The Craig Interpolation Theorem

Hans Halvorson

May 1, 2013

This week we will pick up a couple of topics from the more advanced reaches of logic. The topics we will deal with have to do with whether a theory T has the resources to define a certain concept.

For example, we saw last week that $\mathrm{Th}(\mathbb{N})$ cannot define its own truth predicate. That is, there is no formula $\phi(x)$ of L^* such that $\mathbb{N} \models \phi(\ulcorner \psi \urcorner)$ if and only if $\mathbb{N} \models \psi$.

Theorem 1 (Craig Interpolation)**.** *If $\phi \models \psi$, then there is a formula θ such that:*

1. *All non-logical symbols in θ occur in ϕ or ψ;*
2. *$\phi \models \theta$ and $\theta \models \psi$.*

Proof. Suppose that there is no such θ. We will show then that $\{\phi, \neg\psi\}$ is consistent. We will do so by using the concept of inseparable theories. Given theories T_1 in L_1 and T_2 in L_2, we say that T_1 and T_2 are *separable* just in case there is a sentence θ of $L_1 \cap L_2$ such that $T_1 \models \theta$ and $T_2 \models \neg\theta$. If there is no such sentence, then we say that T_1 and T_2 are *inseparable.*

Let L_1 consist of all non-logical vocabulary in ϕ, let L_2 consist of all non-logical vocabulary in ψ. Let $L = L_1 \cup L_2$ and let $L_0 = L_1 \cap L_2$. Let C be a countably infinite set of constant symbols that do not occur in ϕ or ψ, and let $L_i' = L_i \cup C$.

We claim first that $\{\phi\}$ and $\{\neg\psi\}$ are inseparable as theories in L_1' and L_2'. Suppose for reductio ad absurdum that $\phi \models \theta(c)$ and $\psi \models \neg\theta(c)$. Since c doesn't occur in ϕ, $\phi \models \forall x\theta(x)$, and $\psi \models \neg\forall x\theta(x)$. But $\forall x\theta(x)$ is a sentence in L_0, and we supposed already that there is no such sentence. Therefore $\{\phi\}$ and $\{\neg\psi\}$ are inseparable.

Let $\phi_0, \phi_1, \phi_2, \ldots$ be an enumeration of the sentences in L_1', and let $\psi_0, \psi_1, \psi_2, \ldots$ be an enumeration of the sentences in L_2'. We inductively build two families T_i and U_i of theories that have the following features:

1. $T_0 = \{\phi\}$, $U_0 = \{\neg\psi\}$.
2. If $T_m \cup \{\phi_m\}$ and U_m are inseparable, then $\phi_m \in T_{m+1}$.
3. If $U_m \cup \{\psi_m\}$ and T_{m+1} are inseparable, then $\psi_m \in U_{m+1}$.
4. If $\phi_m = \exists x\sigma(x)$, and $\phi_m \in T_{m+1}$, then $\sigma(c) \in T_{m+1}$ for a new constant symbol c.
5. If $\psi_m = \exists x\sigma(x)$, and $\psi_m \in U_{m+1}$, then $\sigma(c)$ in U_{m+1} for a new constant symbol d.

Let $T_\omega = \bigcup_{i\leq\omega} T_i$ and let $U_\omega = \bigcup_{i\leq\omega} U_i$. Then the compactness theorem entails that T_ω and U_ω are inseparable. We claim now that T_ω is a maximally consistent theory in L_1', U_ω is a maximally consistent theory in L_2', and $T_\omega \cap U_\omega$ is a maximally consistent theory in L_0'. (In fact, we show that these theories have the "C properties" in their respective languages.)

We first show that T_ω is maximally consistent. First, T_ω is clearly consistent, for if it weren't, then it would imply a contradiction, and so would be separable from U_ω. To show that T_ω is maximal consistent, we need to show that for every sentence ϕ_m of L_1', either $\phi_m \in T_\omega$ or $\neg\phi_m \in T_\omega$. Suppose for reductio ad absurdum that neither ϕ_m nor $\neg\phi_m$ is in T_ω. Note that $\neg\phi_m = \phi_n$ for some n. We assume without loss of generality that $n > m$. Since ϕ_m is not in T_ω, it follows that $T_\omega \cup \{\phi_m\}$ and U_ω are separated by some θ in L_0'. That is,

$$T_\omega \models \phi_m \to \theta, \qquad\qquad U_\omega \models \neg\theta.$$

Similarly, $T_\omega \cup \{\phi_n\}$ and U_ω are separated by some θ' in L_0'. That is,

$$T_\omega \models \phi_n \to \theta', \qquad\qquad U_\omega \models \neg\theta'.$$

Since $\phi_m \vee \phi_n$ is a tautology, $T_\omega \models \theta \vee \theta'$, and $U_\omega \models \neg(\theta \vee \theta')$. Thus, T_ω and U_ω are separable, a contradiction. Thus, T_ω is maximal consistent. An analogous argument shows that U_ω is maximal consistent.

Now we observe that $T_\omega \cap U_\omega$ is maximal consistent in L_0'. Given any formula σ of L_0', either $T_\omega \models \sigma$ or $T_\omega \models \neg\sigma$, and similarly for U_ω. But since T_ω and U_ω are inseparable, they agree on σ, i.e. either both imply σ, or both imply $\neg\sigma$. Thus, $T_\omega \cap U_\omega \models \sigma$ or $T_\omega \cap U_\omega \models \neg\sigma$. Therefore $T_\omega \cap U_\omega$ is maximal consistent.

Now we show that $T_\omega \cup U_\omega$ is satisfiable, i.e. we produce a model. First, using the Henkin construction, T_ω has a model M whose domain consists of equivalence classes of constant symbols in C. Similarly, U_ω has a model N whose domain consists of equivalence classes of constant symbols in C.

We claim now that there is an isomorphism $j : N|_{L_0'} \to M|_{L_0'}$. Indeed, both $N|_{L_0'}$ and $M|_{L_0'}$ are models of $T_\omega \cap U_\omega$, which is maximal consistent in L_0'.

Now the isomorphism j induces a model of T_ω on M. Most importantly, this model agrees with M on what it assigns to elements of L_0'. So, putting the pieces together, we have a model M of $T_\omega \cup U_\omega$. Since $\phi \in T_\omega$ and $\neg\psi \in U_\omega$, it follows that $\phi \not\models \psi$. $\square$