Week 10

Notes on natural deduction for first order logic

Instructor: Matteo de Ceglie

28 May 2020

Contents

1 Introduction

The last argument of this first introduction to logic is natural deduction for first order logic. We will cover it in these last two week. This week I will introduce the rules, while next week I will give you some more examples of derivations.

First order semantics and propositional logic semantics are two very different things, they don't operate under the same rules or principles (as we have seen, propositional logic semantics is centred around truth tables, while first order semantics around models). There is no such thing as a truth table for first order quantifiers, just like there is no such thing as a model for propositional logic. On the other hand, natural deduction for propositional logic and first order logic are essentially the same thing: all the rules and methods introduced for propositional logic are still there, and we are just adding some new rules for the quantifiers. How these rules are applied and the general principles of the system are the same (in the same way natural deduction for minimal logic and for classical propositional logic are the same).

In the following notes I will introduce the new rules for this new system, that we call *first order classical logic* (in symbols, I will write $\vdash_{FOL}$ for the proofs carried out in this system). The definition of assumptions, derivation and all the other rules are the one already introduced in week 4 and 5. In particular, we are adding all this rule to our propositional classical logic system, so all the rules of that system are already present (see week 5).

2 Rules

2.1 Universal Elimination

The first rule we add to our system is the one of Universal Elimination. With this rule, we eliminate a universal quantifier from our proof.

Suppose that we know that everything is red. Then, we can infer that a particular thing, let's say a ball, is red:

(1)	Everything is red.
(2)	This ball is red.

This is a valid argument and a valid inference: if we accept the premise, we cannot avoid accepting also the conclusion.

We can start formalising this inference, and say that from "All x are P" we can derive "a is P":

(1)	All x are P.
(2)	a is P.

What we have done here is dropping dropping the universal quantifier ("All") in (1) and substituting the variable x with a constant a. In even more formal terms:

(1)	$\forall x[P(x)]$
(2)	$P(a)$.

The actual rule is written as follows:

$$\frac{\forall x[\varphi(x)]}{\varphi(a)} \langle \forall \mathrm{E} \rangle$$

This rule say that if we have a universal quantifier, we can eliminate it by dropping it and then substituting all the variables that were in his scope in the formula φ (in this case, x) with the same constant (in this case a). So for example if φ is $P(x) \lor Q(x)$, the rule would look like this:

$$\frac{\forall x[P(x) \lor Q(x)]}{P(a) \lor Q(a)} \langle \forall \mathrm{E} \rangle$$

An this could be interpreted, for example, as:

(1)	Everything is a fish or is a man.
(2)	Socrates is a fish or is a man.

Please note that we can substitute only the *free* variables: in $\forall x[P(x) \lor Q(x)]$ there are none, but since we are dropping the quantifier in $[P(x) \lor Q(x)]$ is actually free in both occurrences. However, if there is another variable with another quantifier, we cannot substitute that:

$$\frac{\forall x[\exists y[P(x) \lor Q(x, y)]]}{\exists y[P(a) \lor Q(a, y)]} \langle \forall \mathrm{E} \rangle$$

For the same reason, from the formula $\forall x[\exists y[P(x, y) \land \forall x[Q(x, y, g)]]]$ we can derive $\exists y[P(a, y) \land \forall x[Q(x, y, g)]]$, but not $\exists y[P(a, y) \land \forall x[Q(a, y, g)]]$, since the second x is not free (there is the other universal quantifier).

As another example, we can infer “If Clint is Austrian, then he owns a car" from “Every Austrian own a car":

(1)	Every Austrian owns a car.
(2)	If Clint is Austrian, then he owns a car.

And this, in symbols, would be:

$$\frac{\forall x[A(x) \to C(x)}{A(c) \to C(c)} \langle \forall \mathrm{E} \rangle$$

2.2 Existential Introduction

The second rule we need is the introduction of the existential quantifier. From the assumption that some particular thing is P, we can derive that there exists something that is P. So for example, from “Socrates is immortal" we can infer “Somebody is immortal":

(1)	Socrates is immortal.
(2)	Somebody is immortal.

Or, in more general terms:

(1)	a is P
(2)	There exists x that is P.

From this valid argument we can formulate the following rule:

$$\frac{\varphi(a)}{\exists x[\varphi(x)]}\ \langle\exists\text{I}\rangle$$

Here we substitute all the free occurrences of x with a (but in this case in the opposite direction). Consequently all the restriction that applied in that other case still apply. For example, if φ is $P(x) \vee Q(x, y)$, the following is a correct application of the rule:

$$\frac{P(a) \wedge Q(a)}{\exists x[P(x) \wedge Q(x)]}\ \langle\exists\text{I}\rangle$$

And in natural language we are maybe deriving that “There exists an immortal philosopher" from “Socrates is immortal and is a philosopher".

For another example, we can derive “Somebody loves oneself" from “Narcissus loves himself", with just an application of existential introduction:

$$\frac{L(n, n)}{\exists x[L(x, x)}\ \langle\exists\text{I}\rangle$$

But what if we want to derive "Narcissus loves somebody" from “Narcissus loves himself"? We can still do it, but we need to be cautious. We can do it if and only if φ is $L(n, x)$. So, in this case, we derive:

$$\frac{L(n, n)}{\exists x[L(n, x)]}\ \langle\exists\text{I}\rangle$$

However, from this point there is nothing we can do to derive again $\exists x[L(x, x)]$! If we re-apply the existential introduction, we cannot end up with $\exists x \exists x[L(x, x)]$, but only with $\exists x \exists x[L(n, x)]$ or, less trivially, with $\exists y \exists x[L(y, x)]$. To understand the reason why, remember that the rule asks us to substitute *all* the occurrences of x with a. If we derive $\exists x[L(n, x)]$ from $\exists x[L(n, n)]$, it means that the first n is not the consequences of such a substitution, so it cannot be substituted that way in a subsequent rule!

2.3 A first example

As a first example of a full derivation, suppose we want to prove that “Someone loves somebody else" from "Everyone loves oneself". While this may seem strange, it is possible! First of all we instantiate “Everyone loves oneself":

(1)	Everyone loves oneself.
(2)	Narcissus loves himself.

From this, we can derive that Narcissus loves someone:

(1)	Everyone loves oneself.
(2)	Narcissus loves himself.
(3)	Narcissus loves someone.

And finally, that someone loves somebody else:

(1)	Everyone loves oneself.
(2)	Narcissus loves himself.
(3)	Narcissus loves someone.
(4)	Somebody loves somebody else.

Such a derivation, while "weird" in the real world, makes perfect sense in our natural deduction system, since it is the correct application of the two rules introduces:

$$\dfrac{\dfrac{\dfrac{\forall x[L(x,x)]}{L(n,n)}\langle\forall\mathrm{E}\rangle}{\exists y[L(n,y)]}\langle\exists\mathrm{I}\rangle}{\exists x[\exists y[L(x,y)]]}\langle\exists\mathrm{I}\rangle$$

And this is the derivation $\forall x[L(x,x)] \vdash_{FOL} \exists x[\exists y[L(x,y)]]$.

Note that when applying a rule you do not need to look/remember what a particular constant was: after it is a constant, you can substitute in the following rule as you want (in this case, even though both n derived from x, we derived first an y and after that and x).

2.4 A second example

It is possible to combine these two rules for the quantifiers with the rules for propositional classical logic. So for example, it is possible to prove φ from $F(a)$ and $\neg\exists x[F(x)]$, i.e. $F(a), \neg\exists x[F(x)] \vdash_{FOL} \varphi$:

$$\dfrac{\dfrac{\dfrac{F(a)}{\exists x[F(x)]}\langle\exists\mathrm{I}\rangle \quad \neg\exists x[F(x)]}{\bot}\langle\neg\mathrm{E}\rangle}{\varphi}\langle\bot\mathrm{E}\rangle$$

Note that all the previous rules still function as expected, in particular the ones about the discharge of assumptions.

2.5 Universal Introduction

The next rule we need to discuss is universal introduction. This is a little more complex, and has some restriction. In its general form is very similar to existential introduction: from the assumptions that a particular thing is P, we can infer that everything is P, as before substituting every occurrence of the variable with a constant. Wait, this does not seem right! From "Socrates is a philosopher" I cannot derive "Everyone is a philosopher". This is the reason why some restriction are needed: we need a way to stop deriving "Everyone loves Aristotle" from "Everyone loves themselves".

First of all, consider that the rule formulation is the following:

$$\dfrac{\varphi(a)}{\forall x[\varphi(x)]}\langle\forall\mathrm{I}\rangle$$

It would seem that those derivations we don't want, like:

(1)	Socrates is a philosopher.
(2)	Everyone is a philosopher.

are actually possible! They are possible, from a certain point of view, but the restriction in place to avoid them are so strong that it is not possible to derive unwanted conclusions.

The first restriction is the usual one: when introducing the new quantifier, we substitute every occurrence of the variable x with the constant a. So for example, from $P(a)$ we can derive $\forall x[P(x)]$, but from $P(a, a)$ we cannot derive $\forall x[P(x, a)]$ and then $\forall x \forall x[P(x, x)]$, just like in the case of existential introduction.

The second restriction is the new one and the important one: we cannot substitute a variable x in φ if it occurs *bound* (i.e. in the scope of a quantifier) in any discharged assumption before φ. This means that we can apply this rule only on thing we didn't already mentioned.

In our example above, we cannot do the following:

(1)	Everyone loves Aristotle.
(2)	Aristotle loves himself.
(3)	Everyone loves themselves.

This is because we are trying to substitute "Aristotle", but since he appears in our first assumption, this is not possible. In symbols this is more clear:

$$\dfrac{\dfrac{\forall x[L(x, a)]}{L(a, a)}\langle\forall\mathrm{E}\rangle}{\forall x[L(x, x)]}\langle\forall\mathrm{I}\rangle$$

Here the last application of universal introduction is not a correct one, since we already have that constant a in the undischarged assumption.

2.6 Existential Elimination

The last rule we need for our natural deduction system for first order logic is existential elimination. With this rule we eliminate an existential quantifier from our proof.

This rule is very similar to Disjunction Elimination. Just like in that case, where we could not derive φ from $\varphi \vee \psi$, here we cannot derive $P(a)$ from $\exists x[P(x)]$, since we have no way to know if it is really a that is P, and not b or something else. So how to proceed?

The only way forward is to reason like we did for Disjunction Elimination: if we manage to prove "Socrates is a philosopher" from "Socrates is immortal", and we know that there exists something that is actually immortal, then we can derive that Socrates is a philosopher. In formalised terms, if we prove φ from "a is P", and we have that "Something is P", then we can derive φ without any doubts.

Thus our rule will look like this:

$$\dfrac{\exists x[\varphi(x)] \qquad \begin{array}{c}\underline{[\varphi(a)]^n}\\ \vdots \\ \overline{\psi}\end{array}}{\psi}\langle\exists\mathrm{E}\rangle, n$$

Here, like in all the other rules for propositional logic, we need to replace the dots in the rule with an actual proof, with the correct application of rules.

What this rule means is that if we know that $\exists x[P(x)]$ we can then reason temporarily about an arbitrary element a that is P in order to prove a conclusion that does not depend on y. In natural language, an argument of this form would be "We know that there exists something red and round. Suppose that this glass is red and round. If that is the case, then we can say that it is just red. Then we can infer that there is something red. But since we know that there is something that is red and round, we can further infer that there actually is something red. In conclusion, if there is something red and round, then there is something red." While this could seem convoluted, the formalisation is clearer:

$$
\cfrac{\cfrac{[\exists[P(x) \land Q(x)]]^1 \qquad \cfrac{\cfrac{[P(a) \land P(b)]^2}{P(a)}\ \langle \land \mathrm{E} \rangle}{\exists x[P(x)]}\ \langle \exists \mathrm{I} \rangle}{\exists x[P(x)]}\ \langle \exists \mathrm{E} \rangle{,}2}{\exists x[P(x) \land Q(x)] \to \exists x[P(x)]}\ \langle \to \mathrm{I} \rangle{,}1
$$