Notre Dame Journal of Formal Logic
Volume VIII, Number 3, July 1967

THE CRAIG INTERPOLATION LEMMA

BURTON DREBEN and HILARY PUTNAM

Let A and B be schemata of quantification theory. Assume that A and B are neither valid nor inconsistent and that A implies B. *The Craig Interpolation Lemma* asserts that in this case there exists a schema M *containing only predicate letters common to A and B* such that A implies M and M implies B. The purpose of the present note is to present a short and elementary proof. Familiarity with Quine's *Methods of Logic* (revised edition, including the appendix) is presumed on the part of the reader.

1. *The Interpolation Lemma for propositional calculus.* It is well known that the analogous "interpolation lemma" for propositional calculus is trivial. The following elegant demonstration is due to Kreisel:

Let A and B be schemata of propositional calculus such that A implies B. Let $A = A(p_1, \ldots, p_n)$. Suppose some p_i, say p_1, does not occur in B. Then $A(p_1, p_2, \ldots, p_n) \supset B$ is valid; hence $A(\top, p_2, \ldots, p_n) \supset B$ is valid and $A(\bot, p_2, \ldots, p_n) \supset B$ is valid.

But $A(p_1, p_2, \ldots, p_n) \supset .A(\top, p_2, \ldots, p_n) \vee A(\bot, p_2, \ldots, p_n)$ is valid. Hence $A(p_1, p_2, \ldots, p_n)$ implies $A(\top, p_2, \ldots, p_n) \vee A(\bot, p_2, \ldots, p_n)$ which implies B. The "interpolation formula" $A(\top, p_2, \ldots, p_n) \vee A(\bot, p_2, \ldots, p_n)$ may reduce to just $\top$, but in this case B, being implied by $\top$, must be a tautology. Also, the "interpolation formula" may reduce to just $\bot$, but in this case A, implying $\bot$, must be inconsistent. So, if A and B are neither tautologous nor inconsistent, the "interpolation formula" must reduce to neither $\top$ nor $\bot$, but to a well formed formula of propositional calculus in the usual sense.[1] One of the remaining letters, $p_2, p_3, \ldots, p_n$ may fail to occur in B, but repetition of the method must eventually lead to an "interpolation formula" containing only such p_i as occur in B; hence, p_i common to both A and B.

2. *The Interpolation Lemma for quantification theory.* Assume A implies B and that A and B are neither valid nor inconsistent. Let A be in prenex normal form, and let $\overline{B}$ be a prenex equivalent of $-B$. Since A implies B, $A.-B$ is inconsistent, and hence the set consisting of the two schemata A and $\overline{B}$ is inconsistent. But then, by the appendix of *Methods*, there is a deduction, say,

Received March 15, 1967

$$\begin{array}{lcc} * & (1) & A \\ ** & (2) & \overline{B} \\ ** & & . \\ ** & & . \\ ** & & . \\ ** & & . \end{array}$$

Figure 1

with the properties:

(i) Every line except (1) and (2) comes from a preceding line by either UI or EI.

(ii) The conjunction of all the quantifier-free lines in the deduction is truth-functionally inconsistent.

We shall assume, without loss of generality, that Quine's "rule of thumb" (see *Methods*, p. 164) is conformed to in Fig. 1, i.e., each variable "flagging" a given line is alphabetically later than all the free variables in the line to which the given line is subjoined. (This is equivalent to saying that the "flagged" variable is alphabetically later than all the *other* free variables in the line it flags.) And, of course, no variable is "flagged" more than once; otherwise Fig. 1 would not be a deduction.

Let C_A be the set of all and only the quantifier-free lines in Fig. 1 which come ultimately from A (this is relative to some particular analysis of the deduction shown in Fig. 1, of course, but we shall assume a fixed analysis for the purposes of our argument), and let $C_{\overline{B}}$ be the set of all and only the quantifier-free lines which come ultimately from $\overline{B}$. (Every quantifier-free line in Fig. 1 belongs to either C_A or $C_{\overline{B}}$ and not to both, since the rules UI and EI are the only rules used in Fig. 1, and both rules are one-premise rules.) Let the conjunction of the formulas in C_A be F_A and let the conjunction of the formulas in $C_{\overline{B}}$ be $F_{\overline{B}}$. Since $C_A \cup C_{\overline{B}}$ is the set of all quantifier-free lines in Fig. 1, and this is truth functionally inconsistent, the conjunction $F_A . F_{\overline{B}}$ must also be inconsistent, i.e., F_A implies $-F_{\overline{B}}$ truth-functionally.

By the Interpolation Lemma for propositional calculus, there exists a schema M containing only truth-functional constituents common to F_A and $-F_{\overline{B}}$ such that F_A implies M which implies $-F_{\overline{B}}$. The schema M can contain only predicate letters common to F_A and $-F_{\overline{B}}$, and hence only predicate letters common to A and to B (or $\overline{B}$). Moreover, M can contain only individual variables common to F_A and $-F_{\overline{B}}$, and hence common to C_A and $C_{\overline{B}}$.

Let (Q) be the following string of quantifiers: $(Q_1 v_1)(Q_2 v_2)\ldots(Q_n v_n)$, where $v_1, v_2, \ldots, v_n$ are, in alphabetical order, all the variables common to F_A and $-F_{\overline{B}}$, and Q_i is a universal quantifier (an existential quantifier) if v_i was never introduced (was introduced once) by EI in the course of deriving some line belonging to C_A. We claim:

$(Q)M$ is the desired interpolation formula, i.e., *A implies $(Q)M$ which implies B.*

Proof that A implies $(Q)M$. By just deleting $\overline{B}$ and all steps which come ultimately from $\overline{B}$ in Fig. 1 we obtain a deduction:

*	(1)	A
*	(2)	—
*	(3)	—
*		.
*		.
*		.
*	(k)	—

Figure 2

whose quantifier-free lines are exactly C_A. In this deduction the same variables occuring in C_A are flagged as in Fig. 1, and the "rule of thumb" is, of course, still conformed to.

Since F_A implies M truth-functionally, we can obtain M in one TF step.

But then by n uses of UG and EG we can obtain $(Q)M$ as shown in Fig. 3:

*	(1)	A	
*	(2)	—	
*	(3)	—	
*		.	
*		.	
*		.	
*	(k)	—	
*	(k+1)	M	(1),(2),...,(k), TF
*		.	
*		.	(UG and EG steps)
*		.	
*	(k+n+1)	$(Q)M$	

Figure 3

That the deduction shown in Fig. 3 conforms to Quine's rules may be checked by observing that whenever a variable v_i is universally generalized in Fig. 3 it was *not* flagged in Fig. 2, by the definition of the string of quantifiers (Q), and so no variable is flagged twice in Fig. 3. Also, the free variables in the line flagged by v_i are just $v_1, v_2, \ldots, v_{i-1}$ because the quantifiers (Q) have to be put on in the order $(Q_n v_n)$, $(Q_{n-1} v_{n-1})$, $(Q_{n-2} v_{n-2}), \ldots$ in Fig. 3, if they are to end up in the order $(Q_1 v_1)(Q_2 v_2)\ldots(Q_n v_n)$ and $v_1, v_2, \ldots, v_{i-1}$ are all alphabetically earlier than v_i. So the "rule of thumb" is conformed to. Thus Fig. 3 exhibits a finished deduction of $(Q)M$ from A, and hence, by the soundness of Quine's system, A implies $(Q)M$.

Proof that $(Q)M$ *implies* B. By just deleting A and all steps which come ultimately from A in Fig. 1 we obtain a deduction:

*	(1)	$\overline{B}$
*	(2)	__
*		.
*		.
*		.
*	(j)	__

Figure 4

whose quantifier-free lines are exactly $C_{\overline{B}}$.

If we now adjoin $(Q)M$ as a new premise, we can deduce M by UI and EI in n steps as shown in Fig. 5.

*	(1)	$\overline{B}$	
*	(2)	__	
*	(3)	__	
*		.	
*		.	
*		.	
*	(j)	__	
**	(j+1)	$(Q)M$	
**		.	
**		.	(UI and EI steps)
**		.	
**	(j+n+1)	M	

Figure 5

That the deduction shown in Fig. 5 conforms to Quine's rules may be checked by observing that whenever a variable v_i is existentially instantiated in Fig. 5 it was flagged in Fig. 2, by the definition of the string of quantifiers (Q), and hence *not* flagged in Fig. 4 (since Fig. 2 and Fig. 4 can have no line in common). So no variable is flagged twice in Fig. 5. Also, the free variables other than v_i in the line flagged by v_i are just $v_1, v_2, \ldots, v_{i-1}$ and these are all alphabetically earlier than v_i. So the "rule of thumb" is conformed to. Thus Fig. 5 exhibits a deduction, though not a finished one, with premisses $\overline{B}$ and $(Q)M$.

Since all the lines in $C_{\overline{B}}$ occur in Fig. 5, and M implies $-F_{\overline{B}}$, Fig. 5 contains a truth-functionally inconsistent set of quantifier-free lines, namely $C_{\overline{B}} \cup \{M\}$. Hence the deduction shown in Fig. 5 can be finished by simply adjoining B, by a single TF step:

**	(j+n+2)	B	(1), (2), . . . , (j+n+1), TF

By the soundness of Quine's system, $\overline{B}$ and $(Q)M$ together imply B and hence $(Q)M$ implies B, q.e.d.

If we obtain the "interpolation formula" M by Kreisel's method, it may happen that M reduces to the truth-value $\top$, but in this case B, being implied by $(Q)\top$, has to be a valid formula; and similarly if M reduces to the truth-value $\bot$, then A, implying $(Q)\bot$, has to be a contradictory formula. So if A and B are neither valid nor contradictory M does not reduce to $\top$ or $\bot$, i.e., M is a well formed formula in the usual sense.[1]

We are grateful to W. V. Quine for helpful comments.

FOOTNOTE

[1] *Methods of Logic* (1959) gives a method for eliminating $\top$ and $\bot$ from any formula of the propositional calculus which is neither valid nor contradictory.

Harvard University
Cambridge, Massachusetts