Meta-logic derivation rules

Hans Halvorson

February 19, 2013

Recall that the goal of this course is to learn how to prove things *about* (as opposed to "by means of") classical first-order logic. So, we will be using one system of logical rules, call it $\mathcal{M}$ for "meta-logic", in order to prove things about another system of logical rules, call it $\mathcal{L}$ for "classical first-order logic." In this chapter, we lay down the rules of $\mathcal{M}$. i.e. we answer the questions: "Which arguments in $\mathcal{M}$ are valid? What rules of proof is a user of $\mathcal{M}$ allowed to invoke?"

The system $\mathcal{M}$ has two sorts of rules:

1. Analogues of all the rules of $\mathcal{L}$.

2. Special rules from standard mathematics, e.g. set theory and elementary arithmetic.[1]

Let's first be clear about what we mean by "analogues of the rules of $\mathcal{L}$."

Consider, for example, the rule "→ elimination" of $\mathcal{L}$, which may be better known under the name "modus ponens." This rule has the form

$$\frac{\Gamma \Rightarrow \phi \to \psi \qquad \Delta \Rightarrow \phi}{\Gamma \cup \Delta \Rightarrow \psi}$$

In the system $\mathcal{L}$, monotonicity holds, i.e. if $\Gamma \Rightarrow \phi$ then $\Gamma' \Rightarrow \phi$ for any superset Γ' of Γ. So, we could equally well use the following variant of modus ponens:

$$\frac{\Gamma \Rightarrow \phi \to \psi \qquad \Gamma \Rightarrow \phi}{\Gamma \Rightarrow \psi}$$

By "equally well" here, I mean that if we used the variant rule, then we could prove no more and no less than we could prove with the original rule. In fact, we could equally well use the following even more simple variant:

[1] It can be proven that $\mathcal{M}$ cannot do with just the analogues of the rules of $\mathcal{L}$; there is a precise sense in which $\mathcal{M}$ must be stronger than $\mathcal{L}$ in order to establish the consistency of $\mathcal{L}$. We will return to this issue later in the semester.

$$\frac{\Rightarrow \phi \to \psi \qquad \Rightarrow \phi}{\psi}$$

Exercise 1. Convince yourself that these three rules are equivalent in a system where monotonicity holds.

Now this third form of modus ponens corresponds to the English description:

> From a conditional $\phi \to \psi$ and its antecedent ϕ, one is permitted to infer the consequent ψ.

So, we stipulate that our meta-logic $\mathcal{M}$ allows all analogous inferences. For example:

Example. Suppose that we have established the following two facts:

1. If $\Rightarrow \phi \to \phi$ is a valid sequent, then $\Gamma \Rightarrow \phi \to \phi$ is also a valid sequent, for any set Γ of premises.
2. $\Rightarrow \phi \to \phi$ is a valid sequent.

Then since $\mathcal{M}$ allows modus ponens, we may infer that:

3. $\Gamma \Rightarrow \phi \to \phi$ is a valid sequent, for any set Γ of premises.

$\lrcorner$

Exercise 2. In fact, (1) and (2) are true. How do we know that they are true? If asked to justify the assertion of (1) and (2), what would you say?

Example. Let's prove in the system $\mathcal{M}$ that: $\forall x \neg\phi(x) \models \neg\exists x\phi(x)$: Let M be a structure in which $\forall x\neg\phi(x)$ is true, that is $M \models \forall x\neg\phi(x)$. Suppose for reductio ad absurdum that $M \not\models \neg\exists x\phi(x)$. By the definition of $\models$, it follows that $M \models \neg\neg\exists x\phi(x)$, and hence that $M \models \exists x\phi(x)$. Again by the definition of $\models$, there is an $a \in M$ such that $M \models_x \phi(a)$. However, since $M \models \forall x\neg\phi(x)$, it also follows that $M \models_x \neg\phi(a)$, and so $M \not\models_x \phi(a)$, which is a contradiction. Therefore, $M \models \neg\exists x\phi(x)$. $\lrcorner$

Exercise 3. Rewrite the previous proof line by line, noting each use of an analogue of a rule from $\mathcal{L}$, or a derived rule from $\mathcal{L}$. Make sure that you agree that each step of the argument is warranted by the definitions involved and the rules of first-order logic.

In addition to the rules from first-order logic $\mathcal{L}$, the system $\mathcal{M}$ makes use of a few additional rules — in particular, proofs by induction and proofs using notions that can be defined in the theory of sets. The main objective of the following sections is to introduce those additional rules, and to give some examples of how to use them.

1 Proofs by mathematical induction

Suppose that you want to prove that every member of some set S has some property Φ. i.e. you want to show a statement of the form $\forall x(x \in S \to \Phi(x))$. But suppose further that the set S is defined "inductively", which we schematize as follows:

1. (Base cases) The things $a_1, \dots, a_n$ are in S.

2. (Inductive steps) There are some relations $R_1, \dots R_m$ such that for any $j = 1, \dots, m$ and for any $a_1, \dots, a_{n(j)}$ in S, if $R_j(a_1, \dots, a_{n(j)}, b)$ then b is in S. (Here $n(j)$ is the arity of the relation R_j.)

3. (Exclusion clause) Nothing is in S except things that are admitted by the previous two clauses.

Under these conditions, we stipulate that the following argumentative move is permitted.

$$\begin{array}{l}
\Phi(a_1) \\
\vdots \\
\Phi(a_n) \\
\forall x_1 \dots x_{n(1)} \forall y[\Phi(x_1) \wedge \cdots \wedge \Phi(x_{n(1)}) \wedge R_j(x_1, \dots, x_{n(1)}, y) \longrightarrow \Phi(y)] \\
\vdots \\
\forall x_1 \dots x_{n(m)} \forall y[\Phi(x_1) \wedge \cdots \wedge \Phi(x_{n(m)}) \wedge R_j(x_1, \dots, x_{n(m)}, y) \longrightarrow \Phi(y)] \\
\hline
\forall x(x \in S \to \Phi(x))
\end{array}$$

(Here I have abused notation by borrowing symbols, such as $\forall x$, from the language of first-order logic. But this inference rule belongs to the system $\mathcal{M}$ of meta-logic.)

Example. The natural numbers $\mathbb{N}$ are defined inductively by:

1. (Base case) 1 is in $\mathbb{N}$.

2. (Inductive clause) If n is in $\mathbb{N}$ then $n + 1$ is also in $\mathbb{N}$.

Thus, to show that some predicate Φ holds of all natural numbers, i.e. $\forall x(x \in \mathbb{N} \to \Phi(x))$, it suffices to show that $\Phi(1)$ and that $\forall x(\Phi(x) \to \Phi(x + 1))$. ⌟

Example. Let L be a first-order language. We have already defined three sets, based on L, using induction.

1. The set of L terms is defined inductively: base cases are constant symbols and variables, the inductive clause says that any n-terms can be combined with an n-ary function symbol to make a new term.

2. The set of L formulas is defined inductively.

3. The set of valid sequents of L is defined inductively. Notice here that what is actually defined is a subset of the set of pairs $F(\mathbb{S}) \times \mathbb{S}$, where $\mathbb{S}$ consists of sentences of L, and $F(\mathbb{S})$ consists of finite sequences of sentences of L.

In the days to come, we will use mathematical induction to prove things about each of these three sets. For example, the *soundness theorem* for first-order logic shows that for any sentence ϕ, and any finite sequence Γ of sentences:

If $\Gamma \Rightarrow \phi$ then $\Gamma \models \phi$.

Let V denote the set of valid sequents, and for a pair $x = \langle \Gamma, \phi \rangle$ consisting of a finite set of sentences and a sentence, let $\Phi(x)$ denote that $\Gamma \models \phi$. Then the soundness theorem asserts that:

$$\forall x (x \in V \rightarrow \Phi(x)).$$

Since the set V of valid sequents is defined inductively, it would make perfect sense to prove soundness by induction. ⌟

Exercise 4. The completeness theorem says that: if $\Gamma \models \phi$ then $\Gamma \Rightarrow \phi$. Would it make sense to try to prove completeness by induction?

2 More set theory

We have already discussed some of the basics of set theory. Now we need a bit more. Some of the new notions that we need to understand are:

- What is an equivalence relation? What are equivalence classes?
- When is a function one-to-one? When is a function onto?
- What is a finite set? What is an infinite set?
- What does it mean to say that one infinite set is bigger than another?

Definition. Let A be a set. A relation R on $A \times A$ is called an *equivalence relation* just in case the following hold:

1. (reflexive) $\forall x(\langle x, x\rangle \in R)$

2. (symmetric) $\forall x \forall y(\langle x, y\rangle \in R \rightarrow \langle y, x\rangle \in R)$

3. (transitive) $\forall x \forall y \forall z((\langle x, y\rangle \in R \wedge \langle y, z\rangle \in R) \rightarrow \langle x, z\rangle \in R)$.

An equivalence relation R on $A \times A$ induces a partition of the set A. The cells of this partition are the *equivalence classes* of elements of A, which are defined by:

$$[a] \ = \ \{x \in A : \langle a, x\rangle \in R\}.$$

Since the relation R is reflexive, every element of A is in some cell. We need to see now that the cells have no overlap, i.e. for any two cells $[a], [b]$, either $[a] = [b]$ or $[a] \cap [b] = 0$.

First, let's show that if $[a] \subseteq [b]$ then $[b] \subseteq [a]$. Indeed, supposing that $[a] \subseteq [b]$, let $x \in [b]$, which means that Rbx (where Rbx is shorthand for $\langle b, x\rangle \in R$). Since R is reflexive, $a \in [a]$, and since $[a] \subseteq [b]$, $a \in [b]$. Thus, Rba. Since R is symmetric, Rab, and since R is transitive, Rax. Therefore $x \in [a]$, and it follows that $[b] \subseteq [a]$.

Now we show that if $[a] \cap [b]$ is non-empty, then $[a] \subseteq [b]$. Since $[a] \cap [b]$ is non-empty, there is some $x \in [a] \cap [b]$. Let y be an arbitrary element of $[a]$. Then Rax, Rbx and Ray. Using the symmetry and transitivity of R, it follows that Rby, so $y \in [b]$. Therefore $[a] \subseteq [b]$.

Exercise 5. Let $f : A \rightarrow B$ be a function, and define a relation R on $A \times A$ by

$$Rxy \quad \text{if and only if} \quad f(x) = f(y).$$

Show that R is an equivalence relation.

Definition. Let A and B be sets. A function $f : A \rightarrow B$ is said to be *injective* or *one-to-one* just in case:

$$\forall x \forall y(f(x) = f(y) \rightarrow x = y).$$

Exercise 6. Rewrite the definition of injective using the notation $\langle x, y\rangle \in f$.

Definition. A function $f : A \rightarrow B$ is said to be *surjective* or *onto* just in case:

$$\forall y \exists x(f(x) = y).$$

Definition. A function $f : A \rightarrow B$ is said to be a *bijection* just in case it is injective and surjective.

Exercise 7. Is the projection $\pi_A : A \times B \to A$ injective? Is it surjective? (Be careful: the answer might depend on what A and B look like.)

Exercise 8. Show that if f and g are injective (surjective), then so is $g \circ f$.

Definition. If $f : A \to B$ is a function, a function $g : B \to A$ is called a *right inverse* to f just in case $f \circ g = 1_B$. In this case, f is called a *left inverse* of g. If $f \circ g = 1_B$ and $g \circ f = 1_A$, then g is called a *two-sided inverse*, or simply *inverse*, of f.

Definition. A function $f : A \to B$ is said to be a *isomorphism* just in case it has a two-sided inverse.

Exercise 9. Show that $f : A \to B$ has a left inverse iff f is injective.

Exercise 10. Show that $f : A \to B$ has a right inverse iff f is surjective. (Warning: the proof that a surjective function has a right inverse uses the axiom of choice.)

Exercise 11. Show that $f : A \to B$ is an isomorphism iff f is a bijection.

Note. Sometimes people equate isomorphisms with bijections ... and that would seem justified based on the fact we just stated! But the concept of isomorphism actually generalizes to other "categories" where the concept of bijection doesn't even apply. Moreover, even in categories where the concept of bijection does apply, a bijection is not always an isomorphism. e.g. in the category of groups, in the category of topological spaces. Indeed, as we will soon see, there is a general notion of an isomorphism between structures of a first-order language; but *not all bijections between structures are isomorphisms between those structures.*

Definition (Infinity)**.** A set S is said to be *infinite* just in case there is a function $f : S \to S$ that is injective but not surjective.

Proposition 1. *S is infinite iff there is a function $g : S \to S$ that is surjective but not injective.*

Proof. Suppose that S is infinite. Then there is a function $f : S \to S$ that is injective but not surjective. Thus there is a surjective function g from the image of f onto S. Pick an arbitrary point $x \in S$, and define $g' : S \to S$ by letting g' agree with g on its domain, and $g'(y) = x$ for everything outside of the domain of g. □

Proposition 2. *If S is infinite then for any $x \in S$, $S - \{x\}$ is also infinite.*

Proof. Since S is infinite, there is a function $f : S \to S$ that is injective but not surjective. Since f is not surjective, there is a point y_0 in S that is not in the image of f. Now, we may in fact, suppose that $y_0 \neq x$; because there is a permutation p of S such that $p(x) \neq x$ and $p \circ f$ is also injective but not surjective.

If $f(y_0)$ were in the image of $f \circ f$, then we would have $f(y_0) = f(f(z))$ for some z, and since f is injective $y_0 = f(z)$, contradicting the fact that y_0 is not in the image of f. Therefore, neither y_0 nor $f(y_0)$ is in the image of $f \circ f$. Moreover, $y_0 \neq f(y_0)$ since y_0 is not in the image of f.

Let g be $f \circ f$ restricted to $S - \{x\}$. If x is in the image of g, then replace g with $q \circ g$, where $q : S \to S$ is a permutation that switches x and $f(y_0)$. Call the resulting function g again, but now think of g as a function from $S - \{x\}$ to $S - \{x\}$. Then g is injective, but y_0 is not in the image of g. □

Corollary 3. *If S is infinite then $S - \{x_1, \ldots, x_n\}$ is infinite.*

Exercise 12. Use mathematical induction to prove the corollary.

Definition. Let A and B be sets. We define the exponential B^A to be the set of functions from A into B:

$$B^A = \{f : f \subseteq A \times B, \text{ and } f \text{ is a function}\}.$$

We define on other operation on sets, although we will not use it much in this course.

Definition. If A is a set, we let $P(A)$ denote the set of all subsets of A. That is,

$$P(A) = \{S : S \subseteq A\}.$$

Exercise 13. Show that there is a bijection between $P(A)$ and 2^A, where 2 is a set with two elements, e.g. $2 = \{0, 1\}$. Hint: associate a subset S of A with the function $\chi_S : A \to \{0, 1\}$ defined by

$$\chi_S(a) = \begin{cases} 1 & a \in S, \\ 0 & a \notin S. \end{cases}$$

Definition. Let A and B be sets. We write $|A| \leq |B|$ just in case there is an injective function $f : A \to B$. We write $|A| = |B|$ just in case there is a bijection $f : A \to B$. (Actually it can be shown that if $|A| \leq |B|$ and $|B| \leq |A|$, then there is a bijection $f : A \to B$. But we won't prove that fact here.) We say that a set A is *countable* just in case $|A| \leq |\mathbb{N}|$, where $\mathbb{N} = \{0, 1, 2, \ldots\}$ is the set of natural numbers. If a set is not countable, we say that it is *uncountable.*

Proposition 4. *The set* $\mathbb{N} \times \mathbb{N}$ *is countable.*

Proof. We will give one quick proof, and one slower proof. For the quick proof, we define an injection $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ by

$$g(x, y) \;=\; 2^x 3^y.$$

If $\langle x, y\rangle \neq \langle x', y'\rangle$, then either $x \neq x'$ or $y \neq y'$. In either case, unique factorizability of integers gives $2^x 3^y \neq 2^{x'} 3^{y'}$. Therefore, g is injective. Since there is an injection $g : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$, there is a surjection $h : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$. (For example, h could be defined as the inverse of g on its range, and send everything outside of the range of g to $\langle 0, 0\rangle$.)

We also give a slower proof, because it can be helpful to have a canonical enumeration of $\mathbb{N} \times \mathbb{N}$. Imagine writing down all elements in $\mathbb{N} \times \mathbb{N}$ in an infinte table, whose first few elements look like this:

$$\left(\begin{array}{cccc} \langle 0,0\rangle & \langle 1,0\rangle & \langle 2,0\rangle & \cdots \\ \langle 0,1\rangle & \langle 1,1\rangle & \langle 2,1\rangle & \cdots \\ \langle 0,2\rangle & \langle 1,2\rangle & \langle 2,2\rangle & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{array}\right.$$

Now imagine running a thread diagonally through the numbers: begin with $\langle 0, 0\rangle$, then move down to $\langle 1, 0\rangle$ and up to $\langle 0, 1\rangle$, then over to $\langle 2, 0\rangle$ and down its diagonal, etc.. This process defines a function $f : \mathbb{N} \to \mathbb{N} \times \mathbb{N}$ whose first few values are

$$\begin{aligned} f(0) &= \langle 0, 0\rangle \\ f(1) &= \langle 0, 1\rangle \\ f(2) &= \langle 1, 0\rangle \\ f(3) &= \langle 2, 0\rangle \\ f(4) &= \langle 1, 1\rangle \\ f(5) &= \langle 0, 2\rangle \\ &\vdots \end{aligned}$$

It is not difficult to show that f is surjective, and so $\mathbb{N} \times \mathbb{N}$ is countable. □

Exercise 14. Show that if $A_1, \ldots, A_n$ are countable, then $A_1 \times \cdots \times A_n$ is countable.

Exercise 15. Show that if A and B are countable then $A \cup B$ is countable.

Definition. Let Λ be a set, and for each $\lambda \in \Lambda$, suppose that B_λ is a set. Then we define the set $\bigcup_{\lambda\in\Lambda} B_\lambda$ by

$$x \in \bigcup_{\lambda\in\Lambda} B_\lambda \quad \text{if and only if} \quad \exists\lambda(\lambda \in \Lambda \text{ and } x \in B_\lambda).$$

Proposition 5. *If $\{B_i : i \in \mathbb{N}\}$ is a collection of countable sets, then $\bigcup_{i\in\mathbb{N}} B_i$ is countable.*

Proof. Let $\sqcup_i B_i$ denote the *disjoint* union of the B_i. The only difference between $\sqcup_i B_i$ and $\cup_i B_i$ is that in the former, we do not identify elements that occur in multiple B_i. So clearly $\sqcup_i B_i$ is larger than $\cup_i B_i$, and it will suffice to show that there is an injection f from $\sqcup_i B_i$ to $\mathbb{N} \times \mathbb{N}$. Since B_i is countable, there is an injection $g_i : B_i \to \mathbb{N}$. Define f by setting

$$f(x) \;=\; \langle i, g_i(x)\rangle,$$

where i denotes the index of the set B_i in which x occurs. It remains only to show that f is injective. Suppose then that $f(x) = f(y)$, that is $\langle i, g_i(x)\rangle = \langle j, g_j(y)\rangle$. But then $i = j$, and so $g_i(x) = g_j(y) = g_i(y)$. Since g_i is injective, $x = y$. Therefore f is injective. □

Proposition 6. *Let L be a first-order language with a countable number of symbols, and let $Str(L)$ be the set of all finite strings from the alphabet of L. Then $Str(L)$ is countable.*

Proof. Let A be the alphabet of L. Since L is countable, so is A. Now let $A^n = A \times \cdots \times A$ be the Cartesian product of n copies of A. Thus, A^n corresponds one-to-one with strings of length n from A. The set of all finite-length strings from A thus corresponds to

$$\bigcup_{i\in\mathbb{N}} A^i.$$

It suffices then to show two things. First, if A is countable then A^n is countable (Exercise 14). Second, if $\{B_i : i \in \mathbb{N}\}$ is a collection of countable sets, then $\cup_{i\in\mathbb{N}} B_i$ is countable (Proposition 5). Therefore $Str(L)$ is countable. □

Corollary 7. *If L is countable, then the set of L formulas is countable.*

Exercise 16. If A is a set, we let $\mathbb{F}(A)$ denote the set of all finite subsets of A. Show that (i) if A is finite then $|\mathbb{F}(A)| = |\mathbb{N}|$, and (ii) if A is infinite then $|\mathbb{F}(A)| = |A|$.

Some sets are so big that they are not countable. For example, a famous proof by Georg Cantor shows that the set $\mathbb{R}$ of real numbers is uncountable. The proof goes something like this: consider the set S of all numbers of the form $0.a_1a_2a_3\cdots$ where each a_i is equal to 0 or 1. These numbers all lie in the interval $[0,1]$ in $\mathbb{R}$. If we can show that there is no surjection $\mathbb{N}\to S$, then there is no surjection $\mathbb{N}\to\mathbb{R}$. Suppose for reductio ad absurdum that there is a surjection $s:\mathbb{N}\to S$. So, for each $i\in\mathbb{N}$, s_i is an element of S. We let s_{ij} denote the j-th term in the decimal expansion of s_i. We claim that there is some $z\in S$ that is not in the image of s. Indeed, define

$$z=0.\overline{s}_{11}\overline{s}_{22}\overline{s}_{33}\cdots,$$

where $\overline{s}_{ii}$ is the opposite of s_{ii}. Then $z\neq s_i$ for all i, since the i-th term of z doesn't equal the i-th term of z_i.

Exercise 17. Use a Cantor-style argument to show that for any set A, its powerset $P(A)$ is strictly larger. That is, $|A|<|P(A)|$.

3 Choice axioms

The axiom of choice is a powerful tool for proof, and contemporary mathematicians rarely hesitate to invoke it. Although we will rarely need to use this axiom, it does come up at a couple of crucial junctures — e.g. in the proof of compactness, and in the proof of completeness. (In fact, it is known that these uses of a choice-like axiom are unavoidable: there is no way to prove completeness without using some such axiom.)

There are numerous, provably equivalent, ways of stating the axiom of choice, including:

1. Suppose that $F:I\to S$ is a function such that each $F(i)$ is a non-empty set. Then there is set S' and a function $f:I\to S'$ such that $f(i)\in F(i)$ for all $i\in I$.

2. If $\{F_i:i\in I\}$ is a collection of non-empty sets, then the Cartesian product $\Pi_{i\in I}F_i$ is non-empty.

3. (Zorn's lemma) Suppose that $(P,\leq)$ is a non-empty partially ordered set. By this we mean that $\leq$ is a two-place relation on P that is transitive, reflexive, and anti-symmetric — where the last property means that if $x\leq y$ and $y\leq x$, then $x=y$. A chain in P is a subset C with the further property that any two elements in C are ordered with respect to each other; more precisely, $\forall x\forall y(x,y\in C\to(x\leq y\vee y\leq x))$. An element element x_0 is an *upper bound* for the chain C just in case $\forall x(x\in C\to x\leq x_0)$. Then Zorn's lemma says that if every chain in P has an upper bound, then P has a maximal element.

4. Every surjective function has a section, i.e. a right inverse.

5. (Tychonoff's Product Theorem) A product of compact topological spaces is compact.

It has been shown that these versions of the axiom of choice are equivalent, relative to the background of Zermelo-Frankel (ZF) set theory.

Exercise 18. Show that the first version of the axiom of choice above implies that every surjective function has a section.

But to be clear, there is no question of *proving* the axiom of choice from anything more "elementary" or more obvious. Choice is, after all, an *axiom*; i.e. something that one *postulates* as a valid starting point for proofs.

In logical meta-theory, we need a somewhat weaker consequence of the axiom of choice. In order to state this axiom, we need to define the notion of a filter of sentences.

Definition. Let L be a first-order language. A set F of sentences in L is called a *filter* just in case:

1. If $\phi \in F$ and $\phi \models \psi$ then $\psi \in F$.

2. If $\phi \in F$ and $\psi \in F$, then $\phi \wedge \psi \in F$.

The filter F is called *proper* if it doesn't contain all sentences.

Proposition 8. *A filter F is improper iff F contains an inconsistent sentence.*

Proof. An inconsistent sentence semantically entails any sentence; so, if a filter contains a contradiction, then it contains all sentences and is improper. Conversely, an improper filter contains all sentences, hence an inconsistent sentence. □

Exercise 19. If ϕ is a sentence, we let $\uparrow\phi$ denote the set

$$\uparrow\phi \;=\; \{\psi : \phi \models \psi\}.$$

Show that $\uparrow\phi$ is a filter, and is proper iff ϕ is consistent.

Definition. A proper filter F is called an *ultrafilter* if it is maximal among proper filters, i.e. if F' is a proper filter such that $F \subseteq F'$, then $F = F'$.

Proposition 9. *Let $\{F_i : i \in I\}$ be a nested collection of proper filters, by which we mean that for any F_i, F_j, either $F_i \subseteq F_j$ or $F_j \subseteq F_i$. Then the union of filters $\bigcup_{i \in I} F_i$ is again a proper filter.*

Proof. Suppose that $\phi \in \bigcup_{i \in I} F_i$ and $\phi \models \chi$. Then $\phi \in F_i$ for some $i \in I$, and since F_i is a filter, $\chi \in F_i$. Thus, $\chi \in \bigcup_{i \in I} F_i$.

Suppose now that both $\phi \in \bigcup_{i \in I} F_i$ and $\psi \in \bigcup_{i \in I} F_i$, so that there is some $i \in I$ with $\phi \in F_i$, and some $j \in I$ with $\psi \in F_j$. By assumption, either $F_i \subseteq F_j$ or $F_j \subseteq F_i$. Thus, $F_i \cup F_j = F_i$ or $F_i \cup F_j = F_j$. In either case, $F_i \cup F_j$ is a filter. Furthermore, both ϕ and ψ are in $F_i \cup F_j$, and so $\phi \wedge \psi$ is in $F_i \cup F_j$. Thus, $\phi \wedge \psi \in F_i$ or $\phi \wedge \psi \in F_j$, and in either case, $\phi \wedge \psi \in \bigcup_{i \in I} F_i$.

Finally, $\bigcup_{i \in I} F_i$ is proper, because if it contained an inconsistent sentence, then one of the F_i would contain an inconsistent sentence. □

Exercise 20. Let L be a language, and let M be an L structure. Show that the set

$$Th(M) \;=\; \{\phi : M \models \phi\},$$

is an ultrafilter of sentences of L.

Exercise 21. Let S be any consistent set of sentences. Show that S is contained in an ultrafilter of sentences.

Exercise 22. Let $L = \{p_0, p_1, p_2, \ldots\}$ be a propositional language, and let U be an ultrafilter on the set of sentences of L. Define a function h from sentences of L to $\{0, 1\}$ by setting $h(\phi) = 1$ iff $\phi \in U$. Show that h defines an L-structure (i.e. show that h defines a truth-valuation on the sentences of L).

Proposition 10. *Let F be a proper filter. F is maximal if and only if for any sentence ϕ, either $\phi \in F$ or $\neg\phi \in F$.*

Proof. Let U be an ultrafilter, and suppose that $\neg\phi \notin U$. Let U' consist of the set of semantic consequences of $U \cup \{\phi\}$. We claim that U' is a proper filter. First, if $U \cup \{\phi\} \models \psi$ and $\psi \models \chi$, then $U \cup \{\phi\} \models \chi$. Similarly, the conjunction of semantic consequences is a semantic consequence; so U' is a filter. To see that U' is proper, we need only show that it doesn't contain $\perp$. But if U' contains $\perp$, then $U \cup \{\phi\} \models \perp$, from which it follows that $U \models \neg\phi$. Since U is a filter, $\neg\phi \in U$, a contradiction. Thus, U' is a proper filter. Finally, since U was assumed to be maximal among proper filters, it follows that $U = U'$ and hence that $\phi \in U$.

Conversely, suppose that F is not an ultrafilter. Thus, there is a proper filter F' such that $F \subseteq F'$ and a sentence $\phi \in F' - F$. If $\neg\phi \in F$ then $\neg\phi \in F'$, and since F' is a filter $\phi \wedge \neg\phi \in F'$. But since F' is proper, it cannot contain a contradiction. Therefore, $\phi \notin F$ and $\neg\phi \notin F$. □

Exercise 23. Show that if $\{F_i : i \in I\}$ is a family of filters, then $\bigcap_{i\in I} F_i$ is a filter.

Example. We'll show here that a finitely satisfiable set of sentences gives rise to a proper filter. A set Γ of sentences is said to be *finitely satisfiable* if each finite subset Γ_0 of Γ has a model M_0. Given finitely satisfiable Γ, let

$$Cn(\Gamma) = \{\phi : \Gamma_0 \models \phi, \text{ some finite } \Gamma_0 \subseteq \Gamma\}.$$

We claim that $Cn(\Gamma)$ is a proper filter. First, to see that $Cn(\Gamma)$ is a filter, suppose that $\phi \in Cn(\Gamma)$ and $\phi \models \psi$. Thus, there is a finite $\Gamma_0 \subseteq \Gamma$ such that $\Gamma_0 \models \phi$, and thus $\Gamma_0 \models \psi$. Therefore $\psi \in Cn(\Gamma)$. Suppose now that $\phi, \psi \in Cn(\Gamma)$. Then there is a finite subset Γ_0 of Γ such that $\Gamma_0 \models \phi$, and a finite subset Δ_0 of Γ such that $\Delta_0 \models \psi$. But then $\Gamma_0 \cup \Delta_0$ is a finite subset of Γ such that $\Gamma_0 \cup \Delta_0 \models \phi \wedge \psi$. Therefore, $\phi \wedge \psi \in Cn(\Gamma)$. Finally, to see that $Cn(\Gamma)$ is proper, it will suffice to show that it does not contain a contradiction. But if $\chi \in Cn(\Gamma)$ is a contradiction, then there is a finite subset Γ_0 of Γ such that $\Gamma_0 \models \chi$; which means that Γ_0 is inconsistent. Since Γ is finitely satisfiable, $Cn(\Gamma)$ contains no contradiction. Therefore $Cn(\Gamma)$ is a proper filter. ⌟

Suppose that F is a proper filter of sentences. Could we grow F until it reaches maximal size? That is, can we add sentences to F until it becomes an ultrafilter? Interestingly, the claim:

> **Ultrafilter Extension Lemma (UF)** Every proper filter is contained in an ultrafilter.

cannot be shown to be true by the rules of meta-logic, if by those rules we mean mean the rules of first-order logic supplemented by the axioms of (Zermelo-Frankel) set theory. But one could add UF to ZF set theory as an an axiom. As it turns out, UF is a consequence of ZF+AC (but not vice versa).

Proposition 11. *In ZF set theory, the axiom of choice implies the ultrafilter extension lemma.*

Proof. Let F be a proper filter. We will use Zorn's lemma to show that there is an ultrafilter U containing F. Let P be the set of proper filters containing F, and for x, y in P, we write $x \leq y$ just in case $x \subseteq y$. Clearly, $(P, \leq)$ is a partially ordered set. We claim now that every chain in P has an upper bound. Indeed, if C is a chain in P, let $x_0 = \bigcup\{x : x \in C\}$. By Proposition 9, x_0 is a proper filter, and hence in P. But $x \leq x_0$ for all $x \in C$, so x_0 is an upper bound for C. Since C was an arbitrary chain in P, every chain in P has an upper bound. By Zorn's lemma, P has a maximal element, which we will call U. Of course, U is a proper filter. We need only establish that U is maximal. So, suppose that $U \subseteq U'$, where U' is a proper filter containing F. Then $U' \in P$, and since U is a maximal element for P, $U' \subseteq U$. Therefore, $U = U'$, and U is maximal. □

To show that ZF+UF does not imply AC requires an independence proof, which is beyond the scope of this course.

For our purposes, we really do not need the full power of UF: we are not particularly interested, for example, in languages that are extremely large. In fact, we usually work with finite languages, or at worst, countably infinite languages.

Suppose that L is countably infinite, so that the set of L sentences can be written as $\{\phi_i : i = 0, 1, 2, \ldots\}$. Let F be a proper filter of sentences. Note first that for any sentence ϕ, if $F \cup \{\phi\}$ is inconsistent, then $F \models \neg\phi$, and so $\neg\phi \in F$. Thus, if $\neg\phi \notin F$, then $F \cup \{\phi\}$ is consistent and is contained in a proper filter F'. We define a sequence $\{F_i : i = 0, 1, 2, \ldots\}$ of filters as follows:

- $F_0 = F$.
- If $\neg\phi_i \in F_i$, then let $F_{i+1} = F_i$. If $\neg\phi_i \notin F_i$, let F_{i+1} be the proper filter generated by $F_i \cup \{\phi_i\}$.

By construction, $F_i \subseteq F_j$ when $i \leq j$. Thus, $\bigcup_{i\in\mathbb{N}} F_i$ is a proper filter. We claim that $\bigcup_{i\in\mathbb{N}} F_i$ is maximal. Let ϕ be an arbitrary sentence of L. Then $\phi = \phi_i$ for some $i \in \mathbb{N}$. Thus, either $\neg\phi \in F_{i+1} \subseteq \bigcup_{i\in\mathbb{N}} F_i$ or $\phi \in F_{i+1} = \bigcup_{i\in\mathbb{N}} F_i$.

It may not be immediately obvious what "inference rules" we used in the previous argument. In fact, while we did not use AC, or even UF, we did implicitly use another principle that goes beyond basic ZF set theory.

The data we began with was a proper filter F_0 and an enumeration $\{\phi_0, \phi_1, \ldots\}$ of the sentences of L. Now define a set S of ordered pairs of the form $\langle n, F\rangle$ where $n \in \mathbb{N}$ and F a proper filter containing F_0. First, we let $\langle 0, F\rangle \in S$ just in case $F_0 = F$. Second, for $n = 1, 2, \ldots$, we let $\langle n, F\rangle \in S$ just in case either $\phi_i \in F$ or $\neg\phi_i \in F$ for all $i < n$. Recall now that if F is a proper filter, then either $F \cup \{\phi_i\}$ generates a proper filter, or $F \cup \{\neg\phi_i\}$ generates a proper filter (or both). Define a relation $\prec$ on S by setting $\langle i, F\rangle \prec \langle j, F'\rangle$ just in case $j = i + 1$, and F' is a proper filter that is either generated by $F \cup \{\phi_i\}$ or by $F \cup \{\neg\phi_i\}$. Thus, for each $\langle i, F\rangle \in S$, there is at least one, but not more than two, $\langle j, F'\rangle \in S$ such that $\langle i, F\rangle \prec \langle j, F'\rangle$.

The previous argument asserts that there is a function $f : \mathbb{N} \to S$ such that $f(i) \prec f(i+1)$ for all $i \in \mathbb{N}$. This assertion is an instance of the following set-theoretic axiom.

Axiom of Dependent Choice (DC) If $\prec$ is a total relation on a set S, then there is a function $f : \mathbb{N} \to S$ such that $f(i) \prec f(i+1)$ for all $i \in \mathbb{N}$.

(A relation $\prec$ is *total* just in case $\forall x \exists y (x \prec y)$.) It is known that ZF+DC does *not* imply UF; but we have just seen that ZF+DC entails UF for countable languages.

Exercise 24. Let $L = \{p_0, p_1, \ldots\}$, let ϕ be a consistent sentence of L, and let $\uparrow\phi = \{\psi : \phi \models \psi\}$ be the proper filter generated by ϕ. Show that $\uparrow\phi$ is *not* an ultrafilter. Hint: ϕ contains at most finitely many of $p_0, p_1, p_2, \ldots$. Consider some p_n that does not occur in ϕ.