (1991).

[5]$\mathbb{V}(\cdot)$ and $\models_{\_}$ form an *antitone Galois-connection*, see (Humberstone, 2011), (Hjortland, 2014), (Hardegree, 2005) and (French and Ripley, 2019). In the propositional setting we follow the set-up of (Humberstone, 2011) and (Hjortland, 2014).

[6]The original discussion of *Carnap's Problem* occurs in (Carnap, 1943). Church's review (Church, 1944) contains references to earlier work that was aware of the issue Carnap discussed, though did not address its philosophical dimensions. In the following decades, *Carnap's Problem* was occasionally re-discovered and investigated, though few discussions explicitly addressed the philosophical concerns driving Carnap's treatment of the issue, see (Church, 1954), (McCawley, 1975, 1981), (Leblanc et al., 1977), (Gabbay, 1978), (Shoesmith and Smiley, 1978), (Hart, 1982), (Belnap and Massey, 1990), (Garson, 1990), (Koslow, 1992) and (Humberstone, 1996). See esp. (Humberstone, 2011) for an extensive list of references addressing the propositional versions of *Carnap's Problem.* (Raatikainen, 2008, 283) laments the absence of attention that had been paid to the issues raised by Carnap 60 years earlier and points out several philosophical consequences of Carnap's discovery (see, in this context, especially the responses by (Murzi and Hjortland, 2009) and (Incurvati and Smith, 2010)). Several accounts that have taken up both the formal and the philosophical challenge associated with *Carnap's Problem* include (Smiley, 1996), (Rumfitt, 1997, 2000), (Humberstone, 2000). Recently, *Carnap's Problem* has attracted growing attention among inferentialists and philosophers of logic, see (Koslow, 2010), (Garson, 2010, 2013), (Hjortland, 2014), (Peregrin, 2014), (Woods, 2014), (Bonnay and Westerståhl, 2016), (Button, 2016), (Button and Walsh, 2018), (Haze, 2019), (Brîncuş, 2019; Brîncus, 2024a), (Bonnay and Speitel, 2021), (Murzi and Topey, 2021), (Bonnay and Westerståhl, 2023), (Tong and Westerståhl, 2023), (Tabakçı, 2024), (Westerståhl, 2025), (Picollo, 2025). This renewed attention motivates the current papers' goal of taxonomizing extant approaches and solution-strategies to *Carnap's Problem.*

[7]An *antitone Galois-connection* is a pair of maps $\langle f, g \rangle$ between ordered sets A, B – $f : A \rightarrow B$ and $g : B \rightarrow A$, – such that, for all $a \in A$ and $b \in B$, $b \leq f(a)$ iff $a \leq g(b)$. In the case at hand, $A = \mathcal{P}(\textsc{Val}_{\mathscr{L}})$, $B = \{\vdash_{\mathscr{L}} \mid \vdash_{\mathscr{L}} \subseteq \mathcal{P}(\textsc{Sent}_{\mathscr{L}}) \times \textsc{Sent}_{\mathscr{L}}\}$, f is $\models_{\_}$, g is $\mathbb{V}(\cdot)$, and the order relation is given by $\subseteq$. See (Humberstone, 2011, 0.12 & p. 58 (G2)) for the result that for any set of valuations $\mathcal{V}$, $\mathcal{V} \subseteq \mathbb{V}(\models_{\mathcal{V}})$.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

that, for some pair of sentences φ and ψ, v^* is not boolean, i.e., (a) $v^*(\varphi \wedge \psi) = 1$ but $v^*(\varphi) = 0$ or $v^*(\psi) = 0$; or (b) $v^*(\varphi \wedge \psi) = 0$ yet $v^*(\varphi) = 1$ and $v^*(\psi) = 1$.

Note that, in case (a), v^* would be inconsistent with $\wedge E_1$ or $\wedge E_2$, whereas in case (b) v^* would be inconsistent with $\wedge$I. Clearly, however, $\models_{[\![\wedge]\!]}$ respects $\wedge$I, $\wedge E_1$, and $\wedge E_2$, and v^* is therefore inconsistent with $\models_{[\![\wedge]\!]}$. It follows that $v^* \notin \mathbb{V}(\models_{[\![\wedge]\!]})$ and therefore $[\![\wedge]\!] = \mathbb{V}(\models_{[\![\wedge]\!]})$. In other words, $\wedge$I, $\wedge E_1$ and $\wedge E_2$ *uniquely determine* $[\![\wedge]\!]$.

Now let $\vdash_{CPL}$ be the (single-conclusion) consequence relation of classical propositional logic and consider the valuations $v_\top$ and $v_\vdash$:[8]

(i) $v_\top(\varphi) = 1$ for all $\varphi \in \text{SENT}_{\mathscr{L}}$
(ii) $v_\vdash(\varphi) = 1$ iff $\vdash_{CPL} \varphi$

$v_\top$ does not rule out any sentence and does thus not constitute a counterexample to *any* claim of consequence – it is consistent with every consequence relation and in particular with $\vdash_{CPL}$. Moreover, since anything derivable from a set of theorems in CPL is itself a theorem of CPL, $v_\vdash$ is consistent with $\vdash_{CPL}$ as well. However, note that, for any $\varphi \in \text{SENT}_{\mathscr{L}}$, $v_\top(\varphi) = v_\top(\neg\varphi) = 1$ and, for an arbitrary propositional atom p, $v_\vdash(p) = v_\vdash(\neg p) = 0$ yet $v_\vdash(p \vee \neg p) = 1$.

Since both $v_\top$ and $v_\vdash$ are consistent with $\vdash_{CPL}$ and $\vdash_{CPL}$ is sound and complete with respect to the class of boolean valuations $\mathcal{B}$, we have that $v_\top, v_\vdash \in \mathbb{V}(\models_{\mathcal{B}})$. Furthermore, since $\mathcal{B} \subseteq [\![\neg]\!]$ and $\mathcal{B} \subseteq [\![\vee]\!] \cap [\![\neg]\!]$ for the usual semantic values of $\neg$ and $\vee$ (provided by the standard boolean satisfaction clauses)

$$[\![\neg]\!] = \{v \in \text{VAL}_{\mathscr{L}} \mid \text{for all } \varphi \in \text{SENT}_{\mathscr{L}}\colon v(\varphi) = 1 \text{ iff } v(\neg\varphi) = 0\}$$
$$[\![\vee]\!] = \{v \in \text{VAL}_{\mathscr{L}} \mid \text{for all } \varphi, \psi \in \text{SENT}_{\mathscr{L}}\colon v(\varphi \vee \psi) = 1 \text{ iff } v(\varphi) = 1 \text{ or } v(\psi) = 1\}$$

it follows from the fact that $\mathbb{V}(\cdot)$ and $\models_{\_}$ form a Galois-connection that $\mathbb{V}(\models_{\mathcal{B}}) \subseteq \mathbb{V}(\models_{[\![\neg]\!]})$ and $\mathbb{V}(\models_{\mathcal{B}}) \subseteq \mathbb{V}(\models_{[\![\vee]\!] \cap [\![\neg]\!]})$ (see (Humberstone, 2011, Ch. 1.12)). Hence, $v_\top \in \mathbb{V}(\models_{[\![\neg]\!]})$ and $v_\vdash \in \mathbb{V}(\models_{[\![\vee]\!] \cap [\![\neg]\!]})$. Therefore $\mathbb{V}(\models_{[\![\neg]\!]}) \neq [\![\neg]\!]$ and $\mathbb{V}(\models_{[\![\vee]\!] \cap [\![\neg]\!]}) \neq [\![\vee]\!] \cap [\![\neg]\!]$. In other words, the usual boolean semantic values of '$\neg$' and '$\vee$' are not determined by a (complete) description of their inferential behaviour, as captured by $\vdash_{CPL}$.

This is somewhat surprising given that the class of boolean valuations $\mathcal{B}$ was *sound and complete* with respect to classical propositional consequence $\vdash_{CPL}$, i.e., both $\vdash_{CPL} \subseteq \models_{\mathcal{B}}$ and $\models_{\mathcal{B}} \subseteq \vdash_{CPL}$ hold, hence $\models_{\mathcal{B}} = \vdash_{CPL}$. It means that although the consequence relation of CPL fully and adequately captures the relation of *logical consequence* for the language of classical propositional logic, and the boolean valuations provide sufficiently many counterexamples for any claim of consequence not licensed by CPL, something is left out and cannot be secured at the level of consequence. The usual axiomatizations of CPL fail to provide, in Carnap's words, a *full formalization* of CPL and thus require amendment. Thus, despite the completeness of $\vdash_{CPL}$ for the intended interpretations of the logical constants not every aspect of their meaning is adequately captured by the consequence relation.[9] From the perspective of $\vdash_{CPL}$ there is therefore no reason to think that the boolean clauses are the right way of semantically describing the meanings of the connectives.

This *Carnapian underdetermination* demonstrates two things: on the one hand, we are unable to recover the standard, intended meanings of the connectives from their inferential behaviour. On the other, the usual boolean meanings of these connectives are *unstable* in the sense that they cannot be recovered from the consequence relation they generate – they fail to be uniquely determined on the basis of the inferential patterns they give rise to.

[8]These examples are well-known, see, e.g., (Carnap, 1943) and (Belnap and Massey, 1990).

[9]That not all aspects of the meaning of a connective can be adequately captured at the level of consequence is a claim familiar from debates concerning the relation between classical and non-classical logics: several non-classical logics, such as ST (Cobreros et al., 2012) or *supervaluationist logics* (van Fraassen, 1966), coincide with classical logic at the level of the consequence relation, but differ from it at the level of meta-inferences, cf. (Barrio et al., 2020). Yet, as a reviewer of this paper helpfully points out, meta-inferences have just as legitimate a claim to being relevant to a connective's meaning as inferences do. Hence, not all aspects of a connective's meaning are codified by the consequence relation. For applications of this point to *Carnap's Problem* see Section 3.4 below. Thanks to an anonymous referee of this paper for raising this point.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

How 'bad' or extensive is *Carnap's Problem*? Already in (Carnap, 1943), Carnap provided an exhaustive classification of what can go wrong with valuations at the level of classical propositional consequence. He there showed that what he termed *non-normal valuations*, unintended valuations consistent with $\vdash_{CPL}$, are of two kinds: (i) the single valuation $v_\top$ that interrupts the law of non-contradiction by making everything true; and (ii) valuations that make at least one sentence false and fail to be boolean for at least one connective by violating bivalence and making a sentence and its negation both false (of which $v_\vdash$ is one of many instances).[10]

What is lost through these 'non-normal' valuations is the *truth-functionality* of the semantic values of the connectives. Given a set of valuations $\mathcal{V}$ and a connective c of adicity n, c is *truth-functional* with respect to $\mathcal{V}$ if there exists a function $f_c : \{0,1\}^n \mapsto \{0,1\}$, s.t. for all $\varphi_1, \ldots, \varphi_n \in \textsc{Sent}_{\mathscr{L}}$ and $v \in \mathcal{V}$:[11]

$$v(c(\varphi_1, \ldots, \varphi_n)) = f_c(v(\varphi_1), \ldots, v(\varphi_n))$$

f_c is called a *truth-function* for c. It follows from the truth-functionality of a connective with truth-function f that, whenever $v(\varphi_i) = v(\psi_i)$, then $f(v(\varphi_1), \ldots, v(\varphi_n)) = f(v(\psi_1), \ldots, v(\psi_n))$. Thus, if c is truth-functional, $v(c(\varphi_1, \ldots, \varphi_n)) = v(c(\psi_1, \ldots, \psi_n))$. Consider, once more, the valuation $v_\vdash$ and let $f_\vee$ be the usual binary truth-function of $\vee$. We then have that $v_\vdash(p \vee \neg p) = f_\vee(v_\vdash(p), v_\vdash(\neg p))$. However, $f_\vee(0,0) = 0$ while $v_\vdash(p \vee \neg p) = 1$. Hence, $f_\vee$ cannot be a truth-function for $\vee$ over $\mathcal{B} \cup \{v_\vdash\}$. Similar arguments establish that there can be no truth-function for $\vee$ over $\mathcal{B} \cup \{v_\vdash\}$.[12]

Just as severe as *Carnap's Problem* is for an adequate meaning-theory of the connectives, just as easy it is to address: force the truth-functionality of *any one* of the connectives $\neg, \vee, \rightarrow, \leftrightarrow$ and the standard, intended, truth-functional interpretations will thereby be determined for *all* of them.[13] Thus, what is needed is a way to stabilize the standard boolean meaning of any one of the connectives other than conjunction in order to obtain a solution to *Carnap's Problem* for CPL.[14]

2.2 First-Order Logic (FOL)

Carnap was aware that the underdetermination uncovered by him extended to the quantifiers as well. In (Carnap, 1943) he outlined and discussed the non-normal interpretations affecting the standard first-order universal and existential quantifiers. Arguably, however, Carnap was not yet able to fully appreciate the dimension of underdetermination as it occurs at the level of quantification, since the mature concept of a quantifier only emerged later through the work of (Mostowski, 1957), (Lindström, 1969) and (Montague, 1974). This observation is supported by the (limited) solution he put forward to resolve the underdetermination of the quantifiers by, essentially, reducing universal quantification to infinitary conjunction and existential quantification to infinitary disjunction. Since the concern of the present article lies less with a reconstruction of the historical Carnap we will here treat the issue from the perspective of generalized quantifier theory to emphasize the generality of the issue of *Carnapian underdetermination*.[15]

Let $\mathscr{L}(Q_1, \ldots, Q_n)$ be a relational first-order language[16] with a countably infinite set of individual variables $x_1, x_2, \ldots$, a countably infinite set of relation-symbols $R_1^n, R_2^n, \ldots$ for any adicity n, a full

[10]See (Church, 1944, 493), (Belnap and Massey, 1990, 68), (Raatikainen, 2008, 284) and (Bonnay and Westerståhl, 2016, 727).

[11]See, for example, (Humberstone, 2011, 376), (McCawley, 1975, 412), (Hart, 1982, 132), (Belnap and Massey, 1990, 71).

[12]See (Bonnay and Westerståhl, 2016, 728).

[13]See, e.g., (Carnap, 1943) (McCawley, 1975), (Garson, 2013).

[14]The formal results concerning *Carnap's Problem* are well-known. See (Humberstone, 2011) for an extensive and systematic treatment of the relationship between consequence relations and their valuational presentations. Most, if not all, formal results mentioned in this section can be found there.

[15]Though see (Carnap, 1943) and cf. (Leblanc et al., 1977). Cf. also (Bonnay and Westerståhl, 2016) for criticizing attempts of this kind for resolving *Carnap's Problem*.

[16]For ease of presentation we restrict attention to purely relational first-order languages and signatures in the following, though the results mentioned of course generalize. Bonnay and Westerståhl (2016) consider languages with constant and function symbols. The theorem below is thus a special case of their result which also demands that the base of the principal filter be closed under the interpretation of terms.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

complement of propositional connectives $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$, and quantifier-symbols $Q_1, \dots, Q_n$. Models for $\mathscr{L}(Q_1, \dots, Q_n)$ are standard relational first-order structures $\mathcal{M} = \langle M, \mathcal{R}_1, \mathcal{R}_2, \dots, \rangle$. The definition of *sentence*, *consequence relation* (written: $\vdash_{\mathscr{L}(Q_1,\dots,Q_n)}$) and *truth-in-a-model* are as expected and analogous to the propositional case.[17]

From the perspective of generalized quantifier theory,[18] a quantifier $\mathcal{Q}$ is a second-order predicate 'checking' whether a (sequence of) first-order predicate(s) has the property expressed by the quantifier. A quantifier-expression is of type $\langle k_1, \dots, k_n \rangle$ if it combines with n formulas to form a well-formed expression, binding k_i variables in the i-th formula. The usual universal and existential quantifiers are of type $\langle 1 \rangle$. For simplicity of presentation we consider, unless noted otherwise, only quantifiers of type $\langle 1 \rangle$ in the following.

Let $\mathfrak{M}$ be the class of all first-order models. The semantic value of a (type $\langle 1 \rangle$) quantifier $[\![Q]\!]$ is a class of (first-order) models (Lindström, 1969):[19]

$$[\![Q]\!] = \{\mathcal{M} \in \mathfrak{M} \mid \mathcal{M} = \langle M, X \rangle \text{ and } \Psi(X)\}$$

where Ψ is some (set-theoretic) property. The satisfaction clause for generalized quantifiers is as follows (where $\mathcal{M}$ is a first-order model with domain M and $[\![\varphi(x)]\!]^{\mathcal{M}} = \{a \in M \mid \mathcal{M} \models \varphi(a)\}$):

$$\mathcal{M} \models Qx\varphi(x) \text{ iff } \langle M, [\![\varphi(x)]\!]^{\mathcal{M}} \rangle \in [\![Q]\!]$$

Examples of generalized quantifiers include the usual first-order quantifiers $\forall, \exists$, elementarily definable quantifiers such as 'at least 3' ($\exists_{\geq 3}$), but also quantifiers such as 'infinitely many' (Q_0) and 'uncountably many' (Q_1):

(i) $[\![\forall]\!] = \{\langle M, X \rangle \mid X = M\}$
(ii) $[\![\exists]\!] = \{\langle M, X \rangle \mid X \neq \emptyset\}$
(iii) $[\![\exists_{\geq 3}]\!] = \{\langle M, X \rangle \mid |X| \geq 3\}$
(iv) $[\![Q_0]\!] = \{\langle M, X \rangle \mid |X| \geq \aleph_0\}$
(v) $[\![Q_1]\!] = \{\langle M, X \rangle \mid |X| \geq \aleph_1\}$

Let $\vdash_{FOL}$ be the usual (single-conclusion) consequence relation of classical first-order logic. In a recent paper, Bonnay & Westerståhl (2016) precisely characterized the shape of possible interpretations of $\forall$ consistent with its inferential behaviour in the context of $\vdash_{FOL}$.[20] Let $\mathcal{M}$ be a model with domain M. The *local quantifier* $[\![Q]\!]^{\mathcal{M}}$ over a particular model $\mathcal{M}$ can be obtained from the corresponding global quantifier $[\![Q]\!]$ by restricting attention to the domain M of $\mathcal{M}$ in the following way (for a type $\langle 1 \rangle$ quantifier): $[\![Q]\!]^{\mathcal{M}} = \{X \mid \langle M, X \rangle \in [\![Q]\!]\}$. In keeping with the global characterization of meaning, we say that a quantifier-interpretation $[\![Q]\!]$ is *consistent with a consequence relation* $\vdash$ iff $\vdash$ is sound with respect to the model-theoretic consequence relation generated by $[\![Q]\!]$ (see Appendix for details).[21] Then

[17]Unless noted otherwise, we take a claim of consequence to be, first and foremost, a relation between *sentences* rather than (open) *formulas.* This constitutes a choice point in the debate which, at times, has consequences for resolving *Carnap's Problem* (see esp. the latter part of Section 3.4).

[18]See (Peters and Westerståhl, 2006) for overview.

[19]In analogy with the propositional case, we here identify the meaning of a quantifier-expression with a *global quantifier*, see (Peters and Westerståhl, 2006).

[20]See (Button and Walsh, 2018, Ch. 13) for a characterization of the underdetermination of the first-order quantifiers in terms of Boolean models.

[21]For ease of presentation and consistency with characterizing the meaning of logical expressions globally I here deviate from the formal set-up used by Bonnay and Westerståhl (2016) to establish their result. Whereas they consider so-called *weak models* in the context of languages containing also *predicate-variables*, and assess consistency relative to individual models, I here take the interpretation of a quantifier to be global and, therefore, to depend only on the *domain of a model.* This difference in presentation is largely insubstantial for the result itself. The assumption of predicate-variables in Bonnay & Westerståhl's set-up allows them to avoid having to take definability-facts into account when formulating their main result. The same is achieved by making the interpretation of the quantifier dependent on only the underlying domain of a model, independently of what is and is not definable in that particular model, and adjusting the relevant standard of consistency

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

Theorem (Bonnay and Westerståhl, 2016): An interpretation $[\![\forall]\!]$ of the universal quantifier is consistent with $\vdash_{FOL}$ as long as, for all $\mathcal{M}$, $[\![\forall]\!]^{\mathcal{M}}$ is a principal filter over M.

A *principal filter* $\mathcal{F}_X$ over a set M has the form $\mathcal{F}_X = \{A \subseteq M \mid X \subseteq A\}$ for some $X \subseteq M$.[22] The intended interpretation $[\![\forall]\!]^{\mathcal{M}} = \{M\}$ is of course a principal filter over M (set $X = M$), – it is the *maximal* principal filter over M, – but it is, in general (as long as $|M| > 1$), far from the only one. As a result, $\vdash_{FOL}$ underdetermines the semantic value of $\forall$ and, derivatively, of $\exists$ as well.

How to best understand the scale of underdetermination here? The meaning of $\forall$ according to $\vdash_{FOL}$ essentially boils down to meaning 'for all X' and the meaning of $\exists$ to 'some X', where X is an arbitrary subset of the domain (Bonnay and Westerståhl, 2016). As an example, let X be the base of a principal filter $\mathcal{F}_X$ interpreting $\forall$ over a model $\mathcal{M}$. Due to the duality of $\forall$ and $\exists$ we have that $[\![\exists]\!]^{\mathcal{M}} = \{A \subseteq M \mid X \nsubseteq M - A\}$. Thus, $\mathcal{M} \models \exists x\varphi(x)$ iff $[\![\varphi(x)]\!]^{\mathcal{M}} \in [\![\exists]\!]^{\mathcal{M}}$ iff $X \nsubseteq M - [\![\varphi(x)]\!]^{\mathcal{M}}$ iff $[\![\varphi(x)]\!]^{\mathcal{M}} \cap X \neq \emptyset$, i.e., iff there is *some* X that is φ. Models of first-order logic thus come to resemble inner domains of models of free logic, distinguishing between 'existing' (those in the base of the principal filter) and 'non-existing' (those not present in the base of the principal filter) objects (Bonnay and Westerståhl, 2016). Allowing constant-symbols in the language, the only interpretation guaranteed to be recoverable from first-order inference patterns is the *substitutional interpretation* of the quantifiers.[23]

This is, of course, deeply troubling, for it "undermines the prospects of philosophical ontology construed as the quintessentially armchair project of extracting ontological commitments from the semantic analysis of quantified statements" (Antonelli, 2015, 171). For philosophical projects relying on an objectual interpretation of the usual first-order quantifiers, combined with the idea that this meaning is, ultimately, to be 'read off' their inferential behaviour, *Carnap's Problem* is devastating.

2.3 The Philosophical Significance of Carnap's Problem

Carnap's Problem is more than a mere formal curiosity in the mathematical foundations of propositional and first-order logic. Carnap deemed a *categorical*, i.e., unique, determination of the semantic values of the logical expressions of these systems to be a desideratum on par with soundness and completeness results for the relevant calculi. The underdetermination of semantics by 'syntax' that he uncovered constitutes a stumbling block for many philosophical projects that rely on the formalisms of the respective logical systems. In this brief section I want to provide a few examples of places where *Carnap's Problem* might be taken to cause issues.

Informally, *Carnap's Problem* problematizes the idea that understanding how an expression functions in inference suffices for grasping its truth-conditional content. It therefore constitutes an immediate and sizeable issue for *moderate inferentialism*, a group of positions that attribute an important meta-semantic function to inferential roles in a theory of meaning. According to positions of this type, meanings are (best modelled by) model-theoretic objects (*semantic maximalism*) that can be known or 'gotten to' on the basis of epistemically minimal, and naturalistically acceptable, resources – inferential roles of the relevant expressions. Inherent to moderate inferentialism is the adoption of a *meaning-determination thesis* according to which it is inferential roles that determine, but are not identified with, the meanings of the logical expressions. As such, their success is threatened by *Carnap's Problem*.[24]

Relatedly, *Carnap's Problem* undermines the idea that "soundness and completeness serve to legitimate talk of reference, denotation, semantic value, and the like; these model-theoretic terms derive

accordingly. Thus, Bonnay and Westerståhl's result carries over to the current setting without restriction. In response to criticism of their assumption of second-order variables in stating their result in (Valle-Inclán, 2024) I argued (Speitel, 2025) that the assumption of these second-order variables merely captures the domain-dependency of quantification and thereby serves as a simplifying but, for their purposes, inessential device. See the Appendix for spelling out the differences and correspondences between the current and Bonnay and Westerståhl's setting in more detail. I thank an anonymous referee for urging me to clarify the relationship between the presentation of the result here and in (Bonnay and Westerståhl, 2016).

[22] See, e.g., (Davey and Priestley, 2002, 45) for precise definition.

[23] Cf. also (Garson, 2013, Ch. 14) and (Picollo, 2025) for the underdetermination of quantifiers.

[24] See (Raatikainen, 2008). See (Murzi and Steinberger, 2017) for an overview.

their sense from the connection between models and valid inference" (Ripley, 2013, 142). For it demonstrates that there is a gap between completeness and categoricity (in Carnap's sense): soundness and completeness of a logical system are not sufficient for the categoricity of its logical expressions (and, as will be shown in Section 5.2, neither are they necessary). More, then, is needed to accommodate a grasp of or reference to non-inferential meanings on the basis of inferentially mediated access.

Carnap's Problem severely distorts the idea that the understanding of a logical notion is constituted by an appreciation of its (characteristic) inferential patterns, an idea popular in certain debates in the philosophy of logic and language (see, e.g., (Boghossian, 1996)). Where unique determination is made part of the conditions for the logicality of a notion (see (Feferman, 2015) and (Bonnay and Speitel, 2021)), the need to address Carnapian underdetermination is obvious.

Logic, as a tool for scientific theory-building, is meant to provide a framework that not only ensures safe and reliable inference, but also limits the inevitable underdetermination arising at the level of scientific data (underdetermination of theory by evidence) and at the level of possible models of a theory (due to Löwenheim-Skolem and similar phenomena). Instead, *Carnap's Problem* introduces a further dimension of underdetermination into the logical apparatus used for formalizing scientific theories, thereby increasing indeterminacy at a particularly basic level: "[o]ur student has heard of the difficulties of excluding nonstandard interpretations in the upper stories of mathematics; now he finds the same thing in the basement [of logical theorizing]" (Shoesmith and Smiley, 1978, 3).

A further, philosophically important aspect of *Carnap's Problem* was already briefly mentioned in the previous section. According to Quine's (in)famous *criterion of ontological commitment* (Quine, 1953) the ontological commitments of a theory are determined by its quantified-over variables. Combining this with the view that the meaning of those quantifiers is to be determined in a naturalistically acceptable way by inferential patterns renders this assessment of ontological commitment inadequate: "the possibility of nonstandard interpretations reveals that being the value of a variable is at best a sufficient, but not necessary condition for ontological commitment" (Antonelli, 2013, 657).

Although Carnap's discussion of the eponymously named underdetermination phenomenon in (Carnap, 1943) took place in the context of classical languages and truth-conditional semantics, nothing about *Carnap's Problem* restricts it to this setting. *Carnap's question* can be asked for any language and logical system that possesses a sufficiently formalized syntax and semantics. *Carnap's Problem* arises just as forcefully for other logical systems. What changes when moving to a different language or logic are not just the inferential descriptions of the relevant notions but, more importantly, the semantic space, i.e., the space of values of the logical expressions under consideration. In an intensional setting, for example, it will no longer do to treat propositional meanings as given by functions from sentences to truth-values, and we might have to assess categoricity with respect to a semantic framework identifying meanings with, say, sets of worlds. We will consider applications of *Carnap's Problem* to non-classical and richer settings in sections 4.2 and 4.3 below. At this point it is worth emphasizing the extent and reach of *Carnap's Problem* and the concomitant wealth of philosophical and formal questions revealed by it for debates in logic, philosophical logic and the philosophies of logic and language.

3 Inferential Strategies for Solving Carnap's Problem

Common to inferential strategies addressing *Carnap's Problem* is the belief that the shortcoming revealed by it is to be located in the restricted way inference can be represented in a purely assertion-based, single-conclusion framework. What *Carnap's Problem* shows, according to accounts of this type, is that we should adopt a richer notion of inference to tighten control over the semantic values determined, fixed, or 'pinned down' by rules and inferential patterns. The basic method underlying inferential strategies consists therefore in a strengthening of the proof-theory or language so as to enable it to express further or stronger conditions on the semantic framework with which it is to cohere.

This section briefly surveys four inferential strategies that have been put forward to resolve Carnapian underdetermination and points out some of their (philosophical) weaknesses in the context of addressing *Carnap's Problem*. The first strategy (Section 3.1) takes issue with the fact that inference is considered *single-conclusion*, the second with the fact that inference is taken to be (solely) *assertion-based* (Section 3.2). Nonetheless, both still support the idea that determination of semantic value should be effected by *inferences*, i.e., sequents consisting of premisses and conclusion(s). In contrast, the solution strategies of sections 3.3 and 3.4 shift perspective from sequents to *rules*, and thus from *inferences* to *inferring*, maintaining that the dynamics of drawing inferences have a role to play in the determination of semantic value.

3.1 Enriching Consequence: Multiple Conclusions

Introduced in Gentzen's seminal study of the proof-theory of classical and related systems (Gentzen, 1934), a *multiple conclusion consequence relation* (*mcr*) $\vdash^m_{\mathscr{L}}$ over a language $\mathscr{L}$ is a relation of the form:[25]

$$\vdash^m_{\mathscr{L}} \subseteq \mathcal{P}(\textsc{Sent}_{\mathscr{L}}) \times \mathcal{P}(\textsc{Sent}_{\mathscr{L}})$$

The basic notion of an argument according to the multi-conclusion perspective is thus one with multiple premises and *multiple conclusions*. A valuation v is *consistent* with an mcr $\vdash^m_{\mathcal{L}}$ iff, whenever $\Gamma \vdash^m_{\mathscr{L}} \Delta$ and $v(\Gamma) = 1$, then $v(\delta) = 1$ for *some* $\delta \in \Delta$. The remaining notions are defined analogous to the single-conclusion case.

This 'simple' modification has significant repercussions, for it strengthens the resulting notion of consequence to such a degree that the multiple-conclusion consequence relation for classical logic $\vdash^m_{CPL}$ forces consistent valuations to be boolean.[26] In the context of *mcrs* the usual semantic values of the classical connectives are therefore uniquely determined – the expressive resources of the mcr-framework are sufficiently strong to ensure that the standard meanings of the logical constants possess inferential roles that determine them.

Whence the additional control over semantic values? By way of example, consider the valuations $v_\top$ and $v_\vdash$ from above. $v_\top$ is ruled impermissible by the fact that *mcrs* permit empty succedents. Thus, it holds in particular that $\varphi, \neg\varphi \vdash^m_{CPL} \emptyset$ (Humberstone, 2011, 78). $v_\vdash$, on the other hand, fails to be consistent with $\vdash^m_{CPL} p, \neg p$. The enriched multiple conclusion framework is in fact so expressive that every truth-value assignment to formulas of the language possesses a corresponding statement of deducibility – as a result, any class of valuations can be uniquely described by a set of multi-conclusion inferences.[27] The *mcr*-framework is thereby able to ensure the truth-functionality, and thus booleaneness, of the inferentially characterized classical operators.[28] The "valuational semantics implicit in [a] consequence relation" (Humberstone, 2011, 389) is so tightly constrained and regulated in the setting of mcrs that non-standard interpretations are rendered impossible: the additional means of expression made available by moving to a multi-conclusion framework allow for the enforcement of constraints on consistent valuations that are sufficient to rule out Carnap's non-normal valuations.[29]

For the propositional case, making consequence multi-conclusion thus successfully solves *Carnap's Problem* and *fully formalizes* CPL in the sense of Carnap. Things look less promising for the quantifiers, however. Accounts extending the multiple conclusion strategy to expressions from this grammatical category usually do so by reducing universal and existential quantification to infinite conjunctions and disjunctions, respectively, and stipulating that domains remain countable with every object in them possessing a name.[30] However, "[t]his procrustean strategy shows at best that if quantifiers are reduced to connectives, what works for connectives works for quantifiers as well" (Bonnay and Westerståhl, 2016, 723). In particular, given the infinitary rules that must be adopted for the quantifiers on this

[25]For a systematic study of *mcrs* see (Shoesmith and Smiley, 1978) and (Humberstone, 2011). The multiple conclusion strategy is essentially the strategy adopted by Carnap himself.

[26]See, e.g., (Humberstone, 2011, Theorem 1.16.6). Cf. (Shoesmith and Smiley, 1978, Ch. 17) for further categoricity proofs.

[27]See (Humberstone, 2011, 1.17.3) and (Hjortland, 2014, Theorem 4.6).

[28]See (Humberstone, 2011, Sect. 3.11-3.13).

[29]See (Hjortland, 2014) for an extension of this approach to many-valued logics via multi-sided sequents.

[30]See (Carnap, 1943), (Kneale and Kneale, 1962) or (Hacking, 1979) for this type of strategy.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

conception the result is a rather unattractive formalism coupled with an apparent misconstruction of the grammatical type of quantificational expressions.

The multiple conclusion strategy was, in effect, already applied by Carnap himself (Carnap, 1943). Adopting an mcr-framework to resolve *Carnap's Problem*, however, encounters several philosophical obstacles. Even independently of the problematic treatment of quantifiers, mcrs have been regarded as sufficiently unnatural to lack philosophical motivation, especially in relation to the inferentialist program in the philosophy of logic.[31]

The complaint is that "multiple-conclusion systems represent so marked a departure from our actual practice that they can hardly be said to track that practice even in an idealised sense" (Steinberger, 2011, 335).[32] Given that reliance on inferences was to satisfy the naturalistic demand for non-mysterious determination of logical meanings, such artificiality poses a challenge to the proponent of a multi-conclusion strategy. For the inferences codified in a consequence relation were to capture the usual practice according to which the meaning of the relevant expressions was determined. The move to a distinct and unrelated practice does very little to bridge that gap.

In his review of (Carnap, 1943), Church was less worried about the artificiality of the multi-conclusion framework than he was about the fact that "Carnap's use of them [mcrs] is a concealed use of semantics" (Church, 1944, 496). The worry stems from the way multiple conclusion sequents are interpreted. According to the *disjunctive interpretation*, a (multi-conclusion) sequent $\Gamma \vdash^m_{\mathscr{L}} \Delta$ is equivalent in meaning to the (single-conclusion) sequent

$$\bigwedge_{\gamma\in\Gamma} \gamma \vdash_{\mathscr{L}} \bigvee_{\delta\in\Delta} \delta$$

Understanding a multi-conclusion sequent via the associated single-conclusion sequent, however, seems to violate a fundamental inferentialist tenet in that the "very format of the proof system requires us to have a prior grasp of the meanings of some logical constants" (Steinberger, 2011, 346), namely of (possibly infinite) conjunctions and, more problematically, disjunctions.[33] Similarly, Dummett agrees that "[s]equents with two or more sentences in the succedent [...] have no straightforwardly intelligible meaning, explicable without recourse to any logical constant" (Dummett, 1991, 187). One thus already needs to possess an understanding of disjunction before being able to use the multiple conclusion framework whose purpose was precisely to determine such a meaning.

Without presupposing a prior understanding of disjunction, however, "non-normal interpretations of this 'full formalization' [the mcr formalization] become possible" (Church, 1944, 496),[34] for without fixing the meaning of disjunction a *revenge Carnap's Problem* would affect "$\bigvee$". The disjunctive interpretation of multi-conclusion sequents, built into what it means for a valuation to be consistent with an mcr, therefore appears unable to ground a non-circular explanation of how the semantic values of the propositional connectives are determined in virtue of the inferences they feature in.[35]

3.2 Expanding Language: Bilateralism

The fundamental assumption of *bilateralism* is that there are two types of primitives logical theory has to account for:[36] the speech-acts of *assertion* and *denial*, both falling into the purview of logic,

[31]See esp. (Steinberger, 2011) for sustained criticism of the use of mcrs by inferentialists.

[32]See (Tennant, 1997, 320) and (Rumfitt, 2000, 795) for similar assessments. See (Restall, 2005) for a dissenting opinion.

[33]The situation for conjunction, an understanding of which, it might be alleged, must also already be present in a single-conclusion framework, is disanalogous, as the conjunction of a set of sentences is equivalent to their sequential assertion, unlike in the case of disjunction.

[34]Cf. also (Murzi, 2010, 242): "If commas cease to mean what [...] Carnap takes them to mean, Carnap's non-normal interpretations are not ruled out."

[35]See (Restall, 2005) for an alternative interpretation of multi-conclusion sequents and (Steinberger, 2011) and (Rumfitt, 2008) for comment.

[36]Several different positions fall under the label "bilateralism" in the literature: the *bilateral interpretation* of multi-conclusion consequence relation as in (Restall, 2005), bi- and multi-lateralist interpretations in terms of multiple *consequence relations* (see Wansing and Ayhan (2023)), and the bilateralist position at issue in the current chapter, which relates to a mul-

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

are conceived as “distinct activities on all fours with one another” (Smiley, 1996, 1). Consequently, appropriate axiomatizations of logical systems will have to include two types of rules – those governing the interaction of a constant with the speech-act of assertion, and those governing the interaction of a constant with the speech-act of denial. Notwithstanding the equivalence between the rejection of φ and the assertion of $\neg\varphi$, the bilateral enrichment is not inert: it enables a harmonious formulation of classical logic in a single-conclusion setting (Rumfitt, 2000) and a resolution of *Carnap's Problem*.

To formally express the added dimension of meaning, bilateralists introduce *force-markers* $+$, for assertion, and $-$, for denial, into the language. Just as speech-acts apply to contents, force-markers attach to sentences $\varphi \in \textsc{Sent}_{\mathscr{L}}$ to yield *signed sentences* $+\varphi$ and $-\varphi$. Let the set of all signed sentences of $\mathscr{L}$ be $\textsc{Sent}^*_{\mathscr{L}}$. Force-markers do not “contribute to propositional content, but indicate[...] the force with which that content is promulgated” (Rumfitt, 2000, 803). They are thus, unlike logical constants, non-embeddable and cannot be iterated. Their interaction is governed by *coordination-principles* which constitute *structural rules* of the relevant logical calculi.[37]

What is true or false is of course contents and not the speech-acts themselves, but every valuation v over $\textsc{Sent}_{\mathscr{L}}$ induces a *correctness-valuation* v_c over $\textsc{Sent}^*_{\mathscr{L}}$, s.t. $v_c(+\varphi) = \mathbf{c}$(orrect) if $v(\varphi) = 1$, $v_c(+\varphi) = \mathbf{i}$(ncorrect) if $v(\varphi) = 0$, $v_c(-\varphi) = \mathbf{c}$ if $v(\varphi) = 0$ and $v_c(-\varphi) = \mathbf{i}$ if $v(\varphi) = 1$.[38] Thus, an assertion is correct iff the asserted content is true and a rejection is correct iff the rejected content is false.

Bilateral consequence relations $\vdash^b_{\mathscr{L}}$ are single conclusion consequence relations of the form:

$$\vdash^b_{\mathscr{L}} \subseteq \mathcal{P}(\textsc{Sent}^*_{\mathscr{L}}) \times \textsc{Sent}^*_{\mathscr{L}}$$

tracking the *preservation of correctness*, rather than of truth. This change of focus effects a redefinition of the notion of consistency. A correctness valuation v_c is now *consistent with* $\vdash^b_{\mathscr{L}}$ iff, whenever $\Gamma \vdash^b_{\mathscr{L}} \varphi$ and $v_c(\Gamma) = \mathbf{c}$, then also $v_c(\varphi) = \mathbf{c}$.[39] Despite this shift we may continue to speak of the consistency of a valuation v with a consequence relation $\vdash^b_{\mathscr{L}}$ directly, due to the correspondence between valuations and correctness-valuations: v is consistent with $\vdash^b_{\mathscr{L}}$ iff the induced v_c is consistent with $\vdash^b_{\mathscr{L}}$. Similarly, for a given set of correctness valuations $\mathcal{V}_c$ we say that a (signed) sentence $\varphi \in \textsc{Sent}^*_{\mathscr{L}}$ is a $\mathcal{V}_c$-consequence of a set of (signed) sentences $\Gamma \subseteq \textsc{Sent}^*_{\mathscr{L}}$, $\Gamma \models^b_{\mathscr{L}} \varphi$, if, for all $v_c \in \mathcal{V}_c$, whenever $v_c(\Gamma) = \mathbf{c}$, then also $v_c(\varphi) = \mathbf{c}$. Definitions of related notions can be given analogously to the above.[40]

Smiley (1996) then shows that the bilateralist's framework is sufficiently expressive to uniquely determine the semantic values of the connectives and resolve *Carnap's Problem*: for *any* set of valuations $\mathcal{V}$ and bilateral consequence relation $\models_{\mathcal{V}_c}$ induced by $\mathcal{V}$, it holds that $\mathcal{V} = \mathbb{V}(\models_{\mathcal{V}_c})$ (Smiley, 1996). By way of example, the valuation $v_\top$ is ruled inadmissible due to the fact that $+\varphi \vdash^b_{CPL} -\neg\varphi$.[41] Here, $v_\top$ fails to induce a correctness-preserving correctness valuation, for the correctness valuation induced by $v_\top$ will be such that $v^c_\top(+\varphi) = \mathbf{c}$, whereas $v^c_\top(-\neg\varphi) = \mathbf{i}$ (Hjortland, 2014, 454). Similarly, the valuation $v_\vdash$ is inconsistent with $\vdash^b_{CPL}$ due to the fact that $-p, -\neg p \vdash^b_{CPL} -(p \vee \neg p)$, in which case $v^c_\vdash$ takes us from correct to incorrect (Rumfitt, 2000, 807).

The situation in the bilateral case is analogous to the case of mcrs in that the bilateral formalism is expressive enough to associate every possible truth-value assignment with a statement of deducibility, thereby constraining consistent valuations tight enough so as to only permit boolean valuations

titude of *speech acts* or *attitudes toward contents* that ought to be taken into account in a theory of consequence. For this type of bilateralism see (Smiley, 1996), (Rumfitt, 2000), (Murzi and Hjortland, 2009), (Incurvati and Smith, 2010) and references therein, as well as (Humberstone, 2000) and (Button and Walsh, 2018, Ch. 13.6) for overview and discussion. For recent developments of positions of this type see (Incurvati and Schlöder, 2017, 2019, 2024).

[37] Such as, for example, $+\varphi, -\varphi \vdash^b_{\mathscr{L}} \bot$. See (Humberstone, 2000, 346ff.) and (Rumfitt, 2000, 804ff.).

[38] See (Murzi and Hjortland, 2009, 485), (Hjortland, 2014, 453), (Humberstone, 2000, 345) and (Murzi, 2010, 235).

[39] See, e.g., (Murzi and Hjortland, 2009, 485).

[40] See (Hjortland, 2014, 453/454) for a clean formal treatment.

[41] This inference constitutes an application of the negation-introduction rule for denial of Rumfitt's bilateral natural deduction calculus (Rumfitt, 2000, 802). We can convince ourselves of its correctness as follows: let v_c be a correctness valuation based on boolean valuation v (for the class of which classical bilateral consequence is sound and complete), s.t. $v_c(+\varphi) = c$. Hence, $v(\varphi) = 1$ and $v(\neg\varphi) = 0$. But then $v_c(-\neg\varphi) = c$ as well and the inference $+\varphi \vdash^b_{CPL} -\neg\varphi$ is thus correctness-preserving.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

as semantic values of the classical connectives. This is not surprising, for one can show that, in the classical case, mcrs and bilateral consequence relations are, in fact, intertranslatable.[42] This intertranslatability, however, "should not blind us to what is [...] a crucial philosophical difference" (Rumfitt, 2000, 810) between mcrs and bilateralist consequence. For bilateral systems not only allow for the harmonious formulation of an axiomatization of CPL in a single-conclusion setting, but also resolve *Carnap's Problem* while avoiding the objections brought forward against mcrs.[43]

The resolution of Carnapian underdetermination in the bilateral framework is achieved through the assumption of the force-markers '$+$' and '$-$'. It is therefore not surprising that the status of their meaning is critical to a well-motivated response to *Carnap's Problem*. In many ways, the force-marker of denial behaves just like a negation-operator,[44] raising suspicion that the bilateralist might have made a hidden, and potentially illegitimate, semantic assumption in her usage of '$-$' (see (Murzi and Hjortland, 2009, 486)). Given the (formal) similarity between unilateral negation and bilateral denial, then, it seems unclear how "bilateralism could possibly *hope* to offer any resistance against the semantic underdetermination argument" (Button and Walsh, 2018, 308).

But even if this line of reasoning could be resisted, and a distinction between negation and denial be upheld, a *revenge Carnap's Problem* might arise at the level of correctness valuations. For it might be asked with what justification the bilateralist excludes *non-normal correctness valuations* like $v_c^\top$, where $v_c^\top(\varphi) = \mathbf{c}$ for all $\varphi \in \textsc{Sent}^*_{\mathscr{L}}$ (Murzi and Hjortland, 2009, 486). Bilateral consequence will be consistent with $v_c^\top$, yet $v_c^\top$ will block the reconstruction of the boolean semantic value for negation from $\vdash^b_{CPL}$. *Carnap's Problem* has thus been shifted upwards to the level of correctness valuations, but is by no means resolved.

A proponent of this type of objection will see themselves accused of having failed to appreciate some of the bilateralist's basic assumptions. For the correctness norms governing assertion and denial are part and parcel of the bilateralist framework and not up for re-interpretation (see, e.g., (Incurvati and Smith, 2010, 10)). Taking them to possess the same openness and underdetermination as the logical constants misunderstands the bilateral approach: the question is not what the meaning of assertion and denial is, but "whether the content of the negation sign is fixed by the bilateral inferential practice *when added to a given background of the use of force-markers to construct sentences whose default use is for assertion and rejection*" (Incurvati and Smith, 2010, 10). The bilateral practice, however, is assumed given from the outset – the formal similarity between '$\neg$' and '$-$' belies this crucial difference between the status of the meaning of these symbols.[45] *Still*, it might be objected that the bilateralist approach to *Carnap's Problem* ultimately succeeds because it amounts to the tacit assumption of semantic principles going beyond those that can be established on an inferential basis alone. For some of the philosophical positions sketched in Section 2.3 this will constitute an issue.

Lastly, how does the bilateralist fare with respect to the quantifiers? Results by Button and Walsh suggest that they fare better than the unilateralist, but that preservation of correctness is still insufficient to determine the standard meanings of the first-order universal and existential quantifiers fully (see (Button and Walsh, 2018, Ch. 13)). However, it must be pointed out that a canonical (set-theoretic)

[42]In the sense that to every multi-conclusion sequent there corresponds a bilateral sequent consistent with the same class of valuations, and vice versa. See (Hjortland, 2014, 455), (Rumfitt, 2000, 810) and (Humberstone, 2000, 353ff.). See (Hjortland, 2014) for extension to *multilateral* frameworks for many-valued logics.

[43]For application, discussion and extension of the bilateral approach to resolving Carnapian underdetermination see (Rumfitt, 1997) and (Hjortland, 2014). Note, however, that extensions of the speech-act approach to alternative logics is, philosophically, not straightforward. For the behavior of the speech-acts in the classical two-valued case need not transfer to other logics, whose conception of the relationship of the different speech-acts to one another might be different, see Brendel (2024). Hence, a reformulation of the structural rules governing interactions between speech-acts might be necessary, which might lead to losing the previously attained solution to *Carnap's Problem*. The author thanks Elke Brendel for raising this issue.

[44]A suspicion backed up by various equivalence results, see (Button and Walsh, 2018, Theorem 13.11) and (Murzi and Hjortland, 2009, 486).

[45]The difference between bilateralist framework-assumptions and meaning-determination in their context is reflected at the level of rules: whereas the rules for the logical constants are operational rules, the rules for the force-markers are *structural rules* that articulate constraints on the underlying notion of deducibility itself.

semantics for bilateral quantification is still in its infancy, so that a conclusive judgement on this issue cannot be reached at this point.[46]

3.3 Re-interpreting Inference: Open-Endedness

Open-ended inferentialism is characterized by two general tenets: (i) *rules of inference* determine the meaning of the logical constants; and (ii) such meaning-determining rules "hold *always and without exception*" (Button and Walsh, 2018, 313).[47] The former indicates a shift from consequence relations to *rules* presenting consequence relations (this makes it possible to talk about rules continuing to hold, in full generality, in expansions of a language).[48] The latter is the principle of *open-endedness*, – the idea that "rules of inference are truth preserving within any mathematically possible extension of language" (McGee, 2000, 70), – grounded in the pre-linguistic and language-transcendent nature of inference (see, e.g., (Murzi and Topey, 2021)).[49] The open-ended character of rules of inference is codified through the demand that these rules continue to hold, no matter how the language is expanded.

Button (2016) and Button and Walsh (2018) show that the requirement of open-endedness suffices to pin down the intended semantic space, a two-valued Boolean algebra, among all possible Boolean algebras. Under the assumption that the semantic space providing possible interpretations of the connectives must be a Boolean algebra, the connectives are thereby uniquely determined.[50] Despite these strong results, *Carnap's Problem* remains, at least partially, unaddressed, for the determination succeeds under the assumption that the relevant semantic space forms a Boolean algebra, which precludes certain non-normal but inferentially admissible valuations from the outset. Yet, the restriction to Boolean algebras itself remains inferentially unmotivated.[51]

McGee pursues a different route to unique determination. According to his interpretation of the open-endedness requirement the rules for a connective c must continue to hold when the language is extended with a duplicate c' of that connective, governed by identical rules of inference. A connective c is uniquely determined by its rules if sentences including it are interderivable with sentences that are identical, except that occurrences of c have been replaced with its duplicate c', and vice versa (McGee, 2000, 2015). Such interderivability ensures that the inferential role of a connective c has been so tightly constrained by its rules that there is but a unique candidate filling that role.[52] The usual connectives of propositional logic all possess rules satisfying this requirement (Harris, 1982).

Assuming a further soundness condition for the uniquely characterizing rules with respect to a given semantic space ensures that a constant and its duplicate will possess identical semantic values (McGee, 2000). They are, therefore, not only inferentially, but also semantically uniquely determined. The rules for the usual connectives and quantifiers of FOL, understood as open-ended rules, "create a uniquely defined semantic role for each of the connectives and quantifiers" (McGee, 2000, 68). Note, however, that this type of uniqueness still falls short of solving *Carnap's Problem*. For while inferential uniqueness is sufficient to ensure sameness of semantic values for constants governed by identical sets of rules, it still cannot guarantee that there is only one possible interpretation of the constants with respect to which they are sound (McGee, 2006, 193). It achieves, in other words, identical but not categorical interpretations – the constants are unambiguous, but not unique.[53]

[46] Though see (Incurvati et al., 2019).

[47] There usually is a further implicit assumption active, namely, that every object can, in principle, be named, which I will ignore in the following, however.

[48] Note that this view is compatible with a variety of *rule-formats* which, in turn, influence the possibility of resolving *Carnap's Problem*. To not obscure what I take to constitute the (philosophical) core of the open-ended strategy I will remain non-committal about what format these rules take in the following, though see (Warren, 2020) and (Murzi and Topey, 2021) for concrete implementations.

[49] See esp. (McGee, 2000, 2006, 2015) for motivation and development of the idea of the open-ended nature of rules.

[50] See (Button, 2016) and (Button and Walsh, 2018, Ch. 13.7) for details.

[51] This is of course noticed by (Button, 2016, 13) and (Button and Walsh, 2018, n. 25).

[52] See (Belnap, 1962) for the original suggestion and motivation of this form of *inferential uniqueness*. Cf. (Došen and Schroeder-Heister, 1988) for an in-depth study.

[53] This is a very abbreviated summary of McGee's views. See (McGee, 2000, 2006, 2015) for further elaboration and (Brîncuş, 2019; Brîncus, 2024a,b) for comment and criticism.

How does open-ended inferentialism fare in the case of the quantifiers? Motivated by Quinean and Putnamian concerns regarding indeterminate quantifier meanings and restricted quantification, McGee shows that the possibility of deviant meanings of the quantifiers is undermined by the open-endedness of the quantifier rules (McGee, 2006, 191).[54] Carnapian underdetermination of the quantifiers showed that the standard rules of universal quantification are sound for an interpretation according to which the variables range over a proper subset of the domain, so long as this subset constitutes the base X of a principal filter over the domain M. The meaning of $\forall$ therefore amounts to something like 'for all X' rather than 'for all elements of the domain M'. Under the assumption that the rules for the universal quantifier are open-ended, and thus need to remain valid no matter how the language is extended, they remain applicable when the language is extended by a predicate P, s.t. $[\![P]\!]^{\mathcal{M}} = M$, and a constant c, s.t. $[\![c]\!]^{\mathcal{M}} = \mathbf{a} \notin X$. In this extended language, $\forall x Px$ will be true, yet Pc will be false in $\mathcal{M}$. But this conflicts with the rule of universal instantiation according to which $\forall x Px \vdash_{\mathscr{L}} Pc$ (McGee, 2000, 68). Since this argument can be reiterated as long as the base of the principal filter providing the interpretation of $\forall$ is not maximal, demanding that the rules for the universal quantifier be open-ended appears to force it to take on its standard interpretation: "So, the default value of '$\forall$' [...] is quantification over everything" (McGee, 2000, 69).

The possibility of excluding non-normal interpretations of the quantifiers stems from the unconstrained nature of naming: "singular terms are unconstrained in their taking denotations [...], thereby giving access to the 'dark corners' of the first-order domain where the light of the quantifiers does not shine" (Antonelli, 2013, 638/639). While the rules for the quantifiers by themselves "do not determine the range of quantification", they ensure that "the domain of quantification in a given context includes everything that can be named within that context" (McGee, 2000, 69). Combining this with the idea that the rules must remain valid under any extension of the language, and modulo any artificial restrictions on naming and designation, forces the standard interpretation of the quantifiers. The fact that the reach of singular terms might outstrip the range of quantification was the reason that Bonnay and Westerståhl (2016) had to close the admissible interpretations of the quantifiers under the interpretation of terms. Insistence on open-endedness makes the possibility of naming global and overcomes the remaining local underdetermination, therefore constituting a promising approach to resolving *Carnap's Problem* at the level of quantification.[55]

3.4 Meta-inferential Determinacy: Local Models of Rules

In devising a stable foundation for a moderate inferentialist position that escapes Carnapian underdetermination J.W. Garson, in a series of works (Garson, 1990, 2001, 2010, 2013), slightly shifts the parameters of the way *Carnap's Problem* was construed above, bringing it more in line with traditional inferentialist approaches to meaning.[56] Thus, instead of taking entire consequence relations as basic, Garson considers *sets of rules* instead. Moreover, the rules he considers possess a *meta-inferential* format – they no longer govern transitions between sentences but describe permissible transitions between entire *arguments*. I.e., the rules have the general format of

$$\frac{\Gamma_1 \vdash \varphi_1 \qquad \cdots \qquad \Gamma_n \vdash \varphi_n}{\Delta \vdash \psi}\; R$$

[54]McGee's concerns are slightly different from those of the current paper. While he wishes to rule out Putnamian non-standard models and defend the unrestricted interpretation of universal quantification as the intended and determined interpretation, we are concerned with worries about Carnapian-style underdetermination here.

[55]For applications of the open-endedness strategy to resolve *Carnap's Problem* see (Warren, 2020) and (Murzi and Topey, 2021); cf. also Section 5.3. See (Brîncus, 2024a,b) and (Picollo, 2025) for critical assessment.

[56]Garson, especially (Garson, 2013), pursues a somewhat different project than the one we are engaged in here. Rather than determining the conditions under which a pre-given meaning is determined by a set of associated inferences he asks what, given the latter, an appropriate semantics (the *natural semantics* of an operator) is for the expressions occurring in these inferences, such that such a semantics is uniquely determined by them. His work includes an elegant solution to *Carnap's Problem* in the way it is conceived here. Garson (2013) extends consideration beyond propositional connectives and further formulates and investigates the *natural semantics* for quantifiers and modal operators as well.

The particular rules for disjunction adopted by Garson, for example, are:[57]

$$\frac{\Gamma \vdash \varphi_i}{\Gamma \vdash \varphi_1 \lor \varphi_2} \lor\mathrm{I},\, i \in \{1,2\} \qquad\qquad \frac{\Gamma \vdash \varphi \lor \psi \qquad \Gamma, \varphi \vdash \chi \qquad \Gamma, \psi \vdash \chi}{\Gamma \vdash \chi} \lor\mathrm{E}$$

The shift to a rule-based presentation of transitions between arguments allows for the formulation and implementation of further requirements constraining the relationship between inference and model-theoretic meaning. Carnapian underdetermination, Garson says, is the result of ignoring the way inferential patterns of a constant are given: "The deductive benchmark we are using for what counts as a model of a system is completely insensitive to the rules that are used to formulate it. It has been assumed that all that matters for specifying the inferential relations set up by a logic are the arguments that qualify as provable in the system. However, this view is shortsighted" (Garson, 2013, 15). Admissible models, then, need not just be consistent with the arguments deemed acceptable by the consequence relation, but with the rules themselves: "[a] model of the arguments [...] is insensitive to principles that regulate *how* one deduces new arguments from old ones, and this information matters to the interpretation of the connectives" (Garson, 2010, 166).

For our purposes, the interesting case of *consistency with a rule* is Garson's criterion of *local consistency*:[58] a valuation v is *(locally) consistent with a rule* R if, whenever v is consistent with the rule-premises it must also be consistent with the rule-conclusion (where the notion of a *valuation* and *consistency with a rule-premise-/conclusion* are identical to the notions defined at the beginning of the previous section). A valuation v is thus consistent with a meta-inferential rule R in case v preserves (standard) consistency from the rule-premises to the rule-conclusion. Excluding the valuation $v_\top$ on the basis of a *non-triviality constraint* (see Section 4.1 below), Garson then shows that local consistency with the rules adopted for the classical connectives suffices to establish their boolean meanings (Garson, 2010, 2013). Hence, adopting a meta-inferential, rule-based specification of the inferential behaviour of the classical connectives suffices to solve *Carnap's Problem*.

By way of example, consider, once more, the deviant valuation $v_\vdash$ from above. What, in the current framework, renders $v_\vdash$ inadmissible from the inferential perspective? Recall that, for some propositional letter p, $v_\vdash(p) = v_\vdash(\neg p) = 0$ and assume $v_\vdash$ was consistent with the rules for disjunction. Observe that $v_\vdash$ is consistent with (i) $\vdash_{\mathscr{L}} p \lor \neg p$, (ii) $p \vdash_{\mathscr{L}} p$ and (iii) $\neg p \vdash_{\mathscr{L}} p$. Hence, by the assumed consistency of $v_\vdash$ with $\lor$E, it follows that $v_\vdash(p) = 1$ – contradiction. Hence, $v_\vdash$ is inconsistent with the rules for disjunction. The meta-inferential formulation of the rules coupled with the adoption of a local consistency constraint allows us to make effective use of false sentences in antecedents of rule premises, thereby obtaining access to those rows of a truth-table that were out of reach of simple consequence relations (Garson, 2013, 39).[59]

Despite its success in resolving Carnapian underdetermination Garson ultimately abandons the standard of local consistency. This might be justified on the grounds that local consistency constitutes the wrong standard of consistency for rules that govern transitions not between sentences, but between arguments: while individual inferences must be truth-preserving, meta-inferences – transitions between inferences, – ought to be validity-preserving.[60] Validity, however, is a global phenomenon and should, accordingly, not be assessed 'one valuation at a time', but 'wholesale' with respect to a class of (appropriate) valuations. Thus, whenever a (meta-inferential) rule premise is *valid*, so should be its conclusion. This gives rise to the standard of *global consistency* according to which a *set of valuations* $\mathcal{V}$

[57] See (Garson, 2013, 35ff.) for a full statement of the system he adopts.

[58] Garson ultimately opts for a different standard of consistency, see below.

[59] Garson's notion of *local consistency* leads to issues in the case of the rules for the quantifiers. For generating a *natural semantics* for quantifiers, see (Garson, 2013, Ch. 14).

[60] I am here following the account given in (Picollo, 2025, Sect. 1.1). Garson's own reasoning is different: he rejects local consistency due to the semantics determined by the rules under this standard yielding a meaning for the connectives that outstrips the rules' 'inferential content'. In other words, they determine model-theoretic values that validate inferences which cannot be shown valid on the basis of the rules themselves, they give rise to meanings that are incomplete for the rules that determined them. Garson takes this asymmetry to be undesirable and thus advocates for a stronger standard of consistency to align inferential and model-theoretic meanings more closely, see (Garson, 2013, Ch. 3 & 4).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

is (globally) consistent with a (meta-inferential) rule R just in case whenever all the rule-premisses are consistent with every $v \in \mathcal{V}$, the rule conclusion must also be consistent with every $v \in \mathcal{V}$. If, in other words, the rule preserves ($\mathcal{V}$-)validity (Garson, 2013, 15ff.). Global consistency succeeds in weakening the meanings of the logical constants determined by this standard of consistency so as to remove the asymmetry between what can be established on the basis of the rules and on the basis of the model-theoretic meanings thus determined.[61] However, it fails to remove Carnapian underdetermination and thus to solve *Carnap's Problem* in the way it was constructed in the context of this paper.[62]

Murzi and Topey (2021) develop a variation of Garson's approach that (a) succeeds in resolving Carnapian underdetermination, (b) discharges the unmotivated assumption of the exclusion of $v_\top$ on non-inferentialist friendly grounds, and (c) allows an extension of the approach to quantifier-rules. (a) is, essentially, achieved through the adoption of a (suitably adjusted) standard of local consistency. However, to accommodate the quantifiers and overcome incompleteness phenomena that lead Garson to ultimately abandon the local standard of consistency they generalize his framework significantly. Without fully discussing their proposal in detail, central aspects include the adoption of a calculus of meta-inferential rules featuring single-conclusion arguments that may consist of open formulas in addition to sentences (the inclusion of open formulas makes it necessary to adapt the standard of local consistency to now involve, in addition to models, also variable assignments. The result is what Picollo (2025) terms a *hybrid account*, with models obeying a local, and variable assignments a global standard of consistency).

Moreover, their calculus allows for the possibility of higher-order rules, i.e., the assuming and discharging of *rules*, which makes it possible to interpret *falsum* as a *punctuation mark*, – signaling a dead-end in an argument, – instead of a logical constant, and renders the rule of *reductio ad absurdum* a structural rule (thereby achieving (b) above).[63] They modify the quantifier rules to allow for open formulas to feature in the respective premise- and conclusion-sequents and, further, adopt an open-endedness constraint (see Section 3.3). This allows them to resolve *Carnap's Problem* uniformly for the propositional connectives and the quantifiers (see Section 5.3 for generalizing this strategy to higher-order quantifiers).[64]

Moving from consequence relations to their rule-based presentations, together with the adoption of a specific rule-format, succeeds in avoiding Carnapian underdetermination. Both parameters play an essential role in resolving *Carnap's Problem.* They thus require philosophical justification if they are to feature in a successful defense of the moderate inferentialist position. Even independently of the question whether open formulas are legitimate constituents of (a natural model of) arguments, and whether argumentative practice is best captured in meta-inferential terms,[65] the shift towards rules together with the requirement of a specific rule-format introduces a rather significant, and potentially unwelcome, presentation-dependency: "[t]he upshot of this is that the model of rules criterion is sensitive to details concerning how a system is formulated" (Garson, 2010, 166). Whether or not Carnapian underdetermination is resolved thus seems to depend on particular, and potentially unstable, dynamics of reasoning.

[61]It secures other desirable properties of rule-based specifications of meaning as well, see (Humberstone, 1996).

[62]This is not to the detriment of the project pursued in (Garson, 2013) however, who is engaged in a project different from the one pursued here (see fn. 56). Cf. also Picollo (2025) who develops and argues for a *sentential semantics* of the first-order quantifiers, different from their substitutional and objectual interpretations, as the 'right' semantics of the quantifiers determined on the basis of their rules in an inferentialist setting.

[63]Their calculus is based on Murzi (2020). See Schroeder-Heister (1984) for the framework of higher-order rules.

[64]The approach of Murzi and Topey (2021) has several further nice properties that overcome weaknesses of Garson's setting but that we cannot address in the context of this paper. See Brîncus (2024a) and Picollo (2025) for criticisms of their strategy.

[65]See, e.g., Picollo (2025).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

4 Semantic Strategies for Solving Carnap's Problem I (the propositional case)

The basic idea of *semantic* strategies for solving *Carnap's Problem* consists in restricting the space of possible valuations or models that can serve as interpretations of the logical symbols from the outset. Underlying this is the idea that *Carnap's Problem* "is made artificially difficult by considering all possible interpretations [of the language], no matter how bizarre" (Bonnay and Westerståhl, 2016, 733). Thus, not all consistent models are equally legitimate as some might be excluded on the basis of considerations having to do with general linguistic competence, constraints of logicality, or other factors.

4.1 Valuations vs Interpretations

Do all consistent valuations constitute legitimate *interpretations* of the logical expressions of a language? Is consistency with inferential behaviour, in other words, sufficient for being considered a potential meaning? Bonnay and Westerståhl (2016) argue that it is not.[66] General principles underlying linguistic competence, they claim, narrow down the space of candidate interpretations from the outset, making *Carnap's Problem* thereby easier to track.

One such constraint put forward in (Bonnay and Westerståhl, 2016) is a principle of *non-triviality*:

> (Non-Triv) *Every language contains at least one false sentence.*

Such a principle, they say, is "a very weak requirement, hardly in need of motivation" (Bonnay and Westerståhl, 2016, 725); after all, drawing some kind of distinction between what is true and what is false seems fundamental to the functioning of language. Nonetheless, the adoption of such an obvious constraint on potential interpretations is not inert: it rules out $v_\top$ as inadmissible.

(Non-Triv) is a very natural constraint on a semantic space. Note, though, that if the motivation for adopting it was that a language must be able to draw some kind of distinction to be considered language at all it might best be interpreted as a demand of *non-uniformity*: at least two sentences of the language must take on different truth-values. In this formulation, the constraint of non-uniformity would also rule out the valuation according to which every sentence of the language is false, a valuation violating the classical truth-tables and inconsistent with classical consequence due to the fact that classical logic possesses tautologies.[67]

As soon as we move to a multi-valued setting (see Section 4.2), however, the assumption of non-uniformity becomes problematic. This is the case since the valuation that assigns all atomic sentences the third truth-value in three-valued logics governed by strong Kleene truth-tables extends to a valuation that assigns *all* sentences of the language the third value, thereby violating the constraint of non-uniformity. This valuation, however, plays an important role in the meta-theory of such logics (demonstrating, for example, that K_3 has no theorems). So as simple and natural as non-triviality appears on first view, more might have to be said about the particular shape it takes in different logics.

A more significant requirement defended in (Bonnay and Westerståhl, 2016) and (Westerståhl, 2025) is a constraint of *compositionality*. Compositionality has sometimes been put forward as a *semantic universal*, a principle universally instantiated across languages explaining the otherwise mysterious phenomenon of *linguistic creativity*, the ability of speakers to understand and produce a potential infinitude of meaningful expressions based on finite amounts of data. The principle of compositionality states that[68]

> (Comp) *The meaning of a complex expression is a function of the meanings of its constituting expressions and their way of composition.*

[66] Cf. also (Westerståhl, 2025) for an elaboration of this argument.

[67] Cf. (Tabakçı, 2024, n. 28) for further comment on (Non-Triv).

[68] See (Pagin and Westerståhl, 2010a,b) for precise statement and discussion and (Davidson, 2001) for a defense of a compositionality requirement.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

To be admissible, then, a valuation needs to respect (Comp) – otherwise it won't even be recognized as an acceptable candidate based on rudimentary linguistic competence. In the context of the language of propositional logic in which meanings are truth-values, (Comp) forces the interpretation of a logical constant to be a *truth-function*. *Which* truth-function a particular constant then denotes is determined by the constant's inferential behaviour.

More precisely, *compositionality* requires that the meaning of a constant c is given by a function f_c, s.t. the meaning of a compound expression $c(\varphi_1, \ldots, \varphi_n)$ under a valuation v, $v(c(\varphi_1, \ldots, \varphi_n))$, is a function of the meaning of c, f_c, applied to the values of $\varphi_1, \ldots, \varphi_n$ under v: $v(c(\varphi_1, \ldots, \varphi_n)) = f_c(v(\varphi_1), \ldots, v(\varphi_n))$. Since the possible semantic values of expressions under v are the truth-values 0 and 1, f_c must be a truth-function. A valuation v is said to be c-compositional for a connective c iff there exists a truth-function f_c for c, s.t. $v(c(\varphi_1, \ldots, \varphi_n)) = f_c(v(\varphi_1), \ldots, v(\varphi_n))$ for all $\varphi_1, \ldots, \varphi_n \in \text{SENT}_{\mathscr{L}}$.

Consider, once again, $v_{\vdash}$ to see how the requirement of compositionality suffices to rule out Carnap's non-normal valuations. By (Comp) we know that there must be a truth-function $f_\vee$, s.t. $v(\varphi \vee \psi) = f_\vee(v(\varphi), v(\psi))$ for all $\varphi, \psi \in \text{SENT}_{\mathscr{L}}$ and $v \in \text{VAL}_{\mathscr{L}}$. In particular, then, $v_{\vdash}(p \vee \neg p) = f_\vee(v_{\vdash}(p), v_{\vdash}(\neg p))$. But note that $v_{\vdash}(p) = v_{\vdash}(\neg p)$ and thus $1 = v_{\vdash}(p \vee \neg p) = f_\vee(v_{\vdash}(p), v_{\vdash}(\neg p)) = f_\vee(v_{\vdash}(p), v_{\vdash}(p)) = v_{\vdash}(p \vee p) = 0$ – contradiction (Bonnay and Westerståhl, 2016, 728).[69] Hence, $v_{\vdash}$ is not an admissible valuation, it is not a legitimate candidate for meaning.

Bonnay and Westerståhl (2016) show that (Non-triv) and (Comp) suffice to rule out Carnap's non-normal valuations and to secure the intended values of the propositional connectives. According to (Westerståhl, 2025) this is a favourable and in some sense natural result, for non-compositional valuations shouldn't have even been considered legitimate candidates for interpretations of a language in the first place, as they violate principles characteristic of basic linguistic competence: "absent compositionality, the idea of meanings makes little sense" (Westerståhl, 2025, 1).[70]

4.2 Carnap's Categoricity Problem in Three Values

Westerståhl (2025) demonstrates the effect the assumption of compositionality has on the determination of semantic values of logical constants for a variety of logics and semantics. Here, we are concerned with the possibly most straightforward extension of the semantics of classical propositional logic to a three-valued context, and the question whether compositionality still suffices in this slightly richer setting to uniquely determine the intended semantic values of the logical constants.[71]

A *three-valued logic* shares the language of classical propositional logic. A valuation v is now a total function $v : \text{SENT}_{\mathscr{L}} \to \{0, u, 1\}$, where u is a *third truth-value*. We denote the class of all three-valued valuations by $\text{VAL}^3_{\mathscr{L}}$. The *designated values* of the family of logics we are interested in in the following are the members of the set $\mathcal{D} = \{1, u\}$. For $\mathcal{V} \subseteq \text{VAL}^3_{\mathscr{L}}$, a sentence φ is a $\mathcal{V}$-consequence of a set of sentences Γ, $\Gamma \models^{\mathcal{D}}_{\mathcal{V}} \varphi$, if it preserves designation. If, in other words, for all $v \in \mathcal{V}$, whenever $v(\Gamma) \in \mathcal{D}$, then $v(\varphi) \in \mathcal{D}$ as well. All other notions are analogous to those defined above.[72]

The specific three-valued logics we are concerned with in this section are given by the two classes of valuations $\mathcal{V}_K$ and $\mathcal{V}_{G_3}$, where $v \in \mathcal{V}_K$ iff v obeys the *strong Kleene Schema* and $v \in \mathcal{V}_{G_3}$ iff v obeys the *Gödel Schema*:

Strong Kleene Schema

[69] Bonnay and Westerståhl (2016) attribute this observation to (Carnap, 1943). It can also be found in (Belnap and Massey, 1990).

[70] Other semantic constraints to shrink the space of admissible models could be considered. Johannesson (2022) defends *completeness* as providing a philosophically motivated constraint ruling out non-standard interpretations of the classical connectives. See also Speitel and Tabakçı (2025) for comment and extension of this idea to non-classical logics, as well as for discussion of other semantic principles to achieve categoricity.

[71] The following is based on joint-work with D. Bonnay and S.K. Tabakçı. See (Tabakçı, 2024) for a systematic and detailed investigation of *Carnap's Problem* for K3 and LP, with which the below shares many details.

[72] For a detailed treatment of multi-valued logics see, e.g., (Priest, 2008).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

$\neg$	
1	0
u	u
0	1

$\wedge$	**1**	**u**	**0**
1	1	u	0
u	u	u	0
0	0	0	0

$\vee$	**1**	**u**	**0**
1	1	1	1
u	1	u	u
0	1	u	0

$\rightarrow$	**1**	**u**	**0**
1	1	u	0
u	1	u	u
0	1	1	1

Gödel Schema

$\neg$	
1	0
u	0
0	1

$\wedge$	**1**	**u**	**0**
1	1	u	0
u	u	u	0
0	0	0	0

$\vee$	**1**	**u**	**0**
1	1	1	1
u	1	u	u
0	1	u	0

$\rightarrow$	**1**	**u**	**0**
1	1	u	0
u	1	1	0
0	1	1	1

Note that the strong Kleene and the Gödel Schema agree on the truth-tables for conjunction and disjunction, but disagree on the tables for negation and the conditional.

It can now be shown that $\Gamma \models^{\mathcal{D}}_{\mathcal{V}_{G_3}} \varphi$ iff $\Gamma \vdash_{CPL} \varphi$.[73] In other words, the logic generated by the class of valuations $\mathcal{V}_{G_3}$ with designated values $\mathcal{D} = \{1, u\}$ is a three-valued presentation of classical propositional logic. The logic generated by the class of valuations $\mathcal{V}_K$ and designated values $\mathcal{D} = \{1, u\}$, on the other hand, is the logic LP. Consider now a valuation $v^+ \in \mathcal{V}_{G_3}$, s.t. $v^+(p) = v^+(q) = u$ for some atomic sentences $p, q \in \textsc{Sent}_{\mathscr{L}}$. Then, it is easy to observe the following:

(i) $v^+ \notin \mathcal{V}_K$ since, for example, $v^+(p \rightarrow q) = 1$.
(ii) v^+ is *compositional* as witnessed by the Gödel Schema.
(iii) v^+ is *consistent with* $\models^{\mathcal{D}}_{\mathcal{V}_K}$: let $\Gamma \models^{\mathcal{D}}_{\mathcal{V}_K} \varphi$ and suppose that $v^+(\gamma) \in \mathcal{D} = \{1, u\}$ for all $\gamma \in \Gamma$, but $v^+(\varphi) = 0$. That means that $\Gamma \not\models^{\mathcal{D}}_{\mathcal{V}_{G_3}} \varphi$ and, therefore, $\Gamma \nvdash_{CPL} \varphi$. However, since LP is a sublogic of CPL, – i.e., if $\Gamma \models^{\mathcal{D}}_{\mathcal{V}_K} \varphi$ then $\Gamma \vdash_{CPL} \varphi$, – it follows that $\Gamma \not\models^{\mathcal{D}}_{\mathcal{V}_K} \varphi$, contradiction to the assumption. Hence, v^+ is consistent with $\models^{\mathcal{D}}_{\mathcal{V}_K}$.

The above demonstrates that *compositionality* by itself is insufficient to rule out non-normal valuations for logics moderately richer than classical propositional logic; that there are, in other words, unintended, yet compositional, valuations consistent with the consequence relation of, for example, LP.[74] Tabakçı (2024) investigates in detail just how deep the failure of compositionality in fixing intended interpretations runs for the three-valued logics LP and K3, and discusses possible solution strategies.[75]

Much more would need to be said about the case of three- and multi-valued logics in general to reach a conclusive verdict. However, what the simple example above seems to demonstrate is that the constraint of compositionality would have to be refined further to keep ruling out inadmissible valuations in the case of logics with more than two values. In particular, the above seems to point in the direction of compositionality not being a *local* property of a valuation, but rather a *global* property of a class thereof, for notice that there will not be a single truth-function f interpreting, say, $\rightarrow$, s.t. all $v \in \mathcal{V}_K \cup \{v^+\}$ will be compositional w.r.t. f.[76] We will not pursue the issue further here, but see (Tabakçı, 2024) for a more detailed investigation of *Carnap's Problem* in the three-valued setting.[77]

Foreshadowing some of the concluding remarks of this paper, (Tabakçı, 2024) investigates a further, elegant solution for resolving *Carnap's Problem* for the logics K3 and LP: demanding that the valuational

[73] Cf., e.g., (Open-Logic-Project, 2024).

[74] Note that the above actually provides a *method* for demonstrating the failure of compositionality in fixing intended interpretations for any (properly) sub-classical logic. See (Tabakçı, 2024) for a more elaborate and differentiated investigation of the failure of compositionality for fixing standard interpretations in the three-valued setting, making use of different 'non-normal' matrices in a many-valued SET–SET framework.

[75] In particular, Tabakçı (2024) demonstrates that even in the more expressive SET–SET framework, and under the assumption of additional logical constants, can the non-normal valuations not be excluded unless one is willing to adopt further strong and somewhat ad-hoc semantic assumptions.

[76] See (Tabakçı, 2024, n. 25) for a similar observation.

[77] Cf. also (Speitel and Tabakçı, 2025).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

space obeys the constraint that a complex formula take on the third value u when all its immediate subformulas take on the value u establishes categoricity in a multi-conclusion setting. That such a constraint should be adopted, however, appears to be intimately connected with the role one takes u to play in the semantics of the logic. If a sentence receives the value u because it is *ungrammatical* or otherwise *meaningless*, as, for example, in Bochvar's three-valued logic B_3, it is natural that this property is inherited by more complex sentences involving a constituent with value u, informing the resulting truth-tables of the language. If u is taken to indicate a contingency or indeterminacy, as in Łukasiewicz's three-valued logic $Ł_3$, for example, it becomes less clear why that indeterminacy should be inherited in all cases by more complex sentences and thus whether the semantic constraint is appropriate to impose.[78] Whether it is therefore reasonable to accept such a constraint depends on the philosophical foundations of the logic formalized by means of the relevant matrices.[79]

4.3 Beyond Classical Settings: Intuitionistic Connectives and Modal Operators

Carnap's question can be asked for any logic and logical constant therein. It appears naturally in multi-valued settings and beyond. In (Bonnay and Westerståhl, 2016, Sect. 6) it was established that the classical connectives retain their intended interpretations under the assumption of (Non-Triv) and (Comp) when moving to a possible-world semantics, where sentence-values no longer simply consist of truth-values, but *sets of points* or *possible worlds*.[80] Here, the intended interpretation of '$\neg$' is the complementation operation on the powerset of the set of worlds, '$\wedge$' the union-operation on the same set, and so on. These values are uniquely determined by the classical consequence relation over the enriched semantic space as long as constraints of *non-triviality* and *compositionality* are adhered to.[81]

Of course, once the switch to a more expressive semantics has been made, questions about determinacy don't just arise for the usual logical operators, but also and especially for the novel operators characteristic to the richer settings. In the possible worlds framework this includes, in particular, the modal operator $\Box$.[82] Here, *Carnap's Problem* takes on additional complexities: on the one hand, modal logics admit a plethora of different *types of semantics* – from Kripke, over topological, to neighbourhood semantics, – all with reasonable claim to being an appropriate semantics for the modal language under consideration. On the other hand, the meaning of $\Box$ in modal logic is a different *type of meaning* than that of the other propositional connectives. For while the pointwise truth of a formula without $\Box$ in a model depends only on the values of the subformulas of the formula *at that point*, the pointwise truth of a boxed formula in a model also depends on the values of its subformulas *at other points* of the model (though, usually, not *all* others). Semantically, this special status of the meaning of $\Box$ is captured by including an additional parameter in the model (*accessibility relations* in Kripke frames, the *interior function* in topological frames, *neighbourhoods* in neighbourhood frames) with respect to which truth of a boxed formula at a point in a model is determined.

To fruitfully ask *Carnap's question* in this enriched framework, then, some preliminary issues have to be settled. Relating to the first point raised above, a choice of semantics has to be made: given the different types of semantics, with respect to which are we asking whether the inferential behaviour of $\Box$ determines its intended value? Second, what is characteristic of the intended meaning of the $\Box$ operator within that semantics, and how can this feature best be captured? Ideally, this latter consideration can be translated into semantic constraints restricting the values deemed legitimate for $\Box$ within the chosen

[78] See Malinowski (1993) for interpretations of the third truth-value.

[79] Further ways of resolving Carnapian underdetermination in the three-valued case are explored in (Tabakçı, 2024), where it is demonstrated that fixing the standard behaviour of strong Kleene negation suffices to solve the underdetermination of the other connectives of the logic as well. A similar result had been discovered independently by K. Chatain (unpublished). For proof-theoretic strategies resolving *Carnap's Problem* in the multi-valued case, see also (Rumfitt, 1997) and (Hjortland, 2014).

[80] Cf. also (Westerståhl, 2025).

[81] An interesting fact about this result is that whereas in the case of the two-valued semantics already intuitionistic consequence suffices to fix classical meanings, in this richer setting genuinely classical patterns of inferences are needed to determine the intended interpretations, see (Bonnay and Westerståhl, 2023).

[82] Cf. also (Garson, 2013, Ch. 16) for a discussion of the semantics determined by axioms and rules for modal logics.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

semantics.

Bonnay and Westerståhl (2023) take *possible worlds semantics* to constitute an appropriate compositional semantics for modal languages. The most general type of possible worlds semantics is *neighbourhood semantics*. This settles the first question in a well-motivated manner. What is distinct about the intended interpretation of $\Box$ is that, unlike $\neg$ or $\vee$, it does not receive a unique interpretation, possibly indexed by world domains but, rather, what is codified in *Kripke semantics*: that the meaning of $\Box$ is such as to allow one to recover an accessibility relation between worlds, such that the truth of a boxed formula at a world depends on the truth of its subformulas at accessible worlds. This captures the idea that an intended meaning of $\Box$ is such that the truth-values at a point of formulas involving it may depend on truth-values of its subformulas at different points of the model. This settles the second question concerning $\Box$'s *meaning-type*. A first version of *Carnap's question* then asks under what conditions a modal consequence relation forces consistent neighbourhood interpretations to be, in the informal way described above, *Kripkean* (Bonnay and Westerståhl, 2023, 585).[83]

Unlike propositional or first-order logic, the framework of modal logic does not constitute a single theory, but rather gives rise to a *family of theories* in which the behaviour of $\Box$ can be characterized by different axioms and rules. Since the modal logic K, – containing the characteristic axiom $\Box(p \to q) \to (\Box p \to \Box q)$ and closed under the (meta-)rule of necessitation "from $\vdash \varphi$ infer $\vdash \Box\varphi$" (modal logics satisfying these are *normal modal logics*), – is sound and complete with respect to the class of *all* Kripke frames, it can be taken to articulate some basic adequacy constraints on the meaning of $\Box$. *Carnap's question* thus becomes: under what circumstances do modal consequence relations extending K force a neighbourhood frame to be Kripkean (Bonnay and Westerståhl, 2023, 585)?

For *finite frames* the constraints imposed by K suffice to force the intended interpretation of $\Box$ (Bonnay and Westerståhl, 2023, Theorem 12) – as well as of the remaining constants, – but this result does not generalize to the infinite case (Bonnay and Westerståhl, 2023, Fact 13). Here, further semantic constraints are required to limit the space of consistent interpretations to the class of intended ones. Yet, what sort of semantic constraint is well-motivated based on the choice of semantics for modal languages? What is distinctive of the semantics of the modal elements of these languages?

What is characteristic of the meaning of $\Box$ is that it contributes a *local flavour* to the evaluation of formulas including it: their truth depends on the truth of their subformulas at other 'nearby' points of the model as well.[84] Yet, the set of points that needs to be taken into consideration for the truth of a boxed formula at a point falls (usually) far short of the entire domain of the model. Distinctive of the semantic value of $\Box$ thus seems to be a type of *truth-locality* – some, but not all, points of the model are relevant for the truth of formulas including it. This locality can be semantically expressed by means of *bisimulation invariance*:[85] "[t]he locality of modal logic [...] is often framed in terms of invariance under bisimulation: bisimilar worlds satisfy the same modal formulas, so that only the local features of the structure of the Kripke models [...] matter to modal satisfaction" (Bonnay and Westerståhl, 2023, 598).[86]

Does bisimulation invariance, in conjunction with consistency w.r.t. the consequence relation of modal logics extending K, suffice to yield Kripkeanity?[87] Near enough: together with a further closure condition, bisimulation invariance indeed ensures Kripkeanity, and thus the intended (type of) interpretation of $\Box$ (Bonnay and Westerståhl, 2023, Theorem 44 & Corollary 45). It is worth emphasizing that the adoption of a constraint of bisimulation invariance was not arbitrary. Rather, its acceptability was the result of a reflection on the underlying motivations of the semantic framework itself – what was characteristic of the meaning of the expressions of modal languages was their locality. Once a

[83]We deviate slightly from the terminology in (Bonnay and Westerståhl, 2023). For precise definitions and statements of the results mentioned here we refer the reader to this paper.

[84]Cf. (Bonnay and Westerståhl, 2023, 600): "The distinctive feature of modal logic [...] is the fact that modal evaluation is local; everything is taking place at a world and at worlds reachable from that world."

[85]See, e.g., (Blackburn et al., 2001) for definition.

[86]Bonnay and Westerståhl (2023) also consider a further notion of locality, *subframe invariance*, which we leave out here.

[87]We here present a very simplified picture of the situation and ignore many of the formal subtleties of Bonnay & Westerståhl's account. For precise set-up, definitions, and details see (Bonnay and Westerståhl, 2023).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

particular semantics was chosen, the way that locality could be expressed and captured relative to that framework came with it.

Similar choice-points as in the modal case can also be observed in intuitionistic logic(s): here, too, there is an embarrassment of riches when it comes to adopting a semantics – from Kripke over Beth-, topological, Dragalin and, finally, algebraic semantics there are plenty of candidates to consider with respect to which *Carnap's question* could be asked. Based on results in (Bezhanishvili and Holliday, 2019), Tong and Westerståhl (2023) arrange the above mentioned semantics in a hierarchy, with Kripke-semantics being the most specific, and algebraic semantics being the most general type of semantics for intuitionistic propositional logic.[88]

Carnap's question takes interestingly different shapes when applied to algebraic semantics as opposed to the other, set-based, semantics. Characteristic of the former is that the semantic values of sentences of the language are elements of the algebra, whereas semantic values of sentences in the latter type of semantics are (specific types of) subsets of the domains of the relevant models. This has consequences for the way the meanings of the logical connectives are conceived: in the algebraic case, interpretations of the logical constants are delivered by the respective algebra 'directly'. To informatively ask *Carnap's question* in this case one would therefore have to ask why (and whether it is only) *Heyting algebras* (that) constitute the intended interpretation of the intuitionistic connectives. In the case of non-algebraic semantics the interpretations of the connectives are not part of the model but provided, so to say, from the outside: conjunction is interpreted as set-union, negation as set-complementation, etc. Here, then, we need not consider alternative (types of) models but can ask whether the same class of models, when the connectives are interpreted by different set-theoretic operations, remains consistent w.r.t. intuitionistic propositional consequence $\vdash_{IPC}$.

Tong and Westerståhl (2023) show that *Carnap's Problem* is successfully resolved in the case of intuitionistic propositional logic by assuming compositionality and consistency with $\vdash_{IPC}$ for all the semantics considered. The result for set-based semantics is in fact a special case of the more general result for algebraic semantics. They first recall the fact that an algebraic interpretation of the language of intuitionistic logic is consistent with $\vdash_{IPC}$ iff it is a Heyting algebra (Tong and Westerståhl, 2023, Fact 3.4). They then show that the Heyting-algebra interpretation of the connectives over the domain of an algebra is the unique interpretation over that domain consistent with $\vdash_{IPC}$ (Tong and Westerståhl, 2023, Theorem 3.6).[89] From this it follows that compositionality and consistency with $\vdash_{IPC}$ suffice to uniquely determine the standard interpretations of the intuitionistic connectives over all set-based semantics considered by Tong and Westerståhl, which include Kripke-, Beth-, and topological semantics (Tong and Westerståhl, 2023, Corollary 3.8).

This result is remarkable for a variety of reasons. On the one hand, it demonstrates the surprising robustness of intuitionistic consequence w.r.t. determining the intended extensions for its logical operators across a variety of semantics. On the other hand, the successful determination of the intended interpretations relies essentially on consistency with *full consequence*, consistency with 'mere' theorems is insufficient to achieve the desired determination (Tong and Westerståhl, 2023, Example 3.10). Moreover, the categoricity of the intuitionistic connectives is, in general, *not modular*: while the consequence relation over the full intuitionistic language ensures unique determination for all connectives, this result does not continue to hold when considering arbitrary fragments of the language (Tong and Westerståhl, 2023, 175ff.).

5 Semantic Strategies for Solving Carnap's Problem II (the quantificational case)

Quantifiers introduce additional levels of semantic complexity, for they perform operations on sub-sentential components, thus unveiling a much more fine-grained structure than could be captured and

[88]See (Bezhanishvili and Holliday, 2019) and (Tong and Westerståhl, 2023, 165ff.) for details.

[89]We are very imprecise in our statement of the result and set-up here. For its proper formulation see (Tong and Westerståhl, 2023, 170ff.).

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

expressed in the propositional case. Here, (Comp) and (Non-Triv) prove insufficient to pin down the intended meanings of the usual quantifiers. However, at the level of quantification other constraints emerge as natural candidates for adoption in the attempt to resolve Carnapian underdetermination.

5.1 Invariance

What distinguishes a *quantifier* from a 'qualifier', i.e., a mere run-of-the-mill second-order predicate (a predicate of (first-order) predicates), is that the former should only be sensitive to quantitative, i.e., cardinality-based, properties, of its arguments. Within the set-theoretic framework of the background theory for the definition of quantifiers cardinality is captured in terms of *bijections*.[90] Sets M and N have the same cardinality if there exists a bijection between them. A *permutation* π is a bijection from a set M to itself. A permutation of a set M naturally induces a permutation of objects from the type hierarchy over M.[91] An object o from the type-hierarchy over a domain M is *permutation-invariant* if $\pi[o] = o$ for all permutations π of M. A permutation-invariant object is thus an object that is insensitive to (model-internal) qualitative features – it is only capable of detecting distinctions on the basis of differing cardinality.[92]

Since (local) quantifiers are second-order predicates over domains, demanding that they be permutation-invariant naturally captures their nature as *quantifiers*, thereby distinguishing them from other, non-quantificational predicates like, for example, 'is a colour'. The requirement that quantifiers be permutation- or, more generally, bijection-invariant,[93] is further supported by considerations pertaining to their logicality: invariance has long been regarded as an at least necessary feature of the logicality of a notion, capturing the idea that logical operations are insensitive to the identity of objects and thus uninfluenced by 'empirical' features.[94] Thus, for something to be a quantifier or a logical notion, it ought to be at least *permutation-invariant*:

(Perm) Quantifiers are permutation-invariant.

Bonnay and Westerståhl observe that the only permutation-invariant principal filter over a domain M is the maximal principal filter $\{M\}$ (Bonnay and Westerståhl, 2016, 730). Thus, permutation-invariance, in addition to (Comp) and (Non-Triv), suffices to fix the intended interpretation of the universal, and thereby also the existential, quantifier.[95] This is a very welcome result, given the role played by permutation-invariance in a theory of (logical) quantification. Since being a (logical) quantifier *means* being permutation-invariant, it appears to follow from the very nature of quantification that the standard universal and existential quantifiers of FOL are uniquely determined.[96]

[90] A *bijection* $\beta : M \rightarrow N$ is a function between M and N that is *one-to-one* and *onto*.

[91] Let $\pi : M \rightarrow M$ be a permutation of M. For $X \subseteq M$, $\pi[X] = \{\pi(a) | a \in X\}$; for $X \subseteq \mathcal{P}(M)$, $\pi[X] = \{\pi[A] | A \in X\}$; etc.

[92] Our focus on *permutation*- rather than the more general notion of *bijection-invariance* is solely for presentational purposes. Nothing essential is lost by this.

[93] I.e., closed under isomorphic structures in their global manifestation.

[94] Cf. (Tarski, 1986), (Sher, 1991).

[95] See (Valle-Inclán, 2024) for a recent criticism of Bonnay and Westerståhl's solution strategy.

[96] Note that there are at least two salient ways of distinguishing between permutation-invariant and non-permutation-invariant second-order predicates: those who take permutation-invariance to mark, first and foremost, a distinction between logical and non-logical expressions of a language will distinguish between *logical* (i.e., permutation-invariant) and *non-logical* (non-permutation-invariant) quantifiers. There are thus two types of quantifiers, logical and non-logical ones, and what distinguishes them is a condition that is motivated on the basis of considerations of what makes a denotation or meaning of an expression logical. On the other hand, others maintain that for something to be a *quantifier* simply means for it to be insensitive to all but quantitative features of its arguments. Being permutation-invariant is thus constitutive of an expression to be a quantifier, independently of its logical status. There are, accordingly, not two types of quantifiers, but only one, which is opposed to 'qualifiers' such as 'is a virtue', 'is a property of individual x', etc. Whether or not a specific view on logicality is thus part and parcel of a motivated solution to *Carnap's Problem* depends on the underlying conception of quantification. Elsewhere (Bonnay and Speitel, 2021) we argued that logicality considerations are pertinent to *Carnap's Problem*, and (Bonnay and Westerståhl, 2016) appear to defend Perm on the basis of such considerations. Nonetheless, maintaining that to be a quantifier simply entails permutation-invariance loosens this commitment to specific views on logicality. I thank an anonymous referee for pressing me on this point.

5.2 Generalized Quantification

The universal and existential quantifiers of FOL are at the lower end of a class of possible first-order quantifiers. Bonnay and Westerståhl's result guarantees that considering them as (logical) quantifiers, and thus demanding that they be permutation-invariant, suffices to determine their standard interpretations. This also yields the determinacy of notions definable in terms of them, in particular of finite cardinality quantifiers of the form $[\![\exists_{\leq n}]\!]^{\mathcal{M}} = \{X \subseteq M \mid n \leq |X|\}$. How far beyond $\forall$ and $\exists$ does this strategy generalize? I.e., for which other quantifiers is (PERM) the only constraint needed to uniquely 'fix' their interpretation?

The study of the unique determinability of generalized quantifiers beyond $\forall$ and $\exists$ is still in its beginnings[97] but already allows several interesting observations. We say that a type $\langle 1 \rangle$-quantifier $\mathcal{Q}$ is *generalized elementarily definable*, EC_Δ for short, if there exists a set Δ of sentences of FOL of the form $\varphi(P)$, whose only non-logical symbol is the predicate letter P of adicity 1, s.t. for all models $\mathcal{M} = \langle M, X \rangle$:

$$\mathcal{M} = \langle M, X \rangle \in \mathcal{Q} \text{ iff } \mathcal{M} \models \varphi(P) \text{ for all } \varphi(P) \in \Delta$$

In other words, where Δ is such a set of sentences of FOL and $Mod(\Delta) = \{\mathcal{M} \mid \mathcal{M} \models \Delta\}$, $\mathcal{Q}$ is EC_Δ if there exists Δ, s.t. $\mathcal{Q} = Mod(\Delta)$. When $\mathcal{Q} = Mod(\Delta)$ for some appropriate set of sentences Δ we denote it by $\mathcal{Q}_\Delta$. We then observe the following:

Observation 5.1. *Let $\mathcal{Q} = \mathcal{Q}_\Delta$ for some set of sentences Δ of FOL and assume that $\mathcal{Q}_\Delta$ interprets Q, i.e. $[\![Q]\!] = \mathcal{Q}_\Delta$. Then*

(a) $QxPx \models \varphi$ for all $\varphi \in \Delta$
(b) $\Delta \models QxPx$

Proof: For (a), suppose that $\mathcal{M} \models QxPx$. That means that $\mathcal{M} \in \mathcal{Q}_\Delta = Mod(\Delta)$. Hence, $\mathcal{M} \models \Delta$. For (b), suppose that $\mathcal{M} \models \Delta$. Hence $\mathcal{M} \in Mod(\Delta) = \mathcal{Q}_\Delta$ and therefore $\mathcal{M} \models QxPx$. ■

Let $\mathcal{M} \models^{\mathcal{Q}} \varphi$ mean that $\mathcal{M} \models \varphi$ when Q is interpreted by $\mathcal{Q}$ and designate with $\models_{\mathcal{Q}}$ the resulting model-theoretic consequence relation. From the observation above it then follows that

Proposition 5.2. *If $\mathcal{Q} = \mathcal{Q}_\Delta$ for a set Δ of sentences of FOL, then $\mathcal{Q}$ is uniquely determinable.*

Proof: Suppose that $\mathcal{Q}'$ is consistent with $\models_{\mathcal{Q}}$. Let $\mathcal{M} \in \mathcal{Q}'$. Thus $\mathcal{M} \models^{\mathcal{Q}'} QxPx$. By (a) above we know that $QxPx \models_{\mathcal{Q}_\Delta} \varphi$ for all $\varphi \in \Delta$. Since $\mathcal{Q}'$ is consistent with $\models_{\mathcal{Q}_\Delta}$, it follows that $\mathcal{M} \models^{\mathcal{Q}'} \varphi$ for all $\varphi \in \Delta$. But then $\mathcal{M} \in Mod(\Delta) = \mathcal{Q}_\Delta$. For the other direction, suppose that $\mathcal{M} \in \mathcal{Q}_\Delta$, but $\mathcal{M} \notin \mathcal{Q}'$. Thus, $\mathcal{M} \not\models^{\mathcal{Q}'} QxPx$. Since $\mathcal{M} \in \mathcal{Q}_\Delta = Mod(\Delta)$ we have that $\mathcal{M} \models^{\mathcal{Q}_\Delta} \Delta$. However, since all $\varphi \in \Delta$ are sentences of FOL we also have that $\mathcal{M} \models^{\mathcal{Q}'} \Delta$. Since $\mathcal{Q}'$ is consistent with $\models_{\mathcal{Q}_\Delta}$ it follows, by (b), that $\mathcal{M} \models^{\mathcal{Q}'} QxPx$ – contradiction. Hence, $\mathcal{M} \notin \mathcal{Q}$. Therefore, $\mathcal{Q}' = \mathcal{Q}_\Delta$. ■

Since $\mathcal{Q}_0 = \{\langle M, X \rangle \mid \aleph_0 \leq |X|\} = Mod(\Delta)$ where $\Delta = \{\exists_n x Px \mid n \in \mathbb{N}\}$ it follows that the quantifier *there are infinitely many* is uniquely determinable by its associated consequence relation $\models_{\mathcal{Q}_0}$.[98] Furthermore, it was observed by Dag Westerståhl (p.c.) that the property of unique determinability is preserved under the operation of *complementation*, where the *complement* $\mathcal{Q}^c$ of a (type $\langle 1 \rangle$) quantifier $\mathcal{Q}$ is such that $\langle M, X \rangle \in \mathcal{Q}^c$ iff $\langle M, X^c = M - X \rangle \in \mathcal{Q}$.[99] Since $\mathcal{Q}_{fin} = \{\langle M, X \rangle \mid |X| < \aleph_0\} = \mathcal{Q}_0^c$, the unique determinability of the quantifier *there are finitely many* follows as well. Hence, (PERM) suffices to ensure the unique determinability of quantifiers going beyond FOL.

How much further does the unique determinability of type $\langle 1 \rangle$ quantifiers extend? Several observations suggest that it stops at $\mathcal{Q}_1$: it already follows from results proven in (Keisler, 1970) that a quantifier

[97] See (Bonnay and Speitel, 2021) and (Speitel, 2020) for initial investigation and results.

[98] The unique determinability of the quantifier *there are infinitely many* was first observed by Dag Westerståhl (p.c.) and is stated and proven as following from the more general result above in (Speitel, 2020).

[99] See (Speitel, 2020) for a reproduction of the proof.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

$\mathcal{Q}$ is consistent with the *complete axiomatization* of $\mathscr{L}(Q_1)$ – FOL extended with the quantifier $\mathcal{Q}_1$, – as long as $\mathcal{Q} = \{\langle M, X\rangle \mid \aleph_\alpha \leq |X|\}$ for some regular cardinal $\aleph_\alpha$. Hence, it is immediately apparent that $\mathcal{Q}_1$ is severely underdetermined by its associated consequence relation.[100] This underdetermination continues into higher cardinalities: based on results in (Keisler, 1968) it is possible to show that no quantifier of the form $\mathcal{Q} = \{\langle M, X\rangle \mid \aleph_\alpha \leq |X|\}$ for a strong singular limit cardinal $\aleph_\alpha$ is uniquely determinable by any consequence relation over its language. Making strong set-theoretic assumptions, the underdetermination can be shown to affect further quantifiers: let $\mathcal{Q}_\alpha = \{\langle M, X\rangle \mid \aleph_\alpha \leq |X|\}$. Then, assuming $V = L$, it is possible to show that no quantifier of the form $\mathcal{Q}_{\alpha+1}$ is uniquely determinable by a consequence relation over its language.[101]

Further quantifiers of different types corroborate the failures of unique determination further (see (Speitel, 2020) for examples). What is particularly noteworthy is the coming apart of completeness and unique determinability as demonstrated by $\mathcal{Q}_0$ and $\mathcal{Q}_1$. Whereas the logic of FOL extended with the quantifier $\mathcal{Q}_0$ is incomplete, $\mathcal{Q}_0$ is uniquely determined by $\models_{\mathcal{Q}_0}$. On the other hand, FOL extended with the quantifier $\mathcal{Q}_1$ possesses a complete recursively enumerable axiomatization, yet $\mathcal{Q}_1$ is not uniquely determined by $\models_{\mathcal{Q}_1}$. This not only demonstrates the limits of permutation-invariance in reducing admissible interpretations but also undermines the sometimes implicitly assumed access to reference and denotation by the inferentialist on the basis of completeness and soundness results. It furthermore supports the Carnapian claim that for a 'full formalization of logic' both completeness and categoricity are required, as these results demonstrate that completeness of a logical system is neither necessary nor sufficient for the categoricity of its logical notions.

5.3 Higher-Order Quantification

For the first-order case we are left with an interesting situation: while (Perm) successfully determines the intended quantifier meanings of the quantifiers of FOL and beyond, it does not suffice to resolve underdetermination in general for all generalized quantifiers, no matter how well-behaved their respective logics are. How does the strategy that brought at least partial success in the context of first-order languages fare with respect to second- and higher-order languages?

Murzi and Topey (2021) claim that the first-order strategy generalizes to cover second- as well as higher-order universal and existential quantifiers.[102] The non-standard, unintended, interpretations of the second-order quantifiers are constituted by so-called *general* or *Henkin-interpretations*, in which the quantifiers range over, suitably closed, subsets of the set of all relations over the first-order domain. These interpretations are inferentially indistinguishable from the intended *full* interpretation of the quantifiers according to which they range over *all* relations over the first-order domain.[103] By what mechanism, then, might the full interpretation of the quantifiers be secured?

Murzi and Topey (2021) claim that this can be achieved by means of the same mechanism that ultimately secured the standard interpretations of the first-order quantifiers. In the first-order case, what ensured that standard interpretations of the universal and existential quantifier were determined was their permutation-invariance under *permutations of their range* (i.e., of the first-order domain). In the second-order case, they say, one should, analogously, demand permutation-invariance under the

[100] For a strengthening of this result see (Bonnay and Speitel, 2021) where the unique determinability of a notion by inference was argued to be a necessary condition for its logicality.

[101] See (Speitel, 2020, 328) for this result and its proof. It follows straightforwardly from a result of Jensen (1972), cf. (Schmerl, 1985, Corollary 2.1.7).

[102] Murzi and Topey (2021) explicitly do not advance a *semantic solution strategy* to *Carnap's Problem* in terms of (Perm), but pursue an inferentialist strategy by showing how the *open-endedness* of the rules for the quantifiers (first- and second-order), in combination with a particular rule-format (see Section 3.4 above), suffices to fix their intended interpretations. This they do in a two-step argument: they first demonstrate how permutation-invariance ensures standard interpretations. Then, they show how the (local) validity of open-ended rules implies the permutation-invariance of the quantifiers. The 'detour' via permutation-invariance ensures "that our result is available even to those who don't share our inferentialist assumptions, so long as they accept the topic neutrality of logic" (Murzi and Topey, 2021, 3410). In the context of this paper, we will consider this step separately from Murzi and Topey's overall (inferentialist) goals and project.

[103] See, e.g., Shapiro (1991) or Väänänen (2021). See, e.g., (Putnam, 1980) for pointing out the philosophical significance of the inferential indistinguishability of *Henkin-* and *full* semantics for SOL.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

appropriate range of the second-order quantifiers. This range does not consist of a set M of a model $\mathcal{M}$, however, but rather of the powerset of M, i.e., $\mathcal{P}(M)$:

> Notice that, while the permutation invariance of the interpretation of the first-order quantifier $\forall$ amounts to the invariance of its range under all permutations of the domain M, what the permutation invariance of the interpretation of $\forall_2$ requires is somewhat different. Since we are now quantifying over relations rather than objects, the range of $\forall_2$, when it binds a variable of arity n, must remain invariant under all permutations, not of M itself, but of $\mathcal{P}(M^n)$ – i.e. the set of n-ary relations on M. (Murzi and Topey, 2021, fn. 37)

This, however, essentially reduces the second-order case to the situation of the first-order case. The relations over the domain are treated as objects in their own right that can be mapped directly, so to speak, to other relations. Just as in the first-order case, if any relation is left out of the range of the second-order quantifiers that interpretation will not be permutation-invariant, vis-a-vis the result of (Bonnay and Westerståhl, 2016). Hence, permutation-invariance of the second-order quantifiers ensures their intended, full, interpretation: "insofar as permutation invariance is a necessary condition for logicality, and insofar as the second-order quantifiers are genuinely logical, the rules for the second-order quantifiers are simply incompatible with any restricted interpretation" (Murzi and Topey, 2021, 3411). Moreover, this strategy can be replicated at any finite order, thereby guaranteeing standard interpretations of higher-order quantifiers as well.

Note, however, that the perspective here has, ever so slightly, shifted. For the adoption and application of the permutation-invariance demand has been modified to apply *directly* to objects in $\mathcal{P}(M)$, rather than the permutations being induced via permutations of M. Not only is this a marked departure from the Tarskian picture of logicality (Tarski, 1986) but given the inherent instability of the power-set operation – resulting from its *non-absolute* nature – and related questions concerning the notion of 'all subsets' one might wonder whether some further indeterminacy might be looming in the background here.

6 Outlook: Philosophical Consequences of Carnap's Problem

Carnap (1943) demonstrated that several logical facts about the standard logical constants of propositional and first-order logic remain undecided and underdetermined by the usual formulations of these logics. The question Carnap raises is, however, much broader: *Carnap's question*, the question whether inferential characterizations of a logic uniquely determine that logic's intended semantics, can be asked for any logical system and *Carnap's Problem* arises for most of them. Importantly, it even arises for systems for which the usual adequacy theorems, meant to ensure a match between proof-theoretic and model-theoretic characterizations of consequence, hold. Despite this, some logically relevant aspects might, nonetheless, remain indeterminate. *Carnap's question* therefore reveals an interesting perspective from which to consider what semantic information is contained in inferences, and which sort of facts are left out.

This has philosophical ramifications: the fact that inferential patterns succeed in determining intended values for several constants at once, but not for any of the involved constants in isolation, as was the case for several of the intuitionistic operators, for example, throws into serious doubt the widely held assumption that it is characteristic of logical operators that their meaning is *atomistic*, i.e., can be fully and sufficiently characterized independently of any other non-schematic expressions of the language. Moreover, the fact that some semantic universals succeed in significantly reducing underdetermination suggests that these semantic constraints are part and parcel of the (philosophical) theory underlying the logic. This impression is further strengthened by the observation that constraints that can be motivated on the basis of the philosophical interpretation of a logical theory lead to unique determinability of that logic's operators (see, e.g., Tabakçı (2024)).

Carnap's Problem is not merely a mathematical curiosity in the foundations of logic. It has significant repercussions for theories of meaning that rely on the methods of formal logic. It further impacts philosophical debates at the intersection of logic, mathematics and philosophy. In (Bonnay and Speitel, 2021) a criterion of logicality was motivated which used the insights provided by *Carnap's Problem* to delineate a core of logical operations grounding a particularly stable and reliable set of inferential patterns. (Speitel, 2024) argued, on the basis of uniquely determinable notions, for the possibility of determinate access to the natural number structure. All these direct and indirect repercussions will, we hope, further stimulate interest in *Carnap's Problem*, rehabilitating Carnap's own ambitions to grant the same importance to the unique determinability of logical notions as was given to the completeness of logical systems.

7 Appendix

This short appendix clarifies the connection between Bonnay and Westerståhl's (2016) main result and the way it is stated in the context of this paper.[104]

The following definitions are adaptations of the definitions from (Bonnay and Westerståhl, 2016):

Definition 7.1. *Let $\mathcal{M}$ be a first-order model with domain M and $q_\forall \subseteq \mathcal{P}(M)$. A* weak model *(for the universal quantifier) is a tuple $\langle \mathcal{M}, q_\forall \rangle$.*

Definition 7.2. *Let $\langle \mathcal{M}, q_\forall \rangle$ be a weak model. Then:*

$$\langle \mathcal{M}, q_\forall \rangle \models \forall x \varphi(x) \text{ iff } \{a \in M \mid \langle \mathcal{M}, q_\forall \rangle \models \varphi(a)\} \in q_\forall.$$

Definition 7.3. *A* model $\mathcal{M}$ is consistent with a consequence relation $\vdash$ *iff, whenever $\Gamma \vdash \varphi$ and $\mathcal{M} \models \gamma$ for all $\gamma \in \Gamma$, then $\mathcal{M} \models \varphi$.*

In the following, let $\mathscr{L}^*$ be a *purely relational language* that contains *predicate variables.*[105] Bonnay and Westerståhl (2016) then establish the following result:

Theorem 7.4. *(Bonnay and Westerståhl, 2016) A weak model $\langle \mathcal{M}, q_\forall \rangle$ is consistent with $\vdash_{FOL}$ (over $\mathscr{L}^*$) iff $q_\forall$ is a principal filter over M.*

Arguably, the treatment of quantifier-interpretations via weak models introduces an asymmetry between the treatment of the propositional connectives and quantifiers as 'fixed' expressions in the context of logical languages. For just as the interpretation of the propositional connectives was conceived of *globally*, as consisting of classes of valuations, so quantifiers should be thought of as *Lindström-quantifiers.* For this reason, we adapt the setting as follows:

Definition 7.5. *A* global (type $\langle 1 \rangle$) quantifier *is a class $\mathcal{Q} = \{\langle M, X \rangle \mid M$ is a set and $X \subseteq M\}$.*

Definition 7.6. *Let $\mathcal{Q}$ be a global quantifier. The* local quantifier-on-a-model $\mathcal{Q}^{\mathcal{M}}$*, corresponding to $\mathcal{Q}$, is the set $\mathcal{Q}^{\mathcal{M}} = \{X \mid \langle M, X \rangle \in \mathcal{Q}\}$, where M is the domain of $\mathcal{M}$.*

Note that the interpretation of $\mathcal{Q}^{\mathcal{M}}$ depends solely on $\mathcal{Q}$ and the *domain* of $\mathcal{M}$, and is independent of any further elements of the signature of $\mathcal{M}$. That is:

Observation 7.7. *Let $\mathcal{Q}$ be a global quantifier and $\mathcal{M}_1, \mathcal{M}_2$ be models, s.t. $M_1 = M_2$. Then $\mathcal{Q}^{\mathcal{M}_1} = \mathcal{Q}^{\mathcal{M}_2}$.*

Definition 7.8. *Let $\mathcal{Q}$ be a global quantifier interpreting $\forall$ and $\mathcal{M}$ be a model. $\mathcal{M} \models^{\mathcal{Q}} \forall x \varphi(x)$ iff $\{a \in M \mid \mathcal{M} \models^{\mathcal{Q}} \varphi(a)\} \in \mathcal{Q}^{\mathcal{M}}$.*

Definition 7.9. *$\Gamma \models_{\mathcal{Q}} \varphi$ iff, for all $\mathcal{M}$, whenever $\mathcal{M} \models^{\mathcal{Q}} \gamma$ for all $\gamma \in \Gamma$, then also $\mathcal{M} \models^{\mathcal{Q}} \varphi$.*

[104] See footnote 21. Thanks to an anonymous referee for urging me to clarify the relationship between Bonnay and Westerståhl's set-up and the framework of this paper.

[105] Bonnay and Westerståhl (2016) consider languages that also include singular terms. The following theorem is a special case of their theorem, restricted to languages that do not contain any terms.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

Definition 7.10. *A* (global) quantifier $\mathcal{Q}$ is consistent with a consequence relation $\vdash$ *iff* $\vdash \subseteq \models_{\mathcal{Q}}$.

Lemma 7.11. *Let $\mathcal{Q}$ be a global quantifier interpreting $\forall$ and $\mathcal{M}$ be a model. $\mathcal{M} \models^{\mathcal{Q}} \varphi$ iff $\langle \mathcal{M}, \mathcal{Q}^{\mathcal{M}} \rangle \models \varphi$.*

Proof: The proof proceeds by induction on the complexity of φ. The propositional cases are standard. For $\varphi := \forall x \psi(x)$ we have: $\mathcal{M} \models^{\mathcal{Q}} \forall x \psi(x)$ iff $\{a \in M \mid \mathcal{M} \models^{\mathcal{Q}} \psi(a)\} \in \mathcal{Q}^{\mathcal{M}}$ iff (by the induction hypothesis) $\{a \in M \mid \langle \mathcal{M}, \mathcal{Q}^{\mathcal{M}} \rangle \models \psi(a)\} \in \mathcal{Q}^{\mathcal{M}}$ iff $\langle \mathcal{M}, \mathcal{Q}^{\mathcal{M}} \rangle \models \varphi$. ■

Now let $\mathscr{L}$ be a purely relational first-order language (without predicate variables). Then

Theorem 7.12. *A (global) quantifier $\mathcal{Q}$ interpreting $\forall$ is consistent with $\vdash_{FOL}$ (over $\mathscr{L}$) iff, for all $\mathcal{M}$, $Q^{\mathcal{M}}$ is a principal filter over M.*

Proof: The right-to-left direction follows directly from Bonnay and Westerståhl's original proof: let $\mathcal{M}$ be a model and $Q^{\mathcal{M}}$ a principal filter over M. Let $\Gamma \vdash_{FOL} \varphi$ and assume that $\mathcal{M} \models^{\mathcal{Q}} \gamma$ for all $\gamma \in \Gamma$. By Lemma 7.11 it follows that $\langle \mathcal{M}, \mathcal{Q}^{\mathcal{M}} \rangle \models \gamma$ for all $\gamma \in \Gamma$. Then, by Theorem 7.4, $\langle \mathcal{M}, \mathcal{Q}^{\mathcal{M}} \rangle \models \varphi$ and thus, by Lemma 7.11 again, $\mathcal{M} \models^{\mathcal{Q}} \varphi$ as well. Hence, $\vdash_{FOL} \subseteq \models_{\mathcal{Q}}$.

For the left-to-right direction assume that $\mathcal{Q}$, interpreting $\forall$, is consistent with $\vdash_{FOL}$, yet that there exists $\mathcal{M}$, s.t. $Q^{\mathcal{M}}$ is *not* a principal filter over M. From Bonny and Westerståhl's result we know that this must be due to some set(s) being undefinable over $\mathscr{L}$ (as this possibility is ruled out when all sets are rendered definable through the addition of predicate variables). However, we can easily move to an expansion $\mathcal{M}^*$ of $\mathcal{M}$ where precisely these sets are named by predicate constants of the expanded signature. Since $M = M^*$ it follows from Observation 7.7 that $\mathcal{Q}^{\mathcal{M}} = \mathcal{Q}^{\mathcal{M}^*}$. Yet, as soon as sets that 'interrupt' $\mathcal{Q}^{\mathcal{M}}$ from being a principal filter become definable we can find $\Gamma \cup \{\varphi\}$, s.t. $\Gamma \vdash_{FOL} \varphi$ but $\Gamma \not\models^{\mathcal{Q}} \varphi$ and thus $\vdash_{FOL} \not\subseteq \models_{\mathcal{Q}}$, i.e., $\mathcal{Q}$, interpreting $\forall$, is not consistent with $\vdash_{FOL}$ as assumed.

As a concrete example, assume that $\mathcal{Q}^{\mathcal{M}}$ was not closed under super-sets; i.e. assume there were sets $X, Y \subseteq M$, s.t. $X \subseteq Y$, $X \in \mathcal{Q}^{\mathcal{M}}$, but $Y \notin \mathcal{Q}^{\mathcal{M}}$. Since $\mathcal{Q}^{\mathcal{M}} = \mathcal{Q}^{\mathcal{M}^*}$ we also have that $X \in \mathcal{Q}^{\mathcal{M}^*}$, but $Y \notin \mathcal{Q}^{\mathcal{M}^*}$. Now let $\mathcal{M}^*$ be an expansion of $\mathcal{M}$ containing two additional predicate constants P, R, s.t. $[\![P]\!]^{\mathcal{M}^*} = X$ and $[\![R]\!]^{\mathcal{M}^*} = Y$. Note that $\forall x \varphi(x) \vdash_{FOL} \forall x(\varphi(x) \vee \psi(x))$. But now we have that $\mathcal{M}^* \models^{\mathcal{Q}} \forall x Px$, yet $\mathcal{M}^* \not\models^{\mathcal{Q}} \forall x(Px \vee Rx)$. Hence, $\forall x \varphi(x) \not\models_{\mathcal{Q}} \forall x(\varphi(x) \vee \psi(x))$ and thus $\mathcal{Q}$, interpreting $\forall$, is not consistent with $\vdash_{FOL}$. ■

Thus, instead of internalizing definability facts as Bonnay & Westerståhl do by means of including predicate variables in the language, the same effect is achieved in the current setting by conceiving of quantifier meaning as global and forcing $\mathcal{Q}^{\mathcal{M}}$ to be identical over all models with the same domain.

Acknowledgements The author expresses special thanks to Denis Bonnay and Dag Westerståhl for discussion of and guidance in obtaining several of the results mentioned in this paper, Gila Sher for feedback on the material and its relevance to questions of logicality, Elke Brendel and Constantin Brîncus for discussion, and S. Kaan Tabakçı for fruitful discussions and ongoing work on aspects related to Carnap's Problem in multi-valued and non-classical settings. The author further thanks audiences of the *Bologna-Bonn-Padua Research Seminar* and the *Logic Colloquium 2024* for helpful comments and discussion of the material of the paper.

Funding Work on this paper was supported by an *Argelander Starter-Kit Grant* of the University of Bonn, as well as an *Institute Vienna Circle Research Fellowship* during a stay at the Institute Vienna Circle in 2025.

References

Antonelli, G.A. "On the General Interpretation of First-Order Quantifiers." *Review of Symbolic Logic* 6 (2013): 637–658.

Antonelli, G.A. "Life on the Range." *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics and Language*. Ed. A. Torza. Springer, 2015. 171–190.

Avron, A. "Simple Consequence Relations." *Information and Computation* 92 (1991): 105–139.

Awodey, S. and A. W. Carus. "Carnap, Completeness, and Categoricity: The Gabelbarkeitssatz of 1928." *Erkenntnis* 54 (2001): 145–172.

Barrio, E.A., F. Pailos, and D. Szmuc. "A Hierarchy of Classical and Paraconsistent Logics." *Journal of Philosophical Logic* 49 (2020): 93–120.

Belnap, N. "Tonk, Plonk and Plink." *Analysis* 22 (1962): 130–134.

Belnap, N. and G. Massey. "Semantic Holism." *Studia Logica* 49 (1990): 67–82.

Bezhanishvili, G. and W.H. Holliday. "A Semantic Hierarchy for Intuitionistic Logic." *Indagationes Mathematicae* 30 (2019): 403–469.

Blackburn, P., M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, 2001.

Boghossian, P.A. "Analyticity Reconsidered." *Noûs* 30 (1996): 360–391.

Bonnay, D. and S.G.W. Speitel. "The Ways of Logicality: Invariance and Categoricity." *The Semantic Conception of Logic: Essays on Consequence, Invariance, and Meaning*. Ed. G. Sagi and J. Woods. Cambridge University Press, 2021. 55–79.

Bonnay, D. and D. Westerståhl. "Compositionality Solves Carnap's Problem." *Erkenntnis* 81 (2016): 721–739.

Bonnay, D. and D. Westerståhl. "Carnap's Problem for Modal Logic." *Review of Symbolic Logic* 16 (2023): 578–602.

Brendel, E. "How Classical, Paracomplete and Paraconsistent Logicians (Dis-)Agree." *Global Philosophy* 34 (2024).

Brîncuş, C. "Are the Open-Ended Rules for Negation Categorical?." *Synthese* 198 (2019): 7249–7256.

Brîncus, C. "Categorical Quantification." *Bulletin of Symbolic Logic* 30 (2024): 227–252.

Brîncus, C. "Inferential Quantification and the ω-Rule." *Perspectives on Deduction: Contemporary Studies in the Philosophy, History and Formal Theories of Deduction*. Ed. A. Piccolomini D'Aragona. Springer Verlag, 2024. 345–372.

Button, T. "Knot and Tonk: Nasty Connectives on Many-Valued Truth-Tables for Classical Sentential Logic." *Analysis* 76 (2016): 7–19.

Button, T. and S. Walsh. *Philosophy and Model Theory*. Oxford University Press, 2018.

Carnap, R. *Formalization of Logic*. Harvard University Press, 1943.

Church, A. "Review of Carnap 1943." *Philosophical Review* 53 (1944): 493–498.

Church, A. "Non-Normal Truth-Tables for the Propositional Calculus." *Journal of Symbolic Logic* 19 (1954): 233–234.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

Cobreros, P., P. Egré, E. Ripley, and R. van Rooij. "Tolerant, Classical, Strict." *Journal of Philosophical Logic* 41 (2012): 347–385.

Davey, B.A. and H.A. Priestley. *Introduction to Lattices and Order.* Cambridge University Press, 2002.

Davidson, D. "Theories of Meaning and Learnable Languages." *Inquiries into Truth and Interpretation.* . Oxford University Press, 2001. 3–17.

Došen, K. and P. Schroeder-Heister. "Uniqueness, Definability and Interpolation." *Journal of Symbolic Logic* 53 (1988): 554–570.

Dummett, M. *The Logical Basis of Metaphysics.* Harvard University Press, 1991.

Feferman, S. "Which Quantifiers Are Logical? A Combined Semantical and Inferential Criterion." *Quantifiers, Quantifiers, and Quantifiers: Themes in Logic, Metaphysics and Language.* Ed. A. Torza. Springer, 2015. 19–31.

French, R. and D. Ripley. "Valuations: Bi, Tri, and Tetra." *Studia Logica* 107 (2019): 1313–1346.

Gabbay, D. "What is a Classical Connective?." *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 24 (1978): 37–44.

Garson, J.W. "Categorical Semantics." *Truth or Consequences.* Ed. J. Dunn and A. Gupta. Kluwer Academic Publishers, 1990. 155–175.

Garson, J.W. "Natural Semantics: Why Natural Deduction is Intuitionistic." *Theoria* 67 (2001): 114–139.

Garson, J.W. "Expressive Power and Incompleteness of Propositional Logics." *Journal of Philosophical Logic* 39 (2010): 159–171.

Garson, J.W. *What Logics Mean: From Proof Theory to Model-Theoretic Semantics.* Cambridge University Press, 2013.

Gentzen, G. "Untersuchungen über das logische Schließen." *Mathematische Zeitschrift 39* (1934): 405–431.

Hacking, I. "What is Logic?." *Journal of Philosophy* 76 (1979): 285–319.

Hardegree, G.M. "Completeness and Super-Valuations." *Journal of Philosophical Logic* 34 (2005): 81–95.

Harris, J.H. "What's so Logical about the "Logical" Axioms?." *Studia Logica* 41 (1982): 159–171.

Hart, W.D. "Prior and Belnap." *Theoria* 48 (1982): 127–138.

Haze, T. "A Note on Carnap's Result and the Connectives." *Axiomathes* 29 (2019): 285–288.

Hjortland, O.T. "Speech Acts, Categoricity, and the Meanings of Logical Connectives." *Notre Dame Journal of Formal Logic* 55 (2014): 445–467.

Humberstone, L. "Valuational Semantics of Rule Derivability." *Journal of Philosophical Logic* 25 (1996): 451–461.

Humberstone, L. "The Revival of Rejective Negation." *Journal of Philosophical Logic* 29 (2000): 331–381.

Humberstone, L. *The Connectives.* MIT Press, 2011.

Incurvati, L. and J. Schlöder. "Weak Rejection." *Australasian Journal of Philosophy* 95 (2017): 741–760.

Incurvati, L. and J. Schlöder. "Weak Assertion." *Philosophical Quarterly* 69 (2019): 741–770.

Incurvati, L. and J. Schlöder. *Reasoning with Attitude. Foundations and Applications of Inferential Expressivism.* Oxford University Press, 2024.

Incurvati, L., J. Schlöder, and M. Aloni. Weak Assertion Meets Information States: a Logic for Epistemic Modality and Quantification. presentation, 2019.

Incurvati, L. and P. Smith. "Rejection and Valuations." *Analysis* 70 (2010): 3–10.

Jensen, R.B. "The Fine Structure of the Constructible Hierarchy." *Annals of Mathematical Logic* (1972): 229–308.

Johannesson, E. "Completeness Also Solves Carnap's Problem." *Thought: A Journal of Philosophy* 11 (2022): 192–198.

Keisler, H.J. "Models with Orderings." *Logic, Methodology, and Philosophy of Science III.* Ed. B. van Rootselaar and J.F. Staal. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Company, 1968. 35–62.

Keisler, H.J. "Logic with the Quantifier "There Exist Uncountably Many"." *Annals of Mathematical Logic* 1 (1970): 1–93.

Kneale, W. and M. Kneale. *The Development of Logic.* Clarendon Press, 1962.

Koslow, A. *A Structuralist Theory of Logic.* Cambridge University Press, 1992.

Koslow, A. "Carnap's Problem: What is It Like to Be a Normal Interpretation of Classical Logic?." *Abstracta* 6 (2010): 117–135.

Leblanc, H., J. Paulos, and G.E. Weaver. "Rules of Deduction and Truth-Tables." *Reports on Mathematical Logic* 8 (1977): 71–79.

Leitgeb, H. and A. Carus. "Rudolf Carnap." *The Stanford Encyclopedia of Philosophy.* Ed. E.N. Zalta. Fall 2020 edition. https://plato.stanford.edu/archives/fall2020/entries/carnap/, 2020.

Lindström, P. "On Extensions of Elementary Logic." *Theoria* 35 (1969): 1–11.

Malinowski, G. *Many-Valued Logics.* Oxford University Press, 1993.

McCawley, J.D. "Truth Functionality and Natural Deduction." *Proceedings of the 1975 International Symposium on Multiple-Valued Logic.* Ed. G. Epstein. Indiana University, 1975. 412–418.

McCawley, J.D. *Everything Linguists Have Always Wanted to Know About Logic - But Were Ashamed to Ask.* University of Chicago Press, 1981.

McGee, V. "Everything." *Between Logic and Intuition: Essays in Honor of Charles Parsons.* Ed. G. Sher and R. Tieszen. Cambridge University Press, 2000. 54–79.

McGee, V. "There's a Rule for Everything." *Absolute Generality.* Ed. A. Rayo and G. Uzquiano. Oxford University Press, 2006. 179–202.

McGee, V. "The Categoricity of Logic." *Foundations of Logical Consequence.* Ed. C.R. Caret and O.T. Hjortland. Oxford University Press, 2015. 161–186.

Montague, R. "English as a formal language." *Formal Philosophy: Selected Papers of Richard Montague.* Ed. R.H. Thomason. Yale University Press, 1974. 188–221.

Mostowski, A. "On a Generalization of Quantifiers." *Fundamenta Mathematicae* 44 (1957): 12–36.

Murzi, J. *Intuitionism and Logical Revision.* PhD thesis, University of Sheffield, 2010.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

Murzi, J. “Classical Harmony and Separability.” *Erkenntnis* 85 (2020): 391–415.

Murzi, J. and O.T. Hjortland. “Inferentialism and the Categoricity Problem: Reply to Raatikainen.” *Analysis* 69 (2009): 480–488.

Murzi, J. and F. Steinberger. “Inferentialism.” *Blackwell Companion to Philosophy of Language*. Ed. B. Hale, C. Wright, and A. Miller. Wiley Blackwell, 2017. 197–224.

Murzi, J. and B. Topey. “Categoricity by Convention.” *Philosophical Studies* 178 (2021): 3391–3420.

Open-Logic-Project. “Many-valued Logic.” *http://builds.openlogicproject.org/content/many-valued-logic/many-valued-logic.pdf* (2024).

Pagin, P. and D. Westerståhl. “Compositionality I: Definitions and Variants.” *Philosophy Compass* 5 (2010): 250–264.

Pagin, P. and D. Westerståhl. “Compositionality II: Arguments and Problems.” *Philosophy Compass* 5 (2010): 265–282.

Peregrin, J. *Inferentialism: Why Rules Matter.* Palgrave, 2014.

Peters, S. and D. Westerståhl. *Quantifiers in Language and Logic.* Clarendon Press, 2006.

Picollo, L. “Mysterious Quantifiers.” *Journal of Philosophy* forthcoming (2025): 1–30.

Priest, G. *An Introduction to Non-Classical Logic: From If to Is.* New York: Cambridge University Press, 2008.

Putnam, H. “Models and Reality.” *Journal of Symbolic Logic* 45 (1980): 464–482.

Quine, W.V.O. “On What There Is.” *From a Logical Point of View*. Ed. W.V.O. Quine. Harvard University Press, 1953. 1–19.

Raatikainen, P. “On Rules of Inference and the Meanings of Logical Constants.” *Analysis* 68 (2008): 282–287.

Restall, G. “Multiple Conclusions.” *Logic, Methodology and Philosophy of Science*. Ed. P. Hájek, L. Valdés-Villanueva, and D. Westerståhl. College Publications, 2005.

Ripley, E. “Paradoxes and Failures of Cut.” *Australasian Journal of Philosophy* 91 (2013): 139–164.

Rumfitt, I. “The Categoricity Problem and Truth-Value Gaps.” *Analysis* 57 (1997): 223–236.

Rumfitt, I. “Yes and No.” *Mind* 109 (2000): 781–823.

Rumfitt, I. “Co-ordination Principles: A Reply.” *Mind* 117 (2008): 1059–1063.

Schmerl, J.H. “Transfer Theorems and their Applications to Logic.” *Model-Theoretic Logics*. Ed. J. Barwise and S. Feferman. Springer, 1985. 177–209.

Schroeder-Heister, P. “A Natural Extension of Natural Deduction.” *Journal of Symbolic Logic* 49 (1984): 1284–1300.

Shapiro, S. *Foundations Without Foundationalism: A Case for Second-Order Logic.* Oxford University Press, 1991.

Sher, G. *The Bounds of Logic: A Generalized Viewpoint.* MIT Press, 1991.

Shoesmith, D.J. and T. Smiley. *Multiple-Conclusion Logic.* Cambridge University Press, 1978.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press

Smiley, T. “Rejection.” *Analysis* 56 (1996): 1–9.

Speitel, S.G.W. *Logical Constants between Inference and Reference – An Essay in the Philosophy of Logic.* UC San Diego Dissertation Series, 2020.

Speitel, S.G.W. “Securing Arithmetical Determinacy.” *ERGO. An Open Access Journal of Philosophy* 11 (2024): 1083–1118.

Speitel, S.G.W. “(Global) Quantification & (Local) Definability - Some Comments on Del Valle-Inclán’s Criticism of Bonnay and Westerståhl’s Solution to *Carnap’s Problem.*” *Journal of Philosophical Logic* 54 (2025).

Speitel, S.G.W. and S.K. Tabakçı. Semantic Universals for Moderate Inferentialists – Solutions to the Categoricity Problem. unpublished manuscript, 2025.

Steinberger, F. “Why Conclusions Should Remain Single.” *Journal of Philosophical Logic* 40 (2011): 333–355.

Tabakçı, S.K. “Categoricity Problem for Lp and K3.” *Studia Logica* (2024): 1–35.

Tarski, A. “What Are Logical Notions?.” *History and Philosophy of Logic* 7 (1986): 143–154.

Tennant, N. *The Taming of the True.* Oxford University Press, 1997.

Tong, H. and D. Westerståhl. “Carnap’s Problem for Intuitionistic Propositional Logic.” *Logics* 1 (2023): 163–181.

Väänänen, J. “Second-order and Higher-order Logic.” *The Stanford Encyclopedia of Philosophy.* Ed. E.N. Zalta. Fall 2021 edition. https://plato.stanford.edu/archives/fall2021/entries/logic-higher-order/, 2021.

Valle-Inclán, P. del. “Carnap’s Problem, Definability and Compositionality.” *Journal of Philosophical Logic* 53 (2024): 1321–1346.

Fraassen, B.C.van . “Singular Terms, Truth-Value Gaps, and Free Logic.” *Journal of Philosophy* 63 (1966): 481–495.

Wansing, H. and S. Ayhan. “Logical Multilateralism.” *Journal of Philosophical Logic* 52 (2023): 1603–1636.

Warren, J. *Shadows of Syntax: Revitalizing Logical and Mathematical Conventionalism.* Oxford University Press, 2020.

Westerståhl, D. “Logical Constants and Compositionality.” *Oxford Handbook of Philosophy of Logic.* Ed. E. Brendel, M. Carrara, F. Ferrari, O. Hjortland, G. Sagi, G. Sher, and F. Steinberger. Oxford University Press, 2025.

Woods, J. “Logical Indefinites.” *Logique et Analyse* (2014): 277–307.

https://doi.org/10.1017/bsl.2025.10083 Published online by Cambridge University Press