Introduction to Metalogic

Hans Halvorson

September 21, 2016

Logical grammar

Definition. A **propositional signature** Σ is a collection of items, which we call **propositional constants**. Sometimes these propositional constants are also called **elementary sentences**, or even **atomic sentences**. However, we will be using the word "atomic" for a different concept.

These propositional constants are assumed to have no independent meaning. Nonetheless, we assume a primitive notion of identity between propositional constants; the fact that two propositional constants are equal or non-equal is not explained by any more fundamental fact. This assumption is tantamount to saying that Σ is a **bare set** (and it stands in gross violation of Leibniz's principle of the identity of indiscernibles).

Assumption. The **logical vocabulary** consists of the symbols $\neg, \wedge, \vee, \rightarrow$. We also use two further symbols for punctuation: a left and a right parenthesis.

Definition. Given a propositional signature Σ, we define the set $\mathsf{Sent}(\Sigma)$ of Σ-sentences as follows:

1. If $\phi \in \Sigma$ then $\phi \in \mathsf{Sent}(\Sigma)$.
2. If $\phi \in \mathsf{Sent}(\Sigma)$ then $(\neg\phi) \in \mathsf{Sent}(\Sigma)$.
3. If $\phi \in \mathsf{Sent}(\Sigma)$ and $\psi \in \mathsf{Sent}(\Sigma)$, then $(\phi\wedge\psi) \in \mathsf{Sent}(\Sigma)$, $(\phi\vee\psi) \in \mathsf{Sent}(\Sigma)$, and $(\phi \rightarrow \psi) \in \mathsf{Sent}(\Sigma)$.
4. Nothing is in $\mathsf{Sent}(\Sigma)$ unless it enters via one of the previous clauses.

The symbol ϕ here is a variable that ranges over finite strings of symbols drawn from the alphabet that includes Σ, the connectives $\neg, \wedge, \vee, \rightarrow$, and (when necessary) left and right parentheses "(" and ")". We will subsequently play it fast and loose with parentheses, omitting them when no confusion can result. In particular, we take a negation symbol $\neg$ always to have binding precedence over the binary connectives.

Note that each sentence is, by definition, a finite string of symbols, and hence contains finitely many propositional constants.

Since the set $\mathsf{Sent}(\Sigma)$ is defined inductively, we can prove things about it using "proof by induction." A proof by induction proceeds as follows:

1. Show that the property of interest, say P, holds of the elements of Σ.
2. Show that if P holds of ϕ, then P holds of $\neg\phi$.
3. Show that if P holds of ϕ and ψ, then P holds of $\phi \wedge \psi$, $\phi \vee \psi$, and $\phi \to \psi$.

When these three steps are complete, one may conclude that all things in $\mathsf{Sent}(\Sigma)$ have property P.

Definition. A **context** is essentially a finite collection of sentences. However, we write contexts as sequences, for example $\phi_1, \ldots, \phi_n$ is a context. But ϕ_1, ϕ_2 is the same context as ϕ_2, ϕ_1, and is the same context as ϕ_1, ϕ_1, ϕ_2. If Δ and Γ are contexts, then we let Δ, Γ denote the union of the two contexts. We also allow an empty context.

Proof theory

Definition. We now define the relation $\Delta \vdash \phi$ of derivability that holds between contexts and sentences. This relation is defined **recursively** (aka, **inductively**), with base case $\phi \vdash \phi$ (Rule of Assumptions). Here we use a horizontal line to indicate that if $\vdash$ holds between the things above the line, then $\vdash$ also holds for the things below the line.

Rule of Assumptions $$\frac{}{\phi \vdash \phi}$$

$\wedge$ **elimination** $$\frac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \phi} \qquad \frac{\Gamma \vdash \phi \wedge \psi}{\Gamma \vdash \psi}$$

$\wedge$ **introduction** $$\frac{\Gamma \vdash \phi \qquad \Delta \vdash \psi}{\Gamma, \Delta \vdash \phi \wedge \psi}$$

$\vee$ **introduction** $$\frac{\Gamma \vdash \phi}{\Gamma \vdash \phi \vee \psi} \qquad \frac{\Gamma \vdash \psi}{\Gamma \vdash \phi \vee \psi}$$

$\vee$ **elimination** $$\frac{\Gamma \vdash \phi \vee \psi \qquad \Delta, \phi \vdash \chi \qquad \Theta, \psi \vdash \chi}{\Gamma, \Delta, \Theta \vdash \chi}$$

$\to$ **elimination** $$\frac{\Gamma \vdash \phi \to \psi \qquad \Delta \vdash \phi}{\Gamma, \Delta \vdash \psi}$$

$\to$ **introduction** $$\frac{\Gamma, \phi \vdash \psi}{\Gamma \vdash \phi \to \psi}$$

RAA $$\frac{\Gamma, \phi \vdash \psi \wedge \neg\psi}{\Gamma \vdash \neg\phi}$$

DNE $$\frac{\Gamma \vdash \neg\neg\phi}{\Gamma \vdash \phi}$$

Definition. A sentence ϕ is said to be **provable** just in case $\vdash \phi$. Here $\vdash \phi$ indicates that ϕ stands in the derivability relation with the empty context. We use $\top$ as shorthand for a sentence that is provable, for example, $p \vee \neg p$. We could then add as an inference rule "$\top$ introduction" that allowed us to write $\Delta \vdash \top$. It can be proven that the resulting definition of $\vdash$ would be the same as the original definition. We also sometimes use the symbol $\perp$ as shorthand for $\neg\top$. It might then be convenient to restate RAA as a rule that allows us to infer $\Delta \vdash \neg\phi$ from $\Delta, \phi \vdash \perp$. Again, the resulting definition of $\vdash$ would be the same as the original.

Discussion. The rules we have given for $\vdash$ are sometimes called the **classical propositional calculus** or just the **propositional calculus**. Calling it a "calculus" is meant to indicate that the rules are purely formal, and don't require any understanding of the meaning of the symbols. If one deleted the DNE rule, and replaced it with Ex Falso Quodlibet, the resulting system would be the **intuitionistic propositional calculus**. However, we will not pursue that direction at present.

Semantics

Definition. An **interpretation** (sometimes called a **valuation**) of Σ is a function from Σ to the set $\{\text{true}, \text{false}\}$, i.e. an assignment of truth-values to propositional constants. We will usually use 1 as shorthand for "true", and 0 as shorthand for "false".

Clearly, an interpretation v of Σ extends naturally to a function $v : \mathsf{Sent}(\Sigma) \to \{0, 1\}$ by the following clauses:

1. $v(\neg\phi) = 1$ if and only if $v(\phi) = 0$.
2. $v(\phi \wedge \psi) = 1$ if and only if $v(\phi) = 1$ and $v(\psi) = 1$.
3. $v(\phi \vee \psi) = 1$ if and only if either $v(\phi) = 1$ or $v(\psi) = 1$.
4. $v(\phi \to \psi) = v(\neg\phi \vee \psi)$.

Discussion. The word "interpretation" is highly suggestive, and has caused no small amount of confusion among philosophers. The thought is that elements of $\mathsf{Sent}(\Sigma)$ are uninterpreted, and hence lack meaning, but that an intepretation (in the precise sense defined above) endows these sentences with meaning or content. However, we are wary of this way of talking. An interpretation v is simply a **function**, i.e. a relation between two collections of mathematical objects. It is unclear how v could, in any sense, endow ϕ with meaning.

Definition. A **propositional theory** T consists of a signature Σ, and a set Δ of sentences in Σ. Sometimes we will simply write T in place of Δ, although it must be understood that the identity of a theory also depends on its signature. For example, the theory consisting of a single sentence p is different depending on whether it's formulated in the signature $\Sigma = \{p\}$, or in the signature $\Sigma' = \{p, q\}$.

Definition (Tarski truth)**.** Given an interpretation v of Σ, and a sentence ϕ of Σ, we say that ϕ is **true** in v just in case $v(\phi) = 1$.

Definition. For a set Δ of Σ sentences, we say that v is a **model** of Δ just in case $v(\phi) = 1$, for all ϕ in Δ. We say that Δ is **consistent** if Δ has at least one model.

Any time we define a concept for sets of sentences (e.g. consistency), we can also extend that concept to theories, as long as it's understood that theories are not merely sets of sentences.

Discussion. The use of the word "model" here seems to have its origin in consistency proofs for non-euclidean geometries. In that case, one showed that certain non-euclidean geometries could be represented inside euclidean geometry. Thus, it was argued, if euclidean geometry is consistent, then non-euclidean geometry is consistent.

In our case, it may not be immediately clear what sits on the "other side" of an interpretation, because it's certainly not euclidean geometry. What kind of mathematical thing are we interpreting our logical symbols into? The answer here — as will become apparent in Chapter **??** — is either a Boolean algebra, or a fragment of the universe of sets.

Definition. Let Δ be a set of Σ sentences, and let ϕ be a Σ sentence. We say that Δ **semantically implies** ϕ, written $\Delta \vDash \phi$, just in case ϕ is true in all models of Δ. That is, if v is a model of Δ, then $v(\phi) = 1$.

Exercise 1. Show that if $\Delta, \phi \vDash \psi$, then $\Delta \vDash \phi \to \psi$.

Exercise 2. Show that $\Delta \vDash \phi$ if and only if $\Delta \cup \{\neg\phi\}$ is inconsistent. Here $\Delta \cup \{\neg\phi\}$ is the theory consisting of $\neg\phi$ and all sentences in Δ.

We now state three main theorems of the metatheory of propositional logic.

Soundness Theorem

If $\Delta \vdash \phi$ then $\Delta \vDash \phi$.

The soundness theorem can be proven by an argument directly analogous to the substitution theorem below. We leave the details to the reader.

Completeness Theorem

If $\Delta \vDash \phi$ then $\Delta \vdash \phi$.

The completeness theorem can be proven in various ways. In this book, we will give a topological proof via the Stone Duality Theorem. But that will take a fair amount of background work. You'll have to wait until Chapter **??**.

Compactness Theorem

Let Δ be a set of sentences. If every finite subset Δ_F of Δ is consistent, then Δ is consistent.

The compactness theorem can be proven in various ways. One way of proving it — although, perhaps not the most illuminating — is as a corollary of the completeness theorem. Note that proofs are finite, by definition. Thus, if $\Delta \vdash \phi$, then $\Delta_F \vdash \phi$ for some finite subset Δ_F of Δ. Thus, if Δ is inconsistent, then $\Delta \vdash \perp$, hence $\Delta_F \vdash \perp$ for a finite subset Δ_F of Δ. But then Δ_F is inconsistent.

Definition. A theory T, consisting of axioms Δ in signature Σ, is said to be **complete** just in case Δ is consistent, and for every sentence ϕ of Σ, either $\Delta \vDash \phi$ or $\Delta \vDash \neg\phi$.

Discussion. Note the distinction between completeness of our proof system (which is independent of any theory), and completeness of some particular theory T. Famously, Kurt Gödel proved that the theory of Peano arithmetic is incomplete, i.e. there is a sentence ϕ of the language of arithmetic such that neither $T \vdash \phi$ nor $T \vdash \neg\phi$. However, there are much simpler examples of incomplete theories, as we now see.

Exercise 3. Consider the signature $\Sigma = \{p\}$. How many complete theories are there in this signature?

Definition. Let T be a theory in Σ. The **deductive closure** of T, written $Cn(T)$, is the set of Σ sentences that are implied by T. If $T = Cn(T)$, then we say that T is **deductively closed**.

Example. Let $\Sigma = \{p\}$, and let $\Delta = \{p\}$. Let $\Sigma' = \{p, q\}$, and let $\Delta' = \{p\}$. Here we must think of T and T' as different theories, even though they consist of the same sentences, i.e. $\Delta = \Delta'$. One reason to think of these as different theories: T is complete, but T' is incomplete. Another reason to think of T and T' as distinct is that Δ and Δ' have different deductive closures. For example, $q \vee \neg q$ is in the deductive closure of Δ', but not of Δ.

From a philosophical point of view, we should remember that Carnap and Quine insisted that choosing a theory is not just choosing axioms, it's choosing a language and axioms.

Exercise 4. Show that the theory T' from the previous example is not complete. (You may assume the completeness theorem.)

Exercise 5. Show that $Cn(Cn(T)) = Cn(T)$.

Translating between theories

Discussion. Imagine that here at Princeton, we've been studying the theory $\{p\}$ that consists of one (uninterpreted) proposition. Now, the logic professor at Harvard was a PhD student at Princeton, and was one of my teaching assistants.[1] What's more, before she left for Harvard, she asked me for a copy of my lecture notes, which I gladly shared with her. But before teaching the course at Harvard, she ran a "find and replace" on the notes, replacing each occurence of the propositional constant p with the propositional constant h. Otherwise, her notes are identical to mine.

What do you think: are they studying a different theory at Harvard? Or is it really the same theory, just with different notation?

I would say that the same theory is being taught at Havard and Princeton — that the two theories are **notational variants** of each other. In this section, we make a proposal for how to make precise the notion of "notational variants," or more generally, of **equivalent theories**. But we begin with an even more general notion, of translating one theory into another.

Definition. Let Σ and Σ' be propositional signatures. A **reconstrual** from Σ to Σ' is an assignment of elements in Σ to sentences of Σ'.

A reconstrual f extends naturally to a function $\overline{f} : \mathsf{Sent}(\Sigma) \to \mathsf{Sent}(\Sigma')$, as follows:

1. For p in Σ, $\overline{f}(p) = f(p)$.
2. For any sentence ϕ, $\overline{f}(\neg\phi) = \neg\overline{f}(\phi)$.
3. For any sentences ϕ and ψ, $\overline{f}(\phi \circ \psi) = \overline{f}(\phi) \circ \overline{f}(\psi)$, where $\circ$ stands for an arbitrary binary connective.

When no confusion can result, we use f for $\overline{f}$.

Substitution Theorem. *For any reconstrual $f : \Sigma \to \Sigma'$, if $\phi \vdash \psi$ then $f(\phi) \vdash f(\psi)$.*

Proof. Since the family of sequents is constructed inductively, we will prove this result by induction.

(rule of assumptions) We have $\phi \vdash \phi$ by the rule of assumptions, and we also have $f(\phi) \vdash f(\phi)$.

($\wedge$ intro) Suppose that $\phi_1, \phi_2 \vdash \psi_1 \wedge \psi_2$ is derived from $\phi_1 \vdash \psi_1$ and $\phi_2 \vdash \psi_2$ by $\wedge$ intro, and assume that the result holds for the latter two sequents. That is, $f(\phi_1) \vdash f(\psi_1)$ and $f(\phi_2) \vdash f(\psi_2)$. But then $f(\phi_1), f(\phi_2) \vdash f(\psi_1) \wedge f(\psi_2)$

[1]The story, all names, characters, and incidents portrayed in this production are fictitious. No identification with actual persons, places, buildings, and products is intended or should be inferred.

by $\wedge$ introduction. And since $f(\psi_1) \wedge f(\psi_2) = f(\psi_1 \wedge \psi_2)$, it follows that $f(\phi_1), f(\phi_2) \vdash f(\psi_1 \wedge \psi_2)$.

($\rightarrow$ intro) Suppose that $\theta \vdash \phi \rightarrow \psi$ is derived by conditional proof from $\theta, \phi \vdash \psi$. Now assume that the result holds for the latter sequent, i.e. $f(\theta), f(\phi) \vdash f(\psi)$. Then conditional proof yields $f(\theta) \vdash f(\phi) \rightarrow f(\psi)$. And since $f(\phi) \rightarrow f(\psi) = f(\phi \rightarrow \psi)$, it follows that $f(\theta) \vdash f(\phi \rightarrow \psi)$.

(reductio) Suppose that $\phi \vdash \neg\psi$ is derived by RAA from $\phi, \psi \vdash \bot$, and assume that the result holds for the latter sequent, i.e. $f(\phi), f(\psi) \vdash f(\bot)$. By the properties of f, $f(\bot) \vdash \bot$. Thus, $f(\phi), f(\psi) \vdash \bot$, and by RAA, $f(\phi) \vdash \neg f(\psi)$. But $\neg f(\psi) = f(\neg\psi)$, and therefore $f(\phi) \vdash f(\neg\psi)$, which is what we wanted to prove.

($\vee$ elim) We leave this step, and the others, as an exercise for the reader. $\square$

Definition. Let T be a theory in Σ, let T' be a theory in Σ', and let $f : \Sigma \rightarrow \Sigma'$ be a reconstrual. We say that f is a **translation** or **interpretation** of T into T', written $f : T \rightarrow T'$, just in case:

$$T \vdash \phi \quad \Longrightarrow \quad T' \vdash f(\phi).$$

Note that we have used the word "interpretation" here for a sort of mapping from one theory to another, whereas we previously used it for a sort of mapping from a theory to a different sort of thing, viz. a set of truth values. But there are good reasons to use the same word for the two things. These reasons will become clear in subsequent chapters.

Discussion. Have we been too liberal by allowing translations to map elementary sentences, such as p, to complex sentences, such as $q \wedge r$? Could a "good" translation render a sentence that has no internal complexity as a sentence that does have internal complexity? Think about it.

Definition (equality of translations)**.** Let T and T' be theories, and let both f and g be translations from T to T'. We write $f \simeq g$ just in case $T' \vdash f(p) \leftrightarrow g(p)$ for each atomic sentence p in Σ.

Definition. For each theory T, the identity translation $1_T : T \rightarrow T$ is given by the identity reconstrual on Σ. If $f : T \rightarrow T'$ and $g : T' \rightarrow T$ are translations, we let gf denote the translation from T to T given by $(gf)(p) = g(f(p))$, for each atomic sentence p of Σ. Theories T and T' are said to be **homotopy equivalent**, or simply **equivalent**, just in case there are translations $f : T \rightarrow T'$ and $g : T' \rightarrow T$ such that $gf \simeq 1_T$ and $fg \simeq 1_{T'}$.

Exercise 6. Prove that if v is a model of T', and $f : T \rightarrow T'$ is a translation, then $v \circ f$ is a model of T. Here $v \circ f$ is the interpretation of Σ obtained by applying f first, and then applying v.

Exercise 7. Prove that if $f : T \rightarrow T'$ is a translation, and T' is consistent, then T is consistent.