1 The Soundness of the Fitch Proof System

Early in the course we discussed how, by the method of truth tables, we can decide whether any conclusion \( C \) is a tautological consequence of a set of premises \( P_1, \ldots, P_n \). We only have to make a truth table including all the premises and the conclusion and then check whether the conclusion is true whenever all the premises are. We also noted, however, how much work this can be. If there are, say, 10 different atomic sentences appearing in \( P_1, \ldots, P_n \), and \( C \), then our truth table must have \( 2^{10} = 1024 \) rows.

One reason for introducing the Fitch proof system is to provide a means for generating good (i.e., valid) arguments without having to go through the process of drawing up enormous truth tables like this. You may have asked yourself however, “How can we be sure this proof system really is producing valid arguments?” That is, we have a number of rules that say things like, “If you have \( A \land B \) in your proof somewhere, you can conclude \( A \) by \( \land \) Elimination,” and so on, but is there some way of showing that the result of applying these proof rules any finite number of times must always result in a valid argument? In other words, do these rules really represent inference rules, rather than arbitrary rules for manipulating strings of symbols?

The answer is “Yes,” and a proof that a proof system has this property is said to show that the proof system in question is sound. This is often written as follows

\[ P_1, \ldots, P_n \vdash C \text{ implies } P_1, \ldots, P_n \models C, \]

where ‘\( P_1, \ldots, P_n \vdash C \)’ is read, “There is a proof of \( C \) using \( P_1, \ldots, P_n \) as premises,” and ‘\( P_1, \ldots, P_n \models C \)’ is read, “\( C \) is a tautological consequence of
So, in a sound proof system, if I can prove $C$ from premises $P_1, \ldots, P_n$, then $C$ really does follow validly from $P_1, \ldots, P_n$. How can we prove something like this?

Proofs of this kind of fact all require some version of the principle of mathematical induction, which we’ll discuss more later in the course. The form this principle takes in relation to the proof of soundness offered in the textbook is as follows.

**Proposition 1.1. (Induction Principle)** If $C$ is provable from premises $P_1, \ldots, P_n$ but $C$ doesn’t follow validly from $P_1, \ldots, P_n$, then some step in our proof must produce a sentence that doesn’t follow validly from $P_1, \ldots, P_n$. (The inference that produces $C$, for example, is one such step.) So, there must be a first such misstep.

Call a step in a Fitch proof that produces a conclusion that doesn’t follow validly from the previous steps a bad step. Our proof proceeds by looking at each of the Fitch proof rules and showing that an application of that rule never produces the first bad step. Since there is never a first bad step, there must not be a bad step at all. (If there were, there would have to be a first one by our induction principle: contradiction.)

**Theorem 1.2. (The Soundness Theorem)** $P_1, \ldots, P_n \vdash C$ implies $P_1, \ldots, P_n \models C$.

*Proof.* Suppose our proof has gone along producing only good steps so far. Our job is to make sure that the next step—whatever it may be—is just another good step. This will ensure that there is no first bad step. There are many cases to consider (one for each introduction and elimination rule), but I’ll only go through a few of them to give you the idea of how to do the rest if you’re interested.

Case 1 – $\wedge$ Intro: Suppose our first step after all the good ones that have been taken so far concludes $A \wedge B$ by the $\wedge$ Intro rule. Then $A$ and $B$ must occur earlier in the proof (since we need each of them to be there in order to apply the $\wedge$ Intro rule). Since $A$ and $B$ each appear during the good part of the proof, they each must be tautological consequences of $P_1, \ldots, P_n$. That is, if all of $P_1, \ldots, P_n$ are true, then $A$ and $B$ must both be true too. But that implies that $A \wedge B$ is also true, so $A \wedge B$ is a tautological consequence of $P_1, \ldots, P_n$ if $\wedge$ Intro can be applied. Therefore, an application of $\wedge$ Intro doesn’t produce the first bad step.
Case 2 – $\wedge$ Elim: Suppose that the next step in our proof after the good ones so far concludes $A$ by $\wedge$ Elim. Then something of the form $A \wedge B$ must occur earlier in the proof. And since it occurred in the good part of the proof, whenever all of $P_1, \ldots, P_n$ are true, so is $A \wedge B$. But $A \wedge B$ is only true when $A$ is true as well. So, $A$ follows validly from $P_1, \ldots, P_n$, and $\wedge$ Elim doesn’t produce the first bad step either.

Case 3 – $\rightarrow$ Intro: Suppose the next step in our proof concludes $A \rightarrow B$ by $\rightarrow$ Intro. Then we must have a subproof earlier on that starts with $A$ and ends with $B$. If $P_1, \ldots, P_n$ are all true and $A \rightarrow B$ is false (i.e., if this is a bad step), we must have $A$ true and $B$ false. But then, there must have been an earlier bad step: In fact, the conclusion of $B$ in our earlier subproof must’ve been a bad step since $B$ is not a tautological consequence of the $P_i$ and $A$. But that’s not possible because by assumption all the earlier steps in our proof were good.

Case 4 – $\rightarrow$ Elim: Suppose the next step in our proof concludes $B$ by $\rightarrow$ Elim. Then we must have something of the form $A \rightarrow B$ and $A$ both occurring earlier in the proof. So, whenever $P_1, \ldots, P_n$ are all true, $A \rightarrow B$ and $A$ must be true as well. Suppose $P_1, \ldots, P_n$ are all true and $B$ is false. Then since we already know that $A$ is true, if $B$ is false, then $A \rightarrow B$ is false, which contradicts the fact that $A \rightarrow B$ must be true when $P_1, \ldots, P_n$ are all true. Therefore, $B$ is true as well, and $\rightarrow$ Elim doesn’t make the first bad step.

The rest of the cases use similar reasoning and you should be able to work them out on your own.

2 The Completeness of the Fitch Proof System

The Soundness Theorem says that if we can derive $C$ from the premises $P_1, \ldots, P_n$, then $C$ is a tautological consequence of $P_1, \ldots, P_n$. The next natural question to ask is whether anytime we know that $C$ follows validly from $P_1, \ldots, P_n$ we can actually give a proof of $C$ in our system using only $P_1, \ldots, P_n$ as premises. That is, we want to know if $P_1, \ldots, P_n \models C$ implies $P_1, \ldots, P_n \vdash C$. Showing this to be the case is called showing our proof system to be complete—in other words, it can prove everything we’d hope it can prove. This is a much more difficult question
to answer since we don’t know anything about \( C \) or \( P_1, \ldots, P_n \) besides the fact if all of \( P_1, \ldots, P_n \) are true, \( C \) must be too. How can this information be turned into a proof of \( C \)?

First let’s reduce our problem to a slightly simpler case. We’ll actually be done if we can show that every tautology is provable. That is, suppose that every tautology can be proven—we’ll call this the Tautology Theorem—then we can prove the following.

**Theorem 2.1.** *(The Completeness Theorem)*

\[
P_1, \ldots, P_n \models C \text{ implies } P_1, \ldots, P_n \vdash C.
\]

*Proof.* Suppose that \( C \) follows validly from \( P_1, \ldots, P_n \). Then \( P_1 \rightarrow \ldots \rightarrow P_n \rightarrow C \) is a tautology. Therefore, by the Tautology Theorem, \( \vdash P_1 \rightarrow \ldots \rightarrow P_n \rightarrow C \). Therefore, given \( P_1, \ldots, P_n \) as premises, \( C \) can be proved by \( n \) applications of the \( \rightarrow \) Elim rule.

By this line of reasoning, if we want to show that

\[
P_1, \ldots, P_n \models C \text{ implies } P_1, \ldots, P_n \vdash C,
\]

we can get away with just proving the Tautology Theorem.

There are many ways to attack that problem, but the one that requires the least amount of setup and background knowledge is based on an idea going back to Emil Post that proceeds by showing that any tautology can be shown to be provably equivalent to some other formula that we know how to prove.

**Definition 2.2.** A sentence \( A \) is called a *literal* if it is either an atomic formula or the negation of an atomic formula. A sentence \( A \) is in *disjunctive normal form* (DNF) if it is a disjunction of conjunctions of literals.

**Definition 2.3.** A sentence \( A \) is in *full* disjunctive normal form if it is in DNF and each atomic sentence that occurs in \( A \) occurs in all of the conjuncts making up the disjunction.
Example 2.4. ¬P ∨ ¬Q ∨ (P ∧ Q) is in DNF, but because P and Q don’t both occur in every one of the conjuncts that make up the disjunction, it is not in full DNF. (¬P ∧ Q) ∨ (P ∧ ¬Q) ∨ (P ∧ Q) ∨ (¬P ∧ ¬Q), however, is in full DNF. This sentence also has the same truth table as the first one.

Given any sentence C of our language, there is an algorithm for producing another sentence C* of our language that has the same truth table as C, but is in DNF. The algorithm proceeds by first putting C into negation normal form, where negations only immediately precede atomic sentences; e.g., ¬Cube(a) ∧ Tet(b) is in negation normal form, but ¬(Cube(a) ∧ Tet(b)) is not. We do this by first using the following identity if our sentence contains any →s.

¬P ∨ Q ↔ P → Q

We can then use de Morgan’s laws, which say the following, to push negations inwards.

¬(P ∧ Q) ↔ ¬P ∨ ¬Q

¬(P ∨ Q) ↔ ¬P ∧ ¬Q

Finally, we can use the laws for distributing ∧ over ∨ and vice versa (stated below) to produce something in DNF.

P ∧ (Q ∨ R) ↔ (P ∧ Q) ∨ (P ∧ R)

P ∨ (Q ∧ R) ↔ (P ∨ Q) ∧ (P ∨ R)

Example 2.5. (P ∧ Q) → (R ∧ P) ↔ ¬(P ∧ Q) ∨ (R ∧ P) ↔ ¬P ∨ ¬Q ∨ (R ∨ P)

The sign ↔ here only indicates that the formulas on either side of it are tautologically equivalent, but we can actually prove each of these laws in Fitch. That is, we have each of the following.

⊢ (¬P ∨ Q) ↔ (P → Q)

⊢ ¬(P ∧ Q) ↔ ¬P ∨ ¬Q

⊢ ¬(P ∨ Q) ↔ ¬P ∧ ¬Q

⊢ P ∧ (Q ∨ R) ↔ (P ∧ Q) ∨ (P ∧ R)
So, if we start from some sentence $C$ and apply the alternative definition of the conditional, de Morgan’s laws, and distribution rules to get to a sentence $C^*$ in DNF, we can use the fact that we can prove each of those laws and rules in Fitch to conclude $C^*$ formally in Fitch. That is, $C \vdash C^*$.

Note, however, that each of de Morgan’s laws and the distribution rules has biconditional form. So, if, starting from $C$, we can reach $C^*$ along a chain of these biconditionals, we can also go the other direction to prove $C$ from $C^*$ in Fitch as well. Therefore, $C^* \vdash C$.

Since this is just an overview of the Completeness Theorem, I’ll just state without proof that this process can be extended to put $C$ into full DNF, and that any two formulas with the same truth table and containing the same atomic sentences have the same full DNF.

**Proposition 2.6.** Suppose $C$ and $D$ are sentences of our language with the same truth tables and the same atomic sentences making them up. Then

1. $C \vdash C^*$ and $C^* \vdash C$
2. $D \vdash C^*$ and $C^* \vdash D$,

where $C^*$ is the unique full DNF of both $C$ and $D$.

The following result will let us turn facts about what we can prove with premises into things we can prove without premises.

**Theorem 2.7.** (The Deduction Theorem) If $A \vdash B$, then $\vdash A \rightarrow B$.

*Proof.* $A \vdash B$ means that by taking $A$ as a premise, $B$ is provable. Well, if instead of using $A$ as a premise, we were to make an assumption of $A$ by starting a new subproof, all the same reasoning that allowed for the conclusion $B$ in the proof of $B$ from the premise $A$ can be used to conclude $B$ within the subproof assuming $A$. Therefore, we can conclude $A \rightarrow B$ by the $\rightarrow$ Intro rule.

We can now use all of these facts to finish the proof of the Completeness Theorem by proving the Tautology Theorem.

**Theorem 2.8.** (The Tautology Theorem) If $C$ is a tautology, $\vdash C$. 
Proof. Suppose that $C$ is a tautology and let $A_1, \ldots, A_n$ be a list of all the atomic formulas that occur in $C$. Let $D$ be the sentence $(A_1 \lor \neg A_1) \land \ldots \land (A_n \lor \neg A_n)$. Then $D$ is also a tautology containing the same atomic sentences as $C$, so by Proposition 2.6, there is a sentence $C^*$ in full DNF such that $D \vdash C^*$ and $C^* \vdash C$.

By the Deduction Theorem, $\vdash D \rightarrow C^*$ and $\vdash C^* \rightarrow C$. Since we can easily prove the law of excluded middle in our Fitch system, by using $\land$ Intro, we can conclude,

$$\vdash (A_1 \lor \neg A_1) \land \ldots \land (A_n \lor \neg A_n);$$

that is, $\vdash D$.

Therefore, by two applications of the $\rightarrow$ Elim rule to $D \rightarrow C^*$ and $C^* \rightarrow C$, we have $\vdash C$. So, any tautology is provable in Fitch. $\dashv$