Compositionality Solves Carnap's Problem

Denis Bonnay, Dag Westerståhl

Compositionality solves Carnap's Problem*

[postprint]

Denis Bonnay
Université Paris Ouest Nanterre

Dag Westerståhl
Stockholm University

Abstract

The standard relation of logical consequence allows for non-standard interpretations of logical constants, as was shown early on by Carnap. But then how can we learn the interpretations of logical constants, if not from the rules which govern their use? Answers in the literature have mostly consisted in devising clever rule formats going beyond the familiar what follows from what. A more conservative answer is possible. We may be able to learn the correct interpretations from the standard rules, because the space of possible interpretations is *a priori* restricted by universal semantic principles. We show that this is indeed the case. The principles are familiar from modern formal semantics: compositionality, supplemented, for quantifiers, with topic-neutrality.

Keywords: logical consequence, logical constants, Carnap's Problem, semantic universals, compositionality, topic-neutrality

*We are grateful to an anonymous referee for comments that helped us improve an earlier version of this paper, and we thank Aldo Antonelli, Johan van Benthem, Serge Bozon, Paul Egré, Fredrik Engström, Wes Holliday, Ed Keenan, Sebastiano Moruzzi, Peter Pagin, Stanley Peters, and Shane Steinert-Threlkeld for remarks and suggestions at the various occasions when one of us presented our results about Carnap's Problem. Both authors wish to acknowledge support from the ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL grants. The second author thanks the Center for the Study of Language and Information (CSLI) at Stanford University for hospitality and intellectual stimulation when he was a visiting Research Fellow there for the spring quarter 2015, during which the final version of the paper was written.

1 Introduction: Carnap's Problem

The aim of Carnap's book *Formalization of Logic* (1943) was to point out the existence of what he called non-normal interpretations of classical logic: non-standard interpretations of logical constants which nevertheless validate all the classical laws. He also indicated ways to strengthen the usual proof systems so that such interpretations would not arise. In his review, Church recognized the problem, but was sceptical of Carnap's remedy, arguing that the proposed proof-theoretic revisions were "a concealed use of semantics" [6, p. 496]. Instead, he suggested, the use of semantic notions should be made explicit and vindicated, since no purely syntactic solution would work.

Later logicians and philosophers discussing the issue have not followed Church's advice, however, but mostly engaged in the strengthening of proof rules. In this paper we do exactly what Church recommended. More precisely, we show that assuming a few largely uncontroversial principles of semantic interpretation — the most important of which is the principle of compositionality — Carnap's Problem can be solved. Moreover, in contrast with earlier attempts, which in effect deal only with propositional logic, we show that the result holds for first-order logic as well. Indeed, eliminating non-normal interpretations of the quantifiers is the hard case, in the light of which any tentative solution to Carnap's problem should be evaluated.

But do we really need to worry about Carnap's Problem? Carnap himself took the issue quite seriously, and so do those who have pursued it later. At least we can safely say the following. Logical constants are instrumental to inference, and their usage is captured by means of syntactic rules in a formal system. When the language is given a definition of truth, logical constants receive interpretations representing their contributions to the semantic values of sentences in which they occur. So it certainly is a legitimate question to ask if the proof rules for logical constants determine these interpretations.

One may go further and consider the presence of non-normal interpretations unsatisfactory in principle, at least for a basic system like first-order logic. Carnap seems to have taken the absence of such interpretations to be desirable in its own right, expressing the ability of a syntactic system to adequately reflect semantics, on a par with more familiar properties such as soundness and completeness. Also, from the perspective of a theory of meaning, failure of proof

rules to determine semantic interpretation appears to spell trouble for an account of meaning as use applied to the logical constants; witness the renewed interested in Carnap's Problem in connection with debates about inferentialism (e.g. [16, 8]). At the very least, the presence of non-normal interpretations would make it hard to learn the (classical) meaning of logical constants for someone with access only to their rules of proof.

2 The space of solutions

We shall first further explore the space of possible solutions, in order to compare our explicitly semantic approach to existing solutions. Carnap's Problem is the underdetermination of semantics (the interpretation of logical constants) by syntax (a syntactically defined notion of consequence). Therefore, since the problem concerns the match between syntax and semantics, three different kinds of strategies naturally emerge:

(a) *Syntactic strategy:* one may target the syntax, and strengthen the proof system, so that it imposes additional burdens on the semantics.

(b) *Semantic strategy:* one may target the semantics, and *a priori* constrain the class of possible interpretations, so that making the semantics determinate is easier.

(c) *Strong pairing strategy:* one may target the relationship between syntax and semantics and require more than correctness of provable sequents, so that the same proof system places heavier constraints on the semantics which is to match it.

Carnap himself followed a syntactic strategy and outlined different ways to strengthen proof systems that would eliminate non-normal interpretations. One was to formalize logical contradiction on a par with logical consequence, the other to allow for multiple conclusions. This second option has become popular among proof theorists (see [19]; the idea goes back to Gentzen's notion of a sequent). Alternatively, one may add a primitive notion of rejection, construed as a pragmatic force complementary to assertion [20, 18]. Technically, these systems all succeed in eliminating non-normal interpretations in classical propositional logic. Moreover, [10] shows that categorical characterizations

of connectives can even be given for many-valued propositional logics, using multi-sided sequents. But it is easy to share Church's scepticism about the philosophical significance of these results for Carnap's Problem. Where does the extra expressive power embodied in these proof formats come from? Precisely because they outreach the familiar notion of inference (according to which some proposition follows from some other propositions), such formats may be suspected to rely upon semantic notions which are not part and parcel of our inferential practice. For example, multiple conclusions appear to presuppose a primitive understanding of disjunctive contents, and building the duality of assertion and rejection into the proof system could be argued to presuppose a grasp of negation (see [21] for a detailed rebuttal of multiple conclusions for the inferentialist).

Even more importantly, success in the propositional case for strategy (a) does not easily carry over to the first-order case. The latter is only cursorily dealt with in the literature, and existing treatments either assume a non-standard interpretation of quantifiers ([5, 9]), or relax the standards for what it means to fix the interpretation of quantifiers ([20]). Non-standard treatments construe universal and existential quantification as infinitary conjunction and disjunction (under the assumption that every element in the domain has a name, and that the proof system contains an ω-rule). This procrustean strategy shows at best that if quantifiers are reduced to connectives, what works for connectives works for quantifiers as well. But since at least [15] we recognize $\forall$ and $\exists$ as instances of *generalized quantifiers*, and the real question is whether the proof rules allow any *other* generalized quantifiers than these.

The strong pairing strategy, strategy (c) above, has recently been advocated in [8]. Standardly, the semantics matching a syntactically given consequence relation is simply to be such that the consequence relation is correct with respect to it. Whenever a sequent is derivable, it must be valid, that is, any interpretation which makes its antecedents true is to make its consequent true as well. When rules are given which generate the consequence relation, stronger ties may be demanded. Inference rules are not only a way to produce derivable sequents. They say that if some consequences hold, some other consequence holds. Hence, as Garson suggests, one may ask that the semantics be such that inference rules *preserve validity*. Whenever a sequent can be obtained by means of an inference rule from sequents which are valid with respect to the semantics,

that sequent should also be valid. Garson shows that such a strengthening of the ties between syntax and semantics resolves the underdetermination and yields intuitionistic semantics. The cost for this solution is a rather complex grasp of logical consequence — involving not just the knowledge of valid consequences, but the understanding of validity preserving mechanisms. After all, valid logical reasoning is often from premisses that are not themselves valid.

In the present paper, we wish to make a case for strategy (b). We will show that correctness with respect to classical logical consequence, together with a few principles of semantic interpretation, jointly suffice to fix the classical interpretation of the logical constants. Thus a potential learner does not need to grasp more than the idea of valid consequence — truth preservation, rather than the stronger notion of validity preservation as in (c) — and can get to the classical meanings. Comparing with solutions obtained along the lines of strategy (a), nothing more than standard inferential practice will be needed, and the classical interpretation of quantifiers will be recovered, by the same tactics which crack the propositional case. Thus, strategy (b), which, despite Church's recommendation, has never been put to work, would seem to provide the most satisfactory solution to Carnap's Problem.[1]

Comparison with strategy (c) as implemented in Garson's work draws attention to the importance of the choice of the underlying semantic framework. In [8], the models with respect to which logical connectives get an interpretation are sets of valuations. By contrast, Carnap's Problem is usually phrased in a simpler extensional framework, where models are just valuations [5, 10, 19]. This matters. First, for the question to be well-defined, one needs not only to pick up a certain relation of logical consequence, but also to fix the syntax and semantics of the language in which it is formulated. For logical languages, syntax is unproblematic, but one must choose the semantic values of expressions of various categories. Second, the richer the semantic values are, the more difficult Carnap's Problem becomes. In keeping with Carnap's original framing of the issue, we first adopt a standard extensional setting, solving Carnap's Problem for propositional logic in Section 4 and for first-order logic in Section 5. But maybe we made our task too easy by sticking to an extensional semantics? In Section 6, we ask — and answer — Carnap's question in a different, but also standard, framework for propositional logic, namely, that of possible worlds semantics.

In the next section, we lay down the semantic principles which will guide

us. Finding restrictions on possible interpretations which do the job is not difficult. As noted by Garson, it would suffice, for propositional logic, to only consider valuations in which at least one connective among $\neg$, $\vee$, $\rightarrow$, and $\leftrightarrow$ receives its standard interpretation: the interpretation of all the other is hereby forced to be standard ([8, p. 32]). But such a restriction is clearly *ad hoc*: why assume that the standard interpretation of one connective is known in advance? The difficulty thus lies in finding a *principled* restriction on possible interpretations which does the job. We shall argue in the next section that general and independently motivated semantic principles provide what we need.

3 Three semantic principles

Our semantic strategy is completely standard from the perspective of contemporary formal (model-theoretic) semantics, in which compositionality is a cornerstone. It can also be supported by a learnability argument. Suppose the only empirical evidence available to a learner of the meaning of the logical constants is their behaviour in inferences. Carnap's observation seems to indicate that this is not enough. But if there are semantic principles one can assume to hold for any language, these might sufficiently constrain the range of possible interpretations.

The argument rests on the hypothesis that a competent speaker needs to know the classical meaning of logical constants. This follows from the further assumption that semantic competence encompasses mastery of (classical) truth-conditions (even if meaning is not equated with truth conditions). Logical constants carve out truth-conditional content, and the fact that principles governing their use may suffice to fix their interpretation in advance is a distinctive property of logical constants. The extension of empirical predicates such as "red" or "blue" is not fixed by the functional role of colour concepts alone; it essentially depends on the way the world is. By contrast, we do not expect the world to help us determine which truth-function interprets "and" and which interprets "or", or which more elaborate function is the interpretation of "all" or "some". If this is to be knowable at all, it is knowable by any speaker who masters the appropriate rules of use.

The following three principles, which may be regarded as semantic universals, will suffice:

- *Non-triviality:* The language contains at least one false sentence.
- *Compositionality:* The semantic value of a compound expression is determined by the semantic values of its immediate constituents and the mode of composition.
- *Topic-neutrality* (needed only for the first-order case): Logical constants are permutation invariant.

Non-triviality is a very weak requirement, hardly in need of motivation. The learnability argument for compositionality is well known: If a language can express indefinitely many distinct propositions, compositionality is our currently best explanation of its learnability (see e.g. [17] for discussion). And topic-neutrality, in the precise form of invariance under permutations of the universe, is almost universally agreed to be a *necessary* condition for logicality.[2] It guarantees that the logical core of a language is general enough to carve out content in any conceivable situation of language use, irrespective of what objects are being talked about.

Once again, note that one must choose the semantic values of expressions belonging to a given syntactic category. Only then does compositionality make a definite contribution. In the next two sections, we take for granted a standard extensional framework for propositional and first-order logic, and in the last section we look at propositional logic in an intensional framework where the semantic values of sentences are sets of worlds.

Thus, the learnability argument presupposes that our hypothetical language learner already knows, or guesses, what *kind* of language is to be learnt: what the syntactic categories are, and what kinds of things expressions of these categories stand for. We are not claiming that this framework may itself be derived from a learnability argument, nor that it should be. It could well be adopted just on the basis of its simplicity: if the learner succeeds in making sense of the data available to her using extensional semantic values, she has no incentive to consider richer semantic values. Of course, as linguistic data flows in, she may have to adjust and go for richer values.

Now suppose I know, or assume, that a given expression is of a certain category, though I don't know what it means. Then its interpretation must belong to a certain class of semantic objects, but this class may well be infinite, even uncountable. Can I find out exactly *which* interpretation it has, *merely*

by studying the given relation of logical consequence? This is (our version of) Carnap's question.[3]

4 Non-normal interpretations in propositional logic

We start with classical propositional logic, where non-triviality and compositionality suffice to solve Carnap's Problem. Let *PL* be a standard language for propositional logic, usually with connectives $\neg, \wedge, \vee, \rightarrow$, but others may be added. A *valuation* is a bivalent assignment of truth values 1 or 0 to all sentences.[4] A *consequence relation* is a relation $\vdash$ between sets of sentences and sentences. A valuation v is *consistent* with a consequence relation $\vdash$ if and only if, whenever $\Gamma \vdash \varphi$, if $v(\psi) = 1$ for all $\psi \in \Gamma$, then $v(\varphi) = 1$. Carnap's question, expressed in the current terminology, is about the valuations which are consistent with $\vDash_{PL}$, where $\vDash_{PL}$ is classical logical consequence in *PL*.

Following Carnap, call a valuation v *normal* if it interprets the connectives as the intended truth functions, that is, if v is *#-boolean* for each connective #, where

> v is $\neg$-boolean if for every φ, $v(\neg\varphi) = 1$ iff $v(\varphi) = 0$
>
> v is $\wedge$-boolean if for every φ and ψ, $v(\varphi \wedge \psi) = 1$ iff $v(\varphi) = v(\psi) = 1$
>
> etc.

Carnap showed that there are exactly two kinds of non-normal valuations consistent with $\vDash_{PL}$. The first consists just of the valuation v_{T} which gives every sentence the value 1. It is trivially consistent with $\vDash_{PL}$, but since $v_{\mathrm{T}}(p) = v_{\mathrm{T}}(\neg p) = 1$, it is not $\neg$-boolean (the other connectives can be interpreted normally). The second kind has at least one false sentence and fails to be boolean for at least one binary connective. A typical example is the valuation v^* given by

> $v^*(\varphi) = 1$ iff φ is a tautology

Since logical consequences of tautologies are themselves tautologies, v^* is consistent with $\vDash_{PL}$, but we have $v^*(p) = v^*(\neg p) = 0$ and $v^*(p \vee \neg p) = 1$, so

v^* is neither $\neg$-boolean nor $\vee$-boolean. In fact, Carnap showed that all non-normal valuations of the second kind classify each unary and binary connective as boolean or not in the same way. And all valuations consistent with $\vDash_{PL}$ are $\wedge$-boolean, which points to an asymmetry between $\vee$ and $\wedge$ that Carnap found alarming.

With a consequence relation allowing multiple conclusions, the $\vee$-elimination rule

$$\varphi \vee \psi \vdash \varphi, \psi$$

would restore symmetry, and in fact such rules eliminate non-normal valuations. Our point, however, is semantic: non-normal valuations of the second kind are *not compositional*. To repeat, the compositionality principle says that the semantic value of a compound expression is determined by the semantic values of its immediate constituents and the mode of composition. In formal languages, where the syntactic rules are clear and we don't have to worry about ambiguous expressions, this means that an assignment μ of semantic values to expressions is compositional iff the following holds:

(PC) For every n-ary syntactic rule $\#$ there is a semantic composition function $F_\#$ such that for every well-formed expression $\#(e, \ldots, e_n)$ we have $\mu(\#(e, \ldots, e_n)) = F_\#(\mu(e_1), \ldots, \mu(e_n))$.

For the language PL, where μ is a valuation and the semantic values of sentences are truth values, (PC) says that for every n-ary connective $\#$ there is a function $F_\#$ such that for all sentences $\varphi_1, \ldots, \varphi_n$,

$$v(\#(\varphi_1, \ldots, \varphi_n)) = F_\#(v(\varphi_1), \ldots, v(\varphi_n)) \qquad (\#\text{-compositionality})$$

Thus, it *follows* from compositionality that, in this case, $F_\#$ is an n-ary *truth function*. When v is compositional we can moreover also take it to interpret $\#$ as $F_\#$, and thus write

$$v(\#(\varphi_1, \ldots, \varphi_n)) = v(\#)(v(\varphi_1), \ldots, v(\varphi_n))$$

The valuation v^* above is not $\neg$-compositional: $v^*(p)$ and $v^*(p \wedge \neg p)$ have the same value (0), but $v^*(\neg p)$ and $v^*(\neg(p \wedge \neg p))$ have different values. Now, restricting attention to compositional valuations also restores the symmetry

between $\vee$ and $\wedge$. Here is the general situation, for an n-ary connective $\#$:

(1) A compositional and $\vDash_{PL}$-consistent valuation v is $\#$-boolean if and only if $v(\#)(1,\ldots,1) = 1$.

So $\wedge$, $\vee$, and $\rightarrow$ get their normal interpretations by any such valuation, but not $\neg$ or, say, the Sheffer stroke. In fact, the usual introduction and elimination rules for the first three connectives fix their normal meaning. For example, consider disjunction. We saw that there are valuations v consistent with $\vDash_{PL}$ and sentences φ, ψ such that $v(\varphi) = v(\psi) = 0$ but $v(\varphi \vee \psi) = 1$. But this cannot be if v is compositional, for then

$$v(\varphi \vee \psi) = v(\vee)(v(\varphi), v(\psi)) = v(\vee)(v(\varphi), v(\varphi)) = v(\varphi \vee \varphi) = 0$$

since $\varphi \vee \varphi \vDash_{PL} \varphi$, which is in fact an instance of the usual $\vee$-elimination rule.

The problem with negation (or the Sheffer stroke) comes from the trivial valuation v_{T}, which *is* compositional. However, v_{T} is the *only* problem:

(2) All compositional and $\vDash_{PL}$-consistent valuations are normal except v_{T}. In other words, the classical laws of propositional logic, together with the semantic universals of non-triviality and compositionality, eliminate non-normal interpretations of propositional connectives.

It is quite remarkable that proofs of essentially both (1) and (2) can be found, if one looks carefully, already in [5].[5] Compositionality in the setting of classical propositional logic amounts to truth-functionality, or extensionality as Carnap called it. He duly noted which interpretations are extensional and which are not, but he didn't assign any special role to this property. Church doesn't mention the property in his review. Carnap's idea of semantics in 1943 was very much inspired by Tarski, but one should bear in mind that the modern notion of model-theoretic semantics is of a significantly later date.

5 Non-normal interpretations in first-order logic

Let us now see whether our semantic strategy also cracks the first-order case. To begin, what kind of non-normal interpretations are we to consider for first-order quantifiers? The situation is parallel to the propositional case, where

non-standard compositional interpretations for connectives consist in alternative truth-functions. Guided by the syntactic category of $\forall$ and $\exists$, and more generally by the standard interpretation of noun phrases in formal semantics, we take symbols of this category to denote *unary generalized quantifiers*, that is, sets of subsets of the domain. The standard interpretation for the existential quantifier is the set of all non-empty subsets; for the universal quantifier, it is the singleton of the domain. Accordingly, a non-normal interpretation for the existential or universal quantifier is any set of subsets different from these.

More precisely, consider a first-order language L interpreted over a domain M and the corresponding classical relation of logical consequence $\models_L$ in first-order logic. Since we assume compositionality, our interpretations amount to giving syntactically adequate semantic values to the logical and non-logical vocabulary. Since we furthermore assume non-triviality, we need not worry about the interpretation of connectives, which has to be standard, by (2). Hence, our interpretations can be taken to be pairs of the form $\mathcal{M}, Q$ where $\mathcal{M}$ is a standard L-structure based on M interpreting the non-logical vocabulary of L, and Q is a set of subsets of M, interpreting $\forall$ (we take $\exists$ to be defined as $\neg\forall\neg$). Given $\mathcal{M}, Q$, every sentence of L receives a truth-value by means of a recursive definition of satisfaction. Clauses for atomic formulas and connectives are the standard ones; the clause for $\forall$ now reads:

$$\mathcal{M}, Q \models \forall x\varphi\ \sigma \text{ if and only if } \{a \in M |\ \mathcal{M}, Q \models \varphi\ \sigma[x := a]\} \in Q$$

where σ is an assignment over M and $\sigma[x := a]$ is the assignment which is just like σ except that $\sigma(x) = a$. In keeping with the propositional case, we say that a pair $\mathcal{M}, Q$ is a normal interpretation if and only if $Q = \{M\}$. When $\mathcal{M}, Q$ is normal, the previous satisfaction clause reduces to the more familiar

$$\mathcal{M}, Q \models \forall x\varphi\ \sigma \text{ if and only if for all } a \in M,\ \mathcal{M}, Q \models \varphi\ \sigma[x := a]$$

Under our current working hypotheses, Carnap's question is whether all pairs $\mathcal{M}, Q$ consistent with $\models_L$ are normal. As we shall shortly see, the answer is negative: non-triviality and compositionality do not suffice to eliminate non-normal interpretations. The problem is thus indeed harder for quantifiers than it was for connectives. But, again, an independently motivated universal semantic constraint, in this case topic-neutrality, makes it possible to zero in on normal interpretations, so that Carnap's Problem is solved after all.

We shall now characterize the interpretations of $\forall$ which are consistent with $\vDash_L$, first in general and then when topic-neutrality is assumed. For the sake of simplicity, we will assume that our language contains predicate variables, and not just predicate symbols — without this simplifying assumption, similar results still hold but one must restrict attention to definable sets.[6] Given an L-structure $\mathcal{M}$, we say that a principal filter Q generated from a set A is *closed under the interpretation of terms* in $\mathcal{M}$ iff it is such that, for every term t with n free variables, for every sequence $a_1, \ldots, a_n$ of elements of A, $||t||^{\mathcal{M}}(a_1, \ldots, a_n) \in A$ where $||t||^{\mathcal{M}} : M^n \to M$ is the function interpreting t in $\mathcal{M}$. As a particular case, when t is a term with no free variables, the condition is meant to require that $||t||^{\mathcal{M}} \in A$. We then get the following characterization of possible interpretations for $\forall$ (the proof is given in the Appendix):

(3) An interpretation $\mathcal{M}, Q$ is consistent with $\vDash_L$ if and only if Q is a principal filter closed under the interpretation of terms in $\mathcal{M}$.

As it should be, the standard interpretation $\{M\}$ for $\forall$ is among the consistent ones, but, in general, there are many principal filters which are different from the trivial filter $\{M\}$, so there will be many non-normal interpretations for $\forall$. In view of (3), how wild are these non-normal interpretations? When Q is a principal filter, there is a subset A of M such that a set $B \subseteq M$ is in Q if and only if A is included in B. The satisfaction clause for $\forall$ then simplifies to

$$\mathcal{M}, Q \vDash \forall x\varphi \text{ if and only if for all } a \in A,\ \mathcal{M}, Q \vDash \varphi\ \sigma[x := a]$$

Thus, the quantifiers inhabiting the jungle of non-normal interpretations are still quite well-tamed. "All" means "all A" for some non-empty set of objects A included in the full domain M. Dually, "some" means "some A" for the same set A. The rules of logic do not determine which objects the quantifiers range over, except for the fact that objects with a name are in its range, and its range is also closed under functions named in the language. This is exactly as far as non-normality goes. Objects in the set A generating the filter are the real objects, for which existential import is valid: $\varphi(x) \vDash \exists\varphi(x)$ is satisfied by any $[x := a]$ for $a \in A$, for all formulas $\varphi(x)$. The objects outside A are dummy objects which happen to be in the domain but do not have existential import in the sense just stated. Indeed, the satisfaction clause we get for $\forall$ is nothing but the clause used in some semantics for *free logic*: A is the so-called inner domain

of real objects and its complement in M is the outer domain of non-existing things ([2]). Since our interpretations are consistent with the rules of classical logic, all objects which can be named are to be interpreted in the inner domain, which is guaranteed by the fact that A is closed under the interpretations of terms.

By (3), whenever there is at least one constant symbol in the language, there is a smallest principal filter Q consistent with $\vDash_L$: it is the principal filter generated by the set of objects interpreting constant symbols closed under the interpretations of function symbols. Identifying terms and the objects they denote, this amounts to a *substitutional interpretation* of the quantifiers. Thus, (3) says that the substitutional interpretation has a specific position among all possible interpretations of $\forall$: it is the weakest interpretation consistent with the rules for $\forall$ (weakest in the sense that the smaller the set from which the principal filter is generated, the easier it is to satisfy $\forall x\varphi$).

In non-normal interpretations, quantifiers make a difference between two kinds of objects, depending on whether they belong to the set generating the filter. This difference disappears only in the limiting case of normal interpretations, where this set is the entire domain and no object is left aside. Accordingly, the supplementary assumption that quantifiers treat all objects on a par, formally rendered as invariance under permutation, forces the interpretation of quantifiers to be normal:

(4) A principal filter Q on M is invariant under permutation if and only if $Q = \{M\}$.

Note that this also forces the equality symbol to be interpreted by real identity. The interpretation of the universal quantifier and the connectives being standard, axioms for identity guarantee that the equality symbol is interpreted by a congruence relation. Since the language contains predicate variables, this congruence needs to be the finest.

Permutation invariance for logical constants is thus our last semantic universal, labelled topic-neutrality. Together with non-triviality and compositionality, it ensures that connectives and quantifiers receive their normal interpretations in all interpretations which are coherent with the standard relation of logical consequence. Remarkably, permutation invariance, which is the traditional hallmark of quantifiers *qua* logical constants, was shown along the way not to follow from

quantifier rules. As far as rules or logical consequence are concerned, quantifiers could well not be invariant under permutations; invariance is a supplementary semantic feature which cannot be guessed on the basis of inferential practice.

6 Non-normal interpretations in intensional propositional logic

Carnap's question about the determination of the meaning of the logical symbols can be asked for any logic. We end by showing that in an intensional context, where sentences denote sets of possible worlds, the usual propositional connectives are still determined.

Let an *intensional language* be one built from propositional letters and the connectives $\neg, \wedge, \vee, \rightarrow$, plus possibly intensional propositional operators such as $\Box$. Now $\neg$ and $\Box$ plausibly have the same syntactic category, so they should receive the same kind of semantic values. Clearly, truth values no longer suffice. So let a set W of 'possible worlds' or 'states' be given. We take, as in standard possible world semantics, the semantic values of sentences to be subsets of W. Compositionality (principle (PC) in Section 4) then dictates that the operators must be interpreted as operations on $\mathcal{P}(W)$ (of the appropriate arity). Thus, an *interpretation* I assigns such an operation $I(\#)$ to each operator $\#$.[7] For simplicity, we now treat propositional letters not as symbols to be interpreted, but as *variables* to be assigned values. So an *assignment* f is a function from propositional letters to $\mathcal{P}(W)$. This has the advantage that there is no trivially true interpretation (i.e. one under which every sentence is true), so we actually don't need the non-triviality assumption any more.[8] Let $[\![\varphi]\!]^I_f$ be the value of φ under interpretation I and assignment f. The truth definition, relative to I, becomes:

- $[\![p]\!]^I_f = f(p)$
- $[\![\neg\varphi]\!]^I_f = I(\neg)([\![\varphi]\!]^I_f)$
- $[\![\varphi \wedge \psi]\!]^I_f = I(\wedge)([\![\varphi]\!]^I_f, [\![\psi]\!]^I_f)$

and similarly for all other operator symbols in the language.

Continuing to use Carnap's terminology, the *normal interpretation*, I_{n}, interprets $\neg$ as complement, $\wedge$ as intersection, etc.[9] But if W is infinite, there

are in principle uncountably many possible interpretations of the connectives, and Carnap's question in the current setting is whether the laws of classical propositional logic single out I_{n} as the *only* one.

In the intensional setting, an interpretation I is *consistent* with a consequence relation $\vdash$ if $\Gamma \vdash \varphi$ implies that for all assignments f,

$$\bigcap_{\psi\in\Gamma}[\![\psi]\!]^I_f \subseteq [\![\varphi]\!]^I_f$$

(with the understanding that $\bigcap_{\psi\in\emptyset}[\![\psi]\!]^I_f = W$). Using fairly standard terminology, let us say that an *intensional logic* is a consequence relation, in some intensional language as above, which contains all tautological consequences. Then we can prove the following:

(5) If I is an interpretation consistent with an intensional logic, then I is normal on the connectives $\neg, \wedge, \vee, \rightarrow$.

Note that, modulo the assumption about the propositional variables, the earlier result (2) about $\models_{PL}$ is a special case of (5), namely, when W is a unit set. The compositionality of I is the only assumption required for (5). We give a proof in the Appendix. Interestingly, this result requires more of $\models_{PL}$ than the proof of (2): it is easy to see that in the truth-functional case, already the intuitionistic part of $\models_{PL}$ is enough to fix the classical (!) meaning of the propositional connectives, whereas our proof of (5) requires double negation elimination and other non-intuitionistically valid laws of propositional logic.

In classical possible worlds semantics, all assignments to propositional variables are allowed. This is essential for the proof of (5). To see this, consider the so-called *possibility semantics* of [12]; [11] gives a comprehensive modern treatment. In the language with $\neg$, $\wedge$, and $\rightarrow$ as primitives, but $\varphi \vee \psi$ defined as $\neg(\neg\varphi \wedge \neg\psi)$, and with the standard truth definition clause for $\varphi \wedge \psi$, but the clauses from Kripke semantics for intuitionistic logic for $\neg\varphi$ and $\varphi \rightarrow \psi$, logical consequence turns out to be exactly $\models_{PL}$.[10] It is instructive to see why this is not a counter-example to (5). The reason is that possibility semantics, just as ordinary Kripke semantics for intuitionistic logic, places constraints on the allowed assignments, i.e. on the allowed *models*. Every assignment f must be *persistent*: if $w \in f(p)$ and wRw', then $w' \in f(p)$; possibility semantics adds a further constraint, called *refinability*. Clearly, imposing such constraints can in

principle make room for more interpretations of the connectives being consistent with a given consequence relation. Classical possible worlds semantics, on the other hand, has no such constraints.

7 Conclusion

Our take on Carnap's Problem is that it is made artificially difficult by considering all possible interpretations, no matter how bizarre. As speakers, we know that our language is going to be compositional, that it will have some true and some false sentences, and that its logical constituents will be topic-neutral. Therefore attention may be restricted to interpretations which satisfy these principles. Following Church's advice, this amounts to explicitly factoring out the role of semantic principles and the role of inference rules in fixing the interpretation of logical constants, rather than covertly using semantic notions to make sense of extended inference rules. This strategy proves successful both for propositional connectives and for quantifiers. In the case of classical propositional logic, the mere change of perspective to that of compositional formal semantics shows that the technical solution is in fact given already in [5]. Moreover, in a possible worlds setting, where the connectives are interpreted as operations on sets of worlds, it turns out that compositionality still suffices for the solution. Interestingly, it does not quite suffice to get the standard interpretation of quantifiers from first-order logical consequence. With compositionality as the only semantic assumption, quantifier rules essentially pick up the semantics for free logic. Still, it is rather surprising that nothing more than compositionality is required for the laws of classical logical consequence to fix the interpretation of the logical symbols in these ways. Classical logicians will happily acknowledge topic-neutrality as the extra assumption needed to get to the standard interpretation of the quantifiers.

In addition to the solution of Carnap's original problem, however, we hope to have at least indicated why Carnap's question can be a reasonable and interesting one to ask about any consequence relation in any logical language. In fact, this opens up an area of logical investigation that seems quite promising to us. The most immediate further issue, from the perspective of the present paper, is for which intensional logics also other logical symbols like $\Box$ are determined, but we shall leave this for another occasion.

Appendix

Our first aim in this appendix is to prove (3) and (4) above. Claim (3) states that an interpretation Q for $\forall$ is consistent with $\vDash_L$ iff Q is a principal filter, closed under the interpretation of terms. As announced, we consider a language with predicate variables. Given a set M, we shall say that a set Q of subsets of M is a *commutative filter* if and only if it is a filter and is such that for all relations R on M,

$$\{a \in M |\, R(a) \in Q\} \in Q \text{ iff } \{a \in M |\, R^{-1}(a) \in Q\} \in Q$$

where $R(a)$ is $\{b \in M |\, aRb\}$. We shall prove that the following are equivalent:

(i) $\mathcal{M}, Q$ is consistent with $\vDash_L$,

(ii) Q is a commutative filter on M and, for any term $t(x_1, \ldots, x_n)$, $\mathcal{M}, Q \vDash \forall x Px \to \forall x_1 \ldots \forall x_n Pt(x_1, \ldots, x_n)$

(iii) Q is a principal filter on M which is closed under the interpretation of terms in $\mathcal{M}$.

(3) is the equivalence between (i) and (iii). In a different context, the proof that commutative filters are principal can be found in [22], and the connection with the interpretation of the universal quantifier is already made in [23]. We provide here a somewhat shorter proof for that part. It is inspired by the filter model given by [1], Theorem 6.10, to prove that some valid first-order formulas are not valid in some models where the universal quantifier is interpreted by a filter (of course, in those models, the filter is not principal).[11] And we provide proofs of the other parts.

Proof. (i) implies (ii). The fact that Q is a filter is easily read off from some familiar first-order validities. Closure under finite intersection follows from the fact that $(\forall x\varphi \wedge \forall x\psi) \to \forall x(\varphi \wedge \psi)$ is valid. The validity of $\forall x(\varphi \wedge \psi) \to \forall x\varphi$ ensures closure under supersets. The empty set does not belong to Q because of $\forall x\varphi \to \neg\forall x\neg\varphi$. Moreover, Q is commutative because $\forall x\forall y\varphi \to \forall y\forall x\varphi$ is valid. Finally, $\mathcal{M}, Q \vDash \forall x Px \to \forall x_1 \ldots \forall x_n Pt(x_1, \ldots, x_n)$ simply because the formula is valid in first-order logic.

(ii) implies (iii). First, let Q be a non-principal filter; we show that Q is not commutative. Take any X in Q and consider (using Zorn's Lemma) a maximal

set Q' of subsets of X belonging to Q and totally ordered by inclusion. We define a parallel indexed family $\{\Delta_Y\}_{Y\in Q'}$ by setting $\Delta_Y = X - Y$ for $Y \in Q'$. We shall furthermore consider the sets $Z = \bigcup_{Y\in Q'} \Delta_Y$ and $Z' = X - Z$. The intuition is the following. Q' can be thought of as dropping more and more elements out of X. The indexed family $\{\Delta_Y\}_{Y\in Q'}$ collects the elements which have been dropped at each stage. Z is then the set of elements which have been dropped at some stage and Z' is the set of those that are never dropped. We now define a relation $R \subseteq M \times M$ for which commutativity will fail, by

$$R = (\bigcup_{Y\in Q'} (\Delta_Y \times Y)) \cup (Z' \times X)$$

First, note that $\{a \in M |\, R(a) \in Q\} = Z \cup Z' = X$. It is then sufficient to show that $\{a \in M |\, R^{-1}(a) \in Q\} = Z'$, because Z' does not belong to Q. If it did, Z' would be by construction the smallest element of Q'. Since Q is not principal, Q' cannot have a smallest element. To see this, let C be such a smallest element. Since Q is not principal, there is D in Q such that $C \cap D$ is a proper subset of C, but then, since Q' is assumed to be maximal, $C \cap D$, which belongs to Q, also belongs to Q', contradicting the fact that C is the smallest element of Q'. Thus, it remains only to be shown that $\{a \in M |\, R^{-1}(a) \in Q\} = Z'$. We reason by cases:

- If a is not in X, $R^{-1}(a) = \emptyset$, which is not in Q.
- If a is in Z, $R^{-1}(a) = (\bigcup_{a\in Y} \Delta_Y) \cup Z'$. Since a is in Z, there is a B in Q' such that a is not in B. Moreover, $(\bigcup_{a\in Y} \Delta_Y) \cap B = \emptyset$, since $a \in Y$ implies that B is strictly included in Y, Q' being totally ordered by inclusion. As a consequence, $R^{-1}(a) \cap B = Z'$. Since B is in Q and Z' is not, closure under intersection of Q implies that $R^{-1}(a)$ is itself not in Q.
- If a is in Z', $R^{-1}(a) = Z \cup Z' = X$, which is in Q.

This proves that Q is not commutative. Hence, by contraposition, if Q is commutative, Q is principal. It remains to be shown that Q is closed under the interpretation of terms in $\mathcal{M}$. Let Q be generated by A, let $t(x_1, \ldots, x_n)$ be a term, and let $a_1, \ldots, a_n \in A$. We need to show that $||t(x_1, \ldots, x_n)||^{\mathcal{M}}(a_1, \ldots, a_n) \in A$. First, interpret P by A. Then $\mathcal{M}, Q \vDash \forall x Px$, and therefore $\mathcal{M}, Q \vDash \forall x_1 \ldots \forall x_n Pt(x_1, \ldots, x_n)$ by hypothesis. By the satisfaction clause for $\forall$, it

follows that

$$\{\langle a_1, \ldots, a_n\rangle \mid ||t||^{\mathcal{M}}(a_1, \ldots, a_n) \in A\} \supseteq A^n$$

as desired.

(iii) implies (i). Let Q be a principal filter generated by a subset A of M. Let H be a Hilbert proof system for $\models_L$ as for example the one given by [7], p. 194; predicate symbols and predicate variables being treated on a par. It suffices to establish that if φ is deducible from a set Γ of formulas without free individual variables in H, then $\mathcal{M}, Q \models \Gamma$ implies $\mathcal{M}, Q \models \varphi\ \sigma$ for any assignment σ whose individual variables have values in A. This is shown by induction on the length of proofs in H. The two key steps are the soundness of universal instantiation, which, in H, is the axiom $\forall x\varphi \to \varphi[t/x]$ with t free for x in φ, and the soundness of universal generalization, which, in H, is inferring $\forall x\varphi$ from φ (no restriction on x is needed since we consider only derivability from a set of sentences without free individual variables).

For universal instantiation, we need to show that $\mathcal{M}, Q \models \varphi[t/x]$ for any assignment σ with values in A, for any term t free for x in φ. Assume $\mathcal{M}, Q \models \forall x\varphi\ \sigma$. By the satisfaction clause for $\forall$ and the fact that Q is a filter generated by A, the set of objects $a \in M$ such that $\mathcal{M}, Q \models \varphi\ \sigma[x := a]$ is a superset of A. Hence, for our term t with free variables $x_1, \ldots, x_n$, $\mathcal{M}, Q \models \varphi[t/x]\ \sigma$, because $\sigma(x_1), \ldots, \sigma(x_n)$ are in A and A is closed under the interpretation of t.

For universal generalization, we assume that we have a proof of a formula $\forall x\varphi$ from a set of sentences Γ in H ending with an application of universal generalization. By induction hypothesis, $\mathcal{M}, Q \models \Gamma$ implies $\mathcal{M}, Q \models \varphi\ \sigma$ for any assignment σ with values in A. Assume $\mathcal{M}, Q \models \Gamma$. Let σ be an arbitrary assignment with values in A. By induction hypothesis, the set of objects $a \in M$ such that $\mathcal{M}, Q \models \varphi\ \sigma[x := a]$ is a superset of A. Therefore, since Q is interpreted by a filter generated from A, we have $\mathcal{M}, Q \models \forall x\varphi\ \sigma$, as required. □

Next, we show

(4) A principal filter Q on M is invariant under permutation if and only if $Q = \{M\}$.

Proof. The direction from right to left is immediate. We prove the direction from left to right by contraposition. Assume that Q is a principal filter generated

by a set A different from M. There are $a, b \in M$ such that $a \in A$ but $b \notin A$. Consider the permutation π which swaps a and b and is the identity everywhere else. a is not in $\pi(A)$, so A is not included in $\pi(A)$, so $\pi(A) \notin Q$, contradicting the invariance of Q. □

Finally, we give a proof of

(5) If I is an interpretation consistent with an intensional logic, then I is normal on the connectives $\neg, \wedge, \vee, \rightarrow$.

Let W be a set of worlds, and I an interpretation (over W) consistent with an intensional logic $\vdash$. We must show that $I(\#)$ is normal for $\# \in \{\neg, \wedge, \vee, \rightarrow\}$. Recall that for all formulas φ, ψ, θ,

(6) $\varphi, \psi \vdash \theta$ implies that for each assignment f, $[\![\varphi]\!]^I_f \cap [\![\psi]\!]^I_f \subseteq [\![\theta]\!]^I_f$.

To begin, it is clear that $\wedge$ is normal:

(7) For $X, Y \subseteq W$, we have $I(\wedge)(X, Y) = X \cap Y$.

This follows from the introduction and elimination rules for $\wedge$, in the form

$$p \wedge q \vDash_{PL} p,\ \ p \wedge q \vDash_{PL} q,\ \ p, q \vDash_{PL} p \wedge q$$

Since $\vdash$ is an intensional logic, these rules hold for $\vdash$ too. Let f be such that $f(p) = X$ and $f(q) = Y$. Using (6), (7) follows.

Next, we observe that it suffices to show that negation must be interpreted normally by I.

(8) If $I(\neg)(X) = W - X$ for all $X \subseteq W$, then also $\vee$ and $\rightarrow$ are interpreted normally by I.

Consider $\vee$. From the fact that $p \vdash p \vee q$ and $q \vdash p \vee q$ we obtain from (6) (with a suitable f) that $X \cup Y \subseteq I(\vee)(X, Y)$. Also, since $p \vee q, \neg p \vdash q$ we get, using our assumption,

$$I(\vee)(X, Y) - X \subseteq Y$$

In other words, $I(\vee)(X, Y) \subseteq X \cup Y$, so $I(\vee)(X, Y) = X \cup Y$, as was to be

proved. In a similar way, one shows that $I(\rightarrow)(X,Y) = (W - X) \cup Y$. (The assumption about negation is used to show $W - X \subseteq I(\rightarrow)(X,Y)$, which one gets from $\neg p \vdash p \rightarrow q$.) This proves (8).

Concerning negation, it is easy to see that the following holds:

(9) For all $X \subseteq W$, $X \cap I(\neg)(X) = \emptyset$.

This follows from the fact that $p, \neg p \vdash q$, letting $f(p) = X$ and $f(q) = \emptyset$. Thus, all that remains to prove is:

(10) For all $X \subseteq W$, $X \cup I(\neg)(X) = W$.

First, it follows from (9) that

(11) $I(\neg)(W) = \emptyset$

Also,

(12) $I(\neg)(\emptyset) = W$

This is because $\vdash \neg(p \wedge \neg p)$ (so for all f, $[\![\neg(p \wedge \neg p)]\!]^I_f = W$), with $f(p) = \emptyset$; recall that conjunction is interpreted as intersection.

Now suppose $\emptyset \subsetneq X \subsetneq W$ and, for contradiction, that $X \cup I(\neg)(X) \neq W$. Take $b \in W - (X \cup I(\neg)(X))$, and let $f(p) = X$ and $f(q) = \{b\}$.[12] Then we have

(13) $[\![\neg(p \wedge q)]\!]^I_f = W$

(by (12), since $X \cap \{b\} = \emptyset$), and

(14) $[\![\neg p \wedge q]\!]^I_f = \emptyset$

(since $I(\neg)(X) \cap \{b\} = \emptyset$). Since $\neg(p \wedge q) \vdash \neg p \vee \neg q$, it follows from (13) that

(15) $[\![\neg p \vee \neg q]\!]^I_f = W$

And since $\neg(p \vee \neg q) \vdash \neg p \wedge q$, it follows from (14) that $[\![\neg(p \vee \neg q)]\!]^I_f = \emptyset$, and hence by (12) that $[\![\neg\neg(p \vee \neg q)]\!]^I_f = W$. Thus, since $\neg\neg(p \vee \neg q) \vdash p \vee \neg q$, we have

(16) $\quad [\![p \vee \neg q]\!]^I_f = W$

But we also have

$$p \vee \neg q, \neg p \vee \neg q \vdash \neg q$$

By (15) and (16), this entails that $[\![\neg q]\!]^I_f = I(\neg)(\{b\}) = W$, contradicting the fact (from (9)) that $\{b\} \cap I(\neg)(\{b\}) = \emptyset$. This proves (10), and thereby (5), as desired. □

References

[1] A. Antonelli. On the general interpretation of quantifiers. *Review of Symbolic Logic*, 6:637–658, 2013.

[2] E. Bencivenga. Free logics. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, vol. III: Alternatives to Classical Logic*, pages 373–426. Reidel, Dordrecht, 1986.

[3] D. Bonnay and D. Westerståhl. Consequence mining: constants versus consequence relations. *Journal of Philosophical Logic*, 41(4):671–709, 2012.

[4] Denis Bonnay. Logicality and invariance. *Bulletin of Symbolic Logic*, 14:1:29–68, 2008.

[5] R. Carnap. *Formalization of Logic*. Harvard University Press, Cambridge, Mass., 1943. Vol. 2 of *Studies in Semantics*.

[6] A. Church. Review of Carnap 1943. *Philosophical Review*, 53:493–498, 1944.

[7] R. Cori and D. Lascar. *Mathematical Logic*. Oxford University Press, Oxford, 2000.

[8] J. W. Garson. *What Logics Mean*. Cambridge University Press, Cambridge, 2013.

[9] I. Hacking. What is logic? *Journal of Philosophy*, 76:285–319, 1979.

[10] O. T. Hjortland. Speech acts, categoricity and the meaning of logical connectives. *Notre Dame Journal of Formal Logic*, 55(4):445–467, 2014.

[11] Wesley Holliday. Possibility frames and forcing for modal logic. Manuscript, 2015.

[12] L. Humberstone. From worlds to possibilities. *Journal of Philosophical Logic*, 10(3):313–339, 1981.

[13] L. Humberstone. *The Connectives*. MIT Press, Cambridge MA, 2011.

[14] A. Koslow. Carnap's problem: What is it like to be a normal interpretation of classical logic? *Abstracta*, 6(1):117–135, 2010.

[15] A. Mostowski. On a generalization of quantifiers. *Fundamenta Mathematicae*, 44:12–35, 1957.

[16] J. Murzi and O. Hjortland. Inferentialism and the categoricity problem: Reply to Raatikainen. *Analysis*, 69(3):480–488, 2009.

[17] P. Pagin and D. Westerståhl. Compositionality I (definitions and variants) and II (arguments and problems). *Philosophy Compass*, 5(3):250–282, 2010.

[18] I. Rumfitt. 'Yes' and 'No'. *Mind*, 109:781–823, 2000.

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

[20] T. Smiley. Rejection. *Analysis*, 56(1):1–9, 1996.

[21] F. Steinberger. Why conclusions should remain single. *Journal of Philosophical Logic*, 44:333–355, 2011.

[22] M. van Lambalgen. Independence, randomness and the axiom of choice. *Journal of Symbolic Logic*, 57:1274–1304, 1992.

[23] D. Westerståhl. Self-commuting quantifiers. *Journal of Symbolic Logic*, 61:212–224, 1996.

Notes

[1] [14] also discusses Carnap's problem in the light of Church's criticism, but does not put forward compositionality as a solution.

[2] There is debate about whether it is also a *sufficient* condition; see e.g. [4].

[3] Mustn't we also assume that the learner knows which expressions are *logical* and which are not? This can be a problematic distinction in more complex languages. But for the simple logical languages under discussion in this paper, there is a test our hypothetical learner could in principle perform, again relying only on the given consequence relation: Consider a valid inference in which expression e occurs. Does it remain valid under all replacements of e by another expression of the same category (or better, under all appropriate reinterpretations of e)? If Yes (for all such inferences), e is not a logical expression; if No, it is. For (much) more about this methodology, see [3].

[4] This is the current terminology, e.g. in [13]; in [5] they are called interpretations. We use "interpretation" too in later sections, though in a slightly different sense, since our interpretations are by definition compositional.

[5] Which is why we don't give the proofs here.

[6] See [1] for an adequate framework to carry out this restriction.

[7] We could also treat W simply as a parameter, and let a (global) interpretation be a functor I which to each W assigns a (local) interpretation I^W as defined here. The same results would hold for global interpretations.

[8] Assignments are usually called *valuations* in possible worlds semantics, but we already used that term in Section 4 for mappings from sentences to truth values. There too, we could have avoided the non-triviality assumption by treating propositional letters as variables, but we wanted to stay as close as possible to Carnap's original setting.

[9] Thus, a *model* in the usual sense, for the basic modal language with operators $\neg, \wedge, \vee, \rightarrow, \Box$, has the form (W, f) if we, like Carnap, think of necessity as truth in all worlds, and the form (W, R, f), where $R \subseteq W^2$, in Kripke semantics. Then we have, for $w \in W$,

$$(W, R, f), w \vDash \varphi \;\text{ iff }\; w \in [\![\varphi]\!]^{I_{\mathrm{n}}}_f$$

The normal interpretation of $\Box$ is, in the Carnap setting,

$$I_{\mathrm{n}}(\Box)(X) = \begin{cases} W & \text{if } X = W \\ \emptyset & \text{if } X \neq W \end{cases}$$

for $X \subseteq W$, whereas in the Kripke setting it is

$$I_{\mathrm{n}}(\Box)(X) = \{w \in W : R(w) \subseteq X\}$$

where $R(w) = \{w' : wRw'\}$. Here, however, we are only concerned with the interpretations of $\neg, \wedge, \vee, \rightarrow$, and then it doesn't matter which setting is chosen.

[10] We thank an anonymous referee for drawing our attention to this fact, and for encouraging us to include the proof of (5) in the paper.

[11] Antonelli's result concerns non-standard models in which $\forall$ is interpreted by a (not necessarily principal) filter. Since such non-standard models falsify some first-order validities, it

is natural to ask which further restriction on the interpretation of $\forall$ would be such that it guarantees that valid formulas are exactly those deemed valid by first-order logic, and that it is forced in any model of the validities of first-order logic.

[12]If *persistence* had been a requirement on assignments, this particular f might not be allowed; see the remarks at the end of Section 6.