base. A knowledge-base is composed of sentences in a language with a truth theory (logic), so that someone external to the system can interpret sentences as statements about the world (semantics). Thus, to express knowledge, we need a precise, *declarative* language. By a *declarative language*, we mean that

1. The system believes a statement S *iff* it considers S to be true, since one cannot believe S without an idea of what it means for the world to fulfill S.
2. The knowledge-based must be precise enough so that we must know, (1) which symbols represent sentences, (2) what it means for a sentence to be true, and (3) when a sentence follows from other sentences.

Two declarative languages will be discussed in this chapter: (0 order or) propositional logic and first order logic.

1.2.1 Propositional Logic

Formal logic is concerned with statements of fact, as opposed to opinions, commands, questions, exclamations *etc.* Statements of fact are assertions that are either true or false, the simplest form of which are called *propositions*. Here are some examples of propositions:

The earth is flat.
Humans are monkeys.
$1 + 1 = 2$

At this stage, we are not saying anything about whether these are true or false: just that they are sentences that are one or the other. Here are some examples of sentences that are not propositions:

Who goes there?
Eat your broccoli.
This statement is false.

It is normal to represent propositions by letters like $P, Q, \ldots$. For example, P could represent the proposition 'Humans are monkeys.' Often, simple statements of fact are insufficient to express complex ideas. *Compound statements* can be combining two or more propositions with *logical connectives* (or simply, connectives). The connectives we will look at here will allow us to form sentences like the following:

It is <u>not</u> the case that P
P <u>and</u> Q
P <u>or</u> Q
P <u>if</u> Q

The P's and Q's above are propositions, and the words underlined are the connectives. They have special symbols and names when written formally:

Statement	Formally	Name
It is not the case that P	$\neg P$	Negation
P and Q	$P \wedge Q$	Conjunction
P or Q	$P \vee Q$	Disjunction
P if Q	$P \leftarrow Q$	Conditional

There is, for example, a form of argument known to logicians as the *disjunctive syllogism*. Here is one due to the Stoic philosopher Chrysippus, about a dog chasing a rabbit. The dog arrives at a fork in the road, sniffs at one path and then dashes down the other. Chrysippus used formal logic to describe this:[3]

Statement	Formally
The rabbit either went down Path A or Path B.	$P \vee Q$
It did not go down Path A.	$\neg P$
Therefore it went down Path B.	$\therefore Q$

Here P represents the proposition 'The rabbit went down Path A' and Q the proposition 'The rabbit went down Path B.' To argue like Chrysippus requires us to know how to write correct logical sentences, ascribe truth or falsity to propositions, and use these to derive valid consequences. We will look at all these aspects in the sections that follow.

Syntax

Every language needs a *vocabulary*. For the language of propositional logic, we will restrict the vocabulary to the following:

Propositional symbols:	$P, Q, \ldots$
Logical connectives[4] **:**	$\neg, \wedge, \vee, \leftarrow$
Brackets:	$($, $)$

The next step is to specify the rules that decide how legal sentences are to be formed within the language. For propositional logic, legal sentences or *well-formed formulæ* (wffs for short) are formed using the following rules:

1. Any propositional symbol is a wff;
2. If α is a wff then $\neg\alpha$ is a wff; and

[3]There is no suggestion that the principal agent in the anecdote employed similar means of reasoning.

3. If α and β are wffs then $(\alpha \wedge \beta), (\alpha \vee \beta)$, and $(\alpha \leftarrow \beta)$ are wffs.

Wffs consisting simply of propositional symbols (Rule 1) are sometimes called *atomic* wffs and others *compound* wffs Informally, it is acceptable to drop outermost brackets. Here are some examples of wffs and 'non-wffs':

Formula	Comment
$(\neg P)$	Not a wff. Parentheses are only allowed with the connectives in Rule 3
$\neg\neg P$	P is wff (Rule 1), $\neg P$ is wff (Rule 2), $\therefore\ \neg\neg P$ is wff (Rule 2)
$(P \leftarrow (Q \wedge R))$	P, Q, R are wffs (Rule 1), $\therefore\ (Q \wedge R)$ is a wff (Rule 3), $\therefore\ (P \leftarrow (Q \wedge R))$ is a wff (Rule 3)
$P \leftarrow (Q \wedge R)$	Not a wff, but acceptable informally
$((P) \wedge (Q))$	Not a wff. Parentheses are only allowed with the connectives in Rule 3
$(P \wedge Q \wedge R)$	Not a wff. Rule 3 only allows two symbols within a pair of brackets

One further kind of informal notation is widespread and quite readable. The conditional $(P \leftarrow ((Q_1 \wedge Q_2) \ldots Q_n)))$ is often written as $(P \leftarrow Q_1, Q_2, \ldots Q_n)$ or even $P \leftarrow Q_1, Q_2, \ldots Q_n$.

It is one thing to be able to write legal sentences, and quite another matter to be able to assess their truth or falsity. This latter requires a knowledge of semantics, which we shall look at shortly.

Normal Forms

Every formulae in propositional logic is equivalent to a formula that can be written as a conjunction of disjunctions. That is, something like $(A \vee B) \wedge (C \vee D) \wedge \cdots$. When written in this way the formula is said to be in *conjunctive normal form* or CNF. There is another form, which consists of a disjunction of conjunctions, like $(A \wedge B) \vee (C \wedge D) \vee \cdots$, called the *disjunctive normal form* or DNF. In general, a formula F in CNF can be written somewhat more cryptically as:

$$F = \bigwedge_{i=1}^{n} \left(\bigvee_{j=1}^{m} L_{i,j} \right)$$

and a formula G in DNF as:

$$G = \bigvee_{i=1}^{n} \left(\bigwedge_{j=1}^{m} L_{i,j} \right)$$

Here, $\bigvee F_i$ is short for $F_1 \vee F_2 \vee \cdots$ and $\bigwedge F_i$ is short for $F_i \wedge F_2 \wedge \cdots$. In both CNF and DNF forms above, the $L_{i,j}$ are either propositions or negations of propositions (we shall shortly call these "literals").

Semantics

There are three important concepts to be understood in the study of semantics of well-formed formulæ: *interpretations*, *models*, and *logical consequence*.

Interpretations

For propositional logic, an *interpretation* is simply an assignment of either *true* or *false* to all propositional symbols in the formula. For example, given the wff $(P \leftarrow (Q \wedge R))$ here are two different interpretations:

	P	Q	R
I_1 :	true	false	true
I_2 :	false	true	true

You can think of I_1 and I_2 as representing two different 'worlds' or 'contexts'. After a moment's thought, it should be evident that for a formula with N propositional symbols, there can never be more than 2^N possible interpretations.

Truth or falsity of a wff only makes sense given an interpretation (by the principle of bivalence, any interpretation can only result in a wff being either *true* or *false*). Clearly, if the wff simply consists of a single propositional symbol (recall that this was called an atomic wff), then the truth-value is simply that given by the interpretation. Thus, the wff P is *true* in interpretation I_1 and *false* in interpretation I_2. To obtain the truth-value of compound wffs like $(P \leftarrow (Q \wedge R))$ requires a knowledge the semantics of the connectives. These are usually summarised in a tabular form known as *truth tables*. The truth tables for the connectives of interest to us are given below.

Negation. Let α be a wff[5]. Then the truth table for $\neg\alpha$ is as follows:

α	$\neg\alpha$
false	true
true	false

Conjunction. Let α and β be wffs. The truth table for $(\alpha \wedge \beta)$ is as follows:

[5] We will use Greek characters like α, β to stand generically for any wff.

α	β	$(\alpha \wedge \beta)$
false	false	false
false	true	false
true	false	false
true	true	true

Disjunction. Let α and β be wffs. The truth table for $(\alpha \vee \beta)$ is as follows:

α	β	$(\alpha \vee \beta)$
false	false	false
false	true	true
true	false	true
true	true	true

Conditional. Let α and β be wffs. The truth table for $(\alpha \leftarrow \beta)$ is as follows:

α	β	$(\alpha \leftarrow \beta)$
false	false	true
false	true	false
true	false	true
true	true	true

We are now in a position to obtain the truth-value of a compound wff. The procedure is straightforward: given an interpretation, we find the truth-values of the smallest 'sub-wffs' and then use the truth tables for the connectives to obtain truth-values for increasingly complex sub-wffs. For $(P \leftarrow (Q \wedge R))$ this means:

1. First, obtain the truth-values of P, Q, R using the interpretation;
2. Next, obtain the truth-value of $(Q \wedge R)$ using the truth table for 'Conjunction' and the truth-values of Q and R (Step 1); and
3. Finally, obtain the truth-value $(P \leftarrow (Q \wedge R))$ using the truth table for 'Conditional' and the truth-values of P (Step 1) and $(Q \wedge R)$ (Step 2).

For the interpretations I_1 and I_2 earlier these truth-values are as follows:

	P	Q	R	$(Q \wedge R)$	$(P \leftarrow (Q \wedge R))$
I_1 :	true	false	true	false	true
I_2 :	false	true	true	true	false

Thus, $(P \leftarrow (Q \wedge R))$ is *true* in interpretation I_1 and *false* in I_2.

Models

Every interpretation (that is, an assignment of truth-values to propositional symbols) that makes a well-formed formula *true* is said to be a *model* for that formula. Take for example, the two interpretations I_1 and I_2 above. We have already seen that I_1 is a model for $(P \leftarrow (Q \wedge R))$; and that I_2 is not a model for the same formula. In fact, I_1 is also a model for several other wffs like: P, $(P \wedge R)$, $(Q \vee R)$, $(P \leftarrow Q)$, *etc*. Similarly, I_2 is a model for Q, $(Q \wedge R)$, $(P \vee Q)$, $(Q \leftarrow P)$, *etc*.

As another example, let $\{P, Q, R\}$ be the set of all atoms in the language, and α be the formula $((P \wedge Q) \leftrightarrow (R \rightarrow Q))$. Let I be the interpretation that makes P and R true, and Q false (so $I = \{P, R\}$). We determine whether α is true or false under I as follows:

1. P is true under I, and Q is false under I, so $(P \wedge Q)$ is false under I.
2. R is true under I, Q is false under I, so $(R \rightarrow Q)$ is false under I.
3. $(P \wedge Q)$ and $(R \rightarrow Q)$ are both false under I, so α is true under I.

Since α is true under I, I is a model of α. Let $I' = \{P\}$. Then $(P \wedge Q)$ is false, and $(R \rightarrow Q)$ is true under I'. Thus α is false under I', and I' is not a model of α.

The definition of model can be extended to a set of formulæ; an interpretation I is said to be a model of a set of formulæ Σ if I is a model of all formulæ $\alpha \in \Sigma$. Σ is then said to have I as a model. We will offer an example to illustrate this extended definition. Let $\Sigma = \{P, (Q \vee I), (Q \rightarrow R)\}$, and let $I = \{P, R\}$, $I' = \{P, Q, R\}$, and $I" = \{P, Q\}$ be interpretations. I and I' satisfy all formulas in Σ, so I and I' are models of Σ. On the other hand, $I"$ falsifies $(Q \rightarrow R)$, so I" is not a model of Σ.

At this point, we can distinguish amongst two kinds of formulæ:

1. A wff may be such that *every* interpretation is a model. An example is $(P \vee \neg P)$. Since there is only one propostional symbol involved (P), there are at most $2^1 = 2$ interpretations possible. The truth table summarising the truth-values for this formula is:

	P	$\neg P$	$(P \vee \neg P)$
I_1 :	false	true	true
I_2 :	true	false	true

   $(P \vee \neg P)$ is thus *true* in every possible 'context'. Formulæ like these, for which every interpretation is a model are called *valid* or *tautologies*

2. A wff may be such that *none* of the interpretations is a model. An example is $(P \wedge \neg P)$. Again there is only one propostional symbol involved (P), and thus only two interpretations possible. The truth table summarising the truth-values for this formula is:

	P	$\neg P$	$(P \wedge \neg P)$
I_1 :	false	true	false
I_2 :	true	false	false

$(P \wedge \neg P)$ is thus *false* in every possible 'context'. Formulæ like these, for which none of the interpretations is a model are called *unsatisfiable* or *inconsistent*

Finally, any wff that has at least *one* interpretation as a model is said to be *satisfiable*.

Logical Consequence

We are often interested in establishing the truth-value of a formula given that of some others. Recall the Chrysippus argument:

Statement	Formally
The rabbit either went down Path A or Path B.	$P \vee Q$
It did not go down Path A.	$\neg P$
Therefore it went down Path B.	$\therefore Q$

Here, we want to establish that if the first two statements are true, then the third follows. The formal notion underlying all this is that of *logical consequence*. In particular, what we are trying to establish is that some well-formed formula α is the *logical consequence* of a conjunction of other well-formed formulæ Σ (or, that Σ *logically implies* α). This relationship is usually written thus:

$$\Sigma \models \alpha$$

Σ being the conjunction of several wffs, it is itself a well-formed formula[6]. Logical consequence can therefore also be written as the following relationship between a pair of wffs:

$$((\beta_1 \wedge \beta_2) \ldots \beta_n) \models \alpha$$

It is sometimes convenient to write Σ as the set $\{\beta_1, \beta_2, \ldots, \beta_n\}$ which is understood to stand for the conjunctive formula above. But how do we determine if this relationship between Σ and α does indeed hold? What we want is the following: whenever the statements in Σ are true, α must also be true. In formal terms, this means: $\Sigma \models \alpha$ *if every model of* Σ *is also model of* α. Decoded:

[6]There is therefore nothing special needed to extend the concepts of validity and unsatisfiability to conjunctions of formulæ like Σ. Thus, Σ is valid if and only if every interpretation is a model of the conjunctive wff (in other words, a model for each wff in the conjunction); and it is unsatisfiable if and only if none of the interpretations is a model of the conjunctive wff. It should be apparent after some reflection that if Σ is valid, then all logical consequences of it are also valid. On the other hand, if Σ is unsatisfiable, then any well-formed formula is a logical consequence.

- Recall that a model for a formula is an interpretation (assignment of truth-values to propositions) that makes that formula *true*;
- Therefore, a model for Σ is an interpretation that makes $((\beta_1 \wedge \beta_2) \ldots \beta_n)$ *true*. Clearly, such an interpretation will make each of $\beta_1, \beta_2, \ldots, \beta_n$ *true*;
- Let $I_1, I_2, \ldots, I_k$ be all the interpretations that satisfy the requirement above: that is, each is a model for Σ and there are no other models for Σ (recall that if there are N propositional symbols in Σ and α together, then there can be no more than 2^N such interpretations);
- Then to establish $\Sigma \models \alpha$, we have to check that each of $I_1, I_2, \ldots, I_k$ is also a model for α (that is, each of them make α *true*).

The definition of logical entailment can be extended to the entailment of sets of formulæ. Let Σ and Γ be sets of formulas. Then Γ is said to be a logical consequence of Σ (written as $\Sigma \models \Gamma$), if $\Sigma \models \alpha$, for every formula $\alpha \in \Gamma$. We also say Σ (logically) implies Γ.

We are now in a position to see if Chrysippus was correct. We wish to see if $((P \vee Q) \wedge \neg P) \models Q$. From the truth tables on page 19, we can construct a truth table for $((P \vee Q) \wedge \neg P)$:

	P	Q	$(P \vee Q)$	$\neg P$	$((P \vee Q) \wedge \neg P)$
I_1 :	false	false	false	true	false
I_2 :	false	true	true	true	true
I_3 :	true	false	true	false	false
I_4 :	true	true	true	false	false

It is evident that of the four interpretations possible only one is a model for $((P \vee Q) \wedge \neg P)$, namely: I_2. Clearly I_2 is also a model for Q. Therefore, every model for $((P \vee Q) \wedge \neg P)$ is also a model for Q[7]. It is therefore indeed true that $((P \vee Q) \wedge \neg P) \models Q$. In fact, you will find you can 'move' formulæ from left to right in a particular manner. Thus if:

$$((P \vee Q) \wedge \neg P) \models Q$$

then the following also hold:

$$(P \vee Q) \models (Q \leftarrow \neg P) \quad \text{and} \quad \neg P \models (Q \leftarrow (P \vee Q))$$

These are consequences of a a more general result known as the *deduction theorem*, which we look at now. Using a set-based notation, let $\Sigma = \{\beta_1, \beta_2, \ldots, \beta_i, \ldots, \beta_n\}$. Then, the deduction theorem states: .

[7]Although I_4 is also a model for Q, the test for logical consequence only requires us to examine those interpretations that are models of $((P \vee Q) \wedge \neg P)$. This precludes I_4.

$$\Sigma \models \alpha \text{ if and only if } \Sigma - \{\beta_i\} \models (\alpha \leftarrow \beta_i)$$

Why is this so? Consider first the case that $\Sigma \models \alpha$. That is, every model of Σ is a model of α. Now assume $\Sigma - \{\beta_i\} \not\models (\alpha \leftarrow \beta_i)$. That is, there is some model, say M, of $\Sigma - \{\beta_i\}$ that is not a model of $(\alpha \leftarrow \beta_i)$. That is, β_i is true and α is false in M. That is M is a model for $\Sigma - \{\beta_i\}$ and for β_i, but not a model for α. In other words, M is a model for Σ but not a model for α which is not possible. Therefore, if $\Sigma \models \alpha$ then $\Sigma - \{\beta_i\} \not\models (\alpha \leftarrow \beta_i)$. Now for the "only if" part. That is, let $\Sigma - \{\beta_i\} \models (\alpha \leftarrow \beta_i)$. We want to show that $\Sigma \models \alpha$. Once again, let us assume the contrary (that is, $\Sigma \not\models \alpha$. This means there must be a model M for Σ that is not a model for α. However, since $\Sigma = \{\beta_1, \beta_2, \ldots, \beta_i, \ldots, \beta_n\}$, M is both a model for $\Sigma - \{\beta_i\}$ and a model for each of the β_i. So, M cannot be a model for $(\alpha \leftarrow \beta_i)$. We are therefore in a position that there is a model M for $\Sigma - \{\beta_i\}$ that is not a model of $(\alpha \leftarrow \beta_i)$, which contradicts what was given.
The deduction theorem isn't restricted to propositional logic, and holds for first-order logic as well. It can be invoked repeatedly. Here is an example of using it twice:

$$\Sigma \models \alpha \text{ if and only if } \Sigma - \{\beta_i, \beta_j\} \models (\alpha \leftarrow (\beta_i \wedge \beta_j))$$

With Chrysippus, applying the deduction theorem twice results in:

$$\{(P \vee Q), \neg P\} \models Q \text{ if and only if } \emptyset \models (Q \leftarrow ((P \vee Q) \wedge \neg P))$$

If $\emptyset \models (Q \leftarrow ((P \vee Q) \wedge \neg P))$ then every model for $\emptyset$ must be a model for $(Q \leftarrow ((P \vee Q) \wedge \neg P))$. By convention, every interpretation is a model for $\emptyset$[8]. It follows that every interpretation must be a model for $(Q \leftarrow ((P \vee Q) \wedge \neg P))$. Recall that this is just another way of stating that $(Q \leftarrow ((P \vee Q) \wedge \neg P))$ is valid (page 21)[9].

What is the difference between the concepts of logical consequence denoted by $\models$ and the connective $\rightarrow$ in a statement such as $\Sigma \models \Gamma$? where, $\Sigma = \{(P \wedge Q), (P \rightarrow R)\}$ and $\Gamma = \{P, Q, R\}$? And how do these two notions of implication relate to the phrase 'if....then', often used in propositions or theorems? We delineate the differences below:

1. The connective $\rightarrow$ is a *syntactical symbol* called 'if ... then' or 'implication', which appears within formulæ. The truth value of the formula $(\alpha \rightarrow \xi)$

[8]That is, we take the empty set to denote a distinguished proposition $True$ that is *true* in every interpretation. Correctly then, the formula considered is not $((\beta_1 \wedge \beta_2) \ldots \beta_n)))$ but $(True \wedge ((\beta_1 \wedge \beta_2) \ldots \beta_n))))$.

[9]To translate declarative knowledge into action (as in the case of the dog from Chrysippus's anecdote), one of two possible strategies can be adopted. The first is called 'Negative selection' which involves *excluding any provably futile* actions. The second is called 'Positive selection' which involves *suggesting only actions that are provably safe*. There can be some actions that are neither provably safe nor provably futile.

depends on the *particular* interpretation I we happen to be considering: according to the truth table, $(\alpha \to \xi)$ is true under I if α is false under I and/or ξ is true under I; $(\alpha \to \xi)$ is false otherwise.

2. The concept of 'logical consequence' or '(logical) implication', denoted by '$\models$' describes a *semantical relation* between formulæ. It is defined in terms of *all* interpretations: '$\alpha \models \xi$' is true if every interpretation that is a model of α, is also a model of ξ.

3. The phrase 'if. .. then', which is used when stating, for example, propositions or theorems is also sometimes called 'implication'. This describes a relation between assertions which are phrased in (more or less) natural language. It is used for instance in proofs of theorems, when we state that some assertion implies another assertion. Sometimes we use the symbols ' ' or . ¡=' for this. If assertion A implies assertion B, we say that B is a necessary condition for A (i.e., if A is true, B must necessarily be true), and A is a. Ilufficient condition for B (i.e., the truth of B is sufficient to make A true). In ('tum A implies B, and B implies A, we write "A iff B", where 'iff' abbreviates 'if, and only if'.

Closely related to logical consequence is the notion of *logical equivalence*. A pair of wffs α and β are logically equivalent means:

$$\alpha \models \beta \quad \textit{and} \quad \beta \models \alpha$$

This means the truth values for α and β are the same in all cases, and is usually written more concisely as:

$$\alpha \equiv \beta$$

Examples of logically equivalent formulæ are provided by De Morgan's laws:

$$\neg(\alpha \vee \beta) \equiv \neg\alpha \wedge \neg\beta$$

$$\neg(\alpha \wedge \beta) \equiv \neg\alpha \vee \neg\beta$$

Also, if $True$ denotes the formula that is *true* in every interpretation and $False$ the formula that is *false* in every interpretation, then the following equivalences should be self-evident:

$$\alpha \equiv (\alpha \wedge True)$$

$$\alpha \equiv (\alpha \vee False)$$

More on the Conditional

We are mostly concerned with rules that utilise the logical connective $\leftarrow$. This makes this particular connective more interesting than the others, and it is worth noting some further details about it. Although we will present these here using examples from the propositional logic, the main points are just as applicable to formulæ in the predicate logic.

Recall the truth table for the conditional from page 20:

α	β	$(\alpha \leftarrow \beta)$
false	false	true
false	true	false
true	false	true
true	true	true

There is, therefore, only one interpretation that makes $(\alpha \leftarrow \beta)$ *false*. This may come as a surprise. Consider for example the statement:

$$\textsf{The earth is flat} \leftarrow \textsf{Humans are monkeys}$$

An interpretation that assigns *false* to both 'The earth is flat' and 'Humans are monkeys' makes this statement *true* (line 1 of the truth table). In fact, the only world in which the statement is false is one in which the earth is not flat, and humans are monkeys[10]. Consider now the truth table for $(\alpha \vee \neg\beta)$:

α	β	$\neg\beta$	$(\alpha \vee \neg\beta)$
false	false	true	true
false	true	false	false
true	false	true	true
true	true	false	true

It is evident from these truth tables that every model for $(\alpha \leftarrow \beta)$ is a model for $(\alpha \vee \neg\beta)$ and vice-versa. Thus:

$$(\alpha \leftarrow \beta) \equiv (\alpha \vee \neg\beta)$$

Thus, the conditional:

$$(\textsf{Fred is human} \leftarrow (\textsf{Fred walks upright} \wedge \textsf{Fred has a large brain}))$$

is equivalent to:

[10] The unusual nature of the conditional is due to the fact that it allows premises and conclusions to be completely unrelated. This is not what we would expect from conditional statements in normal day-to-day discourse. For this reason, the $\leftarrow$ connective is sometimes referred to as the *material conditional* to distinguish it from a more intuitive notion.

$$\text{(Fred is human} \lor \neg \text{(Fred walks upright} \land \text{Fred has a large brain))}$$

Or, using De Morgan's Law (page 25) and dropping some brackets for clarity:

$$\text{Fred is human} \lor \neg \text{Fred walks upright} \lor \neg \text{Fred has a large brain}$$

In this form, each of the premises on the right-hand side of the the original conditional (Fred walks upright, Fred has a large brain) appear negated in the final disjunction; and the conclusion (Fred is human) is unchanged. For a reason that will become apparent later we will use the term *clauses* to denote formulæ that contain propositions or negated propositions joined together by disjunction ($\lor$). We will also use the term *literals* to denote propositions or negated propositions. Clauses are thus disjunctions of literals.

It is common practice to represent a clause as a set of literals, with the disjunctions understood. Thus, the clause above can be written as:

$$\{ \text{ Fred is human}, \neg \text{ Fred walks upright}, \neg \text{ Fred has a large brain } \}$$

The equivalence $\alpha \leftarrow \beta \equiv \alpha \lor \neg\beta$ also provides an alternative way of presenting the deduction theorem.

On page 23 the statement of this theorem was:

$$\Sigma \models \alpha \text{ if and only if } \Sigma - \{\beta_i\} \models (\alpha \leftarrow \beta_i)$$

This can now be restated as:

$$\Sigma \models \alpha \text{ if and only if } \Sigma - \{\beta_i\} \models (\alpha \lor \neg\beta_i)$$

The theorem thus operates as follows: when a formula moves from the left of $\models$ to the right, it is negated and disjoined (using $\lor$) with whatever exists on the right. The theorem can also be used in the "other direction": when a formula moves from the right of $\models$ to the left, it is negated and conjoined (using $\land$ or $\cup$ in the set notation) to whatever exists on the left. Thus:

$$\Sigma \models (\alpha \lor \neg\beta) \text{ if and only if } \Sigma \cup \{\neg\alpha\} \models \neg\beta$$

A special case of this arises from the use of the equivalence $\alpha \equiv (\alpha \lor False)$ (page 25):

$$\Sigma \models \alpha \text{ if and only if } \Sigma \models (\alpha \lor False) \text{ if and only if } \Sigma \cup \{\neg\alpha\} \models False$$

The formula $False$ is commonly written as $\square$ and the result above as:

$$\Sigma \models \alpha \text{ if and only if } \Sigma \cup \{\neg\alpha\} \models \square$$

The conditional $(\alpha \leftarrow \beta)$ is sometimes mistaken to mean the same as $(\alpha \land \beta)$. Comparison against the truth table for $(\alpha \land \beta)$ shows that these two formulæ are not equivalent:

α	β	$(\alpha \wedge \beta)$
false	false	false
false	true	false
true	false	false
true	true	true

There are a number of ways in which $(\alpha \leftarrow \beta)$ can be translated in English. Some of the more popular ones are:

If β, then α	α, if β	β implies α
β only if α	β is sufficient for α	α is necessary for β
All β's are α's		

Note the following related statements:

Conditional	$(\alpha \leftarrow \beta)$
Contrapositive	$(\neg\beta \leftarrow \neg\alpha)$

It should be easy to verify the following equivalence:

$$\text{Conditional} \equiv \text{Contrapositive} \qquad (\alpha \leftarrow \beta) \equiv (\neg\beta \leftarrow \neg\alpha)$$

Errors of reasoning arise by assuming other equivalences. Consider for example the pair of statements:

$$S_1 : \text{Fred is an ape} \leftarrow \text{Fred is human}$$
$$S_2 : \text{Fred is human} \leftarrow \text{Fred is an ape}$$

S_2 is the sometimes called the *converse* of S_1. An interpretation that assigns *true* to 'Fred is an ape' and *false* to 'Fred is human' is a model for S_1 but not a model for S_2. The two statements are thus not equivalent.

More on Normal Forms

We are now able to state two properties concerning normal forms:

1. If F is a formula in CNF and G is a formula in DNF, then $\neg F$ is a formula in DNF and $\neg G$ is a formula in CNF. This is a generalisation of De Morgan's laws and can be proved using the technique of mathematical induction (that is, show truth for a formula with a single literal; assume truth for a formula with n literals; and then show that it holds for a formula with $n + 1$ literals.)

2. Every formula F can be written as a formula F_1 in CNF and a formula F_2 in DNF. It is straightforward to see that any formula F can be written as a DNF formula by examining the rows of the truth table for F for which F is true. Suppose F consists of the propositions $A_1, A_2, \ldots, A_n$. Then each such row is equivalent to some conjunction of literals $L_1, L_2, \ldots, L_n$, where L_i is equal to A_i if A_i is true in that row and equal to $\neg A_i$ otherwise. Clearly, the disjunction of each row for which F is true gives the DNF formula for F. We can get the corresponding CNF formula G by negating the DNF formula (using the property above), or by examining the rows for which F is false in the truth table.

It should now be clear that a CNF expression is nothing more than a conjunction of a set of clauses (recall a clause is simply a disjunction of literals). It is therefore possible to convert any propositional formula F into CNF—either using the truth table as described, or using the following procedure:

1. Replace all conditionals statements of the form $A \leftarrow B$ by the equivalent form using disjunction (that is, $A \vee \neg B$). Similarly replace all $A \leftrightarrow B$ with $(A \vee \neg B) \wedge (\neg A \vee B)$.

2. Eliminate double negations ($\neg\neg A$ replaced by A) and use De Morgan's laws wherever possible (that is, $\neg(A \wedge B)$ replaced by $(\neg A \vee \neg B)$ and $\neg(A \vee B)$ replaced by $(\neg A \wedge \neg B)$).

3. Distribute the disjunct $\vee$. For example, $(A \vee (B \wedge C))$ is replaced by $(A \vee B) \wedge (A \vee C)$.

An analogous process converts any formula to an equivalent formula in DNF. We should note that during conversion, formulæ can expand exponentially. However, if only satisfiability should be preserved, conversion to CNF formula can be done polynomially.

Inference

Enumerating and comparing models is one way of determining whether one formula is a logical consequence of another. While the procedure is straightforward, it can be tedious, often requiring the construction of entire truth tables. A different approach makes no explicit reference to truth values at all. Instead, if α is a logical consequence of Σ, then we try to show that we can *infer* or *derive* α from Σ using a set of well-understood rules. Step-by-step application of these rules results in a *proof* that deduces that α follows from Σ. The rules, called *rules of inference*, thus form a system of performing calculations with propositions, called the *propositional calculus*[11]. Logical implication can be mechanized by using a propositional calculus. We will first concentrate on a particular inference rule called *resolution*.

[11] In general, a set of inference rules (potentially including so called logical axioms) is called a *calculus*.

Resolution

Before proceding further, some basic terminology from proof theory may be helpful (this is not specially confined to the propositional calculus). Proof theory considers the *derivability* of formulæ, given a set of inference rules $\mathcal{R}$. Formulæ given initially are called *axioms* and those derived are *theorems*. That formula α is a theorem of a set of axioms Σ using inference rules $\mathcal{R}$ is denoted by:

$$\Sigma \vdash_{\mathcal{R}} \alpha$$

When $\mathcal{R}$ is obvious, this is simply written as $\Sigma \vdash \alpha$. The axioms can be valid (that is, all interpretations are models), or problem-specific (that is, only some interpretations may be models). The axioms together with the inference rules constitute what is called an *inference system*. The axioms together with all the theorems that are derivable from it is called a *theory*. A theory is said to be *inconsistent* if there is a formula α such that the theory contains both α and $\neg\alpha$.

We would like the theorems derived to be logical consequences of the axioms provided. For, if this were the case then by definition, the theorems will be *true* in all models of the axioms (recall that this is what logical consequence means). They will certainly be true, therefore, in any particular 'intended' interpretation of the axioms. Ensuring this property of the theorems depends entirely on the inference rules chosen: those that have this property are called *sound*. That is:

$$\text{if } \Sigma \vdash_{\mathcal{R}} \alpha \text{ then } \Sigma \models \alpha$$

Some well-known sound inference rules are:

$$\begin{array}{ll} \text{Modus ponens:} & \{\beta, \alpha \leftarrow \beta\} \ \vdash \ \alpha \\ \text{Modus tollens:} & \{\neg\alpha, \alpha \leftarrow \beta\} \ \vdash \ \neg\beta \end{array}$$

Theorems derived by the use of sound inference rules can be added to the original set of axioms. That is, given a set of axioms Σ and a theorem α derived using a sound inference rule, $\Sigma \equiv \Sigma \cup \{\alpha\}$.

We would also like to derive *all* logical consequences of a set of axioms and rules with this property at said to be *complete*:

$$\text{if } \Sigma \models \alpha \text{ then } \Sigma \vdash_{\mathcal{R}} \alpha$$

Axioms and inference rules are not enough: we also need a strategy to select and apply the rules. An inference system (that is, axioms and inference rules) along with a strategy is called a *proof procedure*. We are especially interested here in a special inference rule called *resolution* and a strategy called SLD (the meaning of this is not important at this point: we will come to it later). The result is a proof procedure called *SLD-resolution*. Here we will simply illustrate the rule of resolution for manipulating propositional formulæ, and use an unconstrained proof strategy. A description of SLD will be left for a later section.

Suppose we are given as axioms the conditional formulæ (using the informal notation that replaces $\wedge$ with commas):

β_1 : Fred is an ape $\leftarrow$ Fred is human
β_2 : Fred is human $\leftarrow$ Fred walks upright, Fred has a large brain

Then the following is a theorem resulting from the use of resolution:

α : Fred is an ape $\leftarrow$ Fred walks upright, Fred has a large brain

That α is indeed a logical consequence of $\beta_1 \wedge \beta_2$ can be checked by constructing truth tables for the formulæ: you will find that every interpretation that makes $\beta_1 \wedge \beta_2$ *true* will also make α *true*. More generally, here is the rule of resolution when applied to a pair of conditional statements:

$$\{(P \leftarrow Q_1, \ldots, Q_i, \ldots Q_n),\ (Q_i \leftarrow R_1, \ldots R_m)\} \quad \vdash$$
$$P \leftarrow Q_1, \ldots, Q_{i-1}, R_1, \ldots R_m, Q_{i+1}, \ldots, Q_n$$

The equivalence from page 26 ($\alpha \leftarrow \beta \equiv \alpha \vee \neg\beta$) allows resolution to be presented in a different manner (we have taken the liberty of dropping some brackets here):

$$\{(P \vee \neg Q_1 \vee \ldots \neg Q_i \vee \ldots \vee \neg Q_n),\ (Q_i \vee \neg R_1 \vee \ldots \vee \neg R_m)\} \quad \vdash$$
$$P \vee \neg Q_1 \vee \ldots \vee \neg Q_{i-1} \vee \neg R_1 \vee \ldots \vee \neg R_m \vee \neg Q_{i+1} \vee \ldots \vee \neg Q_n$$

On page 27, we introduced the terms clauses and literals. Thus, resolution applies to a pair of 'parent' clauses that contain *complementary* literals $\neg L$ and L. The result (the 'resolvent') is a clause containing all literals from each clause, except the complementary pair. Or, more abstractly, let C_1 and C_2 be a pair of clauses, and let $L \in C_1$ and $\neg L \in C_2$. Then, the resolvent of C_1 and C_2 is the clause:

$$R = (C_1 - \{L\}) \cup (C_2 - \{\neg L\})$$

Resolution of a pair of *unit clauses*—those that contain just single literals L and $\neg L$—results in the the *empty clause*[12], or $\Box$, which means that the parent clauses were inconsistent.

We can show that resolution is a sound inference rule.

Theorem 6 *Suppose R is the resolvent of clauses C_1 and C_2. That is, $\{C_1, C_2\} \vdash R$. The resolution is sound, that is, $\{C_1, C_2\} \models R$.*

Proof:: We want to show that if C_1 and C_2 are true and R is a resolvent of C_1 and C_2 then R is true. Let us assume C_1 and C_2 are true, and that R was obtained by resolving on some literal L in C_1 and C_2. Further, let $C_1 = C \vee L$

[12]Note the difference between an empty clause $\Box$ and empty set of clauses $\{\}$. An interpretation I logically entails C *iff* there exists an $l \in C$ such that $I \models l$. I logically entails Σ if for all $C \in \Sigma$, $I \models C$. Thus, by definition, for all interpretations I, $I \not\models \Box$ and $I \not\models \{\Box\}$, whereas $I \models \{\}$.