Solutions to Exercises 6 for *Introduction to Logic*

Robert Craven and Florian Steinberger

(Note that there's a lot more information here than you'd be expected to write in, for example, a test.)

Exercise 6.1

1. An **interpretation** (of a language for sentential logic) is a function $\mathfrak{I}$ which to each sentential variable A assigns a truth value.

So if $\mathcal{S}$ is the set of sentential variables $\{p_1, p_2, \ldots\}$, then $\mathfrak{I} : \mathcal{S} \to \{\mathsf{t}, \mathsf{f}\}$. For example, an interpretation $\mathfrak{I}$ might 'say that' p_1 is true and p_2 is false, which would be written as $\mathfrak{I}(p_1) = \mathsf{t}$ and $\mathfrak{I}(p_2) = \mathsf{f}$. Interpretations fix which sentential variables are true, and which are false.

2. A **valuation in sentential logic relative to** $\mathfrak{I}$ (where $\mathfrak{I}$ is an interpretation) is a function that assigns a truth value to each formula in the language, in a way that respects the intended meaning of the sentential connectives.

More explicitly: a valuation in sentential logic relative to $\mathfrak{I}$ is a function $\mathfrak{V}_{\mathfrak{I}} : \mathcal{F} \to \{\mathsf{t}, \mathsf{f}\}$ such that:

$$
\begin{aligned}
\mathfrak{V}_{\mathfrak{I}}(p_i) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{I}(p_i) = \mathsf{t} \\
\mathfrak{V}_{\mathfrak{I}}(\neg A) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{f} \\
\mathfrak{V}_{\mathfrak{I}}(A \wedge B) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t} \text{ and } \mathfrak{V}_{\mathfrak{I}}(B) = \mathsf{t} \\
\mathfrak{V}_{\mathfrak{I}}(A \vee B) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t} \text{ or } \mathfrak{V}_{\mathfrak{I}}(B) = \mathsf{t} \\
\mathfrak{V}_{\mathfrak{I}}(A \to B) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{f} \text{ or } \mathfrak{V}_{\mathfrak{I}}(B) = \mathsf{t} \\
\mathfrak{V}_{\mathfrak{I}}(A \leftrightarrow B) = \mathsf{t} &\quad \text{iff} \quad \mathfrak{V}_{\mathfrak{I}}(A) = \mathfrak{V}_{\mathfrak{I}}(B)
\end{aligned}
$$

(This is more explicit, because it makes it clear what it means for $\mathfrak{V}_{\mathfrak{I}}$ to respect the intended meaning of the connectives. Recall that $\mathcal{F}$ is the set of all formulas.)

Thus, a valuation in sentential logic relative to $\mathfrak{I}$ is one where the truth-values of the sentential variables are given by $\mathfrak{I}$, and where the truth-values of the sententially decomposable formulas respect the meanings of '$\neg$', '$\wedge$', '$\vee$', '$\to$' and '$\leftrightarrow$' (the sentential connectives). Valuations give the truth value for *every* formula of sentential logic, in a way that fits $\mathfrak{I}$ and the meanings of the sentential connectives.

3. A **contingent** formula A in $\mathcal{F}$ is one such that there are interpretations $\mathfrak{I}$ and $\mathfrak{I}'$ with $\mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t}$ and $\mathfrak{V}_{\mathfrak{I}'}(A) = \mathsf{f}$.

In other words: there is an assignment of truth-values to sentential variables on which the formula comes out true, and there is another assignment of truth-values to sentential variables on which the formula comes out false. The formula can be true (according to one interpretation), and it can be false (according to another interpretation). Contingent formulas are the ones which have at least one 't' and one 'f' under their main connectives in their truth tables.

4. A **satisfiable** formula A from $\mathcal{F}$ is one which is either contingent or tautological.

Recall that a formula A is tautological just when, for all interpretations $\mathfrak{I}$, we have $\mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t}$. A satisfiable formula is therefore one which can be 'made true' by a choice of an interpretation.

A satisfiable formula is one which has at least one '$\mathtt{t}$' underneath the main connective in its truth table. A formula A is satisfiable iff there is an interpretation $\mathfrak{I}$ with $\mathfrak{V}_{\mathfrak{I}}(A) = \mathtt{t}$ (where $\mathfrak{V}_{\mathfrak{I}}$ is a valuation in sentential logic relative to $\mathfrak{I}$).

5. A formula A **logically entails** a formula B iff there is no interpretation $\mathfrak{I}$ such that $\mathfrak{V}_{\mathfrak{I}}(A) = \mathtt{t}$ and $\mathfrak{V}_{\mathfrak{I}}(B) = \mathtt{f}$. In notation: $A \models B$.

An alternative: a formula A **logically entails** a formula B iff for every interpretation $\mathfrak{I}$ such that $\mathfrak{V}_{\mathfrak{I}}(A) = \mathtt{t}$, we also have $\mathfrak{V}_{\mathfrak{I}}(B) = \mathtt{t}$. (This second definition is equivalent to the first.)

This can also be phrased: B is a *logical consequence* of A. It can also be phrased: *A logically implies B*. (These mean the same. They have the same definition.) The truth of A guarantees the truth of B, in the sense that no matter how (i.e., under what interpretation) A is true, B is true also.

6. The argument form $A_1, \ldots, A_n \therefore C$ is said to be **valid** iff for all interpretations $\mathfrak{I}$ it is the case that *if* $\mathfrak{V}_{\mathfrak{I}}(A_1) = \mathtt{t}, \ldots, \mathfrak{V}_{\mathfrak{I}}(A_n) = \mathtt{t}$, *then* $\mathfrak{V}_{\mathfrak{I}}(C) = \mathtt{t}$.

Thus, the argument form is valid iff its premises logically entail its conclusion. (Since 'validity' seems like an accolade, this is a sign that the logic is going well!) Given the notation for logical entailment (a.k.a. 'logical consequence'), we can say that the argument form is valid iff $A_1, \ldots, A_n \models B$.

Exercise 6.2

(For convenience, I repeat in italics the claims we're asked to adjudicate.)

1. *An argument the conclusion of which is false is invalid.*

False. Why is it false? It states that any argument that has a false conclusion is invalid. An argument is invalid iff it is not valid. Given **6.1**(6), above, we know an argument form is valid iff every interpretation which makes all the premises true, also makes the conclusion true. Therefore, an argument form is invalid iff there is at least one interpretation which makes all premises true and the conclusion *false*. But we are not guaranteed invalidity if the conclusion is *actually* false (which is what the claim says). This is because one of the premises might *actually* be false too.

Here's an example which could help. Consider the argument:

> Aristotle was Greek,
> If Aristotle was Greek then Kant was Greek,
> $\therefore$
> Kant was Greek.

The conclusion here is actually false, yet the argument itself is valid. It has the form $p, p \to q \therefore q$, and the method of truth tables can be used to check that for all $\mathfrak{I}$, if $\mathfrak{V}_{\mathfrak{I}}(p) = \mathtt{t}$ and $\mathfrak{V}_{\mathfrak{I}}(p \to q) = \mathtt{t}$, then $\mathfrak{V}_{\mathfrak{I}}(q) = \mathtt{t}$. This example proves that the italic claim above is false (it's a *counterexample* to the claim).

Note that there are *some* arguments the conclusion of which is false, but which are invalid. It's just that not *every* argument with a false conclusion is like this. For example, consider:

> Aristotle was Greek,
> $\therefore$
> Kant was Greek.

Same conclusion as before, which is false. But this time the argument *is* invalid. It has the form $p \therefore q$, and here there is an interpretation which makes all the premises true, and the conclusion false.

2. *An argument with false premises and a true conclusion is invalid.*

False. We saw in (1) that an argument is invalid iff there is at least one interpretation which makes all premises of its argument form true and the conclusion *false.* But the fact that an argument has false premises and a true conclusion doesn't guarantee this. (If, for example, the premises of the argument form are *always*, under every interpretation, false, then the argument is going to be *valid.*)

Here's a counterexample to the claim again, to help:

> Nietzsche was Russian and if Nietzsche was Russian, then Aristotle was Greek,
> $\therefore$
> Aristotle was Greek.

Here, the premises (there's only one of them, a big long thing) are false, and the conclusion is true. But the argument isn't invalid. It has the form $p \wedge (p \to q) \therefore q$, and this is valid. For any interpretation $\mathfrak{I}$, it is true that if $\mathfrak{V}_{\mathfrak{I}}(p \wedge (p \to q)) = \mathsf{t}$, then $\mathfrak{V}_{\mathfrak{I}}(q) = \mathsf{t}$, as can be checked using truth tables.

Here's another counterexample:

> Boulez was French and Boulez wasn't French,
> $\therefore$
> Stockhausen was German.

This is an example in which, not only are the premises (just one again) actually false, but they are *always* false (for $p \wedge \neg p$ is, of course, contradictory). But the conclusion is true. And here too, the argument is valid: every interpretation which makes the premises true, makes the conclusion true.

Again, like before, there are *some* arguments with false premises and true conclusions which *are* invalid. For example:

> Aristotle was Peruvian,
> $\therefore$
> Kant lived in Königsberg.

False premise, true conclusion, but the form is $p \therefore q$, which is invalid.

3. *An argument with true premises and a false conclusion is invalid.*

True. Recall again that an argument form is *valid* iff every interpretation which makes all the premises true, also makes the conclusion true. So an argument form is invalid iff there is at least one interpretation which makes all premises true and the conclusion *false.* But this is just what the claim in italics says. So the claim is correct.

For example, consider the argument:

> Aristotle was Greek,
> Nietzsche respected Dostoyevsky,
> $\therefore$
> Kierkegaard liked to surf.

The premises are true, and the conclusion is false. The form is just $p, q \therefore r$, and this is certainly invalid: for consider the interpretation $\mathfrak{I}$ where $\mathfrak{I}(p) = \mathfrak{I}(q) = \mathsf{t}$ and $\mathfrak{I}(r) = \mathsf{f}$. This makes all premises true but the conclusion false. So it is *not* the case that every interpretation which makes all the premises true, also makes the conclusion true. So it is not the case that the argument form is valid (for that's just what validity is). So it is invalid.

(Note that here a single example doesn't *prove* that our claim is true.)

4. *An argument that has only true premises and a true conclusion is valid.*

False. The fact that the premises and conclusion are *actually* true doesn't mean that, in *all* circumstances where the premises are true, the conclusion is true as well. (And it's the latter that is required for validity.) There could be an interpretation where the premises are all true and the conclusion is false, which would make the argument invalid.

For example, consider the argument:

> Wittgenstein liked Schubert,
> $\therefore$
> Anscombe was Wittgenstein's student.

The premise and conclusion are true here. But the argument has the form $p \therefore q$, and this is invalid. To see why, consider the interpretation $\mathfrak{I}$ on which $\mathfrak{I}(p) = \mathsf{t}$ and $\mathfrak{I}(q) = \mathsf{f}$. For this interpretation, all the premises are true ($\mathfrak{V}_{\mathfrak{I}}(p) = \mathsf{t}$), but the conclusion is *not* true (we don't have $\mathfrak{V}_{\mathfrak{I}}(q) = \mathsf{t}$).

As before, note that there are *some* arguments which have true premises and a true conclusion, which *are* valid (just, not *all* of them). Here's an example:

> If Obama won the last election, he got a second term in office,
> Obama won the last election,
> $\therefore$
> Obama got a second term in office.

The premises and conclusion are all true. The argument has the form $p \to q, p \therefore q$, which is valid: any interpretation on which the premises are true, is also an interpretation on which the conclusion is true. To see this, note that the only interpretation $\mathfrak{I}$ which makes the premises true is that for which $\mathfrak{I}(p) = \mathfrak{I}(q) = \mathsf{t}$. Clearly, for this $\mathfrak{I}$, we have $\mathfrak{V}_{\mathfrak{I}}(q) = \mathsf{t}$, so the conclusion is true. The argument is therefore valid.

5. *Someone's claim that an argument form is valid, can be disproved by a single counterexample.*

True. For the claim that an argument form is valid, is just the claim that for *all* interpretations $\mathfrak{I}$ on which the premises are true, the conclusion is true too. We can disprove this by giving a counterexample, which will be a case where the premises are all true, and the conclusion false. We saw this in parts (1), (2) and (4) above: as well as giving some abstract reasoning why the claims are false, I also gave some particular counterexamples. Any one of these counterexamples was enough to *prove* that the relevant claim was false.

An example outside logic might make this clearer. Consider the claim:

> All the students are present.

This asserts a general claim about all the students. It can be shown to be false by a single counterexample—a case of a student who *isn't* present. And as we noted above, to *prove* a claim like this, it isn't generally sufficient to give a single example where the claim is true. For suppose there are 50 students. I can't show that 'All the students are present' is true by exhibiting some one particular student who happens to be present: for the other 49 might be flaked out at the beach.

6. *If a valid argument has a false conclusion, then all its premises must be false too.*

False. Take any valid argument, and suppose its conclusion is actually false. The corresponding argument form is valid iff every interpretation which makes all the premises true, also makes the conclusion true. So if all the premises are true in the interpretation which gives the actual truth-values, then the conclusion is true on that interpretation, too. But the conclusion is false on that interpretation. So it must be false that the all the premises are true. *But* from the fact that it's false that all the premises are true, we cannot conclude that all the premises are false. All we can conclude is that at least one of the premises must be false.

Here's an example. Consider the argument:

Wittgenstein liked Schubert,
If Wittgenstein liked Schubert, then Schubert liked Wittgenstein,
$\therefore$
Schubert liked Wittgenstein.

This argument is valid: it has the form $p, p \to q \therefore q$; and the conclusion is false. But it's not the case that all the premises are false: the first premise is true. (But, as the reasoning in the previous paragraph requires, at least one premise—the second, here—*is* false.)

Again, we can note that there are some valid arguments with false conclusions, where all the premises are false. Here is an example:

Aristotle was not Greek,
Plato was born on Mars,
$\therefore$
Aristotle was not Greek and Plato was born on Mars.

This is valid: it has the form $\neg p, q \therefore \neg p \land q$, whose validity can easily be confirmed using truth tables. The conclusion is false, and so are all of the premises.

7. *$A \models B$ if and only if $\models A \to B$.*

True. We can show this by giving a series of steps as follows:

$$\begin{array}{rll} A \models B & \text{iff} & \text{there is no interpretation } \mathfrak{I} \text{ such that } \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t} \text{ and } \mathfrak{V}_{\mathfrak{I}}(B) = \mathsf{f} \\ & \text{iff} & \text{for all interpretations } \mathfrak{I}, \text{ if } \mathfrak{V}_{\mathfrak{I}}(A) = \mathsf{t} \text{ then } \mathfrak{V}_{\mathfrak{I}}(B) = \mathsf{t} \\ & \text{iff} & \text{for all interpretations } \mathfrak{I}, \mathfrak{V}_{\mathfrak{I}}(A \to B) = \mathsf{t} \\ & \text{iff} & \models A \to B \end{array}$$

The first line follows just by the definition of what the notation $A \models B$ means. The right-hand side (RHS) of the first line is true iff the RHS of the second line is true: that's obvious. (The step is like that between "There are no students not in class" to "All students are in class".) The third line RHS matches the second line RHS because of how valuations apply to conditionals: see the answer to **6.1**(2), above. The fourth line RHS must have the same truth value as the third line RHS just by definition of '$\models$'.

8. *The argument form corresponding to the argument 'It is not the case that if Paris is the capital of Belgium, then Paris is the capital of Belgium. Therefore, Paris is the capital of France' is valid.*

True. Let's see why. The argument is:

It is not the case that if Paris is the capital of Belgium,
then Paris is the capital of Belgium,
$\therefore$
Paris is the capital of France.

(Note that this has just the one premise.) The argument form is $\neg(p \to p) \therefore q$. This form is valid if every interpretation which makes all the premises true (just the one), also makes the conclusion true. Which interpretations make the premise true? Look at **6.1**(2). That tells us that for any $\mathfrak{I}$:

$$\begin{array}{rll} \mathfrak{V}_{\mathfrak{I}}(\neg(p \to p)) = \mathsf{t} & \text{iff} & \mathfrak{V}_{\mathfrak{I}}(p \to p) = \mathsf{f} \\ & \text{iff} & \text{it's not the case that: } \mathfrak{V}_{\mathfrak{I}}(p) = \mathsf{f} \text{ or } \mathfrak{V}_{\mathfrak{I}}(p) = \mathsf{t} \\ & \text{iff} & \text{it's not the case that: } \mathfrak{I}(p) = \mathsf{f} \text{ or } \mathfrak{I}(p) = \mathsf{t} \\ & \text{iff} & \mathfrak{I}(p) = \mathsf{t} \text{ and } \mathfrak{I}(p) = \mathsf{f} \end{array}$$

But for no interpretation $\mathfrak{I}$ can it be that the final RHS holds: a single interpretation can't make p both true and false. This means that there is no interpretation which makes the premise true (the premise is a contradiction). So every interpretation which makes the premise true also makes the conclusion true.

We can show the same thing using a truth table, as follows.

p	q	$\neg$	$(p \to p)$	q
t	t	**f**	t	**t**
t	f	**f**	t	**f**
f	t	**f**	t	**t**
f	f	**f**	t	**f**

We must check whether every row which makes the premises true, also makes the conclusion true. But there is no row which makes the premises true, so the check is passed: the argument form is valid (and so, too, is the argument).