Contents

Chapter 2

Logical Consequence

Whereas the previous chapter introduced truth tables, this chapter will present the truth table method which provides a decidable procedure for evaluating the validity of arguments in $\mathcal{L}^{\text{PL}}$. Given the (albeit limited) expressive power of $\mathcal{L}^{\text{PL}}$, this amounts to a mechanical way of evaluating the natural language arguments that admit of reasonably faithful regimentations in $\mathcal{L}^{\text{PL}}$. After reviewing both the advantages and disadvantages of this procedure, the second half of the chapter will present a more versatile alternative.

2.1 Truth-Functional Operators

Any wfs of $\mathcal{L}^{\text{PL}}$ that is not a sentence letter is composed of sentence letters together with the sentential operators. In Chapter 1, we offered truth tables for each operator. The fact that it is possible to give truth tables like this is very significant. It means that our operators are TRUTH-FUNCTIONAL. That is to say, the only thing that matters for determining the truth-value of a given wfs of $\mathcal{L}^{\text{PL}}$ is the truth-values of its constituent. For instance, to determine the truth-value of a sentence $\neg A$, the only thing that matters is the truth-value of A. Given any interpretation, the truth-value of a negation on that interpretation is a *function* of the truth-value of its negand on that interpretation, and likewise for the other operators.

We are using the same notion of a function that you have probably encountered in mathematics. First we may define the CARTESIAN PRODUCT $X \times Y$ of the sets X and Y to be the set of all ordered pairs $\langle x, y\rangle$ where $x \in X$ and $y \in Y$ which we may write in set-builder notation as $X \times Y \coloneqq \{\langle x, y\rangle : x \in X,\ y \in Y\}$.[1] A RELATION from X to Y is any subset $A \subseteq X \times Y$. A FUNCTION $f : D \to R$ from the DOMAIN D to the RANGE R is any relation $f \subseteq D \times R$ from D to R where $f(x) = f(y)$ for any $x, y \in D$ where $x = y$. Intuitively, a function from one set to another associates each member of the first set (the domain) with exactly one member of the second set (the range). Once the first element is fixed, the function uniquely selects an

[1]The '$\coloneqq$' symbol signifies that a definition is being provided.

element of the second set. Instead of always writing $f(x, y)$, it makes sense to write $f(x) = y$ for the unique element in the range y to which the element x in the domain is mapped. For instance, given any numerical value of x, we may unambiguously determine the value of x^2, and so $f(x) = x^2$ is a function. In the same way, the truth-value of A on an interpretation will unambiguously determine the truth-value of $\neg A$ on that interpretation.

Truth-functionality is not inevitable. The syntax of English, for example, permits one to make a new declarative English sentence by prefixing the phrase 'Ted Cruz doesn't care whether' in front of any declarative English sentence. In this respect, 'Ted Cruz doesn't care whether' is syntactically similar to '$\neg$' in $\mathcal{L}^{\text{PL}}$: it is a sentential operator, producing new sentences from old. Nevertheless, it is impossible to give a truth-functional characterization of the operator 'Ted Cruz doesn't care whether' in English that respects its intuitive meaning in English. If you want to know whether Ted Cruz cares about fixing Texas's electrical grid, it's not enough to know whether anyone is fixing Texas's electrical grid. If it is being fixed, he might care or he might not. If it is not be fixed, again he might care or might not. Thus 'Ted Cruz doesn't care whether' is not truth-functional: it operates on more than just the truth-value of its argument. By contrast, the sentential operators included in $\mathcal{L}^{\text{PL}}$ are truth-functional, and it is for this reason that we are able to construct truth tables.

2.2 Complete Truth Tables

A truth table for a sentence may be constructed by writing the sentence in question at the top right of a table, and each of the distinct sentence letters immediately to the left on the top row. We then add 2^n rows below the top row where n is the number of distinct sentence letters. For instance, if there are only two sentence letters, we will need four rows of truth-values. Beginning with the sentence letter furthest to the left, we fill out the column with $2^{(n-1)}$ copies of 1 followed by $2^{(n-1)}$ copies of 0. Moving to the next sentence letter, we fill out the column with $2^{(n-2)}$ copies of 1 followed by $2^{(n-2)}$ copies of 0. We then proceed to the next sentence letter (if there is one), following the same pattern as before but with $2^{(n-3)}$ copies of 1 and 0, respectively. Continue this process until all sentence letters in the table have truth-values below them. This completes the truth table setup.

In order to assign truth-values to complex sentences, consider the following CHARACTERISTIC TRUTH TABLES where instead of particular sentence letters, we will use schematic variables:

φ	$\neg\varphi$
1	0
0	1

φ	ψ	$\varphi \wedge \psi$	$\varphi \vee \psi$	$\varphi \to \psi$	$\varphi \leftrightarrow \psi$
1	1	1	1	1	1
1	0	0	1	0	0
0	1	0	1	1	0
0	0	0	0	1	1

Table 2.1: The characteristic truth tables for the operators of $\mathcal{L}^{\text{PL}}$.

The table above provides a general recipe for calculating truth-values for any sentences φ and ψ, however complex. The characteristic truth table for conjunction, for example, gives the truth conditions for any sentence of the form $\varphi \wedge \psi$. Even if the conjuncts φ and ψ are complex sentences, the conjunction is true if and only if both φ and ψ are true.

Let's construct a truth table for the complex sentence $(H \wedge I) \to H$. We consider all possible combinations of 1 and 0 for H and I, which gives us four rows. We then copy the truth-values for the sentence letters and write them underneath the letters in the sentence.

H	I	$(H$	$\wedge$	$I)$	$\to$	H
1	1	**1**		**1**		**1**
1	0	**1**		**0**		**1**
0	1	**0**		**1**		**0**
0	0	**0**		**0**		**0**

So far, all we have done is duplicate the first two columns. We have written the H column twice— once under each H— and the I column once under the I.

Now consider the subsentence $H \wedge I$ which is a conjunction. Since H and I are both true on the first row and a conjunction is true when both conjuncts are true, we write a 1 underneath the conjunction symbol. In the other three rows, at least one of the conjuncts is false, so the conjunction $H \wedge I$ is false. So we write 0s under the conjunction symbol on those rows:

H	I	$(H$	$\wedge$	$I)$	$\to$	H
1	1	1	**1**	1		1
1	0	1	**0**	0		1
0	1	0	**0**	1		0
0	0	0	**0**	0		0

The entire sentence $(H \wedge I) \to H$ is a conditional. On the second row, for example, $H \wedge I$ is false and H is true. Since a conditional is true when the antecedent is false, we write a 1 in the second row underneath the conditional symbol. Using the truth table for the material conditional where $H \wedge I$ is the antecedent and H as the consequent, we may derive the following values for the material conditional claim as a whole:

H	I	$(H$	$\wedge$	$I)$	$\to$	H
1	1		1		**1**	1
1	0		0		**1**	1
0	1		0		**1**	0
0	0		0		**1**	0

In computing the value for the conditional (the column under $\to$), it is only important to look at the values for its antecedent (the column under $\wedge$) and the value of its consequent (the column under H on the right). The column of 1s underneath the conditional tells us that the sentence $(H \wedge I) \to H$ is true regardless of the truth-values of H and I. They can be true or false in any combination, and the compound sentence still comes out true.

It is crucial that we have considered all of the possible combinations. If we only had a two-line truth table, we could not be sure that the sentence was not false for some other combination of truth-values. Since each row of the truth table represents a different way of interpreting the relevant sentence letters H and I, each row corresponds to a distinct interpretation of the sentence letters in question. Moreover, every possible combination of truth-values for H and I have been included. Since the truth-values of all other sentence letters do not effect the truth-value of the sentence in question, we may conclude that $(H \wedge I) \to H$ is *true in all interpretations* whatsoever. In other words, $(H \wedge I) \to H$ is a *tautology*.

In this example, we have not repeated all of the entries in every successive table, so that it's easier for you to see which parts are new. When actually writing truth tables on paper, however, it is impractical to erase whole columns or rewrite the whole table for every step. Although it is more crowded, the truth table can be written in this way:

H	I	$(H$	$\wedge$	$I)$	$\to$	H
1	1	1	1	1	**1**	1
1	0	1	0	0	**1**	1
0	1	0	0	1	**1**	0
0	0	0	0	0	**1**	0

Most of the columns underneath the sentence are only there for bookkeeping purposes. When you become more adept with truth tables, you will probably no longer need to copy over the columns for each of the sentence letters. In any case, the truth-values for the original sentence is given by the column underneath the main logical operator of the sentence which in this case is the column underneath the conditional, marked in **bold** for clarity.

A COMPLETE TRUTH TABLE has a row for all possible combinations of 1 and 0 for the sentence letters and the characteristic truth tables have been used to write truth-values below all the operators. The size of the complete truth table depends on the number of different sentence letters in the table. A sentence that contains only one sentence letter requires only two rows, as in the characteristic truth table for negation. This is true even if the same letter is repeated many times, as in the sentence $[(C \leftrightarrow C) \to C] \wedge \neg(C \to C)$. The complete truth table requires only two lines because there are only two possible interpretations of C: either it is true or it is false. A single sentence letter can never be marked both 1 and 0 on the same row. The truth table for this sentence looks like this:

C	$[($	C	$\leftrightarrow$	C	$)$	$\to$	C	$]$	$\wedge$	$\neg$	$($	C	$\to$	C	$)$
1		1	1	1		1	1		**0**	0		1	1	1	
0		0	1	0		0	0		**0**	0		0	1	0	

Looking at the column underneath the main operator, we see that the sentence is false on both rows of the table, and so it is false regardless of whether C is true or false. Since the rows of the truth table correspond to the different possible interpretations of the relevant sentence letters, the sentence above is false on all interpretations, and so it is a *contradiction*.

A sentence that contains two sentence letters requires 2^2 lines for a complete truth table as in the other characteristic truth tables and the table for $(H \wedge I) \to I$ above. A sentence that contains three sentence letters requires 2^3 lines. For example:

M	N	P	M	$\wedge$	$(N$	$\vee$	$P)$
1	1	1	1	**1**	1	1	1
1	1	0	1	**1**	1	1	0
1	0	1	1	**1**	0	1	1
1	0	0	1	**0**	0	0	0
0	1	1	0	**0**	1	1	1
0	1	0	0	**0**	1	1	0
0	0	1	0	**0**	0	1	1
0	0	0	0	**0**	0	0	0

This table shows that $M \wedge (N \vee P)$ is true on some interpretations and false on others depending on the truth-values of M, N, and P, and so it is *contingent*.

A complete truth table for a sentence that contains four different sentence letters requires 2^4 lines. This means that the truth table method becomes syntactically unmanageable very quickly. This is a significant limitation of this method.

2.3 Truth Table Definitions

Recall from §0.5.1 that an English sentence was said to be a tautology just in case it is true on all interpretations, a contradiction just in case it is false on all interpretations, and logically contingent just in case it is true on some interpretations and false on others. Similarly, an English sentence logically entails another just in case the latter is true on any interpretation on which the former is true, and two English sentences are logically equivalent just in case they logically entail each other and so are true on exactly the same interpretations. Even in restricting consideration to truth-values in interpreting the sentences of English, there is no well-defined set of English sentences, and so no corresponding definition of an interpretation as an assignment of every English sentence to a unique truth-value. It is for this reason that we introduced $\mathcal{L}^{\text{PL}}$ in Chapter 1, carefully defining the wfss of $\mathcal{L}^{\text{PL}}$.

Given the definition of the wfs of $\mathcal{L}^{\text{PL}}$ together with the characteristic truth tables for the sentential operators of $\mathcal{L}^{\text{PL}}$, we are now in a position to define the interpretations of $\mathcal{L}^{\text{PL}}$ in a way that we could not do so for English. In particular, an INTERPRETATION of $\mathcal{L}^{\text{PL}}$ is any function from the set of wfs of $\mathcal{L}^{\text{PL}}$ (the domain) to the set of truth-values $\{0, 1\}$ (the range) which satisfies the characteristic truth tables given above. For instance, any interpretation that assigns A to 0 will assign $\neg A$ to 1. More generally, an interpretation assigns $\neg\varphi$ to 1 just in case it assigns φ to 0 since this is what is required to satisfy the characteristic truth table for negation. Similarly, an interpretation assigns $\varphi \wedge \psi$ to 1 just in case it assigns both φ and ψ to 1 as specified by the characteristic truth table for conjunction. Given any wfs of $\mathcal{L}^{\text{PL}}$, the characteristic truth table for the main operator of that sentence determines the truth-value of that sentence as a function of the truth-value(s) for its arguments.

Rather than worrying about the truth-value of all wfs of $\mathcal{L}^{\text{PL}}$ whatsoever, constructing a complete truth tables provides a way to interpret a sentence while limiting consideration to its parts. This approach depends on the truth-functionality of the sentential operators. We may then put this method to work to define a TAUTOLOGY of $\mathcal{L}^{\text{PL}}$ to be any sentence of $\mathcal{L}^{\text{PL}}$ whose truth table only has 1s under its main operator. Accordingly, a tautology is a sentence of $\mathcal{L}^{\text{PL}}$ whose truth does not depend on the particular truth-values of the sentence letters from which it was constructed, but rather follows from the *logical form* of that sentence. Similarly, a sentence is a CONTRADICTION in $\mathcal{L}^{\text{PL}}$ just in case the column under its main operator is 0 on every row of its complete truth table. Instead of being true in virtue of its logical form, a contradiction is false in virtue of its logical form. A sentence is LOGICALLY CONTINGENT in $\mathcal{L}^{\text{PL}}$ just in case the column under its main operator includes both 1s and 0s.

2.3.1 Logical Entailment and Equivalence

A sentence in English was said to logically entail another just in case the latter is true on any interpretation on which the former is true. Two sentences were then said to be logically equivalent in English just in case they logically entail each other, and so have the same truth-value on all interpretations. We can now say with greater precision that a sentence of $\mathcal{L}^{\text{PL}}$ LOGICALLY ENTAILS another just in case the complete truth table for these two sentences is such that on any row, the latter has a 1 under its main operator whenever the former has a 1 on that row under its main operator. Consider the following example:

A	B	$\neg$	A	$\neg$	$(A$	$\wedge$	$B)$
1	1	**0**	1	**0**	1	1	1
1	0	**0**	1	**1**	1	0	0
0	1	**1**	0	**1**	0	0	1
0	0	**1**	0	**1**	0	0	0

Whereas before we only considered truth tables for one wfs at a time, the truth table above interprets two wfs of $\mathcal{L}^{\text{PL}}$ at once. By including a column for all sentence letters contained in either wfs, the complete truth table exhausts all possible combinations of truth-values for the sentence letters upon which the truth of the wfss depend. We may the observe that every row in which $\neg A$ has a 1 under its main operator is also a row in which $\neg(A \wedge B)$ has a 1 under its main operator, and so $\neg A$ logically entails $\neg(A \wedge B)$. By contrast, given the second row of truth-values, $\neg(A \wedge B)$ does not logically entail $\neg A$.

In the special case where two wfs of $\mathcal{L}^{\text{PL}}$ logically entail each other, we may say that those sentences are LOGICALLY EQUIVALENT. It follows that any two wfs of $\mathcal{L}^{\text{PL}}$ are logically equivalent just in case the truth-values under their main operators are the same on every row of their complete truth table. For instance, consider the sentences $\neg(A \vee B)$ and $\neg A \wedge \neg B$. In order to find out if they are logically equivalent we may construct their complete truth table:

A	B	$\neg$	$(A$	$\vee$	$B)$	$\neg$	A	$\wedge$	$\neg$	B
1	1	**0**	1	1	1	0	1	**0**	0	1
1	0	**0**	1	1	0	0	1	**0**	1	0
0	1	**0**	0	1	1	1	0	**0**	0	1
0	0	**1**	0	0	0	1	0	**1**	1	0

The columns under the main operators for the two wfs on the right are identical. Since the rows of a complete truth table exhaust the different interpretations of the relevant sentence letters, this amounts to requiring the two wfs of $\mathcal{L}^{\text{PL}}$ to have the same truth-value on every interpretation. It is for this reason that the two wfs of $\mathcal{L}^{\text{PL}}$ are logically equivalent.

2.3.2 Satisfiability

In the previous example, a truth table was constructed for two wfss of $\mathcal{L}^{\text{PL}}$ at once. More generally, we may take a COMPLETE TRUTH TABLE for a set Γ of wfss of $\mathcal{L}^{\text{PL}}$ to be the result of listing each wfs in Γ side-by-side on the top right of a table and then completing the table for each wfs in a similar manner to what was described above.

Recall that a set of sentences in English is satisfiable just in case there is an interpretation which makes them all true. Analogously, we may wish to say that a set Γ of wfss in $\mathcal{L}^{\text{PL}}$ is SATISFIABLE just in case there is a row of a complete truth table including every wfs in the set where the main operator under every sentence is 1, and UNSATISFIABLE otherwise. For instance, look again at the truth table above. We see that $\{\neg(A \vee B), \neg A \wedge \neg B\}$ is satisfiable, because there is at least one row where both sentences have 1 under their main operators.

To take another example, we may ask if the set $\{A, \neg(A \vee \neg B)\}$ is satisfiable. Here the answer is 'No' as the following truth table shows:

A	B	$\neg$	$(A$	$\vee$	$\neg$	$B)$
1	1	**0**	1	1	0	1
1	0	**0**	1	1	1	0
0	1	**1**	0	0	0	1
0	0	**0**	0	1	1	0

Since there is no row in which both of the sentences in the set $\{A, \neg(A \vee \neg B)\}$ have a 1 under their main operator, we may conclude that the set is unsatisfiable.

2.3.3 Logical Consequence and Validity

Whereas Chapter 0 provided an intuitive definition of logical consequence for the sentences of English, we may now appeal to complete truth tables in order to provide a correlate for $\mathcal{L}^{\text{PL}}$. In particular, a wfs φ of $\mathcal{L}^{\text{PL}}$ is a LOGICAL CONSEQUENCE of a set of wfss Γ of $\mathcal{L}^{\text{PL}}$ just in case every row of a complete truth table for the wfs in Γ together with φ is such that φ has a 1 under its main operator whenever every sentence in Γ has a 1 under its main operator. We may then say that an argument in $\mathcal{L}^{\text{PL}}$ with premises Γ and conclusion φ is VALID just in case the conclusion is a logical consequence of the premises.

Instead of saying that an English argument is valid by appealing to the interpretations of English, we may say that an argument in English has a $\mathcal{L}^{\text{PL}}$ regimentation that is valid. Regimenting an English argument in $\mathcal{L}^{\text{PL}}$ and constructing its complete truth table provides a way to identify the logical features which explain why the argument in English is valid. Although there is a precise mathematical definition of validity in $\mathcal{L}^{\text{PL}}$, the same cannot be said for what does and does not count as a faithful regimentation of the argument. Rather, this much remains intuitive, relying on and individual's best judgment. In many cases, there may not be a uniquely best regimentation, or indeed any good regimentation at all.

Given any argument in $\mathcal{L}^{\text{PL}}$ whose premises are unsatisfiable, it follows that the argument is *vacuously valid* since there is no row in a complete truth table for the premises with a 1 under the main operator of every premise, and so vacuously, every row in a complete truth table in which there is a 1 under the main operator of every premises also has a 1 under the main operator for the conclusion. For instance, consider the following argument:

A1. $\neg A \wedge B$
A2. $\underline{\neg B\quad}$
A3. B

Given the truth table method presented above, it is straightforward to show that the argument is indeed valid. In particular, consider the following complete truth table:

A	B	$\neg$	A	$\wedge$	B	$\neg$	B	B
1	1	0	1	**0**	1	**0**	1	**1**
1	0	0	1	**0**	0	**1**	0	**0**
0	1	1	0	**1**	1	**0**	1	**1**
0	0	1	0	**0**	0	**1**	0	**0**

It is easy to see that there is no row in which the premises all have 1 under their main operators, and so trivially, there is no rows in which the premises all have a 1 under their main operators and the conclusion has a 0 under its main operator. Put otherwise, every row in which the premises all have 1 under their main operators— all zero of them— is such that the conclusion has a 1 under its main operator. Thus the argument is valid.

It is worth contrasting the example above with the following argument:

B1. $\neg L \rightarrow (J \vee L)$
B2. $\underline{\neg L}$
B3. J

It is good practice to consult your intuitions about whether an argument is valid before beginning to complete its truth table. For instance, without looking any further, consider whether the argument above is valid. Once you have a guess we may construct the following:

J	L	$\neg$	L	$\rightarrow$	$(J$	$\vee$	$L)$	$\neg$	L	J
1	1	0	1	**1**	1	1	1	**0**	1	**1**
1	0	1	0	**1**	1	1	0	**1**	0	**1**
0	1	0	1	**1**	0	1	1	**0**	1	**0**
0	0	1	0	**0**	0	0	0	**1**	0	**0**

To determine whether the argument is valid, check to see whether there are any rows on which both premises have a 1 under their main operators, but the conclusion has a 0 under its main operator. Notice that unlike the previous argument, there is a row in which both premises have a 1 under their main operators. Nevertheless, the only row in which both the premises have a 1 under their main operators is the second row, and in that row the conclusion also has a 1 under its main operator. It follows that the argument is valid in $\mathcal{L}^{\text{PL}}$, though it is not vacuously valid as before. Rather, vacuous validity is a special case.

Here is another example. Is the following argument valid? Try to check intuitively first.

C1. $P \rightarrow Q$
C2. $\underline{\neg P}$
C3. $\neg Q$

P	Q	P	$\rightarrow$	Q	$\neg$	P	$\neg$	Q
1	1	1	**1**	1	**0**	1	**0**	1
1	0	1	**0**	0	**0**	1	**1**	0
0	1	0	**1**	1	**1**	0	**0**	1
0	0	0	**1**	0	**1**	0	**1**	0

On the third row, each premise has a 1 under its main operator but the conclusion does not, and so the argument is invalid: the conclusion is not a logical consequence of its premises.

2.4 Decidability

Evaluating wfss in $\mathcal{L}^{\text{PL}}$ by constructing complete truth tables provides a simple mechanical procedure that is straightforward to systematically employ. Moreover, since every wfs of $\mathcal{L}^{\text{PL}}$ is of finite length and contains a finite number of sentence letters, constructing a complete truth table for a wfs of $\mathcal{L}^{\text{PL}}$ provides a finite procedure which determines whether that sentence is a tautology, contradiction, or logical contingent. Put otherwise, constructing a complete truth table for a wfs of $\mathcal{L}^{\text{PL}}$ provides an EFFECTIVE METHOD for determining whether that wfs is a tautology or not, and similarly for the other logical properties that sentence may or may not have. Since there is an effective method for determining whether a wfs of $\mathcal{L}^{\text{PL}}$ is a tautology, we may say that it is DECIDABLE whether a wfs of $\mathcal{L}^{\text{PL}}$ is a tautology. It is similarly decidable whether a wfs of $\mathcal{L}^{\text{PL}}$ is a contradiction or logically contingent.

Given a finite set Γ of wfss of $\mathcal{L}^{\text{PL}}$, constructing a complete truth table for those wfss provides an effective method for determining whether Γ is satisfiable or not, and so the question of whether Γ is satisfiable is decidable. However, the same cannot be said for infinite sets of wfss of $\mathcal{L}^{\text{PL}}$. Even though each wfs of $\mathcal{L}^{\text{PL}}$ is of finite length with finitely many sentence letters, an infinite set of wfss of $\mathcal{L}^{\text{PL}}$ may contain infinitely many sentence letters, requiring an infinitely large truth table. Although there is nothing to prevent us from *defining* infinitely large truth tables mathematically, there is of course little hope of *using* such an infinite truth table to determine whether an infinite set of sentences is satisfiable. It is for this reason that the truth table method does not provide an effective procedure for deciding whether a set of wfs of $\mathcal{L}^{\text{PL}}$ is satisfiable. Of course, there could be another effective method for determining whether a set of $\mathcal{L}^{\text{PL}}$ sentences is satisfiable, and so we cannot claim that it is UNDECIDABLE whether a set of sentences is satisfiable just because one method cannot be used.

Although it is beyond the scope of this course, it is in fact undecidable whether an infinite set of wfs of $\mathcal{L}^{\text{PL}}$ is satisfiable or not. That is, there is no effective procedure that we could hope to use to determine whether any infinite set of wfs of $\mathcal{L}^{\text{PL}}$ is satisfiable. However, even in restricting consideration to finite sets, we may observe that it is often infeasible to construct truth tables for a wfs or set of wfss of $\mathcal{L}^{\text{PL}}$ which include too many sentence letters. For instance, a wfs or set of wfss of $\mathcal{L}^{\text{PL}}$ which includes 5 sentence letters would require 32 lines, and double that again for 6 sentence letters. At least for humans using pen and paper, this is at about the limit for what it is possible to use without making mistakes.

Here one might be tempted to respond by appealing to computers. Instead of attempting to write out truth tables by hand, perhaps the method is best developed with computational assistance to avoid making mistakes. This leads into the *Boolean satisfiability problem* in computer science, and also falls outside the scope of this course. Rather, we will be concerned with methods for working out reasoning on paper in a finite amount of time. In later chapters, we will have reason to require these methods to simulate certain natural patterns of reasoning. However, before then it will be important to clean up a few loose ends in order to set the stage for these developments. We will begin by presenting a different procedure.

2.5 Partial Truth Tables

To show that a wfs of $\mathcal{L}^{\text{PL}}$ is a tautology, we need to show that 1 occurs below its main operator on every row of its complete truth table. So we need a complete truth table. However, to show that a wfs is *not* a tautology we only need to complete a row in which 0 is beneath its main operator. Therefore, in order to show that a wfs is not a tautology, it is enough to provide a *partial truth table* regardless of how many sentence letters the wfs might include.

For example, consider $(U \wedge T) \rightarrow (S \wedge W)$. We want to show that it is *not* a tautology by providing a partial truth table. To do so, we begin by writing 0 under the main operator which is a material conditional. In order for the conditional to be false, there must be a 1 under the antecedent and a 0 under the consequent. We fill these in as follows:

S	T	U	W	$(U$	$\wedge$	$T)$	$\rightarrow$	$(S$	$\wedge$	$W)$
					1		**0**		0	

In order for a 1 to occur under $U \wedge T$, a 1 must also occur under both U and T as follows:

S	T	U	W	$(U$	$\wedge$	$T)$	$\rightarrow$	$(S$	$\wedge$	$W)$
	1	1		1	1	1	**0**		0	

Remember that each instance of a given sentence letter must have the same truth-value in a given row of a truth table. You can't have 1 occur under one instance of U and 0 occur under another instance of U in the same row. Thus we put a 1 under each instance of U and T.

Now we just need to work out what follows from the 0 under $S \wedge W$. In particular, a 0 must occur under either S or W, or both. Making an arbitrary decision, we may finish the table:

S	T	U	W	$(U$	$\wedge$	$T)$	$\rightarrow$	$(S$	$\wedge$	$W)$
0	1	1	0	1	1	1	**0**	0	0	0

Although showing that a wfs of $\mathcal{L}^{\text{PL}}$ is a tautology requires a complete truth table, showing that a wfs of $\mathcal{L}^{\text{PL}}$ is not a tautology only requires a partial truth table with a single row where 0 occurs below the main operator of that wfs. That's what we've just done. In just the same way, to show that a wfs of $\mathcal{L}^{\text{PL}}$ is not a contradiction, you only need to construct a single row of a truth table where 1 occurs below the main operator of that wfs. By contrast, to show that a wfs of $\mathcal{L}^{\text{PL}}$ is a contradiction, you must show that a 0 occurs below the main operator on *every* row of a complete truth table, and so you need a complete truth table.

A wfs of $\mathcal{L}^{\text{PL}}$ is contingent just in case its complete truth table has a row in which 1 occurs below the main operator and another row in which 0 occurs below the main operator. Thus to show that a wfs of $\mathcal{L}^{\text{PL}}$ is contingent requires a partial truth table with just two rows. For example, we can show that the sentence above is contingent as follows:

S	T	U	W	$(U$	$\wedge$	$T)$	$\rightarrow$	$(S$	$\wedge$	$W)$
0	1	1	0	1	1	1	**0**	0	0	0
0	1	0	0	0	0	1	**1**	0	0	0

Just as there happens to be more than one combination of truth-values which makes the sentence false, there are even more ways to make the sentence true. However, this is not always the case. For instance, given a sentence letter A, there is exactly two lines in its complete truth table, one in which it is true, and the other in which it is false.

Showing that a wfs of $\mathcal{L}^{\text{PL}}$ is not contingent requires providing a complete truth table. In particular, one must either show that the complete truth table for the wfs has a 1 under its main operator on all rows, or show that the wfs has a 0 under its main operator on all rows. If you do not know whether a particular sentence is contingent or not, then you do not know whether you will need a complete or partial truth table. One way to proceed is to start working on a complete truth table, stopping as soon as you complete rows that show the sentence is contingent. If not, then you must complete the truth table in full.

Showing that two wfss of $\mathcal{L}^{\text{PL}}$ are logically equivalent requires providing a complete truth table. By contrast, showing that two wfss of $\mathcal{L}^{\text{PL}}$ are not logically equivalent only requires a partial truth table with one row in which a 1 occurs below the main operator of one of the wfss and a 0 occurs below the main operator of the other wfs.

Showing that a set of wfss of $\mathcal{L}^{\text{PL}}$ is satisfiable requires a single row of a truth table in which a 1 occurs below the main operator of every wfs in the set. However, to show that a set of wfss of $\mathcal{L}^{\text{PL}}$ is unsatisfiable requires a complete truth table since you must show that on every row of a complete truth table for the set, there is a 0 below the main operator of at least one of the wfss in the set. Of course, we could only hope to succeed if the set of wfss is finite.

Showing that an argument is valid requires a complete truth table. Showing that an argument is invalid only requires providing a single row of a partial truth table in which a 1 occurs below the main operator of every premise and a 0 occurs below the conclusion. Thus we have:

	YES	NO
Tautology	complete truth table	one-line partial truth table
Contradiction	complete truth table	one-line partial truth table
Contingent	two-line partial truth table	complete truth table
Entailment	complete truth table	one-line partial truth table
Equivalent	complete truth table	one-line partial truth table
Satisfiable	one-line partial truth table	complete truth table
Valid	complete truth table	one-line partial truth table

The table above summarizes when a complete truth table is required and when a partial truth table will suffice. If you are trying to remember whether you need a complete truth table or not, the general rule is, if you're looking to establish a claim about *every* interpretation, you need a complete table. Otherwise, a one-line or perhaps two-line truth table may do instead.

2.6 Semantics

Although partial truth tables might help to avoid doing some amount of work, invariably you will end up needing to construct complete truth tables for sometimes long wfs of $\mathcal{L}^{\text{PL}}$ or else for arguments or sets which include many wfss of $\mathcal{L}^{\text{PL}}$. In addition to being extremely tedious and time consuming, constructing large truth tables is also highly prone to human error. These provide some initial reasons that one might hope to devise an alternative.

Complete truth tables also played an important role in defining what it is for a wfs of $\mathcal{L}^{\text{PL}}$ to be a tautology, where something similar may be said for the definitions of a contradiction as well as a logically contingent wfs of $\mathcal{L}^{\text{PL}}$. Similarly, the definitions for logically entailment, logical equivalence, satisfiability, logical consequence, and the validity of arguments given above all appealed to complete truth tables. Whereas the corresponding definitions for the sentences of English could not be made precise given that there is no well-defined sense of what counts as a grammatical sentence of English, the sections above appealed to complete truth tables for the wfss of $\mathcal{L}^{\text{PL}}$ in order to avoid this problem. Nevertheless, the truth table definitions of the logical notions considered above still leave something to be desired.

So far we have identified the rows of a truth table for a relevant wfs or set of wfss of $\mathcal{L}^{\text{PL}}$ with the relevant range of interpretations for that wfs or set of wfss. Of course, the truth tables that we can write down are are finite in size and so cannot specify truth-values for every sentence letter of $\mathcal{L}^{\text{PL}}$. Rather, the rows of a truth table provide PARTIAL INTERPRETATIONS of $\mathcal{L}^{\text{PL}}$ by specifying all combinations of truth-values for some finite set of sentence letters. Insofar as a truth table includes all of the sentence letters that occur in the wfs(s) in question, the truth-values for all other sentence letters do not make a difference, and so may be safely ignored. Nevertheless, we cannot define the interpretations of $\mathcal{L}^{\text{PL}}$ as the rows of any finite truth table since no finite truth table specifies truth-values for every sentence letter. This provides further motivation to present a more general approach.

Next we may consider the definition of a complete truth table itself. Rather than providing a formal definition, we appealed to the table that results from writing the wfs(s) of $\mathcal{L}^{\text{PL}}$ in question at the top right of the table with all of the sentence letters it contains at the top left. We then provided a procedure for adding the appropriate number of rows depending on how many sentence letters were involved and distributing truth-values accordingly. The rest of the values in the table were then to be added by appealing to the characteristic truth tables as a rubric. This left open certain ambiguities like the order of the sentence letters as well as the order of the wfss in a complete truth table for a set of wfss of $\mathcal{L}^{\text{PL}}$. Additionally, at least given what we have said so far, one must rely on an intuitive grasp of the main operator for each subsentence, as well as how to correctly apply the characteristic truth tables.

Although with some ingenuity we could tighten up all of these details by either eliminating ambiguities or else establishing that the ambiguities do not make a difference, the definitions themselves are bound to become even more cumbersome to state precisely. Rather, the truth table definitions given above are best understood as intuitive approximations of the precise definitions to which we will soon turn. Despite their imprecision, the truth table definitions provide a natural and accessible account of the logical notions that we are after, and so may be preserved as helpful heuristics in contemplating the abstract definitions to follow.

Rather than relying on diagrams, officially we will take an INTERPRETATION of $\mathcal{L}^{\text{PL}}$ to be any function $\mathcal{I}$ from the set of sentence letters for $\mathcal{L}^{\text{PL}}$ to the set of truth-values $\{1,0\}$, thereby assigning every sentence letter to exactly one truth-value. Although interpretations only specify the truth-values of sentence letters, we may draw on any interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ in order to define another function which assigns every wfs of $\mathcal{L}^{\text{PL}}$ to exactly one truth-value in accordance with the characteristic truth tables but without appealing to the characteristic truth tables. More precisely, given any interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$, we may recursively define the VALUATION FUNCTION $\mathcal{V}_{\mathcal{I}}$ over the domain of wfss for $\mathcal{L}^{\text{PL}}$ by way of the following semantics:

VALUATION FUNCTION: For any wfss φ and ψ of $\mathcal{L}^{\text{PL}}$:

$\mathcal{V}_{\mathcal{I}}(\varphi) = \mathcal{I}(\varphi)$ *if* φ is a sentence letter of $\mathcal{L}^{\text{PL}}$.

$\mathcal{V}_{\mathcal{I}}(\neg\varphi) = 1$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 0$.

$\mathcal{V}_{\mathcal{I}}(\varphi \vee \psi) = 1$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ or $\mathcal{V}_{\mathcal{I}}(\psi) = 1$ (or both).

$\mathcal{V}_{\mathcal{I}}(\varphi \wedge \psi) = 1$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ and $\mathcal{V}_{\mathcal{I}}(\psi) = 1$.

$\mathcal{V}_{\mathcal{I}}(\varphi \rightarrow \psi) = 1$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 0$ or $\mathcal{V}_{\mathcal{I}}(\psi) = 1$ (or both).

$\mathcal{V}_{\mathcal{I}}(\varphi \leftrightarrow \psi) = 1$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = \mathcal{V}_{\mathcal{I}}(\psi)$.

The clauses above hold for all sentences φ and ψ of $\mathcal{L}^{\text{PL}}$, thereby extending any interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ to $\mathcal{V}_{\mathcal{I}}$ in order to specify a unique truth-value for every wfs of $\mathcal{L}^{\text{PL}}$. Whereas the characteristic truth tables for the operators specify the truth-values for complex sentences visually, the semantic clauses above specify the same information functionally.

In Chapter 1, we specified the primitive symbols for $\mathcal{L}^{\text{PL}}$. These included the sentence letters, punctuation, and sentential operators. When interpreting $\mathcal{L}^{\text{PL}}$, you are not allowed to change the meaning of the sentential operators. For instance, you cannot take the '$\neg$' symbol to mean what '$\wedge$' usually does. Rather, the '$\neg$' symbol will always have the same semantic clause, and so will always express the same truth-function for negation. Since the meanings for the sentential operators are fixed by the semantic clauses given in the definition of the valuation function, it is common to refer to the sentential operators as LOGICAL CONSTANTS.

The sentence letters are sometimes referred to as the NON-LOGICAL VOCABULARY and are interpreted by assigning them to either 1 or 0, where it is the combination of assignments which may differ between interpretations of $\mathcal{L}^{\text{PL}}$. Accordingly, when we translate an argument from English into $\mathcal{L}^{\text{PL}}$, the sentence letter 'M' does not have its meaning fixed as a result. Rather, we rely on interpretations to assign truth-values to sentence letter such as 'M', where the truth-values provide a maximally course-grained way to model what those sentence letters mean, i.e., whether they express a proposition that obtains or does not obtain.

2.7 Formal Definitions

Having provided a definition of the interpretations of $\mathcal{L}^{\text{PL}}$ that is both mathematically precise and simple to state, we may now put this definition to work to redefine the logical properties and relations discussed above. Letting φ and ψ be wfs of $\mathcal{L}^{\text{PL}}$, Γ be a set of wfss of $\mathcal{L}^{\text{PL}}$, and $\mathcal{I}$ and $\mathcal{J}$ be interpretations of $\mathcal{L}^{\text{PL}}$, we may present the following official definitions:

TAUTOLOGY: φ is a *tautology iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ for all $\mathcal{I}$.

CONTRADICTION: φ is a *contradiction iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 0$ for all $\mathcal{I}$.

CONTINGENT: φ is *logically contingent iff* $\mathcal{V}_{\mathcal{I}}(\varphi) \neq \mathcal{V}_{\mathcal{J}}(\varphi)$ for some $\mathcal{I}$ and $\mathcal{J}$.

ENTAILMENT: φ *logically entails* ψ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) \leq \mathcal{V}_{\mathcal{I}}(\psi)$ for all $\mathcal{I}$.

EQUIVALENCE: φ is *logically equivalent* to ψ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = \mathcal{V}_{\mathcal{I}}(\psi)$ for all $\mathcal{I}$.

SATISFIABLE: Γ is *satisfiable iff* there is some $\mathcal{I}$ where $\mathcal{V}_{\mathcal{I}}(\gamma) = 1$ for all $\gamma \in \Gamma$.

CONSEQUENCE: $\Gamma \vDash \varphi$ *iff* $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ for all $\mathcal{I}$ where $\mathcal{V}_{\mathcal{I}}(\gamma) = 1$ for all $\gamma \in \Gamma$.

Note that the double turnstile '$\vDash$' has been introduced for the logical consequence relation which— like the schematic variable 'φ'— is part of the *metalanguage* that we are using to discuss $\mathcal{L}^{\text{PL}}$, and not a part of $\mathcal{L}^{\text{PL}}$ itself. Although officially $\vDash$ takes a set of $\mathcal{L}^{\text{PL}}$ wfss on the left together with a single $\mathcal{L}^{\text{PL}}$ wfs on the right, it is both common and convenient to drop the set notation, writing '$\varphi_1, \ldots, \varphi_n \vDash \psi$' instead of '$\{\varphi_1, \ldots, \varphi_n\} \vDash \psi$'.

As before, we may say that an argument in $\mathcal{L}^{\text{PL}}$ is VALID just in case it's conclusion is a logical consequence of its set of premises. Recall that an argument in $\mathcal{L}^{\text{PL}}$ is a sequence of $\mathcal{L}^{\text{PL}}$ wfs, and so a completely different type of thing than a set of wfs of $\mathcal{L}^{\text{PL}}$. After all, sets do not specify the order of their members. Accordingly, we cannot say that for any set of $\mathcal{L}^{\text{PL}}$ wfss Γ and wfs φ, if $\Gamma \vDash \varphi$, then *the* argument whose premises are the wfss in Γ and whose conclusion is φ is valid. This is because there may fail to be a unique argument that we can construct from a set of $\mathcal{L}^{\text{PL}}$ wfss Γ and a further $\mathcal{L}^{\text{PL}}$ wfs φ. For instance, assuming Γ includes at least two wfss, we can construct different arguments by reversing the order of the premises. Instead, we may claim something more general: $\Gamma \vDash \varphi$ just in case *every* argument where the wfss in Γ are the premises (in some order or other) and φ is the conclusion is valid.

Logical consequence is the most important semantic concept that we will study in this course. In the following chapter, we will introduce a proof theoretic analogue for derivability which we will represent with the single turnstile '$\vdash$'. As we will then go on to show in the metalogical portions of this book, these two relations have the same extension, providing two radically different perspectives on one and the same thing, i.e., formal reasoning. We will then repeat this same methodology for languages with greater expressive power. For the time being, we only pause to indicate the central role that logical consequence will play throughout this text.

2.8 Semantic Proofs

The formal definitions given above are certainly much more elegant than the truth table definitions for the corresponding terms, avoiding all the ambiguities that we mentioned but did not take the time to fully resolve above. Although these new definitions do not provide effective methods for deciding logical questions in the same way as the truth table definitions, they will nevertheless allow us to prove things about the wfss of $\mathcal{L}^{\text{PL}}$ and their various logical properties and relationships. For instance, consider the following set of wfss:

$$\Gamma = \{A \wedge C,\ B \leftrightarrow \neg A,\ C \rightarrow (B \wedge D),\ E\}$$

It turns out that this set is unsatisfiable. However, to show this using a truth table would requiring constructing a complete truth table with five sentence letters and thirty-two rows. In addition to the tedium, the chances of making a mistake cannot be overlooked. Instead of attempting this, we can prove that Γ is unsatisfiable by assuming that it is satisfiable and appealing to the formal definitions in order to derive a contradiction:

> *Proof:* Assume for contradiction that the set Γ is satisfiable. By the definition of satisfiability, $\mathcal{V}_{\mathcal{I}}(A \wedge C) = \mathcal{V}_{\mathcal{I}}(B \leftrightarrow \neg A) = \mathcal{V}_{\mathcal{I}}(C \rightarrow (B \wedge D)) = \mathcal{V}_{\mathcal{I}}(E) = 1$ for some interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$. It follows that $\mathcal{V}_{\mathcal{I}}(A) = \mathcal{V}_{\mathcal{I}}(C) = 1$ by the semantics for conjunction, and so $\mathcal{V}_{\mathcal{I}}(\neg A) = 0$ given the semantics for negation. Since $\mathcal{V}_{\mathcal{I}}(B) = \mathcal{V}_{\mathcal{I}}(\neg A)$ by the semantics for the biconditional, $\mathcal{V}_{\mathcal{I}}(B) = 0$.
>
> Since $\mathcal{V}_{\mathcal{I}}(C \rightarrow (B \wedge D)) = 1$, we know by the semantics for the conditional that either $\mathcal{V}_{\mathcal{I}}(C) = 0$ or $\mathcal{V}_{\mathcal{I}}(B \wedge D) = 1$. Having already shown that $\mathcal{V}_{\mathcal{I}}(C) \neq 0$, we may conclude that $\mathcal{V}_{\mathcal{I}}(B \wedge D) = 1$, and so $\mathcal{V}_{\mathcal{I}}(B) = 1$ and $\mathcal{V}_{\mathcal{I}}(D) = 1$ by the semantics for conjunction. Thus $\mathcal{V}_{\mathcal{I}}(B) \neq 0$, contradicting the above. □

This proof has been made fully explicit so as to make it easy to follow. A proof that is hard to follow is not a very good proof. However, using too many words is not a good thing either, cluttering what might be easier to see otherwise. Writing clear and readable proofs is a skill that requires judgment and practice. If you are new to proof writing, it is best to begin by making everything explicit before tightening things up to write with concision.

Note that the proof stopped once we got a contradiction. Although we might have continued by saying that our assumption must be false given the contradiction that we derived, this much is automatic given the introductory clause "Assume for contradiction…" which sets up this expectation. This is a good example of the kinds of writing strategies that you can use to write informal proofs. By contrast to the formal proofs that we will go on to write in the following chapter, informal proofs are written in the metalanguage mathematical English and are often about wfss of our object language $\mathcal{L}^{\text{PL}}$. In the case above, we established that Γ is unsatisfiable. In order to give you a sense of some of the other methods and phrasing for writing clear and concise informal proofs, it will help to consider a few more examples.

D1. Either the butler is the murderer or the gardener isn't who he says he is.
D2. The gardener is who he says he is.
D3. The butler is the murderer.

In order to determine its validity, let's translate this argument into $\mathcal{L}^{\text{PL}}$ and evaluate the resulting formal argument for validity with a truth table. In particular:

B: The butler is the murderer.
G: The gardener is who he says he is.

E1. $B \vee \neg G$
E2. G
E3. B

It should be pretty clear that this is a valid argument, but to show this we may now write a semantic proof which establishes its validity. Whereas the proof given above proceeded by *reductio ad absurdum*, deriving a contradiction from the negation of what we wanted to prove, it is perhaps easiest to write a direct proof that the argument above is valid.

> *Proof:* Let $\mathcal{I}$ be an arbitrary interpretation of $\mathcal{L}^{\text{PL}}$ for which both of the premises are true, i.e., $\mathcal{V}_{\mathcal{I}}(B \vee \neg G) = \mathcal{V}_{\mathcal{I}}(G) = 1$. We know by the semantics for negation that $\mathcal{V}_{\mathcal{I}}(\neg G) = 0$, where either $\mathcal{V}_{\mathcal{I}}(B) = 1$ or $\mathcal{V}_{\mathcal{I}}(\neg G) = 1$ by the semantics for disjunction, and so may conclude that $\mathcal{V}_{\mathcal{I}}(B) = 1$. Since $\mathcal{I}$ was arbitrary, we may conclude more generally that $\mathcal{V}_{\mathcal{I}}(B) = 1$ for any $\mathcal{I}$ where $\mathcal{V}_{\mathcal{I}}(B \vee \neg G) = \mathcal{V}_{\mathcal{I}}(G) = 1$, and so the argument is indeed valid. $\square$

It is typical to leave this final sentence off assuming that it is sufficiently clear what you are setting out to prove and how you are intended to do so. Nevertheless, it doesn't hurt to include this extra line to make it especially clear in case you are uncertain whether your proof is easy to interpret. In the case above, the key signpost that we used was the construction '*Let $\mathcal{I}$ be an arbitrary interpretation of $\mathcal{L}^{\text{PL}}$ ...*' since this signals that we will establish something general about $\mathcal{L}^{\text{PL}}$ interpretations. We then restrict consideration to those interpretations $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ which make the premises true with 'for which both of the premises are true', checking to see that such an arbitrary interpretation makes the conclusion true.

Whereas the first semantic argument that we gave allowed us to show that Γ is unsatisfiable without having to draw a truth table with thirty-two rows, the winnings in the case above are somewhat less substantial. In particular, the truth table only requires that we include four rows since there are just two sentence letters. Thus we have:

B	G	$(B$	$\vee$	$\neg$	$G)$	G	B
1	1	1	**1**	0	1	**1**	**1**
1	0	1	**1**	1	0	**0**	**1**
0	1	0	**0**	0	1	**1**	**0**
0	0	0	**1**	1	0	**0**	**0**

If the argument were invalid, there would be a row on which the first two bold values are 1 but the third is 0. There is no such row, so the argument is valid. Put otherwise, every row in which the premises have a 1 under their main operators is a row in which the conclusion has a 1 under its main operator. Thus the argument in $\mathcal{L}^{\text{PL}}$ is valid.

In this case, both the semantic proof and truth table methods require a similar amount of work. Given either method, we may conclude that the argument in $\mathcal{L}^{\text{PL}}$ is valid. We may then go on to explain that the English argument is valid by appealing to the fact that the English argument can be regimented by a valid argument in $\mathcal{L}^{\text{PL}}$.

Before pressing on, it will be important to consider one more type of semantic proof which appeals to cases. Consider the following argument:

F1. Either Kat or Lu will with the race.
F2. If Kat wins, then she will celebrate.
F3. If Lu wins, then she will celebrate.
F4. Either Kat or Lu will celebrate.

This argument can be regimented as follows:

K: Kat will win.
L: Lu will win.
C_1: Kat will celebrate.
C_2: Lu will celebrate.

G1. $K \vee L$
G2. $K \to C_1$
G3. $L \to C_2$
G4. $C_1 \vee C_2$

This argument may seem plain enough since this is just the kind of day-to-day reasoning that we are all accustomed to doing. Nevertheless, providing a truth table would require sixteen rows. Although the semantic proof is easier than attempting to fill out so many rows without making any mistakes, there is no way to avoid a *proof by cases*, at least insofar as we are to provide a direct proof that the $\mathcal{L}^{\text{PL}}$ argument is valid.

Writing with slightly more concision than before, we may reason as follows:

Proof: Let $\mathcal{I}$ be an arbitrary interpretation of $\mathcal{L}^{\text{PL}}$ where:

$$\mathcal{V}_{\mathcal{I}}(K \vee L) = \mathcal{V}_{\mathcal{I}}(K \to C_1) = \mathcal{V}_{\mathcal{I}}(L \to C_2) = 1.$$

By the semantics for disjunction, either $\mathcal{V}_{\mathcal{I}}(K) = 1$ or $\mathcal{V}_{\mathcal{I}}(L) = 1$. Consider:

Case 1: Assume $\mathcal{V}_{\mathcal{I}}(K) = 1$. By the semantics for the conditional, it follows from the above that either $\mathcal{V}_{\mathcal{I}}(K) = 0$ or $\mathcal{V}_{\mathcal{I}}(C_1) = 1$, and so $\mathcal{V}_{\mathcal{I}}(C_1) = 1$. Thus $\mathcal{V}_{\mathcal{I}}(C_1 \vee C_2) = 1$ follows from the semantics for disjunction.

Case 2: Assume $\mathcal{V}_{\mathcal{I}}(L) = 1$. By the semantics for the conditional, it follows from the above that either $\mathcal{V}_{\mathcal{I}}(L) = 0$ or $\mathcal{V}_{\mathcal{I}}(C_2) = 1$, and so $\mathcal{V}_{\mathcal{I}}(C_1) = 1$. Thus $\mathcal{V}_{\mathcal{I}}(C_1 \vee C_2) = 1$ follows from the semantics for disjunction.

Thus $\mathcal{V}_{\mathcal{I}}(C_1 \vee C_2) = 1$ in both cases. Since $\mathcal{I}$ was arbitrary, the conclusion is true on any interpretation where the premises true, and so the argument is valid. □

Although the method of proof by cases is an essential tool to have in your toolkit, these proofs are often harder to read and can also be harder to write clearly. Accordingly, proofs by cases are to be avoided whenever possible. In the case above, we could have avoided introducing cases by using a *reductio* argument (short for *reductio ad absurdum*) instead.

Proof: Assume for contradiction that the argument is invalid. Thus the conclusion is not a logical consequence of the premises, and so there is some interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ where $\mathcal{V}_{\mathcal{I}}(K \vee L) = \mathcal{V}_{\mathcal{I}}(K \to C_1) = \mathcal{V}_{\mathcal{I}}(L \to C_2) = 1$ and $\mathcal{V}_{\mathcal{I}}(C_1 \vee C_2) = 0$. By the semantics for disjunction, both $\mathcal{V}_{\mathcal{I}}(C_1) = \mathcal{V}_{\mathcal{I}}(C_2) = 0$.

By the semantics for the material conditional, either $\mathcal{V}_{\mathcal{I}}(K) = 0$ or $\mathcal{V}_{\mathcal{I}}(C_1) = 1$, and similarly, either $\mathcal{V}_{\mathcal{I}}(L) = 0$ or $\mathcal{V}_{\mathcal{I}}(C_2) = 1$. Given the above, $\mathcal{V}_{\mathcal{I}}(K) = \mathcal{V}_{\mathcal{I}}(L) = 0$, and so $\mathcal{V}_{\mathcal{I}}(K \vee L) = 0$ by the semantics for disjunction, contradicting the above. □

This proof doesn't worry about cases and in that way is a bit easier to follow. Nevertheless, direct proofs are often preferable to *reductio* style proofs since deriving a contradiction is not quite as explanatory as what a direct proof shows. Which proof strategy to use is a judgment call and it can take some practice to know which approach is best to use when. Even once you have chosen a basic strategy, the way that you order the steps of your proof can also have a significant effect on the clarity of your resulting proof. These are points that are worth considering as you practice writing clear and concise proofs in this course.

So far, we have only provided semantic proofs for claims that would have required a complete truth table. This is no accident. For instance, in the case where a set of $\mathcal{L}^{\text{PL}}$ wfss is satisfiable, or an argument $\mathcal{L}^{\text{PL}}$ is valid, all we need to do is find a particular interpretation of the relevant sentence letters of $\mathcal{L}^{\text{PL}}$ in order to draw the desired conclusion. Here we may observe that providing a partial truth table does just that, and so is often to be preferred. We may then go on to record a particular row of a truth table by using the notation $\mathcal{I}(\varphi)$ to specify a truth value for each sentence letter φ that occurs in the truth table.

2.9 Tautologies and Weakening

What should we make of the following claim:

$$P \wedge Q \vDash A \leftrightarrow \neg\neg A.$$

Notice that the sentence letters on the left-hand-side are unrelated to the sentence letters on the right-hand-side. So there is a straightforward sense in which the two sides of the logical consequence above have *nothing to do with one another*. Nevertheless, the claim above is true: every interpretation for which the wfs on the left is true is also an interpretation on which the wfs on the right is true for the simple reason that $A \leftrightarrow \neg\neg A$ is true on every interpretation whatsoever. As a result, we could have dropped the wfs on the left entirely, writing:

$$\vDash A \leftrightarrow \neg\neg A.$$

The claim above is short for $\varnothing \vDash A \leftrightarrow \neg\neg A$ where $\varnothing$ is a convenient notation for the empty set $\{\}$, i.e., the set of no wfs of $\mathcal{L}^{\text{PL}}$. Since we said above that we are often going to drop the set notation when stating logical consequences, we do not need to write the empty set at all when a wfs is a logical consequence of the empty set of wfss as above.

It is easy to show that a wfs φ of $\mathcal{L}^{\text{PL}}$ is a tautology just in case $\vDash \varphi$. Since there are no wfss in $\varnothing$, every interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ vacuously makes all of the wfs in $\varnothing$ true, and so $\vDash \varphi$ just in case $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ for every interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$, i.e., φ is a tautology. Given this equivalence, it is common to define the $\mathcal{L}^{\text{PL}}$ tautologies in terms of logical consequence. In particular, we could have said that a wfs φ of $\mathcal{L}^{\text{PL}}$ is a TAUTOLOGY just in case $\vDash \varphi$.

In order to get another perspective on why φ is a tautology just in case $\vDash \varphi$, it will help to consider the set of interpretations that make a wfs φ of $\mathcal{L}^{\text{PL}}$ true. For brevity, we may define the INTERPRETATION SET $|\varphi| := \{\mathcal{I} : \mathcal{V}_{\mathcal{I}}(\varphi) = 1\}$ to be the set of all and only those $\mathcal{L}^{\text{PL}}$ interpretations $\mathcal{I}$ for which $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$. That is, given the set of all interpretations, each wfs φ of $\mathcal{L}^{\text{PL}}$ corresponds to a unique subset of interpretations in which φ is true, i.e., $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$. Or to put it another way, each wfs φ of $\mathcal{L}^{\text{PL}}$ amounts to a *constraint* on the interpretations of $\mathcal{L}^{\text{PL}}$ where only those interpretations $\mathcal{I}$ for which $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ satisfy the constraint.

Given the definition of an interpretation sets for the wfss of $\mathcal{L}^{\text{PL}}$, we may provide a slightly more abstract characterization of logical consequence. By the official definition, $\Gamma \vDash \varphi$ asserts that every $\mathcal{L}^{\text{PL}}$ interpretation $\mathcal{I}$ which makes every $\gamma \in \Gamma$ true also makes φ true. Another way to say this is that the intersection of all of the interpretation sets $|\gamma|$ for $\gamma \in \Gamma$ is a subset of the interpretation set for the conclusion $|\varphi|$. Formally, we may write this as follows:

$$\bigcap\{|\gamma| : \gamma \in \Gamma\} \subseteq |\varphi|.$$

What this says is that every interpretation $\mathcal{I}$ which belongs to every interpretation set $|\gamma|$ for $\gamma \in \Gamma$ is also in the interpretation set $|\varphi|$ for the conclusion. Since an interpretation $\mathcal{I}$ belongs to an interpretation set $|\psi|$ just in case $\mathcal{V}_{\mathcal{I}}(\psi) = 1$, this is equivalent to requiring every interpretation $\mathcal{I}$ where $\mathcal{V}_{\mathcal{I}}(\gamma) = 1$ for all $\gamma \in \Gamma$ to be such that $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$. But this is exactly what is asserted by $\Gamma \vDash \varphi$, only in the language of set theory.

Despite the equivalence, we will maintain our notation and definition for logical consequence $\Gamma \vDash \varphi$ rather than the set theoretic analogue given above. Nevertheless, it can help to take more than one perspective on the same thing, especially for a concept that is as important as logical consequence. The set theoretic analogue given above provides a vivid account of the way that each $\gamma \in \Gamma$ may constrain the interpretations of $\mathcal{L}^{\text{PL}}$ in which φ is said to be true. If there are no wfss in Γ at all, i.e., $\Gamma = \varnothing$, this corresponds to imposing no constraints on the interpretations of $\mathcal{L}^{\text{PL}}$ in which φ is said to be true. Put otherwise, all interpretations belong to $|\varphi|$, and so $\mathcal{V}_{\mathcal{I}}(\varphi) = 1$ for all interpretations $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$, i.e., φ is a tautology.

In order to make this more concrete, consider the following claims:

H1. $\vDash P \vee \neg P$.
H2. $\vDash P \leftrightarrow (P \vee (P \wedge Q))$.
H3. $\vDash (P \wedge \neg P) \rightarrow (A \vee B)$.

Although for different reasons, each of the claims above is true, and so the wfss on the right are all tautologies. It follows that the following claims are also true:

I1. $A, B \ \vDash P \vee \neg P$.
I2. $\neg P \ \vDash P \leftrightarrow (P \vee (P \wedge Q))$.
I3. $C \vee \neg C \ \vDash (P \wedge \neg P) \rightarrow (A \vee B)$.

Given that the wfss on the right are all tautologies, we can add whatever wfss that we like on the left. More generally, it is easy to prove the following principle:

Lemma 2.1 (*Weakening*) If $\Gamma \vDash \varphi$, then $\Gamma \cup \Sigma \vDash \varphi$.

This principle says that whenever φ is a logical consequence of Γ, then φ is also a logical consequence of $\Gamma \cup \Sigma$ for any set of wfss Σ of $\mathcal{L}^{\text{PL}}$ whatsoever where $\Gamma \cup \Sigma$ is the union set including all of the wfss in either Γ or Σ and nothing besides. In the special case where $\Gamma = \varnothing$ is empty, we may conclude that if φ is a tautology (i.e., $\vDash \varphi$), then φ is also a logical consequence of every set Σ of wfss of $\mathcal{L}^{\text{PL}}$ (i.e., $\Sigma \vDash \varphi$). It is an inference of this kind which is what justified drawing I1 – I3 as conclusions from H1 – H3 above.

The reason that weakening holds is easy to see given the set theoretic perspective on logical consequence presented before. Given that a logical consequence $\Gamma \vDash \varphi$ holds, adding further conditions beyond just those included in Γ will only further restrict the interpretations of $\mathcal{L}^{\text{PL}}$ for which φ is said to be true. Since we already know that all interpretations that make every $\gamma \in \Gamma$ true also make φ true given that $\Gamma \vDash \varphi$, then any subset of the interpretations that make every $\gamma \in \Gamma$ true is sure to also make φ true, and so $\Gamma \cup \Sigma \vDash \varphi$ for any set of wfss Σ. To put the point set theoretically, $\bigcap\{|\gamma| : \gamma \in \Gamma \cup \Sigma\} \subseteq \bigcap\{|\gamma| : \gamma \in \Gamma\}$ since $\Gamma \subseteq \Gamma \cup \Sigma$.

2.10 Contradictions and Unsatisfiability

What if a given set Γ of wfss $\mathcal{L}^{\text{PL}}$ imposes so many constraints that there is no interpretation $\mathcal{I}$ of $\mathcal{L}^{\text{PL}}$ which makes all the wfss $\gamma \in \Gamma$ true, i.e., $\mathcal{V}_{\mathcal{I}}(\gamma) = 1$ for all $\gamma \in \Gamma$? Put otherwise, what if Γ is unsatisfiable? In such a case, $\bigcap\{|\gamma| : \gamma \in \Gamma\} = \varnothing$, and since the empty set $\varnothing$ is a subset of every set, we know that $\varnothing \subseteq |\varphi|$ for any wfs φ of $\mathcal{L}^{\text{PL}}$. That is if there are no interpretations that make all of the wfss $\gamma \in \Gamma$ true, then it follows vacuously that all zero of those interpretations are guaranteed to make φ true. More succinctly, every wfs φ of $\mathcal{L}^{\text{PL}}$ is a logical consequence of any unsatisfiable set Γ of wfss of $\mathcal{L}^{\text{PL}}$ whatsoever.

Consider the following examples:

J1. $Q, \neg Q \vDash P$.
J2. $R \to (Q \wedge \neg Q), R \vDash \neg R$.
J3. $\neg(A \to A) \vDash (P \wedge \neg P) \to (A \vee B)$.

Some of the above are more obvious than others. For instance, it is easy to see that there is no interpretation of $\mathcal{L}^{\text{PL}}$ in which both $\mathcal{V}_{\mathcal{I}}(Q) = 1$ and $\mathcal{V}_{\mathcal{I}}(\neg Q) = 1$ since a contradiction follows immediately from the semantics for negation were we to assume otherwise. As a result, J1 is valid for the vacuous reason that no interpretation satisfies $\{Q, \neg Q\}$. Although slightly more work is required, J2 is also valid since $\{R \to (Q \wedge \neg Q), R\}$ is also unsatisfiable.

By similar reasoning, we can show that J3 is valid since $\{\neg(A \to A)\}$ is also unsatisfiable. However, unlike the previous cases, this set has only one member, namely the wfs $\neg(A \to A)$. Since $\neg(A \to A)$ can be shown to be a contradiction, it follows that any set of wfss to which it belongs is unsatisfiable. In particular, the singleton set $\{\neg(A \to A)\}$ is unsatisfiable, and so every wfs φ of $\mathcal{L}^{\text{PL}}$ is a logical consequence of $\{\neg(A \to A)\}$. More generally, any set containing a contradiction is unsatisfiable for the simple reason that there is no interpretation which makes a contradiction true. As a result, every wfs of $\mathcal{L}^{\text{PL}}$ is a logical consequence of a set of wfss of $\mathcal{L}^{\text{PL}}$ which includes a contradiction.

Having defined what it is for a wfs of $\mathcal{L}^{\text{PL}}$ to be a tautology in terms of logical consequence, it is natural to consider how to use the notion of logical consequence to define what it is for a wfs of $\mathcal{L}^{\text{PL}}$ to be a contradiction. It will help to consider the following claim:

$$P \wedge \neg P \vDash Q.$$

This statement is true. It says that Q is true in every $\mathcal{L}^{\text{PL}}$ interpretation in which $P \wedge \neg P$ is true. This follows vacuously since $P \wedge \neg P$ is not true in any $\mathcal{L}^{\text{PL}}$ interpretations, and so Q is true in every interpretation in which $P \wedge \neg P$ is true (all zero of them). Or to take another approach, think about what it would take for the logical consequence above to be false: there would have to be a $\mathcal{L}^{\text{PL}}$ interpretation in which $P \wedge \neg P$ is true and Q is false. But there are no $\mathcal{L}^{\text{PL}}$ interpretations in which $P \wedge \neg P$ is true, and so Q is a logical consequence.

Observe that the reason that the logical consequence above is true has nothing to do with the wfs Q. That is, we do not need to appeal to anything about the logical form of Q in order to explain why $P \wedge \neg P \vDash Q$. Thus the same considerations would demonstrate that $P \wedge \neg P \vDash \varphi$ for any wfs φ of $\mathcal{L}^{\text{PL}}$. Moreover, the same conclusion holds were we to replace $P \wedge \neg P$ with any other $\mathcal{L}^{\text{PL}}$ wfs that is not true on any interpretation. For instance, consider:

K1. $A \wedge \neg A \vDash \neg Q \to R$.
K2. $\neg(P \vee \neg P) \vDash (A_1 \vee A_2) \to \neg(A_3 \leftrightarrow (A_4 \wedge \neg A_2))$.

Since there are no $\mathcal{L}^{\text{PL}}$ interpretations in which the wfss on the left are true, you do not need to examine the wfss on the right to confirm that the logical consequences above are true. Given that the wfss on the right do not matter, it would be nice to have a way to represent that any wfs φ of $\mathcal{L}^{\text{PL}}$ is a logical consequence of the wfss on the left. There are two common ways to do this, but both make use of the same notation: $\Gamma \vDash \bot$. One way to interpret the logical consequence above is as a universal claim that Γ entails every wfs φ of $\mathcal{L}^{\text{PL}}$ whatsoever. Another way to interpret this logical consequence is to take '$\bot$' to abbreviate some particular contradiction of $\mathcal{L}^{\text{PL}}$, though it doesn't matter which. These conventions turn out to amount to the very same thing. For simplicity, we will assume the latter convention where '$\bot$' abbreviates the contradiction '$A \wedge \neg A$' for definiteness.

We have already observed that every wfs φ of $\mathcal{L}^{\text{PL}}$ is a logical consequence of an unsatisfiable set Γ of wfss of $\mathcal{L}^{\text{PL}}$. Given any unsatisfiable set Γ of wfss of $\mathcal{L}^{\text{PL}}$, it follows that $\bot$ in particular is a logical consequence of Γ, i.e., $\Gamma \vDash \bot$. Conversely, we may prove the following:

Lemma 2.2 If $\Gamma \vDash \bot$, then Γ is unsatisfiable.

Proof: Let Γ be an arbitrary set of wfss of $\mathcal{L}^{\text{PL}}$ where $\Gamma \vDash \bot$. Assume that Γ is satisfiable for contradiction. Thus there is a $\mathcal{L}^{\text{PL}}$ interpretation $\mathcal{I}$ where $\mathcal{V}_{\mathcal{I}}(\gamma) = 1$ for all $\gamma \in \Gamma$, and so $\mathcal{V}_{\mathcal{I}}(\bot) = 1$ given the assumption that $\Gamma \vDash \bot$. Given the definition of $\bot$, we may conclude that $\mathcal{V}_{\mathcal{I}}(A \wedge \neg A) = 1$. By the semantics for conjunction $\mathcal{V}_{\mathcal{I}}(A) = \mathcal{V}_{\mathcal{I}}(\neg A) = 1$, and so $\mathcal{V}_{\mathcal{I}}(A) = 0$ by the semantics for negation, resulting in a direct contradiction. Thus Γ is unsatisfiable. □

This shows that a set Γ of wfss of is unsatisfiable if and only if $\Gamma \vDash \bot$. In the special case where $\Gamma = \{\varphi\}$, we may say that φ is a contradiction just in case $\varphi \vDash \bot$. This provides a way to characterize contradictions in terms of logical consequence.

We may close by mentioning a connection between logical consequence and unsatisfiability that we will have occasion to return to later in the metalogical portions of this text.

Lemma 2.3 $\Gamma \vDash \varphi$ just in case $\Gamma \cup \{\neg\varphi\}$ is unsatisfiable.

We will make considerable use of this principle in proving things about the proof system for $\mathcal{L}^{\text{PL}}$ that we will now turn to introduce in the following chapter.