Provability, Soundness and Completeness

Deductive rules of inference provide a mechanism for deriving true conclusions from true premises

Rules of inference

So far we have treated formulas as "given", and have shown how they can be related to a domain of discourse, and how the truth of a set of premises can guarantee (entail) the truth of a conclusion. However, our goal in logic and particularly in AI is to *derive* new conclusions from given facts. For this we need *rules of inference* (and later, strategies for applying such rules so as to derive a desired conclusion, if possible).

In general, a "forward" inference rule consists of one or more premises and a conclusion. Both the premises and the conclusion are generally *schemas*, i.e., they involve metavariables for formulas or terms that can be particularized in many ways (just as we saw in the case of valid formula schemas). We often put a horizontal line under the premises, and write the conclusion underneath the line. For instance, here is the rule of *Modus Ponens*

$$MP: \ \frac{\phi, \ \phi \Rightarrow \psi}{\psi}$$

This says that given a premise formula ϕ, and another formula of form $\phi \Rightarrow \psi$, we may derive the conclusion ψ. It is intuitively clear that this rule leads from true premises to a true conclusion – but this is an intuition we need to verify by proving the rule *sound*, as illustrated below. An example of using the rule is this: from

Dog(Snoopy), Dog(Snoopy) $\Rightarrow$ Has-tail(Snoopy),

we can conclude Has-tail(Snoopy). (Of course, if the premises aren't true, then the conclusion needn't be true either.)

The following are some additional well-known rules of inference, called *Modus Tollens* (MT), *Modus Tollendo Ponens* (MTP, also called *Propositional Resolution*), *Double Negation Elimination* (DN), *And Introduction* (AI), *And Elimination* (AE), and *Or Introduction* (OI):

$$MT: \ \frac{\neg\psi, \ \phi \Rightarrow \psi}{\neg\phi} \qquad MTP: \ \frac{\neg\phi, \ \phi \vee \psi}{\psi}, \ \ \frac{\psi, \ \phi \vee \neg\psi}{\phi}, \ \textit{etc.}$$

$$DN: \frac{\neg\neg\phi}{\phi} \qquad AI: \ \frac{\phi, \ \psi}{\phi \wedge \psi} \qquad AE: \ \frac{\phi \wedge \psi}{\phi}, \ \ \frac{\phi \wedge \psi}{\psi} \qquad OI: \ \frac{\phi}{\phi \vee \psi}$$

All of these rules can be proved *sound* in the sense that the premises *entail* the conclusion:

Soundness of $\frac{\phi_1,...,\phi_n}{\psi}$:
All models of $\{\phi_1, ..., \phi_n\}$ are models of ψ,
i.e., For all models $\mathcal{M}$ such that $\models_{\mathcal{M}} \{\phi_1, ..., \phi_n\}$, $\models_{\mathcal{M}} \psi$.

Soundness is a very desirable property of a rule of inference: it can never lead us to a false conclusion, as long as the facts we started with are correct. Unfortunately, we do not always have the luxury of being able to restrict ourselves to sound inference. For example, if you intend to cross an intersection and the light turns green, you generally proceed on the assumption that the light in the other direction will be red, that this will cause cross-traffic to stop, and it is therefore safe for you to cross. Assuming that you believe that traffic lights can occasionally malfunction, or that drivers occasionally run a red light, this involves unsound (though still reasonably reliable) inferences.

Another kind of example of unsound inference involves *abduction* or *induction*, where we posit, or strengthen our belief in, a general principle based on a few examples. For instance, a small child encountering some neighborhood dogs and finding that each of those dogs (occasionally) barks might abductively conjecture that *all* dogs bark. After some further examples of barking dogs (and no examples of nonbarking dogs), she might inductively confirm this conjecture. It is hard to imagine a sound rule of inference, and a reasonable set of prior beliefs, from which such a conclusion would follow. Yet this kind of unsound reasoning is crucial for learning and coping with the world. We will later study some methods of "defeasible" inference, but right now we continue to examine deductive inference.

Proving soundness: example 1

Let's prove the soundness of Double Negation Elimination (DN). We need to show that any model of $\neg\neg\phi$ is also a model of ϕ. So let $\mathcal{M}$ be any model of $\neg\neg\phi$, i.e.,

$\models_{\mathcal{M}} \neg\neg\phi$.
Hence for all v.a.$\mathcal{U}$, $\models_{\mathcal{M}} \neg\neg\phi[\mathcal{U}]$ (by definition of $\models_{\mathcal{M}} \psi$);
hence for all v.a.$\mathcal{U}$, $\not\models_{\mathcal{M}} \neg\phi[\mathcal{U}]$ (by satisfaction conditions for $\neg$);
hence for all v.a.$\mathcal{U}$, $\models_{\mathcal{M}} \phi[\mathcal{U}]$ (by sat'n cond'ns for $\neg$),
i.e., $\models_{\mathcal{M}}\phi$. $\square$

Proving soundness: example 2

For a slightly more complex case, let's prove the soundness of the following rule (not a commonly used one):

$$\frac{\neg(\phi \Rightarrow \psi)}{\phi \wedge \neg\psi}.$$

Again, let $\mathcal{M}$ be any model of the premise, i.e.,

$\models_{\mathcal{M}} \neg(\phi \Rightarrow \psi)$.
Hence for all v.a.$\mathcal{U}$, $\models_{\mathcal{M}} \neg(\phi \Rightarrow \psi)[\mathcal{U}]$ (by def'n of $\models_{\mathcal{M}} \psi$);

hence for all v.a. $\mathcal{U}$, $\not\models_{\mathcal{M}} (\phi \Rightarrow \psi)[\mathcal{U}]$ (by sat'n cond'n for $\neg$);
i.e., for all v.a. $\mathcal{U}$, it is not the case that $\models_{\mathcal{M}} (\phi \Rightarrow \psi)[\mathcal{U}]$;
hence for all v.a. $\mathcal{U}$, it is not the case that: $\not\models_{\mathcal{M}} \phi[\mathcal{U}]$ or $\models_{\mathcal{M}} \psi)[\mathcal{U}]$
(by sat'n cond'ns for $\Rightarrow$);
i.e., for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} \phi[\mathcal{U}]$ and $\not\models_{\mathcal{M}} \psi)[\mathcal{U}]$;
hence for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} \phi[\mathcal{U}]$ and $\models_{\mathcal{M}} \neg\psi)[\mathcal{U}]$ (by sat'n cond'n for $\neg$);
hence for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} (\phi \wedge \neg\psi)[\mathcal{U}]$ (by sat'n cond'n for $\wedge$);
i.e., $\models_{\mathcal{M}} (\phi \wedge \neg\psi)$. $\square$

Rules involving substitution – equality and quantifiers

Some common rules involving substitution are *Substitution of Equals* (S=), *Universal Instantiation* (UI), *Existential Instantiation* (EI, also called *Skolemization*), and *Existential Generalization* (EG),

$$S=: \ \frac{\tau = \tau',\ \phi(...\tau...)}{\phi(...\tau'...)} \qquad UI: \ \frac{(\forall\nu\phi)}{\phi[\nu/\tau]} \qquad EI: \ \frac{(\exists\nu\phi)}{\phi[\nu/\kappa]} \qquad EG: \ \frac{\phi(...\tau...)}{(\exists\nu\phi(...\nu...))}$$

where τ and τ' are ground (variable-free) terms, ν is a variable, κ is a new constant, $\phi[\nu/\tau]$ is ϕ with *all* free occurrences of ν replaced by τ, and the notation $\phi(...\tau...)$ and $\phi(...\tau'...)$ indicates that τ occurs at least once in ϕ, and we are replacing at least one such occurrence with τ'.

The least obvious of these rules is probably EI, and it deserves some comment (especially because it is used in more general form in resolution theorem proving, the predominant theorem proving technique in AI). The reasoning here is as follows: if there *exists* an object such that ϕ holds for it, then we can invent a name κ for this object (called a Skolem constant), and assert ϕ with the variable replaced by κ.

Interestingly, this rule is not sound, in the strict sense. The reason is that the premise in no way constrains the constant κ (a model of the premise may assign it any value), yet the conclusion *does* constrain the value of κ (its value must be chosen so as to satisfy $\phi[\nu/\kappa]$). As a result some models of the premise fail to be models of the conclusion – they make the "wrong" choice of value for κ. Nonetheless, the rule is "nearly" sound: if we start with premises Δ and perform a series of inferences using sound rules of inference and EI, reaching some conclusion ϕ, then all models of Δ are models of ϕ, provided that no Skolem constants appear in ϕ. In other words, even though EI itself is not sound, it can help us reach conclusions in a sound fashion! In particular, if we deduce a contradiction from Δ, then – even if we have used EI – we can be sure Δ is unsatisfiable (has no models). This fact is also important in resolution-based theorem proving.

The remaining rules are sound, and this is rather easy to see intuitively. In a rigorous proof, one would use induction on formula size and term size. For instance, the basis cases for proving UI sound by induction on formula size are ones of form $\frac{(\forall\nu P(\gamma_1(\nu),...,\gamma_n(\nu)))}{P(\gamma_1(\tau),...,\gamma_n(\tau))}$ and $\frac{(\forall\nu\ \gamma(\nu)=\gamma'(\nu))}{\gamma(\tau)=\gamma'(\tau)}$. The $\gamma_i(\nu)$ and $\gamma(\nu)$ are arbitrarily complex terms containing any

number of free occurrences of ν. But to establish these basis cases, we need an induction on the sizes of the terms. After establishing the soundness of the basis cases, we consider more complex formulas in place of the atomic ones, using the semantics of connectives and quantifiers to make the inductive argument. Much the same applies to the $S=$ and EG rules.

Proof systems for forward proofs

Given a first-order language and some rules of inference, we have most of the ingredients for a **proof system**. However, in some proof systems there is one additional ingredient, a *set of logical axioms* (more accurately, *axiom schemas*). These logical axioms, as the name suggests are logical truths (valid formulas) of the language, and as such can be used as premises in any proof.

For instance, the following is a standard set of axioms for FOL, assuming that we define all connectives syntactically in terms of $\wedge$ and $\neg$ (e.g., we define $(\phi \Rightarrow \psi)$ as $\neg(\phi \wedge \neg\psi)$), and define $\forall\nu\phi$ as $\neg\exists\nu\neg\phi$ (rather than giving them independent semantic definitions, as we did):

A1. $\phi \Rightarrow (\psi \Rightarrow \phi)$ (Implication Introduction)
A2. $(\phi \Rightarrow (\psi \Rightarrow \chi)) \Rightarrow ((\phi \Rightarrow \psi) \Rightarrow (\phi \Rightarrow \chi))$ (Implication Distribution)
A3. $(\neg\psi \Rightarrow \neg\phi) \Rightarrow (\phi \Rightarrow \psi)$ (Contrapositive)
Q1. $(\forall\nu(\phi \Rightarrow \psi)) \Rightarrow ((\forall\nu\phi) \Rightarrow (\forall\nu\psi))$ ($\forall$-Distribution)
Q2. $(\forall\nu\phi) \Rightarrow \phi[\nu/\tau]$ where τ is ground (Universal Instantiation)
Q3. $\phi \Rightarrow (\forall\nu\phi)$ for ν *not* free in ϕ (Universal Generalization)
I1. $\tau = \tau$ for τ a ground term (Self-Indentity)
I2. $(\sigma = \tau) \Rightarrow (\phi \Rightarrow \phi_{\sigma/\tau})$ (Substitution of Equals)

With these axioms, it turns out that we need only one rule of inference, MP. Note that UI is formulated here as an axiom, rather than a rule of inference. But we can see that this axiom, together with MP, gives us the effect of a UI inference rule. In general, there is a "tradeoff" between logical axioms and rules of inference. We can express many axioms as rules or vice versa. Logicians often restrict themselves to as few rules as possible, so that it becomes relatively easy to analyze the theoretical properties of proofs. People who actually want to *use* proof systems (e.g., AI system builders) generally find it much more convenient to design a powerful set of inference rules, with little (or no) reliance on axioms. It can be amazingly hard to prove "obvious" theorems in an axiomatic system. As a small exercise, you might try to fill in the details of the last step of the proof of the deduction theorem in the supplementary Genesereth & Nilsson text (p.59) (on reserve in the Carlson library), which says that from $(\phi \Rightarrow \chi)$ and $\phi \Rightarrow (\chi \Rightarrow \psi)$ we can prove $\phi \Rightarrow \psi$ using MP and an instance of Implication Distribution). (Hint: you need to apply MP twice.)

Proofs

Having introduced the notion of logical axioms, we can now formally define a *proof of* ϕ *from* Δ (in a forward inference system) as a sequence of closed formulas

$\phi_1,\ \phi_2,\ ...,\ \phi_n \qquad (n \geq 1)$

where $\phi_n = \phi$, and for each ϕ_i $(1 \leq i \leq n)$:

(i) $\phi_i \in \Delta$, or
(ii) ϕ_i is an instance of a logical axiom, or
(iii) ϕ_i is the result of applying a rule of inference to a subset of $\{\phi_1, ..., \phi_{i-1}\}$

It is customary to annotate proof steps with the appropriate justifications, as shown in the following example.

Example of proof

Let the premises be

Δ = {(∀x (∃y (Loves(x,y)), (∀x (Loves(John,x) ⇒ x=John))}.

We wish to prove

Loves(John,John).

Proof:

1.	(∀x (∃y (Loves(x,y))	Δ
2.	(∀x (Loves(John,x) ⇒ x = John))	Δ
3.	∃y Loves(John,y)	UI, 1
4.	Loves(John,C)	EI (Skolemization), 3
5.	Loves(John,C) ⇒ C = John	UI, 2
6.	C = John	MP, 4, 5
7.	Loves(John,John)	S=, 6, 4

Provability (derivability), theoremhood, and consistency

We say that ϕ is *provable* from a set of premises Δ, written as

$\Delta \vdash \phi$

iff there is a proof of ϕ from Δ (in the proof system under consideration). The turnstile $\vdash$ is often read as "derives".

A *theorem* ϕ of a logical system is a formula (or schema) derivable from the logical axioms alone, written as

$\vdash \phi$.

For example, all axioms are (trivially) theorems as well, and then there are other theorems such as

$\neg\neg\phi \Rightarrow \phi$, $((\phi \Leftrightarrow \psi) \Leftrightarrow (\neg\phi \Leftrightarrow \neg\psi))$, etc.

A set of formulas is *inconsistent* if it is possible to prove both ϕ and $\neg\phi$ for some formula ϕ. (**NB:** Inconsistency is a *syntactic* property of sets of formulas, relative to a proof system, not a semantic property.) An alternative definition of inconsistency involves introducing a formula $\Box$ (also written as $\bot$), called the empty formula, or falsity, and to add a rule of inference $\frac{\phi, \neg\phi}{\Box}$; then we say Δ is inconsistent iff

$\Delta \vdash \Box$.

If a set of formulas is not inconsistent (i.e., if no contradiction can be derived), then it is *consistent*.

Soundness and Completeness of Proof Systems

We have finally arrived at a very important point, where we can talk about the connection between proof systems and semantics: we can now say what it means for a proof system to be *sound* and *complete*.

The notion of a sound proof system is just a straightforward generalization of the notion of a sound rule of inference. In particular, a proof system is sound iff the only formulas we can derive from premises Δ are logical consequences of Δ. Thus soundness guarantees that we'll never draw false conclusions from true premises. In symbols, a proof system is sound iff for all sets of premises Δ and every possible conclusion ϕ,

$\Delta \vdash \phi$ implies that $\Delta \models \phi$.

Conversely, a proof system is *complete* iff whenever a set of formulas Δ entails (logically implies) ϕ, we can also derive ϕ from Δ. Thus completeness guarantees that we will never fail to derive a logical consequence of a set of premises, assuming that we use some systematic method of trying all possible proofs. In symbols, a proof system is complete iff

$\Delta \models \phi$ implies that $\Delta \vdash \phi$.

Thus for a sound and complete proof system we have

$\Delta \vdash \phi$ if and only if $\Delta \models \phi$.

It turns out that FOL does indeed have sound and complete proof systems. In fact, the axiomatic system mentioned above (with MP as its only rule of inference) is sound and complete, and there are many other sound and complete proof systems for FOL.

A slightly weaker form of completeness is *refutation completeness*. What this means is that for any unsatisfiable set of formulas, we can derive a contradiction:

If Δ is unsatisfiable, then $\Delta \vdash \Box$.

Proving soundness of a proof system is a simple matter, once we have established the soundness of the individual rules of inference. We just need to do a simple induction on

the lengths of proofs.

Proving completeness is quite another matter. It involves a construction of a model where terms denote themselves, i.e., the domain of discourse consists of "objects" like "A", "B", "John", "father-of(John)", "f(A,g(B))", etc. More accurately, when we allow for equality, the terms denote equivalence classes of terms related by equations. This ingenious construction, due to Henkin, is beyond the scope of this course. However, in connection with resolution-principle theorem proving, we may prove refutation completeness for resolution, applied to clauses without equality.

Generalized proof systems with "backward" deduction rules

When people do mathematical proofs, they rarely rely entirely on forward rules of inference. For example, the following is a familiar mode of reasoning: we are given premises Δ and wish to show $(\phi \Rightarrow \psi)$; we do this by *assuming* the antecedent ϕ, adding it to the premises, and then trying to show ψ. Our intuition is that this is sufficient to demonstrate the truth of the desired conditional formula. Symbolically, this rule of *Assumption of the Antecedent* (AA) can be written as follows:

$$AA: \ \frac{\Delta \models (\phi \Rightarrow \psi)}{\Delta, \phi \models \psi}.$$

The double line (not a standard notation!) indicates that if we wish to establish the "numerator", it is sufficient to establish the "denominator". In other words, the problem of demonstrating the relation in the numerator can be "reduced" to the problem of demonstrating the relation in the denominator. (Briefly, the numerator is implied by the denominator; note that this is the opposite as in forward inference rules!)

Some additional examples of backward rules are *And Reduction* (AR), *Or Reduction* (OR), and *Reductio ad Absurdum* (RA):

$$AR: \ \frac{\Delta \models (\phi \wedge \psi)}{\Delta \models \phi, \ \textit{and} \ \Delta \models \psi} \qquad OR: \ \frac{\Delta \models (\phi \vee \psi)}{\Delta, \neg\phi \models \psi} \qquad RA: \ \frac{\Delta \models \phi}{\Delta, \neg\phi \models \Box}$$

AR and OR should be pretty much self-explanatory. The third rule, RA, is very important, both because it is commonly used in mathematics and because it plays an important role in resolution proofs. What is says is that to establish a formula ϕ, we can assume the negation of ϕ, and hence derive a contradiction. Again, it seems intuitively clear that this is a sound method of proof. However, rather than relying on our intuitions, we need to prove such rules sound. This is left as an exercise, but the following illustrates the sort of argument we use.

Example proof of soundness for a backward rule

The following is a "less powerful" – but still sound – version of OR:

$$\frac{\Delta \models (\phi \vee \psi)}{\Delta \models \phi}.$$

Let's prove that this is sound. Since the claim is that establishing the "denominator" is sufficient to establish the "numerator", let's assume the denominator and then prove the numerator.[1] Thus we are assuming that

$\Delta \models \phi$,

i.e.,

For all models $\mathcal{M}$ such that $\models_{\mathcal{M}} \Delta$, $\models_{\mathcal{M}} \phi$.

With this assumption in mind, we now want to show

$\Delta \models (\phi \vee \psi)$,

i.e.,

For all models $\mathcal{M}$ such that $\models_{\mathcal{M}} \Delta$, $\models_{\mathcal{M}} (\phi \vee \psi)$.

So consider an arbitrary model $\mathcal{M}$ such that $\models_{\mathcal{M}} \Delta$. (This is another implicit use of Assumption of the Antecedent!) We want to show that $\models_{\mathcal{M}} (\phi \vee \psi)$. But from the two assumptions we have made, we know that

$\models_{\mathcal{M}} \phi$, i.e.,
for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} \phi[\mathcal{U}]$; hence,
for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} \phi[\mathcal{U}]$ or $\psi[\mathcal{U}]$; hence
for all v.a. $\mathcal{U}$, $\models_{\mathcal{M}} (\phi \vee \psi)[\mathcal{U}]$ (by sat'n cond's for $\vee$);
i.e., $\models_{\mathcal{M}} (\phi \vee \psi)$.

This establishes that, if the "denominator" is true, then any model $\mathcal{M}$ of Δ is a model of $(\phi \vee \psi)$ and hence completes the proof.□

Examples of proof systems

You've already seen an example of a forward-deduction system that relies heavily on axioms, namely the 8-axiom system with MP above. There is also a deduction system with 13 axiom schemas and one rule of inference, MP, in the Genesereth & Nilsson reference (on Carlson reserve, as mentioned before). The two are quite similar, except that the one in G&N is only for FOPC rather than FOL (no allowance for equality); on the other hand, it has extra axioms to deal explicitly with some of the "extra" connectives, like biconditional, rather that relying on definitional replacement of those connectives.

The system of Pelletier in the "Natural Deduction" handout is a good example of a system that relies at least as much on backward deduction as on forward rules. Proofs tend to be much easier to construct in such a system. Note that no axioms at all are used. Also note that the formal definition of a proof becomes more complex than in the

[1] We are *using* Assumption of Antecedent in doing so. This may seem strange – why prove these rules sound if in doing so we need to *use* them? The answer is that we are doing metalevel reasoning *about* a formally defined operation. The formally defined operation, in this case, is the backward inference rule above. Naturally, in proving *anything* – even if it is a claim about formal rules of inference – we have to *use* some rules of inference!

case of a forward deduction system. (We won't attempt such a definition.)

Our next topic will be theorem proving based on the *resolution principle*. In this approach to deduction, we work in two stages, in effect using two different languages. We start off with an unconstrained first-order language for expressing the premises and desired conclusion of a theorem-proving problem. We then apply the *reductio ad absurdum* rule exactly once, adding the denial of the conclusion to the premises. Then we convert to a quantifier-free form called *clause form*, and from that point on we try to derive the empty clause $\square$, using forward inference based *resolution*, plus possibly *factoring* and *paramodulation*. Resolution is a generalization of MTP ("cancelling" contradictory parts of two clauses), while factoring is a kind of simplification and paramodulation is a generalized form of substitution of equals. If there are no occurrences of equality in our premises or desired conclusion, we get by entirely without axioms. If we want to reason with equality, we sometimes need the self-identity axiom

x = x (with the variable x undestood to be universally quantified).

Resolution principle theorem proving is relatively easy to implement, and resolution proofs are also fairly natural and easy to follow, at least compared with proofs in axiomatic systems with just a weak rule like MP. This accounts for its popularity in AI, including its use as a basis for the Prolog programming language.