Notre Dame Journal of Formal Logic
Volume 36, Number 3, Summer 1995

On Gabbay's Proof of the Craig Interpolation Theorem for Intuitionistic Predicate Logic

MICHAEL MAKKAI

Abstract Using the framework of categorical logic, this paper analyzes and streamlines Gabbay's semantical proof of the Craig interpolation theorem for intuitionistic predicate logic. In the process, an apparently new and interesting fact about the relation of coherent and intuitionistic logic is found.

1 Introduction For any logical system with entailment relation $\vdash$, by saying that the Craig interpolation theorem holds we mean the following:

> for any sentences φ_1, φ_2 over the respective vocabularies $\mathcal{L}_1, \mathcal{L}_2$, if $\varphi_1 \vdash \varphi_2$ then there is a sentence φ over $\mathcal{L}_1 \cap \mathcal{L}_2$, the common part of the given vocabularies, such that $\varphi_1 \vdash \varphi \vdash \varphi_2$.

The interpolation theorem for classical predicate logic was found by Craig [4]. Craig's proof used a proof-theoretic approach. Maehara [11] gave the well known proof through Gentzen's sequent calculus. Schütte [16] extended the theorem to intuitionistic logic with a similar proof-theoretical proof. As Bell and Machover point out (see [2], Chapter 9, Section 13), the theorem for classical logic can be deduced from the Intuitionistic version by a use of the double negation interpretation.

The model theory of classical logic gives several illuminating proofs of the interpolation theorem for classical logic. It is customary (see Chang and Keisler [3]) to reduce the result to Robinson's consistency lemma, originally used by Robinson [15] for a proof of Beth's definability theorem (see Beth [1]), a consequence of Craig interpolation. For the proof of the consistency lemma, Robinson used the method of alternating chains of model-extensions. Gabbay's proof of interpolation for intuitionistic logic in Gabbay [5] exploits the same method. Gabbay's proof is based on Kripke's semantic analysis of intuitionistic logic in Kripke [9].

A very elegant and conceptually satisfactory proof of interpolation for intuitionistic logic is given by Pitts [14]; he uses locale-theoretic methods, and his work is

Received September 16, 1991; revised May 29, 1995

inspired by the ground breaking work of Joyal and Tierney [8] (available in preprint form earlier) that introduced locales into topos theory (and thus logic). Pitts's proof is constructive as opposed to Gabbay's, which is model-theoretic and uses non-constructive tools such as the compactness theorem. On the other hand, as we intend to show, Gabbay's proof shows something stronger than the interpolation theorem (we have nothing to say here on the question whether Pitts's methods, or other constructive methods, could yield the same additional conclusion).

Let us recall some basic concepts of categorical logic, intended to capture the notion of theory in intuitionistic, and coherent (see below), predicate logic. A *Heyting category* is a category in which:

1. all finite limits exist;
2. the poset $Sub(A)$ of subobjects of every object A is a lattice (with 0 and 1);
3. for any arrow $f : A \to B$, the map $f^* : Sub(B) \to Sub(A)$ defined as taking pullback (inverse image) is a lattice homomorphism;
4. f^* as in (3) has both left and right adjoints

$$\begin{tikzcd} Sub(A) \arrow[r, shift left=2, "\exists_f"] \arrow[r, shift right=2, "\forall_f"'] & Sub(B) \arrow[l, "f^*"'] \end{tikzcd} : \exists_f \dashv f^* \dashv \forall_f;$$

5. $\exists_f$ is stable under pullback (Beck-Chevalley condition): if

$$\begin{tikzcd} A \arrow[r, "f"] & B \\ A' \arrow[u, "a"] \arrow[r, "g"'] & B' \arrow[u, "b"'] \end{tikzcd}$$ is a pullback square, then

$$\begin{tikzcd} Sub(A) \arrow[r, "\exists_f"] \arrow[d, "a^*"'] & Sub(B) \arrow[d, "b^*"] \\ Sub(A') \arrow[r, "\exists_g"'] & Sub(B') \end{tikzcd}$$ commutes.

(The analogous condition for $\forall_f$s follows.)

The notion of Heyting category embodies that of a theory in multisorted intuitionistic predicate logic with equality. This correspondence is explained in Makkai and Reyes [12], and although Heyting categories are not explicitly introduced there, all the categorical operations involved are. In particular, given a theory $(\mathcal{L}, \mathcal{T})$ in intuitionistic logic with vocabulary $\mathcal{L}$ and set of axioms $\mathcal{T}$, we have the corresponding Heyting category $LT(\mathcal{L}, \mathcal{T})$, or more explicitly $LT_{Heyting}(\mathcal{L}, \mathcal{T})$, with the letters L and T referring to the fact that the construction of $LT(\mathcal{L}, \mathcal{T})$ is like the Lindenbaum-Tarski construction, originally of a Boolean algebra, based on formulas of the theory taken as a basis; see Chapter 8 of [12].

We may, equivalently for the purposes of the Craig interpolation theorem, use Heyting pretoposes instead, as is done in [14]. In fact, every Heyting pretopos is a Heyting category (although not conversely). However, Heyting categories correspond more directly to the usual symbolic logical notion of theory.

A *Heyting functor* is one between Heyting categories, preserving in the usual sense the operations posited in Heyting categories. The various notions of *model* used in intuitionistic logic are *the same*, modulo the correspondence of categories and theories, as Heyting functors with a specific kind of target category. For example Kripke models of $(\mathcal{L}, \mathcal{T})$ ("modified Kripke structures" in Gabbay [6]) correspond to Heyting functors from $LT(\mathcal{L}, \mathcal{T})$ into Set^P, with any poset P. Here Set^P is the category of functors from P (as a category) into Set, the category of (small) sets; Set is a topos, in particular, a Heyting category.

Heyting categories and functors, with all natural transformations between the latter as 2-arrows, form a 2-category called $\mathcal{HEYTING}$; restricting objects to small ones gives the 2-category $\mathcal{H}eyting$. The 2-categorical aspects will play a minor role only; most of the time it is enough to keep in mind the ordinary categories involved.

In fact, the 2-category of Heyting categories that is "good" from an algebraic point of view is the one in which only isomorphism 2-arrows are allowed. By denoting the result of discarding 2-arrows that are not isomorphism by postfixing an asterisk, we are talking about the 2-categories $\mathcal{HEYTING}^*$ and $\mathcal{H}eyting^*$.

Every Heyting category is a *coherent category*. In [12] these were called "logical categories" and played a central role. The definition of a coherent category is obtained from our definition of a Heyting category by deleting reference to $\forall_f$ in (4) and (5). As explained in [12], coherent categories correspond to *coherent theories*. In coherent logic formulas are positive existential (*coherent*), that is, built up of atomic formulas using $\wedge$, $\vee$ and $\exists$; and *axioms* are of the form $\forall\vec{x}(\varphi(\vec{x}) \to \psi(\vec{x}))$ with φ and ψ coherent. In this paper, all the "action" around a Heyting category will take place in the *coherent doctrine*: we will deal with properties of Heyting categories *qua* coherent categories, properties that are not generally true for coherent categories, but hold if the coherent category carries the additional structure of a Heyting category as well. In this way, we "explain" intuitionistic logic via the simpler coherent logic.

A *Boolean category* is a coherent category in which each subobject lattice is complemented (a Boolean algebra); every Boolean category is a Heyting category at the same time. Boolean categories correspond to theories in full classical logic in the usual sense (see [12]).

An important point in our story is that the interpolation property of a logical system such as classical logic, intuitionistic logic, and others as well, is equivalent to an exactness property of the (2-) category of the corresponding kind of categories: for classical logic the 2-category of Boolean categories, for intuitionistic logic the 2-category of Heyting categories. The exactness property in question refers to *pushouts*. For the record we give the definition of the 2-categorical version of pushout, also called bipushout (see Street [17]). However, the reader will not go far wrong if he takes the notion to be the ordinary categorical notion, especially in view of the remarks to follow later.

Given a 2-category and a diagram

$$\begin{tikzcd} S \\ R \arrow[u, "F"] \arrow[r, "G"'] & T \end{tikzcd} \qquad (1)$$

and any object Q, an extension of the form

$$\begin{tikzcd} S \arrow[r, "I"] \arrow[dr, phantom, "\overset{\alpha}{\cong}"] & Q \\ R \arrow[u, "F"] \arrow[r, "G"'] & T \arrow[u, "J"'] \end{tikzcd} \qquad \alpha : IF \xrightarrow{\cong} JG$$

is called a Q-*cocone* on (1). Q-cocones on the given triangle (1) and a fixed Q form a category. With another Q-cocone, and with data denoted by primed letters, an arrow $(Q, I, J, \alpha) \longrightarrow (Q, I', J', \alpha')$ is a pair $(\varphi : I \longrightarrow I', \psi : J \longrightarrow J')$ such that the square of 2-arrows

$$\begin{tikzcd} IF \arrow[r, "\alpha"] \arrow[d, "\varphi F"'] & JG \arrow[d, "\psi G"] \\ I'F \arrow[r, "\alpha'"'] & J'G \end{tikzcd}$$

commutes; composition of these arrows is defined in the expected way. The category of Q-cocones on the fixed (1) is denoted $Cocone_Q$ (with reference to (1) suppressed). Note that for any two objects P and Q, a P-cocone $\Gamma = (P, I, J, \alpha)$ and an arrow $H : P \longrightarrow Q$ gives a Q-cocone $\Gamma^*(h) = (Q, HI, HJ, H\alpha)$, and that, in fact, we obtain a functor:

$$\Gamma^* : \quad \mathrm{Hom}(P, Q) \longrightarrow Cocone_Q.$$

We say that Γ is a *pushout of* (1) if for all objects Q, the functor Γ^* is an *equivalence* of categories. Pushouts, when they exist, are determined up to *equivalence*.

The 2-categories $\mathcal{H}eyting^*$, $\mathcal{B}oole^*$, $Coherent$, are all "good"; they are locally finitely presentable 2-categories (unfortunately, the 2-categorical versions of the well known concepts of Gabriel and Ulmer [7] have not been written down in the literature, despite, or maybe because, of the fact that the work to be done is straightforward), and as a consequence, pushouts exist in them; moreover, pushouts in $\mathcal{H}eyting^*$ are also pushouts in $\mathcal{HEYTING}^*$, etc. In fact, they can be so constructed that the canonical equivalences Γ^* are surjective on objects (although they still fall short of being isomorphisms), which somewhat simplifies the notation when working with them. A further simplification of the notion of pushout is to require that the structure map $\alpha : IF \longrightarrow JG$ in the pushout-cocone be an identity; in particular that $IF = JG$. In all our examples it is in fact possible to achieve this. When this is done a pushout in the 2-category is almost a pushout in the corresponding ordinary category (without the 2-arrows), except that the uniqueness part of the universal condition holds with the qualification "up to an isomorphism 2-arrow." Note however that the definition of the 2-categorical pushout has 2-dimensional aspects that go beyond the defining properties of the ordinary pushout.

In our proofs we will never have the need to use the explicit constructions of pushouts; their universal properties will suffice. Nevertheless, it is important to point out that pushouts have a well-known meaning when our categories are given via theories. Suppose we have the intuitionistic theories $(\mathcal{L}_0, \mathcal{R})$, $(\mathcal{L}_1, \mathcal{S})$, $(\mathcal{L}_2, \mathcal{T})$ with $\mathcal{L}_0 \subset \mathcal{L}_1$, $\mathcal{L}_0 \subset \mathcal{L}_2$, $\mathcal{R} \subset \mathcal{S}$, and $\mathcal{R} \subset \mathcal{T}$; moreover, assume that the vocabularies $\mathcal{L}_1$ and $\mathcal{L}_2$ have nothing more common than the elements of $\mathcal{L}_0$, $\mathcal{L}_1 \cap \mathcal{L}_2 = \mathcal{L}_0$. We then have the pushout diagram:

$$\begin{tikzcd} LT(\mathcal{L}_1, \mathcal{S}) \arrow[r] & LT(\mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{S} \cup \mathcal{T}) \\ LT(\mathcal{L}_1 \cap \mathcal{L}_2, \mathcal{R}) \arrow[u] \arrow[r] & LT(\mathcal{L}_2, \mathcal{T}) \arrow[u] \end{tikzcd} \tag{2}$$

in which all arrows are induced by inclusions of vocabularies. This can be seen by applying the universal property of $LT(\mathcal{L}, \mathcal{T})$ as the free Heyting category in which $(\mathcal{L}, \mathcal{T})$ is interpreted. For the analogous statement of this universal property for the coherent doctrine rather than the Heyting one, see Proposition 8.2.3 on page 247 of [12].

Consider the following property of any 2-category, the *Strong Amalgamation Property*:

(SAP) If in the pushout

$$\begin{tikzcd} S \arrow[r, "I"] \arrow[dr, phantom, "\overset{\alpha}{\cong}"] & Q \\ R \arrow[u, "F"] \arrow[r, "G"'] & T \arrow[u, "J"'] \end{tikzcd}$$

F is conservative then so is J.

(In any of our “concrete” 2-categories, “conservative” means “reflects isomorphisms”; conservative functors correspond to conservative extensions of theories.)

In the main part of this paper, we will be giving a proof of the following theorem.

Theorem 1.1 *Both $\mathcal{B}oole^*$ and $\mathcal{H}eyting^*$ have the SAP.*

It had long been known that the SAP is equivalent to the Craig Interpolation Theorem. For one thing, Robinson’s Lemma, which is used in the usual model-theoretical proofs of interpolation, is a special case of SAP for $\mathcal{B}oole$. With reference to (2), Robinson’s Lemma says that if $(\mathcal{L}_1 \cap \mathcal{L}_2, \mathcal{R})$ is complete, and $(\mathcal{L}_1, \mathcal{S})$, $(\mathcal{L}_2, \mathcal{T})$ are consistent, then $(\mathcal{L}_1 \cup \mathcal{L}_2, \mathcal{S} \cup \mathcal{T})$ is consistent. This follows from the SAP, since by $(\mathcal{L}_1 \cap \mathcal{L}_2, \mathcal{R})$ being complete and $(\mathcal{L}_1, \mathcal{S})$ consistent, F is conservative. By the SAP, J is conservative. Since $(\mathcal{L}_2, \mathcal{T})$ is consistent, its conservative extension is consistent.

Nevertheless, I do not know any place other than the above mentioned paper [14] of Pitts where the SAP is explicitly established, and interpolation is derived from it, for the intuitionistic doctrine. In fact, [14] goes beyond this and shows that any pushout square has the interpolation property in an appropriate sense provided SAP holds. Let me briefly indicate how the interpolation theorem follows from the SAP in a simpler way than through Robinson’s Lemma.

With reference to the notation in the statement of Craig’s Theorem above, consider the following square.

$$\begin{tikzcd}
LT(\mathcal{L}_1, \{\varphi_1\}) \arrow[r] & LT(\mathcal{L}_1 \cup \mathcal{L}_2, \{\varphi_1\}) \\
LT(\mathcal{L}_1 \cap \mathcal{L}_2, Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)) \arrow[r] \arrow[u] & LT(\mathcal{L}_2, Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)). \arrow[u]
\end{tikzcd}$$

Here, $Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)$ denotes the set of consequences of φ_1 over the vocabulary $\mathcal{L}_1 \cap \mathcal{L}_2$, that is, the set of sentences θ over $\mathcal{L}_1 \cap \mathcal{L}_2$ such that $\varphi_1 \vdash \theta$. The arrows are all induced by inclusions of vocabularies. The diagram is a pushout by the above more general situation: note that $\{\varphi_1\} \cup Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)$ has the same deductive strength as $\{\varphi_1\}$. Now, observe that the left vertical arrow is conservative. If the sentence θ over $\mathcal{L}_1 \cap \mathcal{L}_2$ is provable from φ_1, (that is, has become 1 in the left upper corner), then it *belongs to* $Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)$. Thus it is provable from the same set of axioms. (Here we are using the easily seen fact that conservativeness for a map of Heyting categories—but not for coherent categories—is reduced to the following special case: if the monomorphism $U \rightarrowtail 1$ into the terminal object in the domain category becomes an isomorphism in the codomain category, then it was an isomorphism.) By SAP, the right vertical is conservative. Since by the assumption of the interpolation-statement φ_2 is provable from φ_1, it follows that it is provable from $Cn_{\mathcal{L}_1 \cap \mathcal{L}_2}(\varphi_1)$, which clearly implies the desired conclusion.

To continue with the account of the necessary background, let us consider a coherent category T, and let $\mathrm{Mod}(T)$ denote the category of all coherent functors $T \longrightarrow \mathrm{Set}$, with all natural transformations as arrows: $\mathrm{Mod}(T) = \mathcal{COHERENT}(T, \mathrm{Set})$.

We have the evaluation functor:

$$\begin{array}{rccl} e_T: & T & \longrightarrow & \mathrm{Set}^{\mathrm{Mod}(T)} \\ & A & \longmapsto & [M \longmapsto M(A)] \end{array}$$

and also its variant:

$$|e_T|: \quad T \quad \longrightarrow \quad \mathrm{Set}^{|\mathrm{Mod}(T)|}$$

where in the second case we disregard arrows in $\mathrm{Mod}(T)$ (thus, $\mathrm{Set}^{|\mathrm{Mod}(T)|}$ is merely a—large—Cartesian power of Set). Now, the categorical version of the Gödel completeness theorem says that, for a small T (as will be assumed throughout), $|e_T|$ is a conservative functor (see for example Makkai [13], Theorem 2.2.2). In fact, this is obviously equivalent to saying that e_T is conservative. Whereas $\mathrm{Set}^{|\mathrm{Mod}(T)|}$ is a Boolean category, $\mathrm{Set}^{\mathrm{Mod}(T)}$ is not. However, the latter is a Heyting category. A fundamental result of categorical logic is the following theorem.

Theorem 1.2 (Joyal's Theorem) *If T is a (small) Heyting category, then e_T is a Heyting functor.*

Note that $\mathrm{Mod}(T)$ still has the same sense: the category of coherent functors from T to Set. As a consequence of Joyal's and Gödel's theorems, we have a conservative Heyting embedding of T into a very special Heyting category, one of the form $\mathrm{Set}^{\mathbf{I}}$ where $\mathbf{I}$ is a category. Joyal's theorem (combined with Gödel's) is a completeness theorem. In fact, it is essentially equivalent to Kripke's completeness theorem (see his [9]) for intuitionistic logic. From the formulation of Joyal's theorem, it is quite easy to prove the existence of a conservative Heyting embedding of the form $T \longrightarrow \mathrm{Set}^P$ with a poset P, with possible further conditions on P (for example regarding its size, or that it be a forest—"a tree with many roots"). However, for our purposes, the canonical form of Joyal's theorem is quite sufficient.

For later reference, let us review what is involved in the proof of Joyal's theorem (the proof can be found in [12], as Theorem 6.3.5). In the Heyting category $\mathrm{Set}^{\mathbf{I}}$, with $F \xrightarrow{g} G$ and a subfunctor U of F, the subfunctor $\forall_g U$ of G is given by the prescription: for any $i \in |\mathbf{I}|$ and $y \in G(i)$,

$$y \in (\forall_g U)(i) \qquad \Longleftrightarrow \qquad \begin{array}{l} \text{for all arrows } \alpha : i \to j \text{ in } \mathbf{I}, \\ \forall x \in Fj. g_j x = (G\alpha) y \;\Rightarrow\; x \in Uj. \end{array}$$

Let T be a Heyting category, $f : A \to B$ an arrow in T, $X \in Sub(A)$, $\forall_f X \in Sub(B)$. With $\mathbf{I} = \mathrm{Mod}(T)$, $g = e_T(f)$, $F = e_T(A)$, $G = e_T(B)$, and $U = e_T(X)$, to say that e_T preserves the particular $\forall_f X$ is to say that the above equivalence holds when $\forall_g U$ on the left is replaced by $e_T(\forall_f X)$. Let us write $h : M \to N$ for $\alpha : i \to j$. As usual, one direction (left to right) of the equivalence is automatic. Thus, it remains to show that for any $M \in \mathrm{Mod}(T)$, and $y \in M(B)$:

$$y \notin M(\forall_f X) \quad \Rightarrow \quad \begin{array}{l} \text{there are } h : M \to N \text{ and } x \in N(A) \text{ such that} \\ (Nf)(x) = h_A(y) \text{ and } x \notin N(X). \end{array}$$

Let us say that the pair (h, x) *witnesses* $y \notin M(\forall_f X)$ if $(Nf)(x) = h_A(y)$ and $x \notin N(X)$. In [12] it is shown, by an application of the method of diagrams in model theory, that each instance of $y \notin M(\forall_f X)$ is in fact witnessed. We will not repeat the argument here, especially since similar arguments will have to be made below. However, we make an observation on witnessing to be used later.

Let us call a map (natural transformation) $k : N \to P$ in $\mathrm{Mod}(T)$ (for any coherent T for the moment) *pure* if for any object A in T, any $X \in Sub(A)$, and any $x \in N(A)$, we have $k_A(x) \in P(X)$ (if and) only if $x \in N(X)$. Our observation is that, with the above notation, if (h, x) is a witness for $y \notin M(\forall_f X)$, then so is $(kh, k_A(x))$ provided that k is pure.

2 *The new result and other lemmas* Given the arrows $F : R \to S$, $G : R \to T$ in $\mathcal{H}eyting^*$ (or $\mathcal{B}oole^*$), we consider the pushout

$$\begin{tikzcd} S \arrow[r, "I"] \arrow[dr, phantom, "\overset{\cong}{\searrow}\,\alpha"] & P \\ R \arrow[u, "F"] \arrow[r, "G"'] & T \arrow[u, "J"'] \end{tikzcd} \qquad (1)$$

in $\mathcal{H}eyting^*$ (or $\mathcal{B}oole^*$), but also, the pushout

$$\begin{tikzcd} S \arrow[r, "\hat{I}"] \arrow[dr, phantom, "\overset{\cong}{\searrow}\,\hat{\alpha}"] & \hat{P} \\ R \arrow[u, "F"] \arrow[r, "G"'] & T \arrow[u, "\hat{J}"'] \end{tikzcd} \qquad (2)$$

in *Coherent*. (It will be essential that the universal property of $\hat{P}$ refers to arbitrary 2-cells in *Coherent*, not just isomorphisms as in $\mathcal{H}eyting^*$.) Since $\mathcal{H}eyting^*$ is a sub-2-category of *Coherent*, there is a canonical coherent functor $H : \hat{P} \to P$ such that $H\hat{I} = I$, $H\hat{J} = J$, and $H\hat{\alpha} = \alpha$ (H is unique up to isomorphism). Now, if J is to be shown to be conservative, we had better be able to show that $\hat{J}$ is conservative. This we proceed to do now.

Let us pause for a moment. We are claiming that in *Coherent*, the SAP holds for pushouts of Heyting arrows. To be sure, *Coherent* does not satisfy the SAP in general as [14] points out (the first example was given by Reyes). Thus, although we deal with a statement formulated in the coherent doctrine, we will have to use the Heyting character of the data involved.

Let $A \in T$ and $X \in Sub(A)$ with $X \neq 1_A$ (= the top element of the lattice $Sub(A)$). We want to show that $\hat{J}X \neq 1_{\hat{J}A}$. Using Gödel completeness, take any $N \in \mathrm{Mod}(T)$ such that $NX \neq NA (= N(1_A))$. We need a model $L \in \mathrm{Mod}(\hat{P})$ with $L\hat{J}X \neq L\hat{J}A$. But the models L are in an essentially 1-1 correspondence with triples

$(M', N', h' : M'|R \xrightarrow{\cong} N'|R)$ with $M' \in \mathrm{Mod}(S)$, $N' \in \mathrm{Mod}(T)$, and with the notation $M'|R = MF$, $N'|R = NG$; in particular, $L|T =_{def} LJ = N'$, and thus $L\hat{J}X \neq L\hat{J}A$ means $N'X \neq N'A$. (In the above definition of the pushout (2), taking place in $\mathcal{COHERENT}$, let Q be Set; an object of $Cocone_Q$ is an entity (M', N', h') as described; $\mathrm{Hom}(P, Q)$ is $\mathrm{Mod}(\hat{P})$; the equivalence functor Γ^* takes L to (M', N', h') where $N' = LJ$, etc.) Therefore, if we can make (M', N', h') with a *pure* map $n : N \to N'$, we are done. Actually, we first make a careful choice of N with $NX \neq NA$.

Lemma 2.1 *Assume that F is conservative. Suppose that $X \in Sub(A)$, $X \neq 1_A$ in T. Then there are $M \in \mathrm{Mod}(S)$, $N \in \mathrm{Mod}(T)$, and a map $h : M|R \to N|R$, such that $NX \neq NA$.*

Proof: The proof is postponed until Section 3. □

Proposition 2.2 (Main Lemma) *Start with the Heyting functors $F : R \to S$ and $G : R \to T$, without any assumption on conservativeness. Given $M \in \mathrm{Mod}(S)$, $N \in \mathrm{Mod}(T)$ with a map $h : M|R \to N|R$, there are $M' \in \mathrm{Mod}(S)$, $N' \in \mathrm{Mod}(T)$, and maps $m : M \to M'$, $n : N \to N'$ such that n is pure and there is a commutative diagram:*

$$\begin{tikzcd}
M'|R \arrow[r, "h'", "\cong"'] & N'|R \\
M|R \arrow[u, "m|R"] \arrow[r, "h"'] \arrow[ur, phantom, "\circlearrowleft"] & N|R \arrow[u, "n|R"']
\end{tikzcd}$$

with h' an isomorphism as indicated. Schematically:

$$\begin{array}{cc|c}
 & & \begin{array}{ccc} M' & \qquad & N' \\ \uparrow & & \text{pure}\uparrow \\ M & & N \end{array} \\
M \models S & N \models T & \\
 & & \begin{tikzcd} M'|R \arrow[r, "\cong"'] & N'|R \\ M|R \arrow[u] \arrow[r] & N|R \arrow[u] \end{tikzcd} \\
 & \Longrightarrow & \\
\multicolumn{2}{c|}{M|R \longrightarrow N|R} &
\end{array}$$

(here and below, we abbreviate, for arbitrary S and M, "$M \in \mathrm{Mod}(S)$" by "$M \models S$").

The proposition is proved by the help of the following two lemmas, expressed without words in a similar schematic manner.

Lemma 2.3

$M \models S \qquad N \models R$

$M|R \longrightarrow N$

$\Longrightarrow$

$$\begin{tikzcd} M_1 \\ M \arrow[u] \end{tikzcd}$$

$$\begin{tikzcd} M_1|R & \\ M|R \arrow[u] \arrow[r] & N \arrow[ul, "pure"'] \end{tikzcd}$$

Lemma 2.4

$M \models R \qquad N \models T$

$M \xleftarrow{pure} N|R$

$\Longrightarrow$

$$\begin{tikzcd} N_1 \\ N \arrow[u, "pure"] \end{tikzcd}$$

$$\begin{tikzcd} & N_1|R \\ M \arrow[ur] & N|R \arrow[u] \arrow[l] \end{tikzcd}$$

The proofs of the last two lemmas are postponed to Section 3.

Proof of Proposition 2.2 from Lemmas 2.3 and 2.4: For this we build the following diagrams:

$$
\begin{tikzcd}
\vdots & \vdots & \vdots & \vdots \\
M_2 & N_2 & M_2|R \arrow[r] & N_2|R \\
M_1 \arrow[u] & N_1 \arrow[u, "\text{pure}"] & M_1|R \arrow[u] \arrow[r] & N_1|R \arrow[u] \arrow[ul, "\text{pure}"'] \\
M \arrow[u] & N \arrow[u, "\text{pure}"] & M|R \arrow[u] \arrow[r] & N|R \arrow[u] \arrow[ul, "\text{pure}"']
\end{tikzcd}
$$

We start with the given M, N and $M|R \to N|R$. Using Lemma 2.3 with M and N of the statement of the lemma as the present M and $N|R$, we produce $M \to M_1$ and the lowest triangle. Then, applying Lemma 2.4 to $M_1|R$ and N as M and N, we produce $N \xrightarrow{\text{pure}} N_1$ and the next triangle, etc.

With the three infinite diagrams, consider $M' = \mathrm{colim}_{i\to\omega} M_i$, $N' = \mathrm{colim}_{i\to\omega} N_i$. We have the canonical colimit coprojections $M \to M'$, and $N \to N'$. The latter is clearly pure (since all vertical arrows in the diagram of the Ns are pure). Finally, the left to right horizontal arrows induce a map $M' \to N'$, the slanted ones induce one $N' \to M'$, and the commutativity of the infinite diagram shows that these latter two arrows are inverses of each other. The isomorphism $M' \to N'$ clearly satisfies the commutativity required in Proposition 2.2.

Let us return to our proof that $\hat{J}$ is conservative provided F is. Using Lemma 2.1 and Proposition 2.2, we do have N with $NX \neq NA$, and $(M', N', h' : M'|R \xrightarrow{\cong} N'|R)$ with a pure map $N \to N'$, which, according to what was said above, completes the proof. We have thus shown the following proposition. □

Proposition 2.5 *Pushouts in Coherent of pairs of arrows in $\mathcal{H}$eyting satisfy the SAP.*

Let us see that for the case of $\mathcal{B}oole$, this suffices for the proof of the SAP. Of course, it is a special case of Proposition 2.5 that pushouts in *Coherent* of pairs of arrows in $\mathcal{B}oole$ satisfy the SAP. Continuing with our established notation, and assuming that R, S and T are in $\mathcal{B}oole$, we claim that the canonical map $H : \hat{P} \to P$ is conservative; since the composite of conservative maps is conservative, the desired assertion will follow from Proposition 2.5. To demonstrate the claim, consider the left adjoint Φ to the forgetful functor $\Psi : \mathcal{B}oole^* \longrightarrow Coherent^*$; for a coherent category T, ΦT is the free Boolean category extension of T. Since Φ is a left adjoint, it preserves colimits.

Hence it takes a pushout diagram into a pushout diagram. Also, note that if T is itself Boolean, then $\Phi T \simeq T$. The reason is that Ψ is a full inclusion (see Theorem 1 in Section IV.3 of MacLane [10]). The latter statement just says that a coherent functor between Boolean categories is a Boolean functor, which is rather clear. Our remarks so far imply that in the Boolean case, we have an equivalence $K : P \simeq \Phi\hat{P}$, with $K \circ H = \eta_{\hat{P}}$, where $\eta_{\hat{P}}$ is the canonical unit map of the adjunction. However, $\eta_T : T \to \Phi T$ is always conservative: we have the conservative $e_T : T \to \mathrm{Set}^{|\mathrm{Mod}(T)|} =_{def} U$ into a Boolean category; the universal property of η_T implies that there is $L : \Phi T \longrightarrow U$ with $e_T \cong L \circ \eta_T$, from which it follows that η_T is conservative. Our claim follows.

Theorem 2.6 *The canonical coherent functor from the coherent pushout of a pair of Heyting functors to the Heyting pushout of the same pair is always conservative.*

This is the new result of this paper. Of course, together with Proposition 2.5, it implies the SAP for $\mathcal{H}eyting$, and thus Craig for intuitionistic logic. Note however that Theorem 2.6 does not follow (directly at least) from the SAP for $\mathcal{H}eyting$, even for pushouts in which one leg or both are conservative.

Let me point out that this is a result of "pure logic," that is, a result about the syntax of logic.

With reference to (2) in Section 1, let us say that a formula over $\mathcal{L}_1 \cup \mathcal{L}_2$ is relatively coherent if it is built up using $\wedge$, $\vee$, and $\exists$ from $\mathcal{L}_1$-formulas and $\mathcal{L}_2$-formulas. Theorem 2.6 amounts the assertion that if the sentence $\forall\vec{x}(\varphi \to \psi)$, with φ and ψ relatively coherent formulas, is provable from $\mathcal{S} \cup \mathcal{T}$ in intuitionistic logic over $\mathcal{L}_1 \cup \mathcal{L}_2$, then the coherent sequent $\varphi \Rightarrow \psi$ is provable using the rules of coherent logic only (see for example [12]) from assumption entailments each of which is provable either from $\mathcal{S}$ in intuitionistic logic over $\mathcal{L}_1$ or from $\mathcal{T}$ in intuitionistic logic over $\mathcal{L}_2$.

For the proof of Theorem 2.6, we show the following.

Proposition 2.7 *With the notation in (2), the composites:*

$$T \xrightarrow{\hat{J}} \hat{P} \xrightarrow{e_{\hat{P}}} \mathrm{Set}^{\mathrm{Mod}(\hat{P})} \quad \textit{and}$$
$$S \xrightarrow{\hat{I}} \hat{P} \xrightarrow{e_{\hat{P}}} \mathrm{Set}^{\mathrm{Mod}(\hat{P})}$$

are Heyting functors.

Note that the proposition specializes to Joyal's Theorem when both F and G are taken to be identity functors.

Proof: Note that the domain and the codomain of the composite are Heyting categories, although the factors are not necessarily Heyting functors. Looking back at the remarks on the proof of Joyal's Theorem, and using the fact that the objects of $\mathrm{Mod}(\hat{P})$ are essentially triples $(M \models S, N \models T, h : M|R \xrightarrow{\cong} N|R)$ (via the equivalence functor $\Gamma^* : \mathrm{Mod}(\hat{P}) \to \mathit{Cocone}_{set}$; see Section 1), a little thought shows that what is to be shown is this: with $f : A \to B$ and X as before, given (M, N, h) and $y \in NB$, if $y \notin N(\forall_f X)$, then there is an arrow $(m, n) : (M, N, h) \to (M', N', h')$ in Cocone_{set}, and some x such that (n, x) is a witness for $y \notin N(\forall_f X)$ in the original sense introduced above. Now, by (the proof of) Joyal's Theorem, there is a witness $(n^* : N \to N^*, x \in N^*(A))$. Apply Proposition 2.2 for M and N^* as M and N, and

$n^*|R \circ h : M|R \to N^*|R$ as the arrow h. We obtain $m : M \to M'$, $n_1 : N^* \to N'$ such that n_1 is pure, and we obtain the following commutative diagram.

$$\begin{tikzcd}
M' \mid R \arrow[r, "h'", "\cong"'] & N' \mid R \\
M \mid R \arrow[u, "m \mid R"] \arrow[r, "n^* \mid R \circ h"'] & N^* \mid R \arrow[u, "n_1 \mid R"'] \\
M \mid R \arrow[u, "1_{M|R}"] \arrow[r, "h"'] & N \mid R \arrow[u, "n^* \mid R"']
\end{tikzcd}$$

By a remark made above, since n_1 is pure, $(n_1 n^*, (n_1)_A x)$ is a witness, and thus with $n = n_1 n^*$, m, n, and $(n_1)_A(x)$, are as desired. □

Proof of Theorem 2.6: Consider the following diagram:

$$\begin{tikzcd}
 & & & \mathrm{Set}^{\mathrm{Mod}(\hat{P})} \\
S \arrow[urrr, "I^*"] \arrow[rr, "I"] \arrow[dr, "\hat{I}"'] & & P \arrow[ur, "H^*"] & \\
 & \hat{P} \arrow[ur, "H"] & & \\
R \arrow[uu, "F"] \arrow[rr, "G"'] & & T \arrow[ul, "\hat{J}"'] \arrow[uu, "J"'] \arrow[uuur, "J^*"'] &
\end{tikzcd}$$

$$e_{\hat{P}} : \hat{P} \to \mathrm{Set}^{\mathrm{Mod}(\hat{P})}$$

$$\begin{array}{ll}
I^* = e_{\hat{P}} \hat{I}, & J^* = e_{\hat{P}} \hat{J} \\
\alpha : IF \cong JG, & \alpha = H\hat{\alpha} \\
\hat{\alpha} : \hat{I}F \cong \hat{J}G & \\
\alpha^* : I^*F \cong J^*G, & \alpha^* = e_{\hat{P}}\hat{\alpha}.
\end{array}$$

Here, after the familiar ingredients, I^*, J^*, and α^* are defined as shown ($e_{\hat{P}}$ is not drawn). Since by Proposition 2.7 I^* and J^* are Heyting functors, by the universal property of P, we have H^* with $IH^* = I^*$, $JH^* = J^*$, and $\alpha H^* = \alpha^*$. Hence, $\hat{I}H^*H = I^*$, $\hat{J}H^*H = J^*$, and $\hat{\alpha}H^*H = \alpha^*$. By the uniqueness part of the universal property of $\hat{P}$, it follows that $H^*H \cong e_{\hat{P}}$. But, by Gödel, $e_{\hat{P}}$ is conservative; hence, so is H. □

3 Model-theoretical arguments The categorical language is superior for the purposes of expressing the important things (stating results, for instance), but it is not convenient for stating certain arguments, e.g., the model-theoretical ones involving the "method of diagrams." In our view (which reverses the historical order), symbolic logic is the auxiliary language, to be introduced to deal in convenient ways with the concepts primarily given in the categorical framework.

With any coherent category T, we associate a particular coherent theory $(\mathcal{L}_T, \underline{T})$, also written simply as $\underline{T}$, the *internal theory* of T (in [12], T_T is written for $(\mathcal{L}_T, \underline{T})$; see Chapter 3, Section 5, page 128). Here, $\mathcal{L}_T$ has sorts the objects of T, unary sorted operation symbols are the arrows of T. The models of $(\mathcal{L}_T, \underline{T})$ are identical to the models of T (objects of $\mathrm{Mod}(T)$). The (universal algebraic) homomorphisms of models are the same as the arrows in $\mathrm{Mod}(T)$ (see Theorem 3.5.3. of [12]).

For any coherent formula $\varphi(\vec{x})$ with free variables among $\vec{x}$, we have the interpretation of φ, $[\vec{x} \mid \varphi] \in Sub([\vec{x}])$, where $[\vec{x}]$ is the product of the objects that are the respective sorts of the variables in $\vec{x}$ (in [12], $[\varphi]$ is written for $[\vec{x} \mid \varphi]$). Given any subobject $X \in Sub(A)$, there is a canonically selected coherent formula $\underline{X}(x)$ such that $[x \in A \mid \underline{X}(x)] = X$ (here x is a variable of sort A; "$x \in A$" is just a reminder of this). If $X \in Sub(A_1 \times \ldots \times A_n)$, we have $\underline{X}(x_1, \ldots, x_n)(= \underline{X}(\vec{x}))$ such that

$$[\vec{x} \in \prod_{i=1}^{n} X_i \mid \underline{X}(\vec{x})] = X.$$

Let us introduce some auxiliary notation. With $F : R \to S$ a coherent functor, if $\varphi(\vec{x})$ is any formula over $\mathcal{L}_R$ (e.g., in full first-order logic, or even in infinitary logic), $\varphi^F(\vec{x})$ denotes its F-translate, that is, the result of substituting FA for A, and Ff for f, for any $f : A \to B$ in R; the sorting of the variables, both free and bound, is redone accordingly. Thus, $\varphi^F(\vec{x})$ is a formula over $\mathcal{L}_S$. If, in particular, φ is a coherent formula, then $F[\vec{x} \mid \varphi] = [\vec{x} \mid \varphi^F]$; this is precisely to say that the functor F is coherent. With a little abuse of language, we also write φ^S for $\varphi^F (F : R \to S)$.

With $M \in \mathrm{Mod}(R)$, $Diag^+(M)$ denotes the *positive diagram* of M, that is, the set of all coherent (positive existential) sentences, over the language $\mathcal{L}_R(M)$ containing an individual constant $\underline{a}$ of sort A for each pair $(A \in T, a \in M(A))$, that are true in M when $\underline{a}$ is interpreted as a. As is well known, and immediately seen, the arrows (homomorphisms) out of M are in an essentially one-to-one correspondence with the models of $Diag^+(M)$. With $G : R \to T$, say, when writing $Diag^+(M)^T$, we mean the set of G-translates of all members of $Diag^+(M)$, with the new individual constants left alone except that their sortings are redone in the obvious way.

The negative diagram $Diag^-(M)$ is defined similarly, with negated existential positive sentences. Models of $Diag^+(M) \cup Diag^-(M)$ correspond to embeddings, that is, pure maps, out of M.

In this section, we use $\models$ in the sense of semantical consequence in ordinary *classical* predicate logic (many-sorted, with possibly empty sorts).

Proof of Lemma 2.1: Let Σ denote the set of all coherent (positive existential) sentences σ over $\mathcal{L}_R$ such that $\neg\sigma^T$ is a consequence of $\underline{T} \cup \{\neg\forall x \underline{X}(x)\}$. We claim that $\underline{S} \cup \{\neg\sigma^S : \sigma \in \Sigma\}$ is consistent. Otherwise, from the fact that Σ is closed under disjunction, we obtain $\sigma \in \Sigma$ such that $\underline{S} \models \sigma^S$. Since F is conservative, $\underline{R} \models \sigma$; hence,

$\underline{T} \models \sigma^T$. But since $\underline{T} \cup \{\neg\forall x \underline{X}(x)\} \models \sigma^T$ (by the definition of Σ), it follows that $\underline{T} \models \forall x \underline{X}(x)$, contrary to the hypothesis that $X \neq 1_A$.

Let M be any model of $\underline{S} \cup \{\neg\sigma^S : \sigma \in \Sigma\}$; in particular, $M \in \mathrm{Mod}(S)$. We claim that $Diag(M|R)^T \cup \underline{T} \cup \{\neg\forall x \underline{X}(x)\}$ is consistent; once proved, the claim establishes the lemma. If the claim fails, there is $\varphi(\underline{\vec{a}}) \in Diag(M|R)$ with $\underline{T} \cup \{\neg\forall x \underline{X}(x)\} \models \neg\varphi(\underline{\vec{a}})$. Hence, $\exists\vec{x}\varphi \in \Sigma$. Since $M \models (\exists\vec{x}\varphi)^S$, this is a contradiction to the assumption on M. $\square$

Before turning to the proofs of Lemmas 2.3 and 2.4, we discuss how the fact that R is a Heyting category is reflected on the "coherent logic" of R.

First of all, with $A \in R$, and $X, Y \in Sub(A)$, we may consider $\forall_y(X) \in Sub(Y)$ where $y : Y \to A$ is the structure map for $Y \in Sub(A)$. With z the structure map for $\forall_y(X) \in Sub(Y)$, we let $X \to Y$ be the subobject of A represented by the monomorphism $y \circ x$. It is easy to see that $X \to Y$ is the "Heyting implication" (relative pseudo-complement) in the sense that for any $Z \in Sub(A)$, $Z \wedge X \leq Y$ iff $Z \leq X \to Y$. Further, if $f : A \to B$ and $X, Y \in Sub(A)$, then $\forall_f(X \to Y) \in Sub(B)$ satisfies the following condition: for any $Z \in Sub(B)$,

$$Z \leq \forall_f(X \to Y) \iff f^*(Z) \wedge X \leq Y.$$

This is easy to see.

Now, suppose we have the coherent formulas $\varphi(\vec{x}, \vec{y})$, $\psi(\vec{x}, \vec{y})$ over the language $\mathcal{L}_R$. Let $A = [\vec{x}]$ and $B = [\vec{y}]$, thus $A \times B = [\vec{x}, \vec{y}]$. Let $\pi : A \times B \to A$ be the projection. And let $X = [\vec{x}, \vec{y} \mid \varphi]$ and $Y = [\vec{x}, \vec{y} \mid \psi]$. We *define* the subobject $[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]$ to be $\forall_\pi(X \to Y) \in Sub(A)$; note that $\forall\vec{y}(\varphi \to \psi)$ is not a coherent formula, thus the expression $[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]$ is not defined in the general context of a coherent R; $[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]$ is the interpretation in R of the formula $\forall\vec{y}(\varphi \to \psi)$ taken in intuitionistic logic. If there were no difference between classical and intuitionistic logic, we would have:

$$? : \underline{R} \models \forall\vec{x}(\underline{[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]}(\vec{x}) \longleftrightarrow \forall\vec{y}(\varphi(\vec{x}, \vec{y}) \to \psi(\vec{x}, \vec{y}))).$$

The left to right implication and a suitable weakening of the other implication are true:

$$\underline{R} \models \underline{[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]}(\vec{x}) \longrightarrow \forall\vec{y}(\varphi(\vec{x}, \vec{y}) \to \psi(\vec{x}, \vec{y})); \tag{1}$$

and for any coherent $\mathcal{L}_R$-formula $\theta(\vec{x})$,

$$\underline{R} \cup \{\theta(\vec{x})\} \models \forall\vec{y}(\varphi(\vec{x}, \vec{y})) \to \psi(\vec{x}, \vec{y})) \implies \underline{R} \cup \{\theta(\vec{x})\} \models \underline{[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]}(\vec{x}) \tag{2}$$

(in the latter, $\vec{x}$ is treated as a tuple of constants, in the natural way). These assertions are immediately seen on the basis of the definition of $[\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]$.

Furthermore, if $F : R \to S$ is Heyting, then, clearly,

$$F([\vec{x} \mid \forall\vec{y}(\varphi \to \psi)]) = [\vec{x} \mid \forall\vec{y}(\varphi \to \psi)^S].$$

Proof of Lemma 2.3: Let f be the given arrow $M|R \to N$. In forming the diagrams below, we make sure that the set of constants used for elements of M is disjoint from the set of constants used for elements of N.

We modify $Diag^+(N)$ to $Diag^+(N)^*$ by adding all axioms of the form $\underline{a} =_A \underline{b}$ whenever $A \in R$, $a \in (M|R)(A) = (MF)(A)$ and $f_A(a) = b$, and closing the result under conjunction. The assertion of the lemma is equivalent to saying that

$$\underline{S} \ \cup \ Diag^+(M) \ \cup \ (Diag^+(N)^*)^S \ \cup \ (Diag^-(N))^S$$

is consistent. Assume the contrary. Then there is a finite subset Φ of the displayed set which is inconsistent. Let $\underline{\vec{a}}$ be the tuple of M-constants of sorts FA with $A \in R$, $\underline{\vec{c}}$ the rest of the M-constants, and $\underline{\vec{b}}$ the N-constants occurring in Φ. Taking conjunctions, there are $\theta(\underline{\vec{a}}, \underline{\vec{c}}) \in Diag^+(M)$, $\varphi(\underline{\vec{a}}, \underline{\vec{b}}) \in Diag^+(N)^*$, and $\neg\psi(\underline{\vec{b}}) \in Diag^-(N)$ such that

$$\underline{S} \ \models \ \theta(\underline{\vec{a}}, \underline{\vec{c}}) \wedge \varphi^S(\underline{\vec{a}}, \underline{\vec{b}}) \to \psi^S(\underline{\vec{b}}),$$

that is,

$$\underline{S} \cup \{\theta(\underline{\vec{a}}, \underline{\vec{c}})\} \ \models \ \forall \vec{y}(\varphi^S(\underline{\vec{a}}, \vec{y}) \to \psi^S(\vec{y})).$$

Thus, by (2),

$$\underline{S} \cup \{\theta(\underline{\vec{a}}, \underline{\vec{c}})\} \ \models \underline{[\vec{x} \mid \ \forall \vec{y}(\varphi^S(\vec{x}, \vec{y}) \to \psi^S(\vec{y}))]}(\underline{\vec{a}}),$$

which, by $M \models S \cup \{\theta(\underline{\vec{a}}, \underline{\vec{c}})\}$, implies that

$$M \models \underline{[\vec{x} \mid \ \forall \vec{y}(\varphi^S(\vec{x}, \vec{y}) \to \psi^S(\vec{y}))]}(\underline{\vec{a}}).$$

Since $F : R \to S$ is a Heyting functor,

$$M|R \ \models \ \underline{[\vec{x} \mid \ \forall \vec{y}(\varphi(\vec{x}, \vec{y}) \to \psi(\vec{y}))]}(\underline{\vec{a}}),$$

and since the homomorphism f preserves the meaning of coherent formulas,

$$N \ \models \ \underline{[\vec{x} \mid \ \forall \vec{y}(\varphi(\vec{x}, \vec{y}) \to \psi(\vec{y}))]}(\underline{f\vec{a}}). \tag{3}$$

However, $N|R \ \models \ \varphi(\underline{f\vec{a}}, \underline{\vec{b}})$, and $N|R \ \models \ \neg\psi(\underline{\vec{b}})$, hence

$$N \ \models \ \neg\forall \vec{y}(\varphi(\underline{f\vec{a}}, \vec{y}) \to \psi(\vec{y})). \tag{4}$$

(1), (3) and (4) give a contradiction. $\square$

Proof of Lemma 2.4: This is easier; it does not need the Heyting character of the data involved. Let $f : N|R \to M$ be the given map. It suffices to prove that

$$\underline{T} \ \cup \ Diag^+(N) \ \cup \ Diag^-(N) \ \cup \ (Diag^+(M)^*)^T$$

is consistent (here, as before, $Diag^+(M)^*$ is obtained by adding to $Diag^+(M)$ the sentences of the form $\underline{b} =_A \underline{a}$ whenever $A \in R$ and $f_A(b) = a$). Otherwise, there is a finite subset Φ of the displayed set which is inconsistent. Let $\underline{\vec{a}}$ be the tuple of M-constants, $\underline{\vec{b}}$ the N-constants of sorts GA for $A \in R$, and $\underline{\vec{c}}$ the rest of the N-constants occurring in Φ. Taking conjunctions, we have $\theta(\underline{\vec{b}}, \underline{\vec{c}}) \in Diag^+(N)$, $\neg\psi(\underline{\vec{b}}, \underline{\vec{c}}) \in Diag^-(N)$, and $\varphi(\underline{\vec{a}}, \underline{\vec{b}}) \in Diag^+(M)^*$ such that

$$\underline{T} \ \cup \ \{\varphi^T(\underline{\vec{a}}, \underline{\vec{b}})\} \ \models \ \theta(\underline{\vec{b}}, \underline{\vec{c}}) \ \to \ \psi(\underline{\vec{b}}, \underline{\vec{c}}).$$

And as a consequence,

$$\underline{T} \ \cup \ \{\exists\vec{x}\varphi^T(\vec{x}, \vec{\underline{b}})\} \ \models \ \theta(\vec{\underline{b}}, \vec{\underline{c}}) \ \rightarrow \ \psi(\vec{\underline{b}}, \vec{\underline{c}}).$$

Now, since $f : N|R \rightarrow M$ is pure and $M \models \exists\vec{x}\varphi(\vec{x}, f\vec{\underline{b}})$, we have that $N|R \models \exists\vec{x}\varphi(\vec{x}, \vec{\underline{b}})$, that is, $N \models \exists\vec{x}\varphi^T(\vec{x}, \vec{\underline{b}})$. Since also $N \models \underline{T}$, we conclude that $N \models \theta(\vec{\underline{b}}, \vec{\underline{c}}) \rightarrow \psi(\vec{\underline{b}}, \vec{\underline{c}})$. However, this contradicts the choice of the formulas θ, ψ. □

Let us make some remarks comparing our procedures with those of [5]. In that paper, a version of Robinson's Consistency Lemma is proved, and Craig's Theorem is derived from it. The proof of "Robinson's Lemma" is given, essentially, in one piece; no (real) lemmas are formulated; in particular, no statement comparable to Theorem 2.6 is given. Nevertheless, most of the ingredients of our proofs appear in Gabbay's proof. Instead of *models* (in our sense) of T, "complete and saturated extensions of T" appear; this is indeed an equivalent concept. Lemma 2.1, Proposition 2.2 and Proposition 2.5 do not appear in a recognizable form, but Lemmas 2.3 and 2.4 do appear in several places (as essentially the same two arguments repeated). Finally, there is a construction of a tree-based Kripke model, with a tree whose nodes are various tuples consisting of formulas, etc.; this is avoided in our treatment by the use of a Joyal-type canonical Kripke model.

Acknowledgments This paper was supported by NSERC Canada and FCAR Quebec.

REFERENCES

[1] Beth, E., "On Padoa's method in the theory of definition," *Indagationes Mathematicæ*, vol. 15 (1953), pp. 330–339. Zbl 0053.34402 MR 15,385e 1

[2] Bell, J., and M. Machover, *A Course in Mathematical Logic*, North-Holland, Amsterdam, 1977. Zbl 0359.02001 MR 57:12155 1

[3] Chang, C., and H. Keisler, *Model Theory*, North-Holland, Amsterdam, 1973. Zbl 0276.02032 MR 53:12927 1

[4] Craig, W., "Linear reasoning. A new form of the Herbrand-Gentzen theorem," *The Journal of Symbolic Logic*, vol. 22 (1957), pp. 250–268. Zbl 0081.24402 MR 21:3317 1

[5] Gabbay D. M., "Semantic proof of Craig's interpolation theorem for intuitionistic logic and extensions, Part I," pp. 391–402 in *Logic Colloquium '69*, edited by R. O. Gandy and C. M. E. Yates, North-Holland, Amsterdam, 1969. MR 43:3094b 1, 3

[6] Gabbay, D. M., *Semantical Investigations in Heyting's Intuitionistic Logic*, Reidel, Dordrecht, 1981. Zbl 0453.03001 MR 83b:03012 1

[7] Gabriel, P., and F. Ulmer, "Lokal prasentierbare Kategorien," *Lecture Notes in Mathematics*, vol. 221, Springer-Verlag, Berlin, 1971. MR 48:6205 1

[8] Joyal, A., and M. Tierney, "An extension of the Galois theory of Grothendieck," *Memoirs of the American Mathematical Society*, vol. 51, American Mathematical Society, Providence, 1984. Zbl 0541.18002 MR 86d:18002 1

[9] Kripke, S., "Semantical analysis of intuitionistic logic I," pp. 92–130 in *Formal Systems and Recursion Theory*, edited by S. N. Crossley and M. A. E. Dummett, North-Holland, Amsterdam, 1965. Zbl 0137.00702 MR 34:1184 1, 1

[10] MacLane, S., *Categories for the Working Mathematician*, Springer-Verlag, Heidelberg, 1971. Zbl 0705.18001 MR 50:7275 2

[11] Maehara, S., "On the interpolation theorem of Craig," (in Japanese) *Sugaku*, vol. 12 (1960/61), pp. 235–237. Zbl 0123.24601 MR 29:3356 1

[12] Makkai, M., and G. E. Reyes, *First Order Categorical Logic*, Lecture Notes in Mathematics, vol. 611, Springer-Verlag, Heidelberg, 1977. Zbl 0357.18002 MR 58:21600 1, 1, 1, 1, 1, 1, 1, 1, 2, 3, 3, 3

[13] Makkai, M., "Ultraproducts in categorical logic," *Lecture Notes in Mathematics*, vol. 1130 (1985), pp. 222–309. 1

[14] Pitts, A., "An application of open maps to categorical logic," *Journal of Pure and Applied Algebra*, vol. 29 (1983), pp. 313–326. Zbl 0521.03051 MR 85c:03025 1, 1, 1, 1, 2

[15] Robinson, A., "A result on consistency and its application to the theory of definition," *Indagationes Mathematicæ*, vol. 18 (1956), pp. 47–58. Zbl 0075.00701 MR 17,1172d 1

[16] Schütte, K., "Der Interpolationsatz der intuitionistischen Pradikatenlogik," *Mathematische Annalen*, vol. 148 (1962), pp. 192–200. 1

[17] Street, R., "Fibrations in bicategories," *Cahiers Topologie et Geometrie Differentielle*, vol. 21 (1980), pp. 111–160. Zbl 0436.18005 MR 81f:18028 1

Department of Mathematics and Statistics
McGill University
805 Sherbrooke Street West
Montreal, Quebec H3A 2K6
Canada
email: makkai@triples.math.mcgill.ca