PHI-1A02 PHILOSOPHICAL SKILLS
Spring 2010
The Course Handbook

Thomas Forster

January 5, 2010

Contents

0.1 Stuff to fit in

"If you want to eat, there's a lasagne in the oven"

If you are of the kind of literal-minded bent that wants to reply: "Well, it seems that there is lasagne in the oven whether i'm hungry or not!" then you will find logic easy. You don't have to be perverse or autistic to be able to do this: you just have to be self-conscious about your use of language: to not only be able to use language but be able to observe your use of it.

How to use these notes

These notes are the notes from which I shall be giving the lectures. My intention in making them available to you is to make it easier for you to make your own notes of the lectures. Reading them is not an adequate substitute for coming to the lectures and the seminars. In any case they are not in a final state.

Passages in small type are not going to be lectured and may be omitted. You may wish to come back and read them later, after you have mastered the mainstream material.

If you do all the exercises and apply yourself to this course you will learn something about the workings of your own mind.

I will be available to students in my office: 364—next to the departmental office—on mondays after the lecture, on thursday afternoons and on fridays. I shall post exact times on my door, but you are welcome to email me: `tf@dpmms.cam.ac.uk` to make an appointment.

Chapter 1

Introduction to Logic

1.1 What is Logic?

1.1.1 Exercises for the first week: "Sheet 0"

Don't look down on puzzles:

> A logical theory may be tested by its capacity for dealing with puzzles, and it is a wholesome plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.
>
> Bertrand Russell

Exercise 1 *This first bundle is just for you to sharpen your wits.*

1. *Brothers and sisters have I none*
   This man's father is my father's son

   Whom is the speaker referring to?

2. *You are told that every card that you are about to see has a number on one side and a letter on the other. You are then shown four cards lying flat, and on the uppermost faces you see*

   $$E\ K\ 4\ 7$$

   It is alleged that any card with a vowel on one side has an even number on the other. Which of these cards do you have to turn over to check this allegation?

3. *A bag contains a certain number of black balls and a certain number of white balls. (The exact number doesn't matter). You repeatedly do the following. Put your hand in the bag and remove two balls at random: if*

they are both white, you put one of them back and discard the other; if one is black and the other is white, you put the black ball back in the bag and discard the white ball; if they are both black, you discard them both and put into the bag a random number of white balls from an inexhaustible supply that just happens to be handy.

What happens in the long run?

4. *Do the sudoku on page 48.*

5. *Hilary and Jocelyn are married. One evening they invite Alex and Chris (also married) to dinner, and there is a certain amount of handshaking, tho' naturally nobody shakes hands with themselves or their spouse. Later, Jocelyn asks the other three how many hands they have shaken and gets three different answers.*

   How many hands has Hilary shaken? How many hands has Jocelyn shaken?

6. *The next day Hilary and Jocelyn invite Chris and Alex again. This time they also invite Nicki and Kim. Again Jocelyn asks everyone how many hands they have shaken and again they all give different answers. How many hands has Hilary shaken this time? How many has Jocelyn shaken?*

These two are slightly more open-ended.

1. *You are given a large number of lengths of fuse. All you know about each length of fuse is that it will burn for precisely one minute. (They're all very uneven: in each length some bits burn faster than others, so you don't know that half the length will burn in half a minute or anything like that). The challenge is to use the burnings of these lengths of fuse to measure time intervals. You can obviously measure one minute, two minutes, three minutes and so on by lighting each fuse from the end of the one that's about to go out. What other lengths can you measure?*

2. *A Cretan says "Everything I say is false". What can you infer?*

Those exercises might take you a little while, but they are entirely do-able even before you have done any logic. Discuss them with your friends. You might want to devote your first seminar discussion to them. Don't give up on them: persist until you crack them!

If you disposed of all those with no sweat try this one:

You and I are going to play a game. There is an infinite line of beads stretching out in both directions. Each bead has a bead immediately to the left of it and another immediately to the right. A **round** *of the game is a move of mine followed by a move of yours. I move first, and my move is always to point at a bead. All the beads look the same: they are not numbered or anything like that. I may point to any bead I have not already indicated. You then have to give the bead a label, which is one of the letters* `a`–`z`*. The only restriction on your* *moves is that whenever you are called upon to put a label on the neighbour of a bead that already has a label, the new label must be the appropriate neighbour of the bead already labelled, respecting alphabetical order: the predecessor if the new bead is to the left of the old bead, and the successor if the new bead is to the right. For example, suppose you have labelled a bead with* '`p`'*; then if I point at the bead immediately to the right of it you have to label that bead* '`q`'*; were I to point to the bead immediately to the left of it you would have to label it* '`o`'*. If you have labelled a bead* '`z`' *then you would be in terminal trouble were I to point at the bead immediately to the right of it; if you have labelled a bead* '`a`' *then you would be in terminal trouble if I then point at the bead immediately to the left of it. We decide in advance how many rounds we are going to play. I win if you ever violate the condition on alphabetic ordering of labels. You win if you don't lose.*

Clearly you are going to win the one-round version, and it's easy for you to win the two-round version. The game is going to last for five rounds. How do you plan your play? How do you feel about playing six rounds?

1.2 Start Here!

If you start doing Philosophy it's beco's you want to understand. If you want to understand then you certainly want to reason properly the better to stay on the Road to Understanding. This course is going to concentrate on the task of helping you to reason properly. It is, I suppose, not completely obvious that we don't really have a free choice about how you should reason if you want to reason properly: nevertheless there is an objective basis to it, and in this course we will master a large slab of that objective basis.

There is an important contrast with **Rhetoric** here. With rhetoric anything goes that works. Power transactions blah. With reason too, I suppose anything goes that works, but what do we mean by 'works'? What are we trying to do when we reason? The stuff of reasoning is argument and an argument is something that leads us from premisses to conclusions. (An argument, as the Blessed Python said, isn't just contradiction: an argument is a connected series of statements intended to establish a proposition.[1]

Good reasoning will give us true conclusions from true premisses. That is the absolutely minimal requirement!

We are trying to extract new truths from old. And we want to do it reliably. In real life we don't usually expect 100% reliability because Real Life is lived in an Imperfect World. However for the moment we will restrict ourselves to trying to understand reasoning that *is* 100% reliable. Altho' this is only a start on the problem, it is at least a start. The remaining part of the project—trying to classify reasoning that is usually-pretty-reliable or that gives us plausible conclusions from plausible premisses—turns out to be not a project-to-understand-reasoning but actually the same old global project-to-understand-how-the-world-works...

[1] `http://www.mindspring.com/ mfpatton/sketch.htm`

George and Daniel are identical twins;
George smokes and Daniel doesn't.
Therefore George will die before Daniel.

The fact that this is a pretty good inference isn't a fact about *reasoning*; it's a fact about the way the world works. Contrast this with

"It is monday and it is raining; therefore it is monday"

You don't need to know anything about how the world works to know that that is a good inference—a 100% good inference in fact! This illustrates how much easier it is to grasp 100%-reliable inference than moderately-reliable inference.

normative *vs* descriptive

The study of reasoning is nowadays generally known as 'Logic'. Like any study it has a normative wing and a descriptive wing. Modern logic is put to good descriptive use in **Artificial Intelligence** where at least part of the time we are trying to write computer programs that will emulate human ways of thinking. A study with a title like 'Feminist Logic' alluded to below would be part of a descriptive use of Logic. We might get onto that later—next year perhaps—but on the whole the descriptive uses of logic are not nowadays considered part of Philosophy and for the moment we are going to concentrate on the normative rôle of Logic, and it is in its normative rôle that Logic tells us how to reason securely in a truth-preserving way. Interestingly all of that was sorted out in a period of about 50 years ending about 70 years ago. (1880-1930). It's all done and dusted. Logic provides almost the only area of Philosophy where there are brute facts to be learned and tangible skills to be acquired. And it's only a part of Logic that is like that. That's the part you are going to have dinned into you by me.

Digression on Rhetoric:

> Logic is (or at least starts as) the study of *argument* and it is **agent-invariant**. An argument is a good argument or a bad argument irrespective of who is using it: Man or Woman, Black, White, Gay, Asian, Transgendered. . . . Out in the real world there are subtle rules about who is and is not allowed to use what argument—particularly in politics. Those rules are not part of Logic; they are part of *Rhetoric*: the study of how to use words to influence people. That's not to say that they aren't interesting or important—they are. Logicians are often very interested in Rhetoric—I certainly am—but considerations of what kind of argument can be used by whom is no part of our study here. For example "feminist logic" is a misnomer: whether or not a form of reasoning is truth-preserving does not depend on how many X-chromosomes are possessed by the people who use it. People who use the term are probably thinking that it would be a good thing to have a feminist take on rhetoric (agreed!) or that it might be a good idea to study how women reason (ditto).
>
> Even if your primary interest is in rhetoric (and it may be, since we all have to be interested in rhetoric and we don't all have to study logic) logic is an important fragment of rhetoric that can be studied in isolation and as part of a preparation for a fuller study of rhetoric.

Reasoning is the process of inferring statements from other statements. What is a statement? I can give a sort of *contrastive*[2] explanation of what a statement is by contrasting **statements** with **commands** or **questions**. A statement is something that has a **truth-value**, namely `true` or `false`. (We often use the word 'proposition' in philosophy-speak. This is an unfortunate word, because of the connotations of 'proposal' and *embarking on a course of action* but we are stuck with it. This use of the word is something to do with the way in which the tradition has read Euclid's geometry. The *propositions* in Euclid are actually something to do with constructions.) Or performatives

You might find the idea of **evaluation** useful. Evaluation is what you do—in a context—to a statement, or to a question, or a command. In any context a command evaluates to an action; a question evaluates to an answer; a statement evaluates to a truth-value (i.e., to `true` or `false`). That doesn't really give you a definition of what any of these expressions 'context', 'evaluate', 'statement', 'question' etc actually *mean* (that would be too much to ask at this stage, tho' we do later take the concept of *evaluation* seriously) but it tells you something about how they fit together, and that might be helpful.

We are going to start off by analysing the kind of reasoning we can do with some simple gadgets for combining statements—such as 'and' and 'or' and 'not'.

> We are not going to attempt to capture **Conversational Implicature**. "A car was parked near a house; suddenly it moved". You know it's the car that moved, not a house. Also if someone says p' rather than p where p implies p' and p' takes longer to say, you take it that they mean p'-and-not-p. But that is inferred by non-logical means. (See section 1.6.5 on *semantic optimisation*). Logic is not equipped to handle these subtleties. These are known challenges for Artificial Intelligence people (their keyword for it is 'the frame problem') and for people who do natural language processing. (their keyword is 'pragmatics').

From 'It is monday and it is raining' we can infer 'It is raining'. This is a good inference. It's good purely in virtue of the meaning of the word 'and'. Any inference from a compound statement of the form: from 'A and B' infer 'A' is good—in the sense that it is truth-preserving. Every argument has **premisses** and a **conclusion**.

We write premisses above the line, and the conclusion below the line, thus

$$\frac{\textbf{premisses}}{\textbf{conclusion}}$$

$$\frac{\text{It is monday and it is raining}}{\text{It is raining}}$$

$$\frac{\text{It is tuesday and the sun is shining}}{\text{The sun is shining}}$$

[2] A *contrastive* explanation of something is an explanation given by contrasting it with something that it isn't, in the hope that the listener will put two and two together and get the right idea!

There are other similiarly simple inferences around. From 'It is raining' we can infer 'It is monday or it is raining'.

$$\frac{\text{It is raining}}{\text{It is monday or it is raining}}$$

Not very useful, one might think, since the conclusion seems to contain less information than the premisses did but for what it's worth it definitely is a truth-preserving inference: if the stuff above the line is true then sure as eggs is eggs the stuff below the line is true too! And the inference is truth-preserving in virtue of the meaning of the word 'or'.[3]

Introduce some symbols here:

$\wedge$, `&`	which both mean 'and'
$\vee$	which means 'or'
$\neg$	which means 'not',
$\rightarrow$	which means if-then: $P \rightarrow Q$ is "if P then Q".

Connective

week 2?

These things are called **connectives**. (N.B, '$\neg$' is a connective even though it doesn't connect two things: it is a **unary** connective.)

Exercise 2 *Let P abbreviate "I bought a lottery ticket" and Q "I won the jackpot". To what natural English sentences do the following formulæ correspond?*

1. $\neg P$;
2. $P \vee Q$;
3. $P \wedge Q$;
4. $P \rightarrow Q$;
5. $\neg P \rightarrow \neg Q$;
6. $\neg P \vee (P \wedge Q)$.

1.2.1 Truth-functional connectives

Truth-functional

Now we encounter a very important idea: the idea of a **truth-functional** connective. $\wedge$, $\vee$ and $\neg$ are truth-functional. By saying that '$\wedge$' is truth-functional we mean that if we want to know the truth-value of $A \wedge B$ it suffices for us to know the truth values of A and of B; similarly if we want to know the truth-value of $A \vee B$ it suffices for us to know the truth values of A and of B. Similarly if we want to know the truth-value of $\neg A$ it suffices for us to know the truth value of A.

[3] If you doubt this inference read section 1.6.6.

1.2.2 Truth Tables

Take $\wedge$ for example. If I want to know the truth-value of $A \wedge B$ it suffices for me to know the truth values of A and of B. Since $\wedge$ has only two inputs and each input must be true or false and it is only the truth-value of the inputs that matters then in some sense there are only 4 cases (contexts, situations, whatever you want to call them) to consider, and we can represent them in what is called a **truth table** where we write 'F' for 'false' and 'T' for 'true' to save space. Truth Table

A	$\wedge$	B
F	F	F
F	F	T
T	F	F
T	T	T

...sometimes written ...

A	$\wedge$	B
0	0	0
0	0	1
1	0	0
1	1	1

Both the T/F notation and the 1/0 notation are in common use, and you should expect to see them both and be able to cope with both. (Nobody writes out '`false`' and '`true`' in full—it takes up too much space.) I tend to use 0/1 because 'T's and 'F's tend to look the same in the crowd—such as you find in a truth-table.

There are truth-tables for other connectives, $\vee$ and $\neg$:

A	$\vee$	B
0	0	0
0	1	1
1	1	0
1	1	1

A	NOR	B
0	1	0
0	0	1
1	0	0
1	0	1

A	XOR	B
0	1	0
0	0	1
1	0	0
1	1	1

A	$NAND$	B
0	0	0
0	1	1
1	1	0
1	1	1

$\neg$	A
1	0
0	1

The connectives NAND and NOR and XOR are sometimes used, but altho' you will see them in electronics you will never see them in the philosophical literature. In Philosophy we tend to make do with $\wedge$, $\vee$ and $\neg$ and one more, an arrow $\rightarrow$ for if-then which i shall explain soon.

You can also draw Venn diagrams if you like!

I want to flag here a hugely important policy decision. **The only connectives were are going to study are those connectives which can be captured by truth-tables.** We are emphatically *not* going to study connectives that try to capture squishy things like meaning and causation. This might sound excessively restrictive, and suitable only for people who are insensitive to the finer and more delicate things in life, but it is actually very fruitful and is much more sensible than it might sound at first.

One reason why it is sensible is that out there in the real world the kind of reasoning you are interested in exploiting is *reasoning that preserves truth*. Nothing else comes anywhere near that in ultimate importance. Like any other poor metazoan trying to make a living, you need to not get trodden on by dinosaurs, not miss out on desirable food objects or on opportunities to reproduce. It is true you might choose to eschew the odd food object or potential mate from time to time, but you at least want your choice to be informed. Sometimes your detection of a dinosaur or a food morsel or a potential mate will depend on *inference* from lower-level data or on other information supplied by context. If that thing out there really is a dinosaur that might tread on you you need to know that; ditto a food object or potential mate. You will want modes of reasoning to be available to you that will deliver any truth that can be squeezed out of the information available to you. If you have a mode of reasoning available to you that reliably gives true conclusions from correct information then you cannot afford to turn your nose up at it merely on the grounds that it doesn't preserve meaning or that your colour therapist doesn't like it. Your competitor who is satisfied merely with truth-preservation will avoid the dinosaur and get the forbidden fruit and you won't. Truth-preserving inference is what it's all about!

That's not to say that we will never want to study modes of inference that preserve more than truth, but it does mean that it is very sensible to start with truth-preservation!

1.3 The Language of Propositional Logic

Let's mow get straight what the gadgetry is that we are going to use. I shall use lower case Roman letters for propositional letters ('atomic' formulæ) and upper case letters for compound ('molecular') formulæ. There are several different traditions that use this gadgetry of formal Logic, and they have different habits. Philosophers tend to use lower case Roman letters ('p', 'q' *etc.*); Other communities use upper case Roman letters or even Greek letters. We will stick to Roman letters.

We are going to have two symbols '$\top$' and '$\bot$' which are propositional letters of a special kind: '$\bot$' always takes the value `false` and '$\top$' always takes the value `true`.

We have symbols '$\wedge$', '$\vee$' and '$\rightarrow$' which we can use to build up compound expressions.

We need the notion of the **principal connective** of a formula: And of **immediate subformula**.

Some illustrations needed here

Definition 1 *A formula whose principal connective is a*

$\wedge$ *is a* **conjunction** *and its immediate subformulæ are its* **conjuncts**
$\vee$ *is a* **disjunction** *and its immediate subformulæ are its* **disjuncts**
$\rightarrow$ *is a* **conditional** *and its immediate subformulæ are its* **antecedent** *and its* **consequent**
$\longleftrightarrow$ *is a* **biconditional**

Thus
$A \wedge B$ is a conjunction, and A and B are its **conjuncts**;
$A \vee B$ is a disjunction, and A and B are its **disjuncts**;
$A \rightarrow B$ is a conditional, where A is the **antecedent** and B is the **consequent.**

for second week

Exercise 3 *What are the principal connectives and the immediate subformulæ of the formulæ below?*

1. $P \vee \neg P$
2. $\neg(A \vee \neg(A \wedge B))$
3. $(A \vee B) \wedge (\neg A \vee \neg B)$
4. $A \vee (B \wedge (C \vee D))$;
5. $\neg(P \vee Q)$
6. $P \rightarrow (Q \rightarrow P)$
7. $(P \longleftrightarrow Q) \wedge (P \vee Q)$
8. $(P \longleftrightarrow Q) \longleftrightarrow (Q \longleftrightarrow P)$
9. $P \rightarrow (P \vee Q)$

10. $P \to (Q \vee P)$

11. $P \to \bot$

12. $(P \to Q) \vee (Q \to P)$

13. $(P \to Q) \to \neg(Q \to P)$

14. $P \to (P \wedge Q)$

15. $A \to [(A \to C) \to ((B \to C) \to C)]$

16. $B \to [(A \to C) \to ((B \to C) \to C)]$

17. $(A \vee B) \to [(A \to C) \to ((B \to C) \to C)]$.

Truth-tables are a convenient way of representing/tabulating all the valuations a formula can have. If a formula has n propositional letters in it, there are precisely 2^n ways of evaluating each propositional letter in it to `true` or to `false`. This is why the truth-table for $\wedge$ has four rows, and the truth-table for $A \vee (B \wedge C)$ has eight.

for third week

Exercise 4 *Can you see why it's* 2^n*?*

(each time you add a new propositional letter you double the number of possible combinations)

Usually we can get by with propositional connectives that have only two arguments (or, in the case of $\neg$, only one!) but sometimes people have been known to consider connectives with three arguments, for example

`if`	p	`then`	q	`else`	r
1	1		1		1
$\vdots$	1		0		1
$\vdots$	1		1		0
$\vdots$	1		0		0
$\vdots$	$\vdots$		$\vdots$		$\vdots$
$\vdots$	$\vdots$		$\vdots$		$\vdots$

for third week

Exercise 5 *You might like to write out the rest of the truth-table for this connective, putting the truth-value of the compound formula under the '*`if`*' as I have done. What I have written in the top row means that* `if` p `then` q `else` r *is true when* p*,* q *and* r *are all true. What do we want to say about the truth of* `if` p `then` q `else` r *when* p *and* r *are true but* q *is false?*

(How many other rows will the truth-table have?)

might want to make a point about lazy evaluation here

1.3.1 Truth-tables for Compound Expressions

How to fill in truth-tables for compound expressions: a couple of worked examples

$\neg(A \vee B)$:

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

We can fill in the column under the '$\vee$' ...

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

and then the column under the '$\neg$' ...

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

$A \wedge (B \vee C)$: The truth table for $A \wedge (B \vee C)$ will have 8 rows because there are 8 possibilities. The first thing we do is put all possible combinations of 0s and 1s under the A, B and C thus:

A	$\vee$	$(B$	$\wedge$	$C)$
0		0		0
0		0		1
0		1		0
0		1		1
1		0		0
1		0		1
1		1		0
1		1		1

Then we can put in a column of 0s and 1s under the $B \wedge C$ thus:

A	$\vee$	$(B$	$\wedge$	$C)$
0		0	0	0
0		0	0	1
0		1	0	0
0		1	1	1
1		0	0	0
1		0	0	1
1		1	0	0
1		1	1	1

Then we know what to put under the '$\vee$'

A	$\vee$	$(B$	$\wedge$	$C)$
0	0	0	0	0
0	0	0	0	1
0	0	1	0	0
0	1	1	1	1
1	1	0	0	0
1	1	0	0	1
1	1	1	0	0
1	1	1	1	1

...by combining the numbers under the 'A' with the numbers under the '$B \wedge C$': for example, the first row has a '0' under the 'A' and also a '0' under the '$B \wedge C$' and $0 \vee 0$ is 0.

There will be more examples life in the lectures: they take up more space than is available in a written text.

These worked exercise i have just gone through illustrate how the truth-value that a complex formula takes in a row of a truth-table can be calculated from the truth-value taken by its subformulæ in that row. This phenomenon has the grand word: **compositional**.

The bundle of rows of the truth-table exhaust all the possibilities that a truth-functional connective can see. Any truth-functional connective can be characterised by a truth-table.

1.3.2 Logical equivalence

Two complex formulæ with the same truth-table are said to be **logically equivalent**

for second ?third? week

Exercise 6 *In the following table*

(1)	$A \wedge A$	A
(2)	$A \vee A$	A
(3)	$\neg(A \vee B)$	$(\neg A) \wedge (\neg B)$
(4)	$A \vee B$	$\neg((\neg A) \wedge (\neg B))$
(5)	$\neg(A \wedge B)$	$(\neg A) \vee (\neg B)$

(6) $A \wedge B$	$\neg((\neg A) \vee (\neg B))$
(7) $A \vee \bot$	A
(8) $A \wedge \bot$	$\bot$
(9) $A \vee \top$	$\top$
(10) $A \wedge \top$	A
(11) $\bot \vee A$	A
(12) $A \vee (B \vee C)$	$(A \vee B) \vee C$
(13) $A \wedge (A \vee B)$	A
(14) $A \vee (A \wedge B)$	A
(15) $A \vee (B \vee C)$	$(A \vee B) \vee C$
(16) $A \wedge (B \wedge C)$	$(A \wedge B) \wedge C$
(17) $A \vee (B \wedge C)$	$(A \vee B) \wedge (A \vee C)$
(18) $(A \wedge B) \vee ((\neg A) \wedge C)$	$(A \to B) \wedge ((\neg A) \to C)$

we find that, in each line, the two formulæ are logically equivalent.

Write out truth-tables to prove it.

Items (3–6) are sometimes called the **de Morgan laws**. We will see more of them later.

Say something about leaving out brackets.

EXERCISE 7 *Write out truth-tables for the firt five formulæ in exercise 3.*

for third week

EXERCISE 8 *Identify the principal connective of each formula below. In each pair of formulæ, say whether they are (i) logically equivalent or are (ii) negations of each other or (iii) neither.*

$(\neg A \wedge \neg B)$;	$\neg(A \vee B)$
$(\neg A \vee \neg B)$;	$\neg(A \vee B)$
$(\neg A \wedge \neg B)$;	$\neg(A \wedge B)$
$(\neg A \vee \neg B)$;	$\neg(A \wedge B)$

DEFINITION 2

Associativity: *We have seen that $(A \vee B) \vee C$ is logically equivalent to $A \vee (B \vee C)$. Also we can see that $(A \wedge B) \wedge C$ is logically equivalent to $A \wedge (B \wedge C)$; we say that $\vee$ and $\wedge$ are* **associative**.

Idempotence. *$A \wedge A$ is logically equivalent to A; $A \vee A$ is equivalent to A: we say $\wedge$ and $\vee$ are* **idempotent**.

Distributivity. *We capture the fact that $A \vee (B \wedge C)$ and $(A \vee B) \wedge (A \vee C)$ are logically equivalent by saying that $\vee$* **distributes over** *$\wedge$.*

The alert reader will have noticed that i have been silent on the subject of if-then while discussing truth-tables. The time has come for us to do somethign about this omission.

We write '$\to$' for the connective that we use to formalise inference. It will obey the rule "from P and $P \to Q$ infer Q":

$$\frac{P \qquad P \to Q}{Q}$$

which we now know is called **modus ponens**. Q is the **conclusion** or the **consequent**; P is the **minor premiss** and $P \to Q$ is the **major premiss**. I know I haven't given you a truth-table for '$\to$' yet. All in good time! There is some other gadgetry we have to get out of the way first.

1.3.3 Non truth functional connectives

Causation and necessity are not truth-functional. Consider

1. Labour will lose the next election *because* unemployment has been rising throughout 2009;
2. *Necessarily* Man is a rational animal.

The truth-value of (1) depends on more than the truth values of "unemployment will be rising throughout 2009" and "Labour will lose the next election"; the truth-value of (2) depends on more than the truth value of "Man is a rational animal". Necessity and causation are not truth-functional and accordingly cannot be captured by truth-tables.

Counterfactuals

Can we say anything intelligent about the difference? Digression here to explain intensional-extensional distinction:

1.4 Intension and Extension

The intension-extension distinction is a device of mediæval philosophy which was re-imported into the analytic tradition by Church [5] and Carnap [3] in the middle of the last century, probably under the influence of Franz Brentano.

The standard illustration in the philosophical literature concerns the two properties of being *human* and being a *featherless biped*—a creature with two legs and no feathers. There is a perfectly good sense in which these concepts are the same (one can tell that this illustration dates from before the time when the West had encountered Australia with its kangaroos! It actually goes back to Aristotle), but there is another perfectly good sense in which they are different. We name these two senses by saying that 'human' and 'featherless biped' are the same **property in extension** but different **properties in intension**.

Intensions are generally finer than extensions. Lots of different properties-in-intension correspond to the property-in-extension that is the class of human. Not just Featherless Bipeds and Rational Animals but Naked Apes. Possessors of language? Tool makers?

The intension–extension distinction is not a formal technical device, and it does not need to be conceived or used rigorously, but as a piece of logical slang it is very useful.

One reason why it is useful is captured by an *aperçu* of Quine's ([9] p 23): "No entity without identity". What this *obiter dictum* means is that if you wish to believe in the existence of a suite of entities—numbers, ghosts, properties-in-intension or whatever it may be—then you must be able to tell when two numbers (ghosts, propertiess-in-intension) are the same number (ghost, etc.) and when they are different numbers (ghosts, etc). If we are to reason reliably about entities from a particular suite we need **identity criteria** for them.

Clouds give us quite a good illustration of this. There are two concepts out there: cloud as *stuff* and clouds as *things*. There's not much mystery about clouds-as-stuff: it's lots of water droplets of a certain size (the right size to scatter visible light) suspended in air. The concept of cloud-as-object is not well-defined at all. "This is a cloud"; "That patch is two clouds not one". You will notice that the weather people never tell us how many clouds there will be in the sky tomorrow, but they might tell us what percentage of the sky they expect to be covered in cloud. That's cloud-as-stuff of course. We don't have good identity criteria for when two clouds are the same cloud: we don't know how to *individuate* them.

What has this last point got to do with the intension/extension distinction? The point is that we have a much better grasp of identity criteria for extensions than for intensions. Propositions are intensions, and the corresponding extensions are truth-values: there are two of them, the `true` and the `false`.

You might think there are more. Wouldn't it be a sensible precaution to have also a `don't-know` up our sleeve as a third truth-value? [4]

The trouble is that although 'don't-know' is third possibility, it's not a third truth-value for the proposition: it's a third possible state of your relation to that proposition: a relation of not-knowing. What is it you don't know? You don't know which of the two(!) mutually-exclusive and jointly-exhaustive possibilities for that proposition (truth vs falsity) holds.

There are various things that might tempt you into thinking that the third possibility is a third truth-value. If you don't know the truth-value of the proposition you are evaluating it may be merely that you are unsure which proposition it is that you are evaluating. To argue for a third truth-value you have to be sure that none of the plausible cases can be accounted for in this way. There are tricks you can play with three-valued truth-tables—and we shall see some of them later—but the extra truth-values generally don't seem to have any real meaning.

The difference between the `true` and the `false` is uncontroversial but it's not clear when two propositions are the same proposition. To say—as some people do—that two propositions are the same proposition iff they are true in the same possible world appeals to some pretty dodgy machinery and it's not

[4] And why stop there? On at least one reading of a text (The Heart Sutra) in the Classical Buddhist literature there are no fewer than *five* truth-values: `true` and `false` as usual of course, but also `both`, `neither` and finally `none-of-the-above`.

clear that it explains anything. What is a possible world after all? (Properties, too, are intensions: the corresponding extensions are sets, and it's much easier to see when two sets are the same or different than it is to see when two properties are the same or different. We are not going to do much set theory here and the only reason why I am bringing it in at this stage is to illustrate the intension/extension distinction.)

The fact that it is not always clear when two propositions-(in-intension) are the same proposition sabotages all attempts to codify reasoning with propositions-in-intension. If it is not clear to me whether or not p implies q this might be because in my situation there are two very similar salient propositions, p and p', one of which implies q and the other doesn't—and I am equivocating unconsiously between them. If we had satisfactory identity criteria for propositions then fallacy of equivocation would be less of a danger, but we don't! So what we want to do in logic—at least to start with—is study relations beween *propositions-in-extension*. This sounds as if all we are going to do is study the relationship between the `true` and the `false`—which would make for a rather short project. However if we think of propositions-in-extension as *things-that-have-been-evaluated-to-*`true`*-or-to-*`false` then we have a sensible programme. We can combine propositions with connectives, $\wedge$, $\vee$, $\neg$ and so on, and the things that evaluate them to `true` and `false` are **valuations**: a valuation is a row in a truth-table.

DEFINITION 3 *A valuation is a function that sends each propositional letter to a truth-value.*

As remarked earlier, the connectives we want are **truth-functional**.

There is a long tradition of trying to obtain an understanding of intensions by tunneling towards them through the corresponding extensions. Hume's heroic attempt to understand causation (between event-types) by means of constant conjunction between the corresponding event tokens is definitely in this spirit.

There is a certain amount of coercion going on in the endeavour to think only in terms of extensional (truth-functional) connectives: we have to make do with extensional mimics of the intensional connectives that are the first things that come to mind. The best extensional approximation to "p unless q" seems to be $p \vee q$. But even this doesn't feel quite right: disjunction is symmetrical: $p \vee q$ has the same truth-value as $q \vee p$, but 'unless' doesn't feel symmetrical. Similarly 'and' and 'but' are different intensionally but both are best approximated by '$\wedge$'. Notice that Strawson's example: 'Mary got married and had a baby' $\neq$ 'Mary had a baby and got married' doesn't show that 'and' is intensional, but rather that our word 'and' is used in two distinct ways, logical conjunction and temporal succession.

Statements, too, have intensions and extensions. The intension of a statement is its meaning. Mediæval writers tended to think that the meaning of a piece of language was to be found in the intention of the speaker, and so the word 'intention' (or rather its latin forbears) came to mean *content* or *meaning*. 'Extension' seems to be a back-formation from 'intention': the extension of a statement is its truth-value, or—better perhaps—a tabulation of its truth-value in contexts: its *evaluation behaviour*.

Connectives that are truth-functional are extensional. The others (such as “implies” “because”) are intensional. Everything we study is going to be truth-functional. This is a policy decision taken to keep things simple in the short term. We may get round to studying non-truth-functional (“intensional”) systems of reasoning later, but certainly not in first year.

I talked about intensions and extensions not just because they are generally important but because the intension-extension distinction is the way to cope with the difficulties we will have with `implies`. The connectives `and` and `or` and `not` are truth-functional, but `implies` and `because` and `necessarily` are not.

1.4.1 If–then

Clearly we are going to have to find a way of talking about implication, or something like it. Given that we are resolved to have a purely truth-functional logic we will need a truth-functional connective that behaves like implies. (‘Necessarily’ is a lost cause but we will try to salvage if-then). Any candidate must at least obey *modus ponens*

$$\frac{A \qquad A \to B}{B}$$

A conditional is a binary connective that is an attempt to formalise a relation of implication. The word ‘conditional’ is also used (in a second sense) to denote a formula whose principal connective is a conditional (in the first sense). Thus we say both that ‘$\to$’ is a conditional and that $A \to B$’ is a conditional. The conditional $\neg B \to \neg A$ is the **contrapositive** of the conditional $A \to B$, and the **converse** is $B \to A$. (cf., converse of a relation). A formula like $A \longleftrightarrow B$ is a **biconditional**.

The two components glued together by the connective are the **antecedent** (from which one infers something) and the **consequent** (which is the something that one infers). In *modus ponens* one *affirms* the antecedent and *infers* the consequent, thus:

$$\frac{A \to B \qquad A}{B}$$

Modus tollens is the rule:

$$\frac{A \to B \qquad \neg B}{\neg A}$$

where one affirms the conditional and denies the consequent.

Affirming the consequent and inferring the antecedent:

$$\frac{A \to B \qquad B}{A}$$

is a **fallacy** (= defective inference). This is an important fallacy, for reasons that will emerge later. This particular fallacy is the **fallacy of affirming the consequent**.

In fact—because it is only truth-functional logic we are trying to capture—we will stipulate that $P \to Q$ will be equivalent to '$\neg(P \wedge \neg Q)$' or to '$\neg P \vee Q$'. $\to$ is the **material conditional**. $P \to Q$ evaluates to `true` unless P evaluates to `true` and Q evaluates to `false`.

So we have a conditional that is defined on extensions. So far so good. Reasonable people might expect that what one has to do next is solve the problem of what the correct notion of conditional is for intensions. We can make a start by saying that P implies Q if—for all valuations—what P evaluates to materially implies what Q evaluates to. This does not solve the problem of identifying the intensional conditional, but it gets us a surprisingly long way. However this is a very hard problem, since it involves thinking about the internal structure of intensions and nobody really has a clue about that. (This is connected to the fact that we do not really have robust criteria of identity for intensions, as mentioned earlier.) It has spawned a vast and inconclusive literature which beginners can safely ignore.

Once we've got it sorted out ...

A	$\to$	B
F	T	F
F	T	T
T	F	F
T	T	T

or, in 0/1 notation:

A	$\to$	B
0	1	0
0	1	1
1	0	0
1	1	1

It's sometimes written $\supset$ (particularly in the older philosophical literature) and sometimes with a double shaft: $\Rightarrow$.

Going for the material conditional means we don't have to worry ourselves sick about whether or not $A \to (B \to A)$ captures a correct principle of inference. If we take the arrow to be a material conditional then it is! (If the arrow is intensional then it is not at all clear that $A \to (B \to A)$ is a good principle of inference).

Exercise 9 *In the following table*

(1)	$\neg(A \wedge \neg B)$	$A \to B$
(2)	$A \to A$	$\top$
(3)	$A \to B$	$\neg A \vee B$

(4)	$A \to \bot$	$\neg A$
(5)	$\top \to A$	A
(6)	$\bot \to A$	$\top$
(7)	$A \to \top$	$\top$
(8)	$A \to B$	$\neg B \to \neg A$
(9)	$A \to \neg A$	$\neg A$
(10)	$\neg A \to A$	A

we find that, in each line, the two formulæ are logically equivalent. Write out truth-tables to prove it.

1.4.2 Logical Form and Valid Argument

Now we need the notion of **Logical Form** and **Valid Argument**. An argument is valid if it is truth-preserving in virtue of its form.

For example the following argument (from page 11) is truth-preserving because of its form.

$$\frac{\text{It is tuesday and the sun is shining}}{\text{The sun is shining}}$$

The point is that there is more going on in this case than the mere fact that the premisses are true and that the conclusion is also true. The point is that the argument is of a *shape* that guarantees that the conclusion will be true if the premisses are. The argument has the form

$$\frac{\text{A and B}}{\text{B}}$$

and all arguments of this form with true premisses have a true conclusion.

To express this concept snappily we will need a new bit of terminology.

1.4.3 The *Type-Token* Distinction

The terminology 'type-token' is due to the remarkable nineteenth century American philosopher Charles Sanders Peirce. (It really is 'e' before 'i' . . . Yes i know, but then we've always known that Americans can't spell.) The expression

$$((A \to B) \to A) \to A$$

is sometimes called *Peirce's Law*. Do not worry if you can't see what it means: it's quite opaque. But do try constructing a truth-table for it!

The two ideas of token and type are connected by the relation "is an instance of". Tokens are **instances** of types.

It's the distinction we reach for in situations like the following

- (i) "I wrote a *book* last year"
  (ii) "I bought two **books** today"

  In (ii) the two things I bought were physical objects, but the thing I wrote in (i) was an abstract entity. What I wrote was a *type*. The things I bought today with which I shall curl up tonight are *tokens*. This important distinction is missable because we typically use the same word for both the type and the token.

- A best seller is a book large numbers of whose *tokens* have been sold. There is a certain amount of puzzlement in copyright law about ownership of tokens of a work versus ownership of the type. James Hewitt owns the tokens of Diana's letters to him but not the type (the copyright)

- I read somewhere that " ...next to Mary Woollstonecroft was buried Shelley's heart, *wrapped in one of his poems*." To be a bit more precise, it was wrapped in a *token* of one of his poems.

- You have to write an essay of 5000 words. That is 5000 word *tokens*. On the other hand, there are 5000 words used in this course material that come from Latin. Those are word *types*.

- Grelling's paradox: a **heterological** word is one that is not true of itself.[5] 'long' is heterological: it is not a *long* word. 'English' is not heterological but *homological*, for it is an English word. Notice that it is word *types* not word *tokens* that are heterological (or homological!) It doesn't make any sense to ask whether or not 'italicised' is heterological. Only word *tokens* can be italicised!

- Genes try to maximise the number of tokens of themselves in circulation. We attribute the intention to the gene *type* because it is not the action of any *one* token that invites this mentalistic metaphor, but the action of them all together. However it is the number of *tokens* that the type appears to be trying to maximise.

- *First diner*:

  > "Isn't it a bit cheeky of them to put "vegetable<u>s</u> of the day" when there is nothing but carrots in the way of vegetables?"

  Second diner:

  > "Well, you did get more than one carrot so perhaps they're within their rights!"

The type-token distinction is important throughout Philosophy.

[5] This is a sleeper for next year's logic course—or rather a *non*-sleeper, since it is intended to deprive you of sleep: is the word 'heterological' heterological?

- People who do æsthetics have to be very careful about the difference between things and their representations—and related distinctions. I can't enjoy being unhappy, so how can I enjoy reading Thomas Hardy? There is an important difference between the fictional disasters that befall Jude the Obscure (to which we have a certain kind of relation) and the *actual* disasters that befall the *actual* Judes of this world—to which these fictional disasters allude—and to which we have (correctly) an entirely different reaction. The type/token/representation/etc. distinction is not just a plaything of logicians: it really matters.

- In Philosophy of Mind there are a variety of theories called *Identity Theories*: mental states are just physiological states of some kind, probably mostly states of the brain. But if one makes this identification one still has to decide whether a particular mental state *type*—thinking-about-an-odd-number-of-elephants, say—is to be identified with a particular *type* of physiologicial state? Is it is just that every time I think about an odd number of elephants (so I am exhibiting a token of the type of that mental state, then there is a token of physiological state I must be in—but the states might vary (be instances of different physiological state-types) from time to time? These two theories are *Type Identity* and *Token Identity*.

1.4.4 Copies

Buddhas

It is told that the Buddha could perform miracles. But—like Jesus—he felt they were vulgar and ostentatious, and they displeased him.

> A merchant in a city of India carves a piece of sandalwood into a bowl. He places it at the top of some bamboo stalks which are high and very slippery, and declares that he will give the bowl to anyone who can get it down. Some heretical teachers try, but in vain. They attempt to bribe the merchant to say they had succeeded. The merchant refuses, and a minor disciple of the Buddha arrives. (His name is not mentioned except in this connection). The disciple rises through the air, flies six times round the bowl, then picks it up and delivers it to the merchant. When the Buddha hears the story he expels the disciple from the order for his frivolity.

But that didn't stop him from performing them himself when forced into a corner. In [2] (from which the above paragraph is taken) J. L. Borges procedes to tell the following story, of a miracle of *courtesy*. The Buddha has to cross a desert at noon. The Gods, from their thirty-three heavans, each send him down a parasol. The Buddha does not want to slight any of the Gods, so he turns himself into thirty-three Buddhas. Each God sees a Buddha protected by a parasol he sent.[6]

[6] As is usual with Borges, one does not know whether he has a source for this story in the literature, or whether he made it up. And—again, as usual—it doesn't matter.

Apparently he did this routinely whenever he was visiting a city with several gates, at each of which there would be people waiting to greet him. He would make as many copies of himself as were needed for him to be able to appear at all the gates simultaneously—and thereby not disappoint anyone.

Minis

Q: How many elephants can you fit in a mini?
A: Four: two in the front and two in the back.

Q: How many giraffes can you fit in a mini?
A: None: it's full of elephants.

Q: How can you tell when there are elephants in the fridge?
A: Footprints in the butter.

Q: How can you tell when there are *two* elephants in the fridge?
A: You can hear them giggling when the light goes out.

Q: How can you tell when there are *three* elephants in the fridge?
A: You have difficulty closing the fridge door.

Q: How can you tell when there are *four* elephants in the fridge?
A: There's a mini parked outside.

1.5 Tautology and Validity

1.5.1 Valid Argument

Now that we are armed with the type-token distinction we can give a nice snappy definition of the important concept of **Valid Argument**.

Definition 4 *A* **valid** *argument (type) is one such that any argument of that form (any token of it) with true premisses has a true conclusion.*

And while we are about it, we'll give a definition of a related concept as a spin-off.

Definition 5 *A* **sound** *argument (token) is a token of a valid argument type all of whose premisses are true.*

The final example on page 29 is an example of a valid argument. It is a matter of debate whether or not it is *sound*!

Arguments that are valid are valid in virtue of their structure. That is what makes Logic possible!

The idea that the reliability of an argument relies at least in part on its shape or form is deeply embedded in everyday rhetoric. Hence the rhetorical device of the *tu quoque* and the rhetorical device of argument by analogy

Explain this

Need more illustrations before we ask them to do these exercises

Exercise 10 *Abbreviate "Jack arrives late for lunch" etc etc., to single letters, and use these abbreviations to formalise the arguments below. (To keep things simple you can ignore the tenses!)*

week?

1. *If Jill arrives late for lunch, she will be cross with Jack. Jack will arrive late. Therefore Jill will be cross with Jack.*

2. *If Jill arrives late for lunch, Jack will be cross with her. Jill will arrive late. Therefore Jill will be cross with Jack.*

3. *If Jill arrives late for lunch, Jack will be cross with her. Jack will arrive late. Therefore Jill will be cross with Jack.*

4. *If Jack arrives late for lunch, Jill will be cross with him. Jack will arrive late. Therefore Jill will be cross with Jack.*

5. *If George is guilty he'll be reluctant to answer questions; George is reluctant to answer questions. Therefore George is guilty.*

6. *If George is broke he won't be able to buy lunch; George is broke. Therefore George will not be able to buy lunch.*

7. *Assuming that the lectures are dull, if the text is not readable then Alfred will not pass.*

8. *If Logic is difficult Alfred will pass only if he concentrates.*

9. *If Alfred studies, then he receives good marks. If he does not study, then he enjoys college. If he does not receive good marks then he does not enjoy college. [Therefore Alfred receives good marks]*

10. *If Herbert can take the flat only if he divorces his wife then he should think twice. If Herbert keeps Fido, then he cannot take the flat. Herbert's wife insists on keeping Fido. If Herbert does not keep Fido then he will divorce his wife—at least if she insists on keeping Fido. [Therefore Herbert should think twice]*

11. *If Herbert grows rich, then he can take the flat. If he divorces his wife he will not receive his inheritance. Herbert will grow rich if he receives his inheritance. Herbert can take the flat only if he divorces his wife.*

12. *If God exists then He is omnipotent. If God exists then He is omniscient. If God exists then He is benevolent. If God can prevent evil then—if He knows that evil exists—then He is not benevolent if He does not prevent it. If God is omnipotent, then He can prevent evil. If God is omniscient then He knows that evil exists if it does indeed exist. Evil does not exist if God prevents it. Evil exists. [Therefore God does not exist]*

This last one is a bit of a mouthful! But it's made of lots of little parts. Do not panic!

(3) onwards are taken from [7]. Long out of print, but you can sometimes find second-hand copies. If you find one, buy it. I'll tell you later who Alfred is. (I don't know about Herbert: I am making enquiries).

The cat sat on the mat and the dog sat in front of the fire.

This will be abbreviated to

$$A \wedge B$$

where blah. It is true that A and B both have internal structure (subject-verb-object etc) but that internal structure is not of the kind that can be expressed using and/or/not.

(The concept of a valid argument is not the only thing that matters from the rhetorical point of view, from the point of view of transacting power relations: there are other things to worry about, but as far as we are concerned, arguments that are useful in power-transactions without being valid are not of much concern to us. Logic really has nothing to say about arguments in terms of the rights of the proponents of various sides to say what they say: it concerns itself only with what they say, not with their right to say it.)

In a valid argument the premisses **imply** the conclusion. We can **infer** the conclusion from the premisses. People often confuse these two words, and use 'infer' when they mean 'imply'. *You* mustn't! You are Higher Life Forms.

Then we can replace the propositions in the argument by letters. This throws away the content of the argument but preserves its structure. You no longer know which token you are looking at, but you do know the type.

Some expressions have in their truth-tables a row where the whole formula comes out false. '$A \vee B$' is an example; in the row where A and B are both false $A \vee B$ comes out false too. Such formulæ are said to be **falsifiable**.

Some expressions—'$A \vee \neg A$' is an example—come out true in all rows. Such an expression is said to be **tautology**.

DEFINITION 6

- *A Tautology is an expression which comes out true under all valuations (= in all rows of its truth table).*
- *A tautology is also said to be* **logically true**.
- *The negation of a tautology is said to be* **logically false**.
- *A formula that is not the negation of a tautology is said to be* **satisfiable**.

I sometimes find myself writing 'truth-table tautology' instead of mere 'tautology' because of the possibility of other uses of the word.

7

These two ideas, (i) of valid argument, and (ii) *tautology* are closely related, and you might get them confused. But it's easy:

Definition 7 *An argument*

$$\frac{P_1, P_2, \ldots P_n}{C}$$

is valid if and only if the conditional

$$(P_1 \wedge P_2 \wedge \ldots P_n) \to C$$

(whose antecedent is the conjunction of its premisses and whose consequent is its conclusion) is a tautology.

In order to be happy with the idea of a valid argument you really have to have the idea of there being **slots** or **blanks** in the argument which you can fill in. The two miniature arguments over here [it is monday and it is raining therefore it is monday, and: the cat sat on the mat and the dog in front of the fire therefore the cat sat on the mat] are two tokens of the one argument-type.

We will be going into immense detail later about what form the slots take, what they can look like and so on. You're not expected to get the whole picture yet, but i would like you to feel happy about the idea that these two arguments are tokens of the same argument-type.

which bits can you not rewrite? The LOGICAL VOCABULARY: the truth-functional part

Exercise 11 *Which of the arguments in exercise 10 are valid?*

1.5.2 $\bigwedge$ and $\bigvee$ *versus* $\wedge$ and $\vee$

The connectives $\wedge$ and $\vee$ are *associative* (it doesn't matter how you bracket $A \vee B \vee C$; we saw this on page 19) so we can omit brackets .... This looks like a simplification but it brings a complication. If we ask what the principal connective is of '$A \vee B \vee C$' we don't know which of the two '$\vee$'s to point to. We could write

$$\bigvee \{A, B, C\}$$

to make sure that there is only one '$\vee$' to point to. This motivates more complex notations like

$$\bigvee_{i \in I} A_i \tag{1.1}$$

[7] We use the word 'tautology' in popular parlance too—it's been borrowed from Logic and misused (surprise surprise). Once my ex (an EFL teacher) threatened to buy me a new pair of trousers. When i said that I would rather have the money instead she accused me of tautology (thinking of the repetition in 'rather' and 'instead'). She's wrong: it's not a tautology, the repetition makes it a *pleonasm*.

... since there it is obvious that the '$\bigvee$' is the principal connective. However this notation looks rather mathematical and could alarm some people so we would otherwise prefer to avoid it![8] We won't use it.

However we can't really avoid it entirely: we do need the notion of the disjunction of a set of formulæ (and the notion of the conjunction of a set of formulæ). We will return to those two ideas later. For the moment just take on board the idea that '$A \vee B \vee C$' is a disjunction, that its principal connective is '$\vee$' and its immediate subformulæ are 'A', 'B' and 'C'...

The empty conjunction and the empty disjunction

Since a conjunction or disjunction can have more than two disjuncts, it's worth asking if it can have fewer...

As we have just seen, '$\vee$' and '$\wedge$' have uppercase versions '$\bigvee$' and '$\bigwedge$' that can be applied to sets of formulæ: $\bigvee\{A, B\}$ is obviously the same as $A \vee B$ for example, and $\bigwedge\{A, B\}$ is $A \wedge B$ by the same token.

Slightly less obviously $\bigwedge\{A\}$ and $\bigvee\{A\}$ are both A. But what is $\bigvee \emptyset$? (the disjunction of the empty set of formulæ). Does it even make sense? Yes it does, and if we are brave we can even calculate what it is.

If X and Y are sets of formulæ then $\bigvee(X \cup Y)$ had better be the same as $\bigvee X \vee \bigvee Y$. Now what if Y is $\emptyset$, the empty set? Then

$$\bigvee X$$

$$= \bigvee(X \cup \emptyset)$$

(because $X = X \cup \emptyset$)

$$= (\bigvee X) \vee (\bigvee \emptyset)$$

so

$$(\bigvee X) \vee (\bigvee \emptyset) \;=\; (\bigvee X) \tag{1.2}$$

and this has got to be true for all sets X of formulæ. This compels '$\bigvee \emptyset$' to always evaluate to `false`. (If it were anything else then a situation might arise in which the left-hand side of (1.2) were true and the right-hand-side false.) In fact we could think of '$\bot$' as an abbreviation for '$\bigvee \emptyset$'.

Similarly '$\bigwedge \emptyset$' must always evaluate to `true`. In fact we could think of '$\top$' as an abbreviation for '$\bigwedge \emptyset$'.

1.5.3 Conjunctive and Disjunctive Normal Form

Each row of a truth-table for a formula records the truth-value of that formula under a particular valuation: each row corresponds to a valuation and *vice versa*. The Disjunctive Normal Form of a formula A is simply the disjunction of the

[8] If $I = \{1, 2, 3, 4\}$ then $\bigvee\limits_{i \in I} A_i$ is $(A_1 \vee A_2 \vee A_3 \vee A_4)$.

rows in which A comes out true, and each row is thought of as the conjunction of the atomics and negatomics that come out true in that row. Let us start with a simple example:

A	$\longleftrightarrow$	B
1	1	1
1	0	0
0	0	1
0	1	0

is the truth-table for '$\longleftrightarrow$'. It tells us that $A \longleftrightarrow B$ is true if A and B are both true, or if they are both false (and not otherwise). That is to say, $A \longleftrightarrow B$ is logically equivalent to $(A \wedge B) \vee (\neg A \wedge \neg B)$.

A slightly more complicated example:

$(A$	$\longleftrightarrow$	$B)$	$\longleftrightarrow$	C
1	1	1	1	1
1	0	0	0	1
0	0	1	0	1
0	1	0	1	1
1	1	1	0	0
1	0	0	1	0
0	0	1	1	0
0	1	0	0	0

This says that $(A \longleftrightarrow B) \longleftrightarrow C$ comes out true in the row where A, B and C are all true, and in the row where ... in fact in those rows where an even number of A, B and C are false. (Check it!)

So $(A \longleftrightarrow B) \longleftrightarrow C$ is logically equivalent to

$$(A \wedge B \wedge C) \;\vee\; (A \wedge \neg B \wedge \neg C) \;\vee\; (\neg A \wedge \neg B \wedge C) \;\vee\; (\neg A \wedge B \wedge \neg C)$$

(Notice how much easier this formula is to read once we have left out the internal brackets!)

Say something about how one is tempted to put '$\wedge$' between them ...

DEFINITION 8

A formula is in **Disjunctive Normal Form** *if the only connectives in it are '$\wedge$', '$\vee$' and '$\neg$' and there are no connectives within the scope of any negation sign and no '$\vee$' within the scope of a '$\wedge$';*

A formula is in **Conjunctive Normal Form** *if the only connectives in it are '$\wedge$', '$\vee$' and '$\neg$' and there are no connectives within the scope of any negation sign and no '$\wedge$' within the scope of a '$\vee$'.*

Using these definitions it is not blindingly obvious that a single propositional letter by itself (or a disjunction of two propositional letters, or a conjunction of two propositional letters) is a formula in both CNF and DNF, though this is in fact the case.[9]

We cannot describe CNF in terms of rows of truth-tables in the cute way we can describe DNF.

fourth week

Exercise 12 *Recall the formula "`if` p `then` q `else` r" from exercise 5. Put it into CNF and also into DNF.*

fourth week

Exercise 13 *For each of the following formulæ say whether it is in CNF, in DNF, in both or in neither.*

(i)	$\neg(p \wedge q)$
(ii)	$p \wedge (q \vee r)$
(iii)	$p \vee (q \wedge \neg r)$
(iv)	$p \vee (q \wedge (r \vee s))$
(v)	p
(vi)	$(p \vee q)$
(vii)	$(p \wedge q)$

Theorem 9 *Every formula is logically equivalent both to something in CNF and to something in DNF.*

Proof:

We force everything into a form using only $\wedge$, $\vee$ and $\neg$, using equivalences like $A \to B \longleftrightarrow \neg A \vee B$.

Then we use the following equivalences to "import" $\neg$, so that the '$\neg$' sign appears only attached to propositional letters.

We saw earlier (exercise 6) that

> $\neg(A \wedge B)$ and $\neg A \vee \neg B$ are logically equivalent;
>
> and
>
> $\neg(A \vee B)$ and $\neg A \wedge \neg B$ are logically equivalent;

So $\neg(A \wedge B)$ can be rewritten as $\neg A \vee \neg B$ and $\neg(A \vee B)$ can be rewritten as $\neg A \wedge \neg B$.

There is also:

> $\neg(A \to B)$ is logically equivalent to $A \wedge \neg B$

so $\neg(A \to B)$ can be rewritten as $A \wedge \neg B$;

The effect of these rewritings is to "push the negations inwards".

Then we can use use the two distributive laws to turn formulæ into CNF or DNF

$$A \vee (B \wedge C) \longleftrightarrow (A \vee B) \wedge (A \vee C) \tag{1.1}$$

[9] It doesn't much matter since the question hardly ever arises. I think Wikipædia gives a different definition.

which means that $A \vee (B \wedge C)$ are logically equivalent $(A \vee B) \wedge (A \vee C)$ so $A \vee (B \wedge C)$ can be rewritten as $(A \vee B) \wedge (A \vee C)$. We use this to "push $\vee$ inside $\wedge$" if we want the formula in CNF

or

$$A \wedge (B \vee C) \longleftrightarrow (A \wedge B) \vee (A \wedge C) \tag{1.2}$$

which means that $A \wedge (B \vee C)$ and $(A \wedge B) \vee (A \wedge C)$ are logically equivalent so $A \wedge (B \vee C)$ can be rewritten as $(A \wedge B) \vee (A \wedge C)$

We use this to "push $\wedge$ inside $\vee$" if we want the formula in DNF

Two further simplifications are allowed:

1. We can replace $B \wedge (A \vee \neg A)$ by B;
2. We can replace $B \vee (A \wedge \neg A)$ by B.

(because $B \wedge (A \vee \neg A)$ is logically equivalent to B, and $B \vee (A \wedge \neg A)$ is logically equivalent to B). ■

We should have a couple of illustrations.

In DNF inconsistencies vanish: the empty disjunction is the `false`, and In CNF tautologies vanish: the empty conjunction is the `true`. (Recall what we were saying on page 32 about the empty conjunction and the empty disjunction.)

Here are some examples:

1. $$(p \vee q) \to r$$
   convert the '$\to$':
   $$\neg(p \vee q) \vee r$$
   import '$\neg$'
   $$(\neg p \wedge \neg q) \vee r$$
   and it is now in DNF. Then distribute '$\vee$' over '$\wedge$' to obtain
   $$(\neg p \vee r) \wedge (\neg q \vee r)$$
   which is in CNF.

2. $$p \to (q \wedge r)$$
   convert the '$\to$':
   $$\neg p \vee (q \wedge r)$$
   and it is now in DNF. Then distribute '$\vee$' over '$\wedge$' to obtain
   $$(\neg p \vee q) \wedge (\neg p \vee r)$$
   which is now in CNF.

3.
$$p \wedge (q \rightarrow r)$$

convert the '$\rightarrow$':
$$p \wedge (\neg q \vee r)$$

which is now in CNF. Then distribute '$\wedge$' over '$\vee$' to obtain

$$(p \wedge \neg q) \vee (p \wedge r)$$

which is in DNF.

4.
$$(p \wedge q) \rightarrow r$$

convert the '$\rightarrow$':
$$\neg(p \wedge q) \vee r$$

de Morgan
$$(\neg p \vee \neg q) \vee r$$

Drop the brackets because '$\vee$' is associative ...

$$\neg p \vee \neg q \vee r$$

which is in both CNF and DNF.

5.
$$p \rightarrow (q \vee r)$$

convert the '$\rightarrow$'
$$\neg p \vee (q \vee r)$$

Drop the brackets because '$\vee$' is associative ...

$$\neg p \vee q \vee r$$

which is in both CNF and DNF.

6.
$$(p \vee q) \wedge (\neg p \vee r)$$

is in CNF. To get it into DNF we have to distribute the '$\wedge$ over the '$\vee$'. (Match 'A' to '$p \vee q$', match B to '$\neg p$' and 'C' to 'r' in '$A \wedge (B \vee C) \longleftrightarrow ((A \wedge B) \vee (A \wedge C))$.)

$$((p \vee q) \wedge \neg p) \ \vee \ ((p \vee q) \wedge r)$$

and then distribute again in each disjunct:

$$((p \wedge \neg p) \vee (q \wedge \neg p)) \vee ((p \wedge r) \vee (q \wedge r))$$

Now $p \wedge \neg p$ is just $\perp$...

$$((\perp \vee (q \wedge \neg p)) \vee ((p \wedge r) \vee (q \wedge r))$$

and $\perp \vee (q \wedge \neg p)$ is just $q \wedge \neg p$:

$$((q \wedge \neg p) \vee ((p \wedge r) \vee (q \wedge r))$$

finally dropping backets because '$\vee$' is associative ...

$$(q \wedge \neg p) \vee (p \wedge r) \vee (q \wedge r)$$

Note that in CNF (DNF) not every conjunct (disjunct) has to contain every letter Finally, by using CNF and DNF we can show that any truth-functional connective whatever can be expressed in terms of $\wedge$, $\vee$ and $\neg$. Any formula is equivalent to the disjunction of the rows (of the truth-table) in which it comes out true. To illustrate, consider the expression $P \longleftrightarrow Q$. If you write out the truth-table for this formula you will see that the two rows in which it comes out true are (i) the row in which both P and Q are true, and (ii) the row in which they are both false. Therefore $P \longleftrightarrow Q$ is equivalent to $(P \wedge Q) \vee (\neg P \wedge \neg Q)$.

1.6 Further Useful Logical Gadgetry

We've already encountered the intension/extension distinction and the type-token distinction. There are a few more.

1.6.1 The Analytic-Synthetic Distinction

More detail needed here

(*For the moment—until I get this section sorted out—read the articles in The Stanford Online Encyclopædia or in Wikipædia.*)

This is one of a trio of distinctions collectively sometimes known as **Hume's wall**. They are the Analytic-synthetic distinction, the *a priori-a posteriori* distinction and the necessary-contingent distinction. It is sometimes alleged that they are all the same distinction—specifically

$$\text{Analytic} = \text{necessary} = \textit{a priori}$$

and

$$\text{Synthetic} = \text{contingent} = \textit{a posteriori}.$$

Hume's wall indeed. Quine famously claimed that the analytic-synthetic distinction is spurious. Kant thought there were assertions that were synthetic but *a priori*. Kripke claims there are necessary truths that are *a posteriori*.

The cast of philosophical pantomime includes the analytic truth "All bachelors are unmarried".[10] You can see that this allegation is true merely by analysing it—hence *analytic*.

[10] "Oh no they aren't!!"

The analytic/synthetic distinction seems to be connected with the intension/extension distinction[11]: facts about intensions are analytic and facts about extensions are synthetic; equations between intensions are analytic and equations between extensions are synthetic. To illustrate

1. The equation

   bachelor = *unmarried man*

   expressing the identity of the two properties-in-intension *bachelor* and *unmarried man* is an analytic truth;

2. The inequation

   human ≠ *featherless biped*

   expressing the distinctness of the two properties-in-intension *human* and *featherless biped* is also an analytic truth;

3. The equation

   human = *featherless biped*

   expressing the identity of the two properties-in-extension *human* and *featherless biped* is a synthetic truth;

4. It's analytic that *Man is a rational animal*...

5. ... but purely synthetic that the set of humans is coextensive with the set of featherless bipeds. (Sets are unary [one-place] properties-in-extension)

explain the equivocation on 'featherless biped'

1.6.2 Necessary and Sufficient Conditions

If $A \to B$ is true then we often say that A *is a sufficient condition for* B. And indeed, that is all there is to it. If A is a sufficient condition for B then $A \to B$: the two forms of words are synonymous.

A *is a* necessary *condition for* B is a related idea. That means that if B holds, it must be because A holds. B can only be true of A is. That is to say, if B then A.

Say something about unfortunate overloading of 'neccesary'

Thus: A **is a necessary condition for** B if and only if B **is a sufficient condition for** A.

[11]See for example Carnap *Meaning and Necessity*, [3].

1.6.3 The Use-Mention Distinction

We must distinguish words from the things they name: the word 'butterfly' is not a butterfly. The distinction between the word and the insect is known as the "use-mention" distinction. The word 'butterfly' has nine letters and no wings; a butterfly has two wings and no letters. The last sentence *uses* the word 'butterfly' and the one before that *mentions* it. Hence the expression 'use-mention distinction'. (It is a bit more difficult to illustrate the difference between using and mentioning butterflies!)

Haddocks' Eyes

As so often the standard example is from [4].

> [...] The name of the song is called 'Haddock's eyes'."
>
> "Oh, that's the name of the song is it", said Alice, trying to feel interested.
>
> "No, you don't understand," the Knight said, looking a little vexed. "That's what the name is *called*. The name really is *'The agèd, agèd man'*."
>
> "Then I ought to have said, 'That's what the *song* is called'?" Alice corrected herself.
>
> "No you oughtn't: that's quite another thing! The *song* is called '*Ways and means*', but that's only what it is *called*, you know!"
>
> "Well, what *is* the song, then?" said Alice, who was by this time completely bewildered.
>
> "I was coming to that," the Knight said. "The song really is *'A-sitting on a Gate'* and the tune's my own invention".

The situation is somewhat complicated by the dual use of single quotation marks. They are used both as a variant of ordinary double quotation marks for speech-within-speech (to improve legibility)—as in "Then I ought to have said, 'That's what the *song* is called'?"—and also to make names of words—'butterfly'. Even so, it does seem clear that the White Knight has got it wrong. At the very least if the name of the song is "The agèd agèd man" as he says then clearly Alice was right to say that was what the song was called, It might have more names than just that one—such as 'Ways and means'—but that was no reason for him to tell her she had got it wrong. And again, if his last utterance is to be true he should leave the single quotation marks off the title, or failing that (as Martin Gardner points out in *The Annotated Alice*) burst into song. These mistakes must be mistakes of the White Knight not Lewis Carroll, but it is hard to see what purpose these errors serve, beyond multiplying in Alice's head the sense of nightmare and confusion that she already feels ... Perhaps he had the reader in his sights too.

'Think'

> "If I were asked to put my advice to a young man in one word, Prestwick, do you know what that word would be?"
>
> "No" said Sir Prestwick.
>
> " 'Think', Prestwick, 'Think' ".
>
> "I don't know, R.V. 'Detail'?"
>
> "No, Prestwick, 'Think'."
>
> "Er, 'Courage'?"
>
> "No! 'Think'!"
>
> "I give up, R.V., 'Boldness'?"
>
> "For heavan's sake, Prestwick, what is the matter with you? '*Think*'!"
>
> " 'Integrity'? 'Loyalty'? 'Leadership'?"
>
> " 'Think', Prestwick! 'Think', 'Think', 'Think' 'Think'!"

Michael Frayn: *The Tin Men.* Frayn has a degree in Philosophy.

Ramsey for Breakfast

In the following example Ramsey[12] uses the use-mention distinction to generate something very close to paradox: the child's last utterance is an example of what used to be called a "self-refuting" utterance: whenever this utterance is made, it is not expressing a truth.

> PARENT: Say 'breakfast'.
> CHILD: Can't.
> PARENT: What can't you say?
> CHILD: Can't say 'breakfast'.

The use-mention distinction is a rich source of jokes. One of my favourites is the joke about the compartment in the commuter train, where the passengers have travelled together so often that they have long since all told all the jokes they know, and have been reduced to the extremity of numbering the jokes and reciting the numbers instead. In most versions of this story, an outsider arrives and attempts to join in the fun by announcing "*Fifty-six!*" which is met with a leaden silence and he is told "It's not the joke, it's the way you tell it". In another version he then tries "*Forty-two!*" and the train is convulsed with laughter. Apparently they hadn't heard that one before.

We make a fuss of this distinction because we should always be clear about the difference between a thing and its representation. Thus, for example, we distinguish between numerals and the numbers that they represent.

Notice that bus "numbers" are typically numerals not numbers. Not long ago, needing a number 7 bus to go home, I hopped on a bus that had the

[12] You will be hearing more of this chap.

string '007' on the front. It turned out to be an entirely different route! Maybe this confusion in people's minds is one reason why this service is now to be discontinued.

A good text to read on the use-mention distinction is the first six paragraphs (that is, up to about p. 37) of Quine's [8]. However it deos introduce stubtleties we will not be respecting.

Related to the use-mention distinction is the error of attributing powers of an object to representations of that object. I tend to think that this is a use-mention confusion. But perhaps it's a deliberate device, and not a confusion at all. So do we want to stop people attributing to representations powers that strictly belong to the things being represented? Wouldn't that spoil a lot of fun? Perhaps, but on the other hand it might help us understand the fun better. There was once a famous English stand-up comic by the name of *Les Dawson* who (did mother-in-law jokes but also) had a routine which involved playing the piano *very badly*. I think Les Dawson must in fact have been quite a good pianist: if you want a sharp act that involves playing the piano as badly as he seemed to be playing it you really have to know what you are doing.[13]. The moral is that perhaps you only experience the full *frisson* to be had from use-mention confusion once you understand the use-mention distinction properly.

1.6.4 Language-metalanguage distinction

The language-metalanguage distinction is important for rhetoric. A debate is going to be conducted in some language or other: there will be a specified or agreed vocabulary and so on. (it will be part of what the literary theorists call a *discourse*.)Let us suppose the debate is about widgets. The people commenting on, or observing the debate will have a different language (discourse) at their disposal. This language will provide the commentators with means for discussing and analysing the motives and strategies of the participants in the debate, and all sorts of other things beyond widgets. All sorts of things, in fact, which the chairman of the debate would rule to be irrelevant to the debate *about widgets*.

(This is connected to ideas in literary theory and elsewhere about the difference between an observer and a participant). Participants in a debate will attempt to represent themselves as expert witnesses who are above the fray whereas they are in fact interested parties. If you speak metalanguage you have the last word—and that of course is what every debater wants.

The language-metalanguage distinction is related to the use-mention distinction in the following way. If I am going to discuss someone else's discourse, I need a lexicon (a vocabulary) that has words to denote items in (the words in) their discourse. One standard way to obtain a name for a word is to put single quotation marks round a token of that word. So if you are discussing the activities of bird-watchers you will need words to describe the words they use. They talk about—for example—*chaffinches* and so they will have a word for this bird. That word is 'chaffinch'. (Note single quote) The people who discuss

[13]Wikepædia confirms this: apparently he was an accomplished pianist

the linguistic behaviour of twitchers will have a name for that word, and that name will be ‘ ‘chaffinch’ ’. (Observe: *two* single quotes!)

There are some intellectual cultures that make great use of the device of always putting tokens of their opponents’ lexicon inside quotation marks. This serves to express distaste for the people they are discussing, to make it look ridiculous, and to make it clear that the offending words are not part of their language. This is not quite the same move, since the quotes here are “scare-quotes” rather than naming quotes, but the device is related.

(The language-metalanguage distinction will be useful later in connection with sequent calculus.)

1.6.5 Semantic Optimisation and the Principle of Charity

When a politician says “We have found evidence of weapons-of-mass-destruction programme-related activities”, you immediately infer that that have *not* found weapons of mass destruction (whatever they are). Why do you draw this inference?

Well, it’s *so* much easier to say “We have found weapons of mass destruction” than it is to say “We have found evidence of weapons-of-mass-destruction-related programme-related activities” that the only conceivable reason for the politician to say the second is that he won’t be able to get away with asserting the first. After all, why say something longer and less informative when you can say something shorter and more informative? One can see this as a principle about maximising the amount of information you convey while minimising the amount of energy you expend in conveying it. If you were a first-year economics student you would probably be learning some elementary optimisation theory at this stage, and you might like to learn some on the fly: economists have had some enlightening things to say about philosophy of language. It’s not difficult to learn enough optimisation theory to be able to see where it could usefully lead. (It’s not a bad idea to think of ourselves as generally trying to minimise the effort involved in conveying whatever information it is that we want to convey.)

Quine used the phrase “The Principle of Charity” for the assumption one makes that the people one is listening to are trying to minimise effort in this way. It’s a useful principle, in that by charitably assuming that they are not being unneccessarily verbose it enables one to squeeze a lot more information out of one’s interlocutors’ utterances than one otherwise might, but it’s dangerous. Let’s look at this more closely.

Weapons of Mass Destruction

Suppose I hear you say

> We have found evidence of weapons-of-mass-destruction programme-related activities. (1)

Now you *could* have said

> We have found weapons of mass destruction. (2)

...which is shorter; so why did you not say it? The principle of charity tells me to infer that you were not in a position to say (2), which means that you have *not* found weapons of mass destruction. However, you should notice that (1) emphatically does *not* imply that

We have *not* found weapons of mass destruction. (3)

After all, had you been lucky enough to have found weapons of mass destruction then you have most assuredly found evidence of weapons-of-mass-destruction programme-related activities: the best possible evidence indeed. So what is going on?

What's going on is that (1) does not imply (3), but that (4) does!

We had no option but to say "We have found evidence of weapons-of-mass-destruction programme-related activities" rather than "We have found weapons of mass destruction ". (4)

Of course (1) and (4) are not the same!

The principle of charity is what enables us to infer (4); and to infer it not from (3) but from the fact that they said (3) instead of (2).

Perhaps a better example—more enduring and more topical—is...

Wrong Kind of Snow

80% of our trains arrive within 5 minutes of their scheduled time. (A)

Note that (A) does *not* imply:

20% of our trains are more than 5 minutes late. (B)

The claim (A) is certainly not going to be falsified if the train company improves its punctuality. So what is going on when people infer (B) from (A)?

What is going on is that although (A) doesn't imply (B), (C) certainly does imply (B).

The train company has chosen to say "80% of our trains arrive within 5 minutes of their scheduled time", and the train companies wish to put themselves in the best possible light. (C)

...and the second conjunct of (C) is a safe bet.

Now the detailed ways in which this optimisation principle is applied in ordinary speech do not concern us here—beyond one very simple consideration. I want you to understand this optimisation palaver well enough to know when you are tempted to apply it, and to lay off. The languages of formal logic are languages of the sort where this kind of subtle reverse-engineering of interlocutors' intentions is a hindrance not a help. Everything has to be taken literally.

See also the beautiful discussion of the Rabbinical tradition in [11] starting on p. 247.

1.6.6 Inferring A-or-B from A

You might be unhappy about inferring A-or-B from A because you feel that anyone who says A-or-B is claiming knowledge that at least one of them is true but (since they are not saying A and not saying B) are—and you get this by the principle of charity—denying knowledge of A and denying knowledge of B. Whereas of course the person who says A is claiming knowledge of A!

If that is what is going on in your head you are being too subtle and not literal enough!

1.6.7 Fault-tolerant pattern-matching

Explain what it is!

Fault-tolerant pattern matching is very useful in everyday life but absolutely no use at all in the lower reaches of logic. Too easily fault-tolerant pattern matching can turn into overenthusiastic pattern matching—otherwise known as *syncretism*: the error of making spurious connections between ideas. A rather alarming finding in the early days of experiments on sensory deprivation was that people who are put in sensory deprivation tanks start hallucinating: their receptors expect to be getting stimuli, and when they don't, they wind up their sensitivity until they start getting positives. Since they are in a sensory deprivation chamber, those positives are one and all spurious.

1.6.8 The conjunction fallacy

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which is more probable?

1. *Linda is a bank teller;*
2. *Linda is a bank teller and is active in the feminist movement.*

Thinking that (2) is more probable than (1) is the **conjunction fallacy**—the mistake of attaching a higher probability to $P \wedge Q$ than to P. See [10] (from which this comes) and also the Wikipædia article.

1.7 Fallacy of Equivocation

verb: to equivocate.

Bronze is a metal, all metals are elements, so bronze is an element.

here we equivocate on the word 'metal'. It appears twice, and the two occurrences of it bear different meanings—at least if we want both the premisses to be true. But if both premisses are true then the conclusion must be true—and it isn't!

Chapter 2

Proof systems for Propositional Logic

2.1 Arguments by LEGO

The arguments I've used as illustrations so far are very simple. Only two premisses and one conclusion. Altho' it's true that all the arguments we are concerned with will have only one conclusion, many of them will have more than two premisses. So we have to think about how we obtain the conclusion of an argument from its premisses. This we do by manipulating the premisses according to certain rules, which enable us to take the premisses apart and reassemble them into the conclusions we want. These rules have the form of little **atomic** arguments, which can be assembled into **molecular** arguments which are the things we are actually interested in.

We know what a valid expression of propositional logic is. We know how to detect them by using truth tables. In this chapter we explore a method for generating them.

2.2 The Rules of Natural Deduction

In the following table we see that for each connective we have two rules: one to introduce the connective and one to eliminate it. These two rules are called the **introduction rule** and the **elimination rule** for that connective.

Richard Bornat calls the elimination rules "use" rules because the elimination rule for a connective $\mathcal{C}$ tells us how to **use** the information wrapped up in a formula whose principal connective is $\mathcal{C}$.

(The idea that everything there is to know about a connective can be captured by an elimination rule plus an introduction rule has the same rather operationalist flavour possessed by the various *meaning is use* doctrines one encounters in philosophy of language. In this particular form it goes back to references?

Prawitz, and possibly to Gentzen.)

The rules tell us how to use the information contained in a formula (Some of these rules come in two parts.)

$\vee$-int: $\dfrac{A}{A \vee B}$; $\dfrac{B}{A \vee B}$;

$\vee$-elim(1): $\dfrac{A \vee B \quad \begin{array}{c}[A]^1\\ \vdots \\ C\end{array} \quad \begin{array}{c}[B]^1\\ \vdots \\ C\end{array}}{C}$

$\wedge$: $\dfrac{A \quad B}{A \wedge B}$;

$\wedge$-elim: $\dfrac{A \wedge B}{A}$; $\dfrac{A \wedge B}{B}$

$\to$-int(1) $\dfrac{\begin{array}{c}[A]^1\\ \vdots \\ B\end{array}}{A \to B}$

$\to$-elim: $\dfrac{A \quad A \to B}{B}$

Ex falso sequitur quodlibet; $\dfrac{\bot}{A}$

Double negation $\dfrac{\neg\neg A}{A}$

Some small print:

N.B.: in $\to$-introduction you don't have to cancel all occurrences of the premiss: it is perfectly all right to cancel only some of them .

The Latin expression *ex falso* ... means: "From the `false` follows whatever you like".

We have no rules for '$\neg$'. '$\neg A$' is logically equivalent to '$A \to \bot$' so whenever we have '$\neg A$' we thnk of it as '$A \to \bot$' and use the rules for '$\to$' instead.

Some of these rules look a bit daunting so let's start by cutting our teeth on some easy ones.

Probably for week 5

EXERCISE 14 *Using just the two rules for $\wedge$, the rule for $\vee$-introduction and $\to$-elimination see what you can do with each of the following sets of formulæ:*[1]

1. $A,\ A \to B$;
2. $A,\ A \to (B \to C)$;
3. $A,\ A \to (B \to C),\ B$;
4. $A,\ B,\ (A \wedge B) \to C$;
5. $A,\ (A \vee B) \to C$;
6. $A \wedge B,\ A \to C$;
7. $A \wedge B,\ A \to C,\ B \to D$;
8. $A \to (B \to C),\ A \to B,\ B \to C$;
9. $A,\ A \to (B \to C),\ A \to B$;

[1] Warning: in some cases the answer might be "nothing!".

10. A, $\neg A$.

You will probably notice in doing these questions that you use one of your assumptions more than once, and indeed that you have to *write it down* more than once (= write down more than one token!) This is particularly likely to happen with $A \wedge B$. If you need to infer both of A and B then you will have to write out '$A \wedge B$' *twice*—once for each application of $\wedge$-elimination.

If you try writing down only one token you will find that you want your sheet of paper to be made of lots of plaited ribbons. Ugh.

The two rules of *ex falso* and *double negation* are the only rules that specifically mention negation. Recall that $\neg B$ is logically equivalent to $B \to \bot$, so the inference

$$\frac{A \qquad \neg A}{\bot} \tag{2.1}$$

—which *looks* like a new rule—is merely an instance of $\to$-elimination.

Finally we need the identity rule:

$$\frac{A\ B\ C \ldots}{A} \tag{2.2}$$

(where the list of extra premisses may be empty) which records the fact that we can deduce A from A. Not very informative, one might think, but it turns out to be useful. After all, how else would one obtain a proof of the undoubted tautology $A \to (B \to A)$, otherwise known as 'K'? (You established that it was a truth-table tautology in exercise 7.) One could do something like

$$\cfrac{\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A \wedge B}\ \wedge\text{-int}}{A}\ \wedge\text{-elim}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{2.3}$$

but that is grotesque: it uses a couple of rules for a connective that doesn't even appear in the formula being proved! The obvious thing to do is

$$\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A}\ \text{identity rule}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{2.4}$$

If we take seriously the observation above concerning the rule of $\to$-introduction—namely that you are not required to cancel *all* occurrences of an assumption—then you infer that you can cancel none of them, and that suggests that you can cancel assumptions that aren't there—then we will not need this rule. This means we can write proofs like 2.5 below. To my taste, it seems less bizarre to discard assumptions than it is to cancel assumptions that aren't there, so I prefer 2.4 to 2.5. It's a matter of taste.

$$\dfrac{\dfrac{[A]^1}{B \to A}\ \to\text{-int}}{A \to (B \to A)}\ \to\text{-int (1)} \tag{2.5}$$

It is customary to connect the several occurrences of a single formula at introductions (it may be introduced several times) with its occurrences at elimination by means of superscripts. Square brackets are placed around eliminated formulæ, as in the formula displayed above.

There are funny logics where you are not allowed to use an assumption more than once: in these **resource logics** assumptions are like sums of money. This also gives us another illustration of the difference between an argument (as in logic) and a debate (as in rhetoric). In rhetoric it may happen that a point—even a good point—can be usefully made only once ... in an ambush perhaps.

Do some very simple illustrations of compound proofs here

2.2.1 What do the rules *mean*??

One way in towards an understanding of what the rules do is to dwell on the point made by my friend Richard Bornat that elimination rules are **use** rules:

The rule of →-elimination

The rule of →-elimination tells you how to use the information wrapped up in '$A \to B$'. '$A \to B$' informs us that if A, then B. So the way to use the information is to find yourself in a situation where A holds. You might not be in such a situation, and if you aren't you might have to assume A with a view to using it up later—somehow. We will say more about this.

The rule of ∨-elimination

The rule of ∨-elimination tells you how to **use** the information in '$A \vee B$'. If you are given $A \vee B$, how are you to make use of this information without supposing that you know which of A and B is true? Well, **if** you know you can deduce C from A, and you ALSO know that you can deduce C from B, **then** as soon as you are told $A \vee B$ you can deduce C. One could think of the rule of ∨-elimination as a function that takes (1) $A \vee B$, (2) a proof of C from A and (3) a proof of C from B, and returns a proof of C from $A \vee B$.

Here is an example, useful to those of you who fry your brains doing sudoku.

	3	8						
	1	6		4		9	7	
4		7	1					6
		2	8		7			5
	5			1			8	
8			4			2		
7		5			1	8		4
	4	3		5		7	1	
						6		

There is a '5' in the top right-hand box—somewhere. But in which row? The '5' in the top left-hand box must be in the first column, and in one of the top two rows. The '5' in the fourth column must be in one of the two top cells. (It cannot be in the fifth row because there is already a '5' there, and it cannot be in the last three rows because that box already has a '5' in it.) So the '5' in the middle box on the top must be in the first column, and in one of the top two rows. These two '5's must of course be in different rows. So where is the '5' in the rightmost of the three top boxes? Either the '5' in the left box is on the first row and the '5' in the middle box is on the second row or the 5 in the middle box is in the first row and the '5' in the left box is in the second row. We don't know which of the possibilities is the true one, but it doesn't matter: either way the '5' in the rightmost box must be in the bottom (third) row.

There is a more general form of $\vee$-elimination:

$$\frac{\begin{array}{cccc} [A_1]^1 & [A_2]^1 & \ldots & [A_n]^1 \\ \vdots & \vdots & & \vdots \\ C & C & & C \end{array} \qquad A_1 \vee A_2 \vee \ldots A_n}{C} \vee\text{-elim (1)} \tag{2.1}$$

where we can cancel more than one assumption. That is to say we have a *set* $\{A_1 \ldots A_n\}$ of assumptions, and the rule accepts as input a list of proofs of C: one proof from A_1, one proof from A_2, and so on up to A_n. It also accepts the disjunction $A_1 \vee \ldots A_n$ of the set $\{A_1 \ldots A_n\}$ of assumptions, and it outputs a proof of C.

The rule of $\vee$-elimination is a hard one to grasp so do not panic if you don't get it immediately. However, you should persist until you do.

EXERCISE 15

1. *Deduce C from $(A \vee B) \to C$ and A;*

2. *Deduce B from $(A \to B) \to A$ and $A \to B$;*

3. *Deduce R from $P \to (Q \to R)$ and $P \to Q$;*

4. *Deduce* C *from* A *and* $((A \to B) \to B) \to C$*;*

5. *Deduce* $\neg P$ *from* $\neg(Q \to P)$*;*

6. *Deduce* $(A \to B) \to B$ *from* A*;*

7. *Deduce* A *from* $B \vee C$*,* $B \to A$ *and* $C \to A$*;*

8. *Deduce* $\neg A$ *from* $\neg(A \vee B)$*;*

9. *Deduce* Q *from* P *and* $\neg P \vee Q$*;*

10. *Deduce* Q *from* $\neg(Q \to P)$*;*

2.2.2 Worries about *reductio* and hypothetical reasoning

Many people are unhappy about hypothetical reasoning of the kind used in the rule of $\to$-introduction. I am not entirely sure why, so I am not 100% certain what to say to make the clouds roll away. However here are some thoughts.

Part of it may arise from the failure to distinguish between "If A then B" and "A, therefore B". If you do not distinguish between these you won't be inclined to see any difference between the act-of-assuming-A-and-deducing-B (in which you assert A) and the act-of-deducing-$A \to B$ (in which you do *not* assert A)

Another unease about argument by *reductio ad absurdum* seems to be that if I attempt to demonstrate the falsity of p by assuming p and then deducing a contradiction from it then—if I succeed—I have somehow not so much proved that p was false but instead contrived to explode the machinery of deduction altogether: if p was false how could I have sensibly deduced anything from it in the first place?! I have somehow sawn off the branch I was sitting on. I *thought* I was deducing something, but I couldn't have been. This unease then infects the idea of hypothetical reasoning: reasoning where the premisses are—if not actually known to be false—at least not known to be true. No idea is so crazy that no distinguished philosopher can ever be found to defend it (as Descartes said, and he should know!) and one can indeed find a literature in which this idea is defended.

Evert Beth said that Aristotle's most important discovery was that the same processes of reasoning used to infer new truths from propositions previously known to be true are also used to deduce consequences from premises not known to be true and even from premises known to be false.[2]

[2] See [6]. Spinoza believed hypothetical reasoning to be incoherent, but that's because he believed *all* truths to be necessary, and even people who are happy about counterfactual reasoning are nervous about attempting to reason from premisses known to be necessarily false! This may be why there is no very good notion of explanation in Mathematics or Theology. They both deal with necessary truth, and counterfactuals concerning necessary truths are problematic. Therefore explanation in these areas is obstructed to the extent that explanation involves counterfactuals.

But it's not hard to see that life would be impossible without hypothetical reasoning. Science would be impossible: one would never be able to test hypotheses, since one would never be able to infer testable predictions from them! Similarly, as one of my correspondents on the philosophy-in-europe mailing list pointed out, a lawyer cross-examining a hostile witness will draw inferences from the witness's testimony in the hope of deducing an absurdity. Indeed if one were unwilling to imagine oneself in the situation of another person (which involves subscribing to their different beliefs) then one would be liable to be labelled as autistic.

This may be related to the phenomenon of people not seeing clearly the difference between "A, therefore B" and "if A then B". The person who says "A, therefore B" is asserting A (as well as asserting B). The person who says "If A then B" is NOT asserting A! Despite this, the relation-between-A-and-B that is being asserted is the same in the two cases: *that's* not where the difference lies.

Finally one might mention the Paradox of the Unexpected Hanging in this connection. There are many things it seems to be about, and one of them is hypothetical reasoning. ("If he is to be hanged on the friday then he would know this by thursday so it can't be friday ..." Some people seem to think that altho' this is a reasonable inference the prisoner can only use it once he has survived to thursday: he cannot use it *hypothetically*...)[3]

2.2.3 Goals and Assumptions

When you set out to find a proof of a formula, that formula is your **goal**. As we have just mentioned, the obvious way to attack a goal is to see if you can obtain it as the output of (a token of) the introduction rule for its principal connective. If that introduction rule is $\to$-introduction then this will generate an **assumption**. Once you have generated an assumption you will need—sooner or later—to extract the information it contains and you will do this by means of the *elimination* rule for the principal connective of that assumption. It's actually idiotically simple:

1. Attack a **goal** with the introduction rule for its principal connective;

2. Attack an **assumption** with the elimination rule for its principal connective.

Consider (1). We have the goal $((A \to B) \to A) \underline{\to} ((A \to B) \to B)$. The principal connective of this formula is the arrow in the middle that I underlined. So we **assume** the antecedent (which is $(A \to B) \to A$) and then the consequent (which is $(A \to B) \to B$) becomes our new goal. So we have traded the old

[3]However, this is almost certainly not what is at stake in the Paradox of the Unexpected Hanging. A widespread modern view—with which I concur—is that the core of the puzzle is retained in the simplified version where the judge says "you will be hanged tomorrow and you do not believe me".

goal $((A \to B) \to A) \underline{\to} ((A \to B) \to B)$ for the new goal $((A \to B) \to B)$ and generated the new assumption $((A \to B) \to A)$.

I have noticed that beginners often treat assumptions as if they were goals. Perhaps this is because they encounter goals first and they are *perseverating*. In the example of the preceding paragraph we generated the assumption $(A \to B) \to A$. How are you going to use this assumption? Do not attempt to *prove* it; you must *use* it! And the way to use it is to whack it with the elimination rule for its principal connective—which is $\to$. The only way you can do this is if you have somehow got hold of $A \to B$—and this gives you the new goal of $A \to B$

define 'counterfactual'; first appearance here

Your first step—when challenged to find a natural deduction proof of a formula—should be to identify the principal connective. (That was the point of exercise 3.) For example, when challenged to find a proof of $(A \wedge B) \to A$, the obvious gamble is to expect that the last step in the proof was a $\to$-introduction rule applied to a proof of A with the assumption $A \wedge B$.

2.2.4 The small print

It isn't always true that you should attack an assumption (or goal) with the elimination (introduction) rule for its main connective. It might be that the goal or assumption you are looking at is a propositional letter and therefore *does not have* a principal connective! In those circumstances you have to try something else. Your assumption might be P and if you have in your knapsack the formula $(P \vee Q) \to R$ it might be a good idea to whack the 'P' with a $\vee$-introduction to get $P \vee Q$ so you can then do a $\to$-elimination and get R. And of course you might wish to refrain from attacking your assumption with the elimination rule for its principal connective. If your assumption is $P \vee Q$ and you already have in your knapsack the formula $(P \vee Q) \to R$ you'd be crazy not to use $\to$-elimination to get R. And in so doing you are not using the elimination rule for the principal connective of $P \vee Q$.

And even when a goal or assumption does have a principal connective attacking it with the appropriate rule for that principal connective is not absolutely *guaranteed* to work. Consider the task of finding a proof of $p \vee \neg p$. (p here is a propositional letter, not a complex formula). If you attack the principal connective you will of course use $\vee$-int and generate the attempt

$$\frac{p}{p \vee \neg p}\ \vee\text{-int} \tag{2.1}$$

or the attempt

$$\frac{\neg p}{p \vee \neg p}\ \vee\text{-int} \tag{2.2}$$

and clearly neither of these is going to turn into a proof of $p \vee \neg p$, since we are not going to get a proof of p (nor a proof of $\neg p$). It turns out you have to use the rule of double negation: assume $\neg(p \vee \neg p)$ and get a contradiction. There is a pattern to at least some of these cases where attacking-the-principal-connective is not the best way forward, and we will say more about it later.

The moral of this is that finding proofs is not a simple join-up-the-dots exercise: you need a bit of ingenuity at times. Is this because we have set up the system wrongly? Could we perhaps devise a system of rules which was completely straightforward, and where short tautologies had short proofs[4] which can be found by blindly following rules like always-use-the-introduction-rule-for-the-principal-connective-of-a-goal? You might expect that, the world being the kind of place it is, the answer is a resounding 'NO!' but curiously the answer to this question is not known. I don't think anyone expects to find such a system, and i know of no-one who is trying to find one, but the possibility has not been excluded.

In any case the way to get the hang of it is to do lots of practice!! So here are some exercises. They might take you a while.

2.2.5 Some Exercises

to be continued ...

First some warnings that might save you from tripping yourself up ...

You can cancel an assumption only if it appears in the branch above you!

(If you try to prove $(A \to (B \vee C)) \to ((A \to B) \vee (A \to C))$ you will find that you have to use the rule of double negation. If you think you have a proof that doesn't use it then you have probably tried to cancel an assumption in another branch)

There is a danger of ellipsis with $\vee$-elimination: one of my students wrote

$$\frac{A \to C \qquad B \to C \qquad A \vee B}{C}\ \vee\text{-elim} \tag{2.1}$$

I can see what she meant! It was

$$\frac{\dfrac{[A]^1 \qquad A \to C}{C}\ \to\text{-elim} \qquad \dfrac{[B]^1 \qquad B \to C}{C}\ \to\text{-elim} \qquad A \vee B}{C}\ \vee\text{-elim}\ (1) \tag{2.2}$$

for week 5

Exercise 16 *Find natural deduction proofs of the following tautologies:*

1. $(P \to Q) \to ((Q \to R) \to (P \to R))$;
2. $(A \to C) \to ((A \wedge B) \to C)$;
3. $((A \vee B) \to C) \to (A \to C)$;
4. $P \to (\neg P \to Q)$;
5. $A \to (A \to A)$ *(you will need the identity rule)*;
6. $(((P \to Q) \to Q) \to Q) \to (P \to Q)$;
7. $((A \to B) \to A) \to ((A \to B) \to B)$

[4]'short' here can be given a precise meaning.

8. $A \to ((((A \to B) \to B) \to C) \to C)$;
9. $(P \vee Q) \to (((P \to R) \wedge (Q \to S)) \to (R \vee S))$;
10. $(P \wedge Q) \to (((P \to R) \vee (Q \to S)) \to (R \vee S))$;
11. $\neg(A \vee B) \to (\neg A \wedge \neg B)$;
12. $A \vee \neg A$; *(*)*
13. $\neg(A \wedge B) \to (\neg A \vee \neg B)$; *(hard!) (*)*
14. $(A \wedge (B \vee C)) \to ((A \wedge B) \vee (A \wedge C))$;
15. $(((A \wedge B) \vee (A \wedge C)) \to (A \wedge (B \vee C))$;
16. $(A \vee (B \wedge C)) \to ((A \vee B) \wedge (A \vee C))$;
17. $((A \vee B) \wedge (A \vee C)) \to (A \vee (B \wedge C))$;
18. $A \to [(A \to C) \to ((B \to C) \to C)]$; *(for this and the next you will need the identity rule);*
19. $B \to [(A \to C) \to ((B \to C) \to C)]$; *then put these last two together to obtain a proof of*
20. $(A \vee B) \to [(A \to C) \to ((B \to C) \to C)]$;
21. $(B \vee (B \to A)) \to A \to A$; *(*)*
22. $(A \wedge B) \ \vee \ (A \wedge \neg B) \ \vee (\neg A \wedge B) \ \vee \ (\neg A \wedge \neg B)$. *(Hard! For enthusiasts only) (*)*

You should be able to do the first eight without breaking sweat. If you can do the first dozen without breaking sweat you may feel satisfied. The starred items will need the rule of double negation. For the others you should be able to find proofs that do not use double negation. The æsthetic into which you are being inducted is one that says that proofs that do not use double negation are always to be preferred to proofs that do. Perhaps it is a bit belittling to call it an æsthetic: there is a principled philosophical position that denies the rule of double negation, and one day you might want to engage with it.

If you want to get straight in your mind the small print around the $\to$-introduction rule you might like to try the next exercise. In one direction you will need to cancel two occurences of an assumption, and in the other you will need the identity rule, which is to say you will need to cancel zero occurences of the assumption.

EXERCISE 17

Get these in something like increasing order of difficulty

1. *Provide a natural deduction proof of* $A \to (A \to B)$ *from* $A \to B$;

2. *Provide a natural deduction proof of* $A \to B$ *from* $A \to (A \to B)$.

To make quite sure you might like to try this one too

EXERCISE 18

1. *Provide a natural deduction proof of* $A \to (A \to (A \to B))$ *from* $A \to B$;

2. *Provide a natural deduction proof of* $A \to B$ *from* $A \to (A \to (A \to B))$.

Which week?

EXERCISE 19 *Life is complicated on Planet Zarg. The Zarglings believe there are three truth-values:* `true`, `intermediate` *and* `false`. *Here we write them as 1, 2 and 3 respectively. Here is the truth-table for the connective* $\to$ *on planet Zarg:*

P	$\to$	Q
1	1	1
1	2	2
1	3	3
2	1	1
2	1	2
2	3	3
3	1	1
3	1	2
3	1	3

On Zarg the truth-value of $P \vee Q$ *is simply the smaller of the truth-values of* P *and* Q; *the truth-value of* $P \wedge Q$ *is the larger of the truth-values of* P *and* Q.

Write out Zarg-style truth-tables for

1. $P \vee Q$;

2. $P \wedge Q$;

3. $((P \to Q) \to P) \to P$;

4. $P \to (Q \to P)$;

5. $(P \to Q) \to Q)$;

[Brief reality check: What is a tautology on Planet Earth?]

What might be a good definition of tautology *on Planet Zarg?*

According to your definition of a tautology-on-planet-Zarg, is it the case that if P *and* Q *are formulæ such that* P *and* $P \to Q$ *are both tautologies, then* Q *is a tautology?*

There are two possible negations on Zarg:

P	$\neg^1 P$	$\neg^2 P$
1	3	3
2	2	1
3	1	1

Given that the Zarglings believe $\neg(P \wedge \neg P)$ to be a tautology, which negation do they use?

Using that negation, do they believe the following formulæ to be tautologies?

1. $P \vee \neg P$?
2. $(\neg\neg P) \vee \neg P$?
3. $\neg\neg(P \vee \neg P)$?
4. $(\neg P \vee Q) \to (P \to Q)$?

weeks 5 or 6

Exercise 20 *Annotate the following proofs, indicating which rules are used where and which premisses are being cancelled when.*

$$\dfrac{\dfrac{\dfrac{P \qquad P \to Q}{Q}}{(P \to Q) \to Q}}{P \to ((P \to Q) \to Q)} \tag{2.3}$$

$$\dfrac{\dfrac{\dfrac{P \wedge Q}{Q}}{P \vee Q}}{(P \wedge Q) \to (P \vee Q)} \tag{2.4}$$

$$\dfrac{\dfrac{\dfrac{P \qquad \neg P}{\bot}}{Q}}{P \to Q} \tag{2.5}$$

$$\dfrac{\dfrac{P \vee Q \qquad \dfrac{P \qquad P \to R}{R} \qquad \dfrac{Q \qquad Q \to R}{R}}{R}}{(P \vee Q) \to R} \tag{2.6}$$

$$\dfrac{\dfrac{\dfrac{A \qquad B}{A \wedge B}}{B \to (A \wedge B)}}{A \to (B \to (A \wedge B))} \tag{2.7}$$

$$\dfrac{\dfrac{\dfrac{(A \to B) \to B \qquad A \to B}{B}}{((A \to B) \to B) \to B}}{(A \to B) \to (((A \to B) \to B) \to B)} \tag{2.8}$$

2.3 Sequent Calculus

Imagine you are given the task of finding a natural deduction proof of the tautology

$$(A \to (B \to C)) \to ((A \to B) \to (A \to C)).$$

Obviously the first thing you do is to attack the principal connective, and claim that $(A \to B) \to (A \to C)$ is obtained by an $\to$-introduction as follows:

$$\dfrac{\begin{array}{c} A \to (B \to C) \\ \vdots \end{array}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{2.1}$$

in the hope that we can fill the dots in later. Notice that we don't know at this stage how many lines or how much space to leave .... At the second stage the obvious thing to do is try $\to$-introduction again, since '$\to$' is the principal connective of '$(p \to q) \to (p \to r)$'. This time my proof sketch has a conclusion which looks like

$$\dfrac{\dfrac{\vdots}{A \to C} \to\text{-int}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{2.2}$$

and we also know that floating up above this—somewhere—are the two premisses $A \to (B \to C)$ and $A \to B$. But we don't know where on the page to put them!

This motivates a new notation. Record the endeavour to prove

$$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$$

by writing

$$\vdash (A \to (B \to C)) \to ((A \to B) \to (A \to C)).$$

using the new symbol '$\vdash$'.[5] Then stage two (which was formula 2.1) can be described by the formula

$$A \to (B \to C) \vdash ((A \to B) \to (A \to C)).$$

[5] For some reason this symbol is called 'turnstile'.

which says that $(A \to B) \to (A \to C)$ can be deduced from $A \to (B \to C)$. Then the third stage [which I couldn't write down and which was formula 2.2, which said that $A \to C$ can be deduced from $A \to B$ and $A \to (B \to C)$] comes out as

$$A \to (B \to C),\ A \to B\ \vdash\ A \to C$$

This motivates the following gadgetry.

A **sequent** is an expression $\Gamma \vdash B$ where Γ is a set of formulæ and B is a formula. $\Gamma \vdash B$ says that there is a deduction of B from Γ. In sequent calculus one reasons not about formulæ—as one did with natural deduction—but instead about sequents, which are assertions about deductions between formulæ. Programme: sequent calculus is natural deduction with control structures! A sequent proof is a program that computes a natural deduction proof.

Capital Greek letters denote sets of formulæ.

We accept any sequent that has a formula appearing on both sides. Such sequents are called **initial sequents**. Clearly the allegation made by an initial sequent is correct!

There are some obvious rules for reasoning about these sequents. Our endeavour to find a nice way of thinking about finding a natural deduction proof of

$$(A \to (B \to C)) \to ((A \to B) \to (A \to C))$$

gives us something that looks in part like

$$\cfrac{\cfrac{\cfrac{A \to (B \to C), (A \to B), A \vdash C}{A \to (B \to C), (A \to B) \vdash (A \to C)}}{A \to (B \to C) \vdash (A \to B) \to (A \to C)}}{\vdash (A \to (B \to C)) \to ((A \to B) \to (A \to C))}$$

and this means we are using a rule

$$\frac{\Gamma, A\ \vdash\ B}{\Gamma\ \vdash\ A \to B} \to \mathrm{R} \tag{2.3}$$

Of course there are lots of other rules, and here is a summary of them:

$\vee L: \quad \dfrac{\Gamma, B \vdash \Delta \quad \Gamma', A \vdash \Delta'}{\Gamma \cup \Gamma', \underline{B \vee A} \vdash \Delta \cup \Delta'}$ $\qquad \vee$ R: $\dfrac{\Gamma \vdash \Delta, A}{\Gamma \vdash \Delta, \underline{B \vee A}}$

$\vee$ R: $\dfrac{\Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, \underline{B \vee A}}$

$\wedge L: \quad \dfrac{\Gamma, B, A \vdash \Delta}{\Gamma, \underline{B \wedge A} \vdash \Delta}$ $\qquad \wedge R: \quad \dfrac{\Gamma \vdash \Delta, B \quad \Gamma' \vdash \Delta', A}{\Gamma \cup \Gamma' \vdash \Delta \cup \Delta', \underline{B \wedge A}}$

$\neg L: \quad \dfrac{\Gamma \vdash \Delta, B}{\Gamma, \underline{\neg B} \vdash \Delta}$ $\qquad \neg R: \quad \dfrac{\Gamma, B \vdash \Delta}{\Gamma \vdash \Delta, \underline{\neg B}}$

$\to L: \quad \dfrac{\Gamma \vdash \Delta, A \quad \Gamma', B \vdash \Delta'}{\Gamma \cup \Gamma', \underline{A \to B} \vdash \Delta \cup \Delta'}$ $\qquad \to R: \quad \dfrac{\Gamma, B \vdash \Delta, A}{\Gamma \vdash \Delta, \underline{B \to A}}$

Weakening-L: $\dfrac{\Gamma \vdash \Delta}{\Gamma, A \vdash \Delta}$; Weakening-R: $\dfrac{\Gamma \vdash \Delta}{\Gamma \vdash \Delta, B}$;

Contraction-L: $\dfrac{\Gamma, B, B \vdash \Delta}{\Gamma, \underline{B} \vdash \Delta}$; Contraction-R: $\dfrac{\Gamma \vdash \Delta, B, B}{\Gamma \vdash \Delta, \underline{B}}$;

Cut: $\dfrac{\Gamma \vdash \Delta, \underline{B} \quad \Gamma', \underline{B} \vdash \Delta'}{\Gamma \cup \Gamma' \vdash \Delta, \Delta'}$.

In this box I have followed the universal custom of writing 'Γ, B' for '$\Gamma \cup \{B\}$; I have not so far followed the similarly universal custom of writing 'Γ, Δ' instead of '$\Gamma \cup \Delta$' but from now on I will.

You might find useful the terminology of **eigenformula**. The eigenformula of an application of a rule is the formula being attacked by that application. In each rule in the box above I have underlined the eigenformula.

There is no rule for the biconditional: we think of a biconditional $A \longleftrightarrow B$ as a conjunction of two conditionals $A \to B$ and $B \to A$.

Now that we have rules for $\neg$ we no longer have to think of $\neg p$ as $p \to \bot$.

The two rules of $\vee$-R give rise to a derived rule which makes good sense when we are allowed more than one formula on the right. it is

$$\underline{\Gamma \vdash \Delta, A, B}$$

$$\Gamma \vdash \Delta, A \vee B$$

$$\frac{\Gamma \vdash \Delta, A, B}{\Gamma \vdash \Delta, A \vee B}$$

I shall explain soon (section 2.3.3) why this is legitimate.

A word is in order on the two rules of contraction. Whether one needs the contraction rules or not depends on whether one thinks of the left and right halves of sequents as sets or as multisets. Both courses of action can be argued for. If one thinks of them as multisets then one can keep track of the multiple times one exploits an assumption. If one thinks of them as as sets then one doesn't need the contraction rules. It's an interesting exercise in philosophy of mathematics to compare the benefits of the two ways of doing it, and to consider the sense in which they are equivalent. Since we are not hell-bent on rigour we will equivocate between the two approaches: in all the proofs we consider it will be fairly clear how to move from one approach to the other and back.

A bit of terminology you might find helpful. Since premisses and conclusion are the left and right parts of a sequent, what are we going to call the things above and below the line in a sequent rule? The terminology **precedent** and **succedent** is sometimes used. I'm not going to expect you to know it: I'm offering it to you here now because it might help to remind you that it's a different distinction from the premiss/conclusion distinction. I think it is more usual to talk about the **upper sequent** and the **lower sequent**.

You will notice that I have cheated: some of these rules allow there to be more than one formula on the right! There are various good reasons for this, but they are quite subtle and we may not get round to them. If we are to allow more than one formula on the right, then we have to think of $\Gamma \vdash \Delta$ as saying that every valuation that makes everything Γ true also makes something in Δ true. We can't correctly think of $\Gamma \vdash \Delta$ as saying that there is a proof of something in Δ using premisses in Γ because:

$$A \vdash A$$

is an initial sequent. so we can use $\neg$−R to infer

$$\vdash A, \neg A.$$

So $\vdash A, \neg A$ is an OK sequent. Now it just isn't true that there is always a proof of A or a proof of $\neg A$, so this example shows that it similarly just isn't true that a sequent can be taken to assert that there is a proof of something on the right using only premisses found on the left—unless we restrict matters so that there is only one formula on the right. This fact illustrates how allowing two formulæ on the right can be useful: the next step is to infer the sequent

$$\vdash A \vee \neg A$$

and we can't do that unless we allow two formulæ on the right.

However, it does help inculcate the good habit of thinking of sequents as metaformulæ, as things that formalise facts about formulæ rather than facts of the kind formalised by the formulæ.

One thing you will need to bear in mind, but which we have no space to prove in this course, is that sequent proofs with more than formula on the right correspond to natural deduction proofs using the rule of double negation. N.B.: commas on the left of a sequent mean 'and' while commas on the right-hand side mean 'or'! This might sound odd, but it starts to look natural quite early, and you will get used to it easily.

probably weeks 5 or 6

A summary of what we have done so far with Natural Deduction and Sequent Calculus.

- A sequent calculus proof is a log of attempts to build a natural deduction proof.
- So a sequent is telling you that there is a proof of the formula on the right using as premisses the formulæ on the left.
- But we muck things up by allowing more than one formula on the right so we have to think of a sequent as saying if everything on the left is true then something on the right is true.
- Commas on the left are **and**, commas on the right are **or**.

Display this properly

Exercise 21 *Now find sequent proofs for the formulæ in exercise 16 (page 53). For the starred formulæ you should expect to have to have two formulæ on the right at some point.*

Be sure to annotate your proofs by recording at each step which rule you are using. That makes it easier for you to check that you are constructing the proofs properly.

if sequents are week 7 then this should be week 7

2.3.1 Soundness of the Sequent Rules

If we think of a sequent $\Gamma \vdash \Delta$ as an allegation that there is a natural deduction proof of something in Δ using assumptions in Γ, then we naturally want to check that all basic sequents are true and that all the sequent rules are truth-preserving. That is to say, in each rule, if the sequent(s) above the line make true allegations about the existence of deductions, then so does the sequent below the line

To illustrate, think about the rule $\wedge$-L:

$$\frac{A, B \ \vdash \ C}{A \wedge B \ \vdash \ C} \wedge \mathrm{L} \tag{2.1}$$

It tells us we can infer "$A \wedge B \vdash C$" from "$A, B \vdash C$". Now "$A, B \vdash C$" says that there is a deduction of C from A and B. But if there is a deduction of C from A and B, then there is certainly a deduction of C from $A \wedge B$, because one can get A and B from $A \wedge B$ by two uses of $\wedge$-elim.

The $\rightarrow$-L rule can benefit from some explanation as well.

$$\frac{\Gamma \ \vdash \ \Delta, A \qquad \Gamma, B \ \vdash \ \Delta}{\Gamma, A \rightarrow B \ \vdash \ \Delta} \rightarrow \mathrm{L} \tag{2.2}$$

Assume the two sequents above the line. We want to use them to show that there is a derivation of something in Δ from $A \to B$ and all the premisses in Γ. The first sequent above the line tells us that there is either a deduction of something in Δ using premisses in Γ (in which case we are done) or there is a deduction of A. But we have $A \to B$, so we now have B. But then the second sequent above the line tells us that we can infer something in Δ.

Need exercises here

Explain "effective"

In fact it is easy to check that not only are they truth-preserving they are *effective*. Consider $\wedge$-L, for example. Assume $\Gamma, A, B \vdash \Delta$. This tells us that there is a deduction $\mathcal{D}$ of some D in Δ assuming only assumptions in Γ plus possibly A or B or both. We have several cases to consider. (i) If $\mathcal{D}$ does not use A or B then it is a witness to the truth of $\Gamma, A \wedge B \vdash \Delta$;

'witness'

(ii) If it uses either A or B (or both) then we can append[6] one (or two) applications of $\wedge$-elimination to it to obtain a new proof that is a witness to the truth of $\Gamma, A \wedge B \vdash \Delta$

finish this off, with a picture

The one exception is $\neg$-R. ($\neg$-L is OK because of *ex falso*.) If we think of the rule of $\neg$-R as telling us something about the existence blah

This illustrates how

> - sequent rules on the **right** correspond to natural-deduction **introduction** rules; and
> - sequent rules on the **left** correspond to natural-deduction **elimination** rules.

The sequent rules are all sound. Given that the sequent $\Gamma \vdash A$ arose as a way of saying that there was a proof of A using only assumptions in Γ it would be nice if we could show that the sequent rules we have are sound in the sense that we cannot use them to deduce any false allegations about the existence of proofs from true allegations about the existence of proofs. However, as we have seen, this is sabotaged by our allowing multiple formulæ on the right.

However, there is a perfectly good sense in which they are sound even if we do allow multiple formulæ on the right. If we think of the sequent $\Gamma \vdash \Delta$ as saying that every valuation making everything in Γ true makes something in Δ true then all the sequent rules are truth-preserving.

All this sounds fine. There is however a huge problem:

Insert discussion of $\neg$-R rule here

2.3.2 The rule of cut

It's not hard to check that if the two upper sequents in an application of the rule of cut make true allegations about valuations, then the allegation made by the lower sequent will be true too,

$$\frac{\Gamma \vdash \Delta, A \quad A, \Gamma' \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'}$$

[*hint*: consider the two cases: (i) A true, and (ii) A false.] Since it is truth-preserving ("sound") and we want our set of inference rules to be exhaustive

[6] The correct word is probably 'prepend'!

("complete") we will have to either adopt it as a rule or show that it is derivable from the other rules.

There is a very powerful argument for not adopting it as a rule if we can possibly avoid it: it wrecks the **subformula property**. If—without using cut—we build a sequent proof whose last line is $\vdash A$ then any formula appearing anywhere in the proof is a subformula of A. If we are allowed to use the rule of cut then, well ...

Imagine yourself in the following predicament. You are trying to prove a sequent $A \vdash B$. Now if cut is not available you have to do one of two things: you can use the rule-on-the-right for the chief connective of B, or you can use the rule-on-the-left for the chief connective of A. There are only those two possibilities. (Of course realistically there may be more than one formula on the left and there may be more than one formula on the right, so you have finitely many possibilities rather than merely two, but that's the point: the number of possibilities is finite.) If you are allowed cut then the task of proving $A \vdash B$ can spawn the two tasks of proving the two sequents

$$A \vdash B, C \quad \text{and} \quad C, A \vdash B$$

and C could be anything at all. This means that the task of finding a proof of $A \vdash B$ launches us on an infinite search. Had there been only finitely many things to check then we could have been confident that whenever there is a proof then we can be sure of finding it by searching systematically. If the search is infinite it's much less obvious that there is a systematic way of exploring all possibilities.

If we take the other horn of the dilemma we have to show that the rule of cut is unnecessary, in the sense that every sequent that can be proved with cut can be proved without it. If we have a theory T in the sequent calculus and we can show that every sequent that can be proved with cut can be proved without it then we say we have proved **cut-elimination** for T. Typically this is quite hard to do, and here is why. If we do not use cut then our proofs have the subformula property. (That was the point after all!). Now consider the empty sequent:

$$\boxed{\quad\vdash\quad}$$

The empty sequent[7] claims we can derive the empty conjunction (the thing on the right is the empty conjunction) from the empty disjunction (the thing on the left is the empty disjunction). So it claims we can derive $\bot$ from $\top$. This we certainly cannot do, so we had better not have a proof of the empty sequent! Now any cut-free proof of the empty sequent will satisfy the subformula property, and clearly there can be no proof of the empty sequent satisfying the subformula

[7] I've put it into a box, so that what you see—in the box—is not just a turnstile with nothing either side of it but the empty sequent, which is not the same thing at all ... being (of course) a turnstile with nothing either side of it.

property. Therefore, if we manage to show that every sequent provable in the sequent version of T has a cut-free proof then we have shown that there is no proof of the empty sequent in T. But then this says that there is no proof of a contradiction from T: in other words, T is consistent.

So: proving that we can eliminate cuts from proofs in T is as hard as showing that T is free from contradiction. As it happens there is no contradiction to be derived from the axioms we have for predicate calculus but proving this is quite hard work. We can prove that all cuts can be eliminated from sequent proofs in predicate calculus but I am not going to attempt to do it here.

2.3.3 Two tips

2.3.3.1 Keep a copy!!

One of things to bear in mind is that one can always *keep a copy* of the eigenformula. What do I mean by this? Well, suppose you are challenged to find a proof of the sequent

$$\Gamma \vdash A \to B \tag{2.1}$$

You could attack a formula in Γ but one thing you can do is attack the formula on the right, thereby giving yourself the subordinate goal of proving the sequent

$$\Gamma, A \vdash B \tag{2.2}$$

However, you could also generate the goal of proving the sequent

$$\Gamma, A \vdash B, A \to B \tag{2.3}$$

The point is that if you do a $\to$-R to sequent 2.3.3 you get sequent 2.3.3. Thus you get the same result as if you had done a $\in$-R to sequent 2.3.3. Sometimes *keeping a copy* of the eigenformula in this way is the only way of finding a proof.

Fainthearts may skip this section: it won't be examined

For example, there is a proof of the sequent

$$(A \to B) \to B \vdash (B \to A) \to A$$

but you have to keep copies of eigenformulæ to find it. That's a hard one!

In both these illustrations the extra copy you are keeping is a copy on the right. I should try to find an illustration where you need to keep a copy on the left too.

numbering not working properly

Exercise 22 *Find a proof of the sequent:*

$$(A \to B) \to B \vdash (B \to A) \to A$$

Another reason why keeping copies can be useful. You might be wondering why the $\vee$-R rule is not of the form

$$\frac{\Gamma \vdash A, B}{\Gamma \vdash A \vee B}$$

The answer is we can justify that as a derived rule by the following inference:

$$\cfrac{\cfrac{\cfrac{\Gamma \ \vdash \ A, B}{\Gamma \ \vdash \ A \vee B, B}\ \vee\text{ R}}{\Gamma \ \vdash \ A \vee B, A \vee B}\ \vee\text{ R}}{\Gamma \ \vdash \ A \vee B}\ \text{contraction-R} \tag{2.4}$$

...keeping an extra copy of '$A \vee B$'

2.3.3.2 Keep checking your subgoals for validity

It sounds obvious, but when you are trying to find a sequent proof by working upwards from your goal sequent, you should check at each stage that the goal-sequents you generate in this way really are valid in the sense of making true claims about valuations. After all, if the subgoal you generate doesn't follow from the assumptions in play at that point then you haven't a snowflake in hell's chance of proving it, have you? It's usually easy to check by hand that if everything on the left is true then something on the right must be true.

As I say, it sounds obvious but lots of people overlook it!

And don't start wondering: "if it's that easy to check the validity of a sequent, why do we need sequent proofs?". The point is that one can use the sequent gadgetry in more complex settings where simple tautology-checking of this kind is not available. See section 2.5 p 70.

2.3.4 Exercises

Which week?

You can now attempt to find sequent proofs for all the formulæ in exercise 16 page 53.

week 7 or 8

If you are a first-year UEA student who is not interested in pursuing Logic any further you can skip the rest of this chapter and go straight to chapter 3.

2.4 Interpolation

You do not need to know the proof of the theorem in this section, but I want you to be able to state it correctly, and to understand why it matters.

By now the reader will have had some experience of constructing natural deduction proofs. If they examine their own practice they will notice that if they are trying to prove a formula that has, say, the letters 'p', 'q' and 'r' in it, they will never try to construct a proof that involves letters other than 'p', 'q' and 'r'. There is a very strong intuition of irrelevance at work here. It's so strong and so natural that you probably didn't notice even that you were using it. The time has now come to discuss it. But we need a bit more gadgetry first.

The following puzzle comes from Lewis Carroll.

> Dix, Lang, Cole, Barry and Mill are five friends who dine together regularly. They agree on the following rules about which of the two condiments—salt and mustard—they are to have with their beef. Each of them has precisely one condiment with their beef. Carroll tells us:
>
> 1. If Barry takes salt, then either Cole or Lang takes only *one* of the two condiments, salt and mustard. If he takes mustard then either Dix takes neither condiment or Mill takes both.
> 2. If Cole takes salt, then either Barry takes only *one* condiment, or Mill takes neither. If he takes mustard then either Dix or Lang takes both.
> 3. If Dix takes salt, then either Barry takes neither condiment or Cole takes both. If he takes mustard then either Lang or Mill takes neither.
> 4. If Lang takes salt, then either Barry or Dix takes only *one* condiment. If he takes mustard then either Cole or Mill takes neither.
> 5. If Mill takes salt, then either Barry or Lang takes both condiments. If he takes mustard then either Cole or Dix takes only one.

As I say, this puzzle comes from Lewis Carroll. The task he sets is to ascertain whether or not these conditions can in fact be met. I do not know the answer, and it would involve a lot of hand-calculation—which of course is the point! However I am using it here to illustrate a different point.

Let's consider the first item:

"If Barry takes salt, then either Cole or Lang takes only *one* of the two condiments, salt and mustard. If he takes mustard then either Dix takes neither condiment or Mill takes both."

If Barry takes salt then either Cole or Lang takes only *one* of the two condiments, salt and mustard.

If Barry does not take salt then either Dix takes neither condiment or Mill takes both.

Now we do know that either Barry takes salt or he doesn't, so we are either in the situation where Barry takes salt (in which case either Cole or Lang takes only *one* of the two condiments, salt and mustard) or we are in the situation where Barry does not take salt (in which case either Dix takes neither condiment or Mill takes both).

This illustrates a kind of splitting principle. If we have some complex combination of information, wrapped up in a formula A, say, and p is some atomic piece of information (a propositional letter) in A, then we can *split on* p as it were, by saying to ourselves:

> "Either p holds—in which case we can simplify A to A' (which 'p' doesn't appear in) or p does not hold—in which case A simplifies to something different, call it A' in which, again 'p' does not appear.
>
> So A is equivalent to $(p \wedge A') \vee (\neg p \wedge A'')$, where '$p$' does not appear in A' or in A''"

How do we obtain A' and A'' from A? A' is what happens when p is true, so just replace all occurrences of 'p' in A by '$\top$'. By the same token, replace all occurences of 'p' in A by '$\bot$' to get A''. That's sort-of all right, but it would be nice to get rid of the '$\bot$'s and the '$\top$'s as well to make things simpler. We saw in exercise 6 that

$p \vee \top$ is logically equivalent to $\top$
$p \vee \bot$ is logically equivalent to p
$p \wedge \top$ is logically equivalent to p
$p \wedge \bot$ is logically equivalent to $\bot$

and in exercise 9 that

$p \to \top$ is logically equivalent to $\top$
$\top \to p$ is logically equivalent to p
$\bot \to p$ is logically equivalent to $\top$
$p \to \bot$ is logically equivalent to $\neg p$

We can use these equivalences to simplify complex expressions and get rid of all the '$\top$'s and '$\bot$'s.

Let's have some illustrations:

- $p \to (q \vee r)$ There are two cases to consider.
    1. The case where p is true. Then we infer $q \vee r$. So in this case we get $p \wedge (q \vee r)$.
    2. The case where p is false. In this case the $p \to (q \vee r)$ that we started with tells us nothing, so all we get is $\neg p$.
- $(p \vee q) \to (r \wedge s)$
    1. In the case where p is true this becomes

    $$(\top \vee q) \to (r \vee s)$$

    and $\top \vee q$ is just $\top$ so

    $$(p \vee q) \to (r \wedge s)$$

becomes

$$\top \to (r \wedge s)$$

which is just

$$r \wedge s.$$

So we get

$$p \wedge (r \wedge s).$$

2. In the case where p is false this becomes

$$(\bot \vee q) \to (r \wedge s)$$

and $\bot \vee q$ is just q so we get

$$q \to (r \wedge s)$$

and

$$\neg p \wedge (q \to (r \wedge s))$$

So $(p \vee q) \to (r \wedge s)$ is equivalent to

$$(p \wedge (r \wedge s)) \vee (\neg p \wedge (q \to (r \wedge s)))$$

By this means we can prove:

Theorem 10 *The Splitting Principle.*
Suppose A is a propositional formula and 'p' is a letter appearing in A. There are formulæ A_1 and A_2 not containing 'p' such that A is logically equivalent to $(A_1 \wedge p) \vee (A_2 \wedge \neg p)$.

Definition 11 *Let $\mathcal{L}(P)$ be the set of propositional formulæ that can be built up from the propositional letters in the alphabet P.*

Let us overload this notation by letting $\mathcal{L}(A)$ be the set of propositional formulæ that can be built up from the propositional letters in the formula A.

Suppose $A \to B$ is a tautology, but A and B have no letters in common. What can we say? Well, there is no valuation making A true and B false. But, since valuations of A and B can be done independently, it means that either there is no valuation making A true, or there is no valuation making B false. With a view to prompt generalisation, we can tell ourselves that, despite A and B having no letters in common, $\mathcal{L}(A)$ and $\mathcal{L}(B)$ are *not* disjoint because $\top$ is the conjunction of the empty set of formulæ and $\bot$ is the disjunction of the empty set of formulæ and therefore both '$\top$' and '$\bot$' belong to the language over the empty alphabet—which is to say to $\mathcal{L}(A) \cap \mathcal{L}(B)$. We established that either $A \to \bot$ is a tautology (so A is the negation of a tautology) or $\top \to B$ is a tautology (so B is a tautology). But, since $A \to \top$ and $\bot \to B$ are always tautologies (as we saw in exercise 9) we can tell ourselves that what we have established is that there is some formula C in the common vocabulary (which must be either '$\top$' or '$\bot$') such that both $A \to C$ and $C \to B$ are tautologies.

If we now think about how to do this "with parameters" we get a rather more substantial result.

Insert Chapter break here . . . ?

THEOREM 12 *(The interpolation lemma)*

Let A and B be two expressions such that we can deduce B from A. (Every valuation making A true makes B true). Then we can find an expression C containing only those propositional letters common to A and B such that we can deduce C from A, and we can deduce B from C.

Proof: We have seen how to do this in the case where A and B have no letters in common. Now suppose we can do it when A and B have n letters in common, and deduce that we can do it when they have $n+1$ letters in common. Suppose 'p' is a letter they have in common. The we can split A and B at p to get

$(p \wedge A') \vee (\neg p \wedge A'')$ which is equivalent to A

and

$(p \wedge B') \vee (\neg p \wedge B'')$ which is equivalent to B

So any valuation making $(p \wedge A') \vee (\neg p \wedge A'')$ true must make $(p \wedge B') \vee (\neg p \wedge B'')$ true. So that means that any valuation making $(p \wedge A')$ true must make $(p \wedge B')$ true and any valuation making $(\neg p \wedge A'')$ true must make $(\neg p \wedge B'')$ true. Indeed any valuation making A' true must make B' true, and any valuation making A'' true must make B'' true. But observe that A' and B' have only n propositional letters in common so we can find C' containing only those letters they have in common, such that every valuation making A' true makes C' true and evey valuation making C' true makes B' true, and similarly A'' and B'' have only n propositional letters in common so we can find C'' containing only those letters they have in common, such that every valuation making A'' true makes C'' true and evey valuation making C'' true makes B' true. So the interpolant we want is

$$(p \wedge C') \vee (\neg p \wedge C'')$$

■

EXERCISE 23 *Find an interpolant Q for*

$$(A \wedge B) \vee (\neg A \wedge C) \quad \vdash \quad (B \to C) \to (D \to C)$$

and supply proofs (in whatever style you prefer) of

$$(A \wedge B) \vee (\neg A \wedge C) \quad \to \quad Q$$

and

$$Q \quad \to \quad ((B \to C) \to (D \to C))$$

"with parameters"?

Say something about interpolation equiv to completeness but much more appealing: humans have strong intuitions of irrelevance from having to defend ourselves from conmen over many generations

2.5 Why do we need proof systems for propositional Logic. . . ??

. . . given that we can check the validity of any inference by means of truth-tables?? You may well be asking this question.

There are several reasons, but the general theme common to them is that there are more complex kinds of Logic where there is no handy and simple analogue of the truth-table gadgetry. This being the case we need more complicated gadgets, and the process of mastering those gadgets is greatly helped by practising on propositional calculus in the first instance—using the toy versions of the gadgets in question.

In particular we will be studying predicate calculus, where truth-tables are not sensitive enough for us to be able to use them and them alone for checking the validity of inferences.

2.6 Some advanced exercises for enthusiasts

Life on Planet Zarg taught us that Peirce's law does not follow from K and S alone: we seem to need the rule of double negation. In fact Peirce's law, in conjunction with K and S, implies all the formulæ built up only from $\to$ that we can prove using the rule of double negation.

Exercise 24 *We saw in exercise 6 page 18 part (8) that $(P \to Q) \to Q$ has the same truth-table as $P \vee Q$.*

Construct a natural deduction proof of R from the premisses $(P \to Q) \to Q$, $P \to R$ and $Q \to R$. You may additionally use as many instances of Peirce's law as you wish.[8]

[8] I am endebted to Tim Smiley for this amusing fact.

Chapter 3

Predicate (first-order) Logic

3.1 Towards First-Order Logic

We saw earlier the following puzzle from Lewis Carroll (page 66).

1. If Barry takes salt, then either Cole or Lang takes only *one* of the two condiments, salt and mustard. If he takes mustard then either Dix takes neither condiment or Mill takes both.

2. If Cole takes salt, then either Barry takes only *one* condiment, or Mill takes neither. If he takes mustard then either Dix or Lang takes both.

3. If Dix takes salt, then either Barry takes neither condiment or Cole takes both. If he takes mustard then either Lang or Mill takes neither.

4. If Lang takes salt, then either Barry or Dix takes only *one* condiment. If he takes mustard then either Cole or Mill takes neither.

5. If Mill takes salt, then either Barry or Lang takes both condiments. If he takes mustard then either Cole or Dix takes only one.

As I say, this puzzle comes from Lewis Carroll. The task he sets is to ascertain whether or not these conditions can in fact be met. I do not know the answer, and it would involve a lot of hand-calculation—which of course is the point! I don't suppose for a moment that you want to crunch it out (I haven't done it and I have no intention of doing it—I have a life) but it's a good idea to think a bit about some of the preparatory work.

The way to do this would be to create a number of propositional letters, one each to abbreviate each of the assorted assertions "Barry takes salt", "Mill takes mustard" and so on. How many propositional letters will there be? Obviously 10, co's you can count them: each propositional letter corresponds to a choice of one of {Dix, Lang, Cole, Barry, Mill}, and one choice of {salt, mustard} and $2 \times 5 = 10$. We could use propositional letters 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x'

and 'y'. But notice that using ten different letters—*mere* letters—in this way fails to capture certain relations that hold between them. Suppose they were arranged like:

'p': Barry takes salt	'u': Barry takes mustard
'q': Mill takes salt	'v': Mill takes mustard
'r': Cole takes salt	'w': Cole takes mustard
's': Lang takes salt	'x': Lang takes mustard
't': Dix takes salt	'y': Dix takes mustard

Then we see that two things in the same row are related to each other in a way that they aren't related to things in other rows; ditto things in the same column. This subtle information cannot be read off just from the letters 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x' and 'y' themselves. That is to say, there is *internal structure* to the propositions "Mill takes salt" etc, that is not captured by reducing each one to a single letter. The time has come to do something about this.

A first step would be to replace all of 'p', 'q', 'r', 's', 't', 'u', 'v', 'w', 'x' and 'y' by things like 'ds' and 'bm' which will mean 'Dix takes salt' and 'Barry takes mustard'. (Observe that 'ds' is a *single* character.) Then we can build truth-tables and do other kinds of hand-calculation as before, this time with the aid of a few mnemonics. If we do this, the new things like 'bm' are really just propositional letters as before, but slightly bigger ones. The internal structure is visible to *us*—*we* know that 'ds' is really short for 'Dix takes salt'—but it is not visible to the logic. The logic regards 'ds' as a single propositional letter. To do this satisfactorily we must do it in a way that makes the internal structure explicit.

What we need is **Predicate Logic**. It's also called **First-Order Logic** and sometimes **Predicate Calculus**. In this new pastime we don't just use suggestive mnemonic symbols for propositional letters but we open up the old propositional letters that we had, and find that they have internal structure. "Romeo loves Juliet" will be represented not by a single letter 'p' but by something with suggestive internal structure like $L(r, j)$. We use capital Roman letters as **predicate** symbols (also known as **relation** symbols). In this case the letter 'L' is a *binary* relation symbol, co's it relates *two* things. The 'r' and the 'j' are **arguments** to the relation symbol. They are **constants** that denote the things that are related by the (meaning of the) relation symbol.

The obvious way to apply this to Lewis Carroll's problem on page 66 is to have a two-place predicate letter 'T', and symbols 'd', 'l', 'm', 'b' and 'c' for Dix, Lang, Mill, Barry and Cole, respectively. I am going to write them in lower case beco's we keep upper case letters for predicates—relation symbols. And we'd better have two constants for the condiments salt and mustard: 's' for salt and—oops!—can't use 'm' for mustard co's we've already used that letter for Mill! Let's use 'u'. So, instead of 'p' and 'q' or even 'ds' etc we have:

'$T(b, s)$': Barry takes salt	'$T(b, u)$': Barry takes mustard

'$T(m,s)$': Mill takes salt	'$T(m,u)$': Mill takes mustard
'$T(c,s)$': Cole takes salt	'$T(c,u)$': Cole takes mustard
'$T(l,s)$': Lang takes salt	'$T(l,u)$': Lang takes mustard
'$T(d,s)$': Dix takes salt	'$T(d,u)$': Dix takes mustard

And now the symbolism we are using makes it clear what it is that two things in the same row have in common, and what it is that two things in the same column have in common. I have used here a convention that you always write the relation symbol first, and then put its arguments after it, enclosed within brackets. We don't write 'mTs'. Tho' we do write "Hesperus = Phosphorous" (the two ancient names for the evening star and the morning star) and when we write the relation symbol between its two arguments we say we are using **infix** notation. (Infix notation only makes sense if you have two arguments not three: If you had three aguments where would you put the relation symbol if not at the front?)

What you should do now is look at the question on page 29, the one concerning Herbert's love life, pets and accommodation arrangements.

> If Herbert can take the flat only if he divorces his wife then he should think twice. If Herbert keeps Fido, then he cannot take the flat. Herbert's wife insists on keeping Fido. If Herbert does not keep Fido then he will divorce his wife—at least if she insists on keeping Fido.

You will need constant names 'h' for Herbert, 'f' for Fido, and 'w' for the wife. You will also need a few binary relation symbols: K for *k*eeps, as in "Herbert keeps Fido". Some things might leave you undecided. Do you want to have a binary relation symbol 'T' for *t*akes, as in $T(h,f)$ meaning "Herbert takes the flat"? If you do you will need a constant symbol 'f' to denote the flat. Or would you rather go for a unary relation symbol 'TF' to be applied to Herbert? No-one else is conjectured to take the flat after all ... If you are undecided between these, all it means is that you have discovered the wonderful flexibility of predicate calculus.

Rule of thumb: We use Capital Letters for *properties* and *relations*; on the whole we use small letters for *things*. (We do tend to use small letters for functions too). The capital letters are called **relational symbols** or **predicate letters** and the lower case letters are called **constants**.

Exercise 25 *Formalise the following, using a lexicon of your choice*

1. *Romeo loves Juliet; Juliet loves Romeo.*
2. *Balbus loves Julia. Julia does not love Balbus. What a pity.* [1]
3. *Fido sits on the sofa; Herbert sits on the chair.*

[1] I found this in a latin primer: *Balbus amat Juliam; Julia not amat Balbum ....*

4. *Fido sits on Herbert.*

5. *If Fido sits on Herbert and Herbert is sitting on the chair then Fido is sitting on the chair.*

6. *The sofa sits on Herbert.* [just because something is absurd doesn't mean it can't be said!]

7. *Alfred drinks more whisky than Herbert; Herbert drinks more whisky than Mary.*

8. *John scratches Mary's back. Mary scratches her own back.*

   [A binary relation can hold between a thing and itself. It doesn't have to relate two distinct things.]

3.1.1 The Syntax of First-order Logic

All the apparatus for constructing formulæ in propositional logic works too in this new context: If A and B are formulæ so are $A \vee B$, $A \wedge B$, $\neg A$ and so on. However we now have new ways of creating formulæ, new gadgets which we had better spell out:

Constants and variables

Constants tend to be lower-case letters at the start of the Roman alphabet ('a', 'b' ...) and variables tend to be lower-case letters at the end of the alphabet ('x', 'y', 'z' ...). Since we tend to run out of letters we often enrich them with subscripts to obtain a larger supply: 'x_1' etc.

Predicate letters

These are upper-case letters from the Roman alphabet, usually from the early part: 'F' 'G' .... They are called *predicate* letters because they arise from a programme of formalising reasoning about predicates and predication. '$F(x, y)$' could have arisen from 'x is fighting y'. Each predicate letter has a particular number of terms that it expects; this is the **arity** of the letter. 'loves' has arity 2 (it is binary) 'sits-on' is binary too. If we feed it the correct number of terms—so we have an expression like $F(x, y)$—we call the result an **atomic formula.**

The equality symbol '$=$' is a very special predicate letter: you are not allowed to reinterpret it the way you can reinterpret other predicate letters. (The Information Technology fraternity say of strings that cannot be assigned meanings by the user that they are **reserved**). It is said to be **part of the logical vocabulary**. The equality symbol '$=$' is the only relation symbol that is reserved. In this respect it behaves like '$\wedge$' and '$\forall$' and the connectives, all of which are reserved in this sense.

Unary predicates have one argument, **binary** predicates have two; n-**ary** have n. Similarly functions.

Atomic formulæ can be treated the way we treated literals in propositional logic: we can combine them together by using '$\wedge$' '$\vee$' and the other connectives.

lots of illustrations here please

Finally we can bind variables with **quantifiers**. There are two: $\exists$ and $\forall$. We can write things like

$$(\forall x)F(x)$$

everything is a frog;

$$(\forall x)(\forall y)L(x, y)$$

everybody loves everyone

we might write this second thing as

$$(\forall xy)L(x, y)$$

to save space

The syntax for quantifiers is variable-preceded-by quantifier enclosed in brackets, followed by stuff inside brackets:

$$(\exists x)(\ldots) \text{ and } (\forall y)(\ldots).$$

We sometimes omit the pair of brackets to the right of the quantifier when no ambiguity is caused thereby.

The difference between variables and constants is that you can bind variables with quantifiers, but you can't bind constants. The meaning of a constant is fixed.

...free

complete this explanation; quantifiers are connectives too

For example, in a formula like

$$(\forall x)(F(x) \to G(x))$$

the letter 'x' is a variable: you can tell because it is bound by the universal quantifier. The letter 'F' is not a variable, but a predicate letter. It is not bound by a quantifier, and cannot be: the syntax forbids it. In a first-order language you are not allowed to treat predicate letters as variables: you may not bind them with quantifiers. Binding predicate letters with quantifiers (treating them as variables) is the tell-tale sign of **second-order** Logic.

We also have

Function letters

These are lower-case Roman letters, typically 'f', 'g', 'h' .... We apply them to variables and constants, and this gives us **terms**: $f(x)$, $g(a, y)$ and suchlike. In fact we can even apply them to terms: $f(g(a, y))$, $g(f(g(a, y), x))$ and so on. So a term is either a variable or a constant or something built up from variables-and-constants by means of function letters. What is a function? That is, what sort of thing do we try to capture with function letters? We have seen an example: *father-of* is a function: you have precisely one father; *son-of* is not a function. Some people have more than one, or even none at all.

3.1.2 Warning: Scope ambiguities

Better examples are "All that glisters is not gold"; "all is not lost"...

Years ago when I was about ten a friend of my parents produced a German quotation, and got it wrong. (I was a horrid child, and I blush to recall the episode). I corrected him, and he snapped "All right, everybody isn't the son of a German Professor") (My father was Professor of German at University College London at the time). Quick as a flash I replied "What you mean is 'Not everybody is the son of a professor of German'.".

I was quite right. (Let's overlook the *German professor/professor of German* bit). He said that Everybody Isn't the son of a professor of German. That's not true. Plenty of people are; I am, for one. What he meant was "Not everybody is ...". It's the difference between "$(\forall x)(\neg \ldots)$" and "$\neg(\forall x)(\ldots)$"—the difference is real, and it matters.

The difference is called a matter of **scope**. 'Scope'? The point is that in "$(\forall x)(\neg \ldots)$" the "scope" of the '$\forall x$' is the whole formula whereas in the '$\neg(\forall x)(\ldots)$ it isn't.

For you, the moral of this story is that you have to identify with the annoying ten-year old rather than with the adult that he annoyed: it's the annoying 10-year-old that is your rôle model here!

It is a curious fact that humans using ordinary language can be very casual about getting the bits of the sentence they are constructing in the right order so that each bit has the right scope. We often say things that we don't literally mean. ("Everybody isn't the son of ..." when we mean "Not everybody is ...") On the receiving end, when trying to read things like $(\forall x)(\exists y)(x$ loves $y)$ and $(\exists y)(\forall x)(x$ loves $y)$, people often get into tangles because they try to resolve their uncertainty about the scope of the quantifiers by looking at the overall meaning of the sentence rather than by just checking to see which order they are in!

3.2 Exercises

probably week 8

Exercise 26

This batch of exercises may take you several hours, and you may wish to spread it over two weeks not one. Fainthearts may omit the starred items.

In the first of these exercises group the argument (tokens) that you see into argument types.

Guidance: This first bunch involve monadic predicates only and no nested quantifiers. Do not be distracted by the observation that some of these arguments are invalid; what use would Logic be if we could formalise valid arguments only? We'd never be able to use Logic to tell the difference!

1. *Every good boy deserves favour; George is a good boy. Therefore George deserves favour.*

2. *All cows eat grass; Daisy eats grass. Therefore Daisy is a cow.*

3. *Socrates is a man; all men are mortal. Therefore Socrates is mortal.*
4. *Daisy is a cow; all cows eat grass. Therefore Daisy eats grass.*
5. *Daisy is a cow; all cows are mad. Therefore Daisy is mad.*
6. *No thieves are honest; some dishonest people are found out. Therefore Some thieves are found out.*
7. *No muffins are wholesome; all puffy food is unwholesome. Therefore all muffins are puffy.*
8. *No birds except peacocks are proud of their tails; some birds that are proud of their tails cannot sing. Therefore some peacocks cannot sing.*
9. *No fossil can be crossed in love; an oyster may be crossed in love. Therefore oysters are not fossils.*
10. *All who are anxious to learn work hard; some of these students work hard. Therefore some of these students are anxious to learn.*
11. *Some lessons are difficult; what is difficult needs attention. Therefore some lessons need attention.*
12. *All humans are mammals; all mammals are warm blooded. Therefore all humans are warm-blooded.*
13. *Louis is King of France; all Kings of France are bald. Therefore Louis is bald.*

EXERCISE 27
Now render all the arguments in exercise 27 into predicate calculus, using a lexicon of your choice.

EXERCISE 28 *Match up the formulæ on the left with their English equivalents on the right.*

(i) $(\forall x)(\exists y)(x \text{ loves } y)$	*(a) Everyone loves someone*
(ii) $(\forall y)(\exists x)(x \text{ loves } y)$	*(b) There is someone everyone loves*
(iii) $(\exists y)(\forall x)(x \text{ loves } y)$	*(c) There is someone that loves everyone*
(iv) $(\exists x)(\forall y)(x \text{ loves } y)$	*(d) Everyone is loved by someone*

EXERCISE 29 *Given the following lexicon*

$S(x)$*: x is a student.*
$L(x)$*: x is a lecturer.*
$C(x)$*: x is a course.*
$T(x, y, z)$*: (lecturer) x lectures (student) y for (course) z.*
$A(x, y)$*: (student) x attends (course) z.*
$F(x, y)$*: x and y are friends.*
$Z(x)$*: x lives in a ziggurat.*
$M(x)$*: x has measles.*

*Turn the following into English. (***normal** *English: no xs and ys.)*

1. $(\forall x)(F(\texttt{Kim}, x) \to F(\texttt{Alex}, x))$
2. $(\forall x)(\exists y)(F(x, y) \wedge M(y) \wedge Z(y))$
3. $(\forall x)(F(\texttt{Kim}, x) \to Z(x))$
4. $(\forall x)((Z(x) \wedge M(x)) \to F(\texttt{Kim}, x))$
5. $(\forall x)(Z(x) \to (\exists y)(F(x, y) \wedge M(y)))$
6. $(\forall x)(S(x) \to (\exists yz)(T(y, x, z))$
7. $(\exists x)(S(x) \wedge (\forall z)(\neg A(x, z))$
8. $(\exists x)(C(x) \wedge (\forall z)(\neg A(z, x))$
9. $(\exists x)(L(x) \wedge (\forall yz)(\neg T(x, y, z))$
10. $(\forall x_1 x_2)[(\forall z)(A(x_1, z) \longleftrightarrow A(x_2, z)) \to x_1 = x_2]$
11. $(\forall x_1 x_2)[(\forall z)(A(z, x_1) \longleftrightarrow A(z, x_2)) \to x_1 = x_2]$
12. $(\forall y)(\exists x, z)(T(x, y, z)) \to (\forall u, v)(\neg T(y, u, v))]$
13. $(\forall xy)(x \neq y \to (\exists z)(F(z, x) \longleftrightarrow \neg F(z, y)))$

EXERCISE 30 *What are the principal connectives of the following formulæ of first-order logic? (Don't forget that quantifiers are connectives too!)*

$(\exists y)(F(y) \vee P)$
$(\exists y)(F(y)) \vee P$
$A \to (\forall x)(B(x))$

EXERCISE 31 *Render the following into Predicate calculus, using a lexicon of your choice. (These involve nestings of more than one quantifier, polyadic predicate letters, equality and even function letters. The flag "(=)" in the ight-hand margin opposite an item means you will need the equality symbol in you answer)*

1. *Anyone who has forgiven even one person is a saint.*
2. *Nobody in the logic class is cleverer than everybody in the history class.*
3. *Jane saw a bear, and Roger saw one too.*
4. *Jane saw a bear and Roger saw it too.*
5. *None but the brave deserve the fair.*
6. *Some students are not taught by every teacher;*
7. *No student has the same teacher for every subject.*

8. *Everone who loves is loved;*
9. *Everyone loves a lover;*
10. *You are loved only if you yourself love someone;*
11. *The enemy of an enemy is a friend;*
12. *The friend of an enemy is an enemy;*
13. *Any friend of George's is a friend of Alex's;*
14. *Jack and Jill have at least two friends in common;*
15. *Two people who love the same person do not love each other;*
16. *If there is anyone in the residences with measles then anyone who has a friend in the residences will need a measles jab;*
17. *Everyone likes Mary—except Mary herself;*[2] (=)
18. *Everybody loves My Baby, but My Baby loves nobody but me.* (=)
19. *There is at most one king of France;* (=)
20. *Alex knows (at least) two pop stars;* [3] (=)
21. *Hilary knows no more than two pop stars;* (=)
22. *There is precisely one king of France;* (=)
23. *If there is anyone in the residences with measles then anyone who has a friend in the residences will need a measles jab.*
24. *Brothers and sisters have I none; this man's father is my father's son;*
25. *Anyone who is between a rock and a hard place is also between a hard place and a rock.*
26. *There are two islands in New Zealand;* (=)
27. *There are* three[4] *islands in New Zealand;* (=)
28. *Jocelyn knows three FRS's and one of them is bald;* (=)
29. *You are loved only if you yourself love someone [other than yourself!]* (=)

[2]Beware: do not attempt to formalise this as the conjunction of 'Everybody likes Mary' and 'Mary doesn't like Mary': that would be a contradiction; in contrast "Everyone likes Mary—except Mary herself" is perfectly consistent!

[3]You must resist the temptation to express this as a relation between Alex and a plural object consisting of two pop stars coalesced into a kind of plural object like Jeff Goldblum and the Fly. You will need to use '=', the symbol for equality.

[4]The third is Stewart Island.

3.2.1 Russell's Theory of Descriptions

'There is precisely one King of France and he is bald' can be captured satisfactorily in predicate calculus/first-order logic by anyone who has done the preceding exercises. We get

$$(\exists x)(K(x) \wedge (\forall y)(K(y) \to y = x) \wedge B(x)) \qquad \text{(A)}$$

Is the formulation we arrive at the same as what we would get if we were to try to capture (B)?

"The King of France is bald" (B)

Well, if (A) holds then the unique thing that is King of France and is bald certainly sounds as if it is going to be *the* King of France, and it is bald, and so if (A) is true then the King of France is bald. What about the converse (or rather its contrapositive)? If (A) is false, must it be false that the King of France is bald? It might be that (A) is false because there is more than one King of France. In those circumstances one might want to suspend judgement on (B) on the grounds that we don't yet know which of the two prospective Kings of France is the real one, and one of them might be bald. Indeed they might *both* be bald. Or we might simply feel that we can't properly use expressions like "the King of France" unless we know that there is precisely one. If there isn't precisely one then allegations about the King of France simply lack truth-value—or so it is felt.

What's going on here is that we are trying to add to our language a new quantifier, a thing like '$\forall$' or '$\exists$'—which we could write '$(Qx)(\ldots)$ so that '$(Qx)(F(x))$' is true precisely when the King of France has the property F. The question is: can we translate expressions containing this new quantifier into expressions that do not contain it? The answer depends on what truth-value you attribute to (B) when there is no King of France. If you think that (B) is false under those circumstances then you may well be willing to accept (A) as a translation of it, but you won't if you think that (B) lacks truth-value.

If you think that (A) is the correct formalisation of (B), and that in general you analyse "The F is G" as

$$(\exists x)((F(x) \wedge (\forall y)(F(y) \to y = x) \wedge G(x))) \qquad \text{(C)}$$

then you are a subscriber to **Russell's theory of descriptions.**

3.3 First-order and second-order

We need to be clear right from the outset about the difference between first-order and second-order. In first-order languages predicate letters and function letters cannot be variables. The idea is that the variables range only over individual inhabitants of the structures we consider, not over sets of them or properties of them. This idea—put like that—is clearly a semantic idea. However it can be (and must be!) given a purely syntactic description.

In propositional logic every wellformed expression is something which will evaluate to a truth-value: to `true` or to `false`. These things are called **booleans** so we say that every wellformed formula of propositional logic is of type `bool`. Explain this idiom

In first order logic it is as if we have looked inside the propositional letters 'p', 'q' etc that were the things that evaluate to `true` or to `false`, and have discovered that the letter—as it might be—'A' actually, on closer inspection, turned out to be '$F(x, y)$'. To know the truth-value of this formula we have to know what objects the variables 'x' and 'y' point to, and what binary relation the letter 'F' represents.

First-order logic extends propositional logic in another way too. Blah quantifiers.

3.3.1 Higher-order vs Many-Sorted

Predicate modifiers

A predicate modifier is a second-order function letter. They are sometimes called *adverbial* modifiers. For example we might have a predicate modifier $\mathcal{V}$ whose intended meaning is something like "a lot" or "very much', so that if $L(x, y)$ was our formalisation of x *loves* y then $\mathcal{V}(L(x, y))$ means x *loves* y *very much*. In the old grammar books I had at school we were taught that adjectives had three forms: *simple* ("cool") *comparative* ("cooler") and *superlative* ("coolest"). These could be represented in higher order logic by two predicate modifiers. Another predicate modifier is *too*.

> No woman is too thin or too rich.

We will not consider them further.

Many-sorted

If you think the universe consists of only one kind of stuff then you will have only one domain of stuff for your variables to range over. If you think the universe has two kinds of stuff (for example, you might think that there are two kinds of stuff: the mental and the physical) then you might want two domains for your variables to range over. If you are a cartesian dualist trying to formulate a theory of mind in first-order logic you would want to have variables of two *sorts*: for mental and for physical entities. If you are a geometer you might want to have one sort of variable for points and another sort of variable for lines.

3.3.2 First-person and third-person

Natural languages have these wonderful gadgets like 'I' and 'you'. These connect the denotation of the expressions in the language to the *users* of the language. This has the effect that if A is a formula that contains one of these pronouns then different tokens of A will have different meanings! This is completely unheard-of in the languages of formal logic: it's formula *types* that the semantics gives meanings to, not formula-*tokens*. Another difference between formal languages and natural languages is that the users of formal languages (us!) do not belong to the world described by the expressions in those languages. (Or at least if we do then the semantics has no way of expressing this fact.) Formal languages do have *variables*, and variables function grammatically like pronouns, but the pronouns they resemble are *third person* pronouns not first- or second-person pronouns. This is connected with their use in science: no first- or second-person perspective in science. This is because science is agent/observer-invariant. Connected to *objectivity*. The languages that people use/discuss in Formal Logic do not deal in any way with speech acts/formula tokens: only with the types of which they are tokens.

3.4 Validity

Once you've tangled with a few syllogisms you will be able to recognise which of them are good and which aren't. 'Good'? A syllogism (or any kind of argument in this language, not just syllogisms) is **valid** if the truth of the conclusion follows from the truth of the premisses simply by virtue of the logical structure of the argument. Recall the definition of valid argument from propositional logic. You are a valid argument if you are a token of an argument type such that every token of that type with true premisses has a true conclusion. We have exactly the same definition here! The only difference is that we now have a slightly more refined concept of argument type.

We can use the expressive resources of the new language to detect that

Socrates is human
All humans are mortal
Socrates is mortal

. . . is an argument of the same type as

Daisy is a cow
all cows are mad
Daisy is mad

Both of these are of the form:

$$\frac{M(s) \qquad (\forall x)(M(x) \to C(x))}{C(s)}$$

We've changed the letters but that doesn't matter. The overall *shape* of the two formulæ is the same, and it's the shape that matters.

The difference between the situation we were in with propositional logic and the situation we are in here is that we don't have a simple device for testing validity the way we had with propositional logic. There we had truth tables. To test whether an argument in propositional logic is valid you form the condition whose antecedent is the conjunction of the premisses of the argument and whose consequent is the conclusion. The argument is valid iff the conditional is a tautology, and you write out a truth-table to test whether or not the conditional is a tautology.

I am not going to burden you with analogues of the truth-table method for predicate logic. For the moment what I want is merely that you should get used to rendering English sentences into predicate logic, and then get a nose for which of the arguments are valid.

There is a system of natural deduction we can set up to generate all valid arguments capturable by predicate calculus but for the moment I want to use this new gadget of predicate calculus to describe some important concepts that you can't capture with propositional logic. supply reference

DEFINITION 13

- *A relation R is* **transitive** *if*

$$\forall x \forall y \forall z((R(x,y) \wedge R(y,z)) \to R(x,z))$$

- *A relation R is* **symmetrical** *if*

$$\forall x \forall y (R(x,y) \longleftrightarrow R(y,x))$$

- $(\forall x)(R(x,x))$ *says that R is* **reflexive***; and* $(\forall x)(\neg R(x,x))$ *says that R is* **irreflexive***.*

- *A relation that is transitive, reflexive and symmetrical is an* **equivalence relation***.*

A binary relation R is **extensional** if $(\forall x)(\forall y)(x = y \longleftrightarrow (\forall z)(R(x,z) \longleftrightarrow R(y,z))$. Notice that a relation can be extensional without its converse being extensional: the relation $R(x,y)$ defined by "x is the mother of y" is extensional (two women with the same children are the same woman) but its converse isn't (two distinct people can have the same mother).

There is a connection of ideas between 'extensional' as in 'extensional relation' and 'extension' as contrasted with 'intension'.

It's worth noting that

$$x \text{ is bigger than } y;\ y \text{ is bigger than } z. \text{ Therefore } x \text{ is bigger than } z. \qquad \text{(S)}$$

is not valid. (S) would be a valid argument if

$$(\forall xyz)((\texttt{bigger-than}(x,y) \wedge \texttt{bigger-than}(y,z)) \to \texttt{bigger-than}(x,z)). \quad \text{(T)}$$

were a logical truth. However (T) is not a logical truth. (S) is truth-preserving all right, but it's not truth-preserving in virtue of its logical structure.

It's truth-preserving once we have nailed down (as we noted, the Computer scientists would say "reserved") the words 'bigger-than'. Another way of making the same point is to say that the transitivity of bigger-than is not a fact of logic: it's a fact about the bigger-than relation. It's not true of the relation *is-a-first-cousin-of* nor of the relation *is-a-half-sibling-of*.

One way of putting this is to say that (T) is not a logical truth because there are other things with the same logical structure as it which are not true. If you replace 'is bigger than' in (T) by 'is the first cousin of' you obtain a false statement.

Notice in contrast that

$$(\forall x \forall y \forall z)(x = y \land y = z \to x = z)$$

is a logical truth! This is because '=' is part of the logical vocabulary and we are not allowed to substitute things for it.

Beginners often assume that symmetrical relations must be reflexive. They are wrong, as witness "rhymes with", "conflicts with", "can see the whites of the eyes of", "is married to", "is the sibling of" and many others.

Observe that equality is transitive, reflexive and symmetrical and is therefore an equivalence relation.

Not all relations are unary or binary

It may be worth making the point that not all relations are unary or binary relations. There is a natural three-place relation of *betweenness* that relates points on a line, but that doesn't concern us much as philosophers. Another example (again not of particular philosophical interest but cutely everyday) is the three-place relation of "later than" between hours on a clock. We cannot take this relation to be binary because, if we do, it will simply turn out to be the universal relation—every time on the clock is later than every other time if you wait long enough:

3 o'clock is later than 12 o'clock. (A)

and

12 o'clock is later than 3 o'clock. (B)

(A) and (B) are both true, which is not what we want. However, with a three-place relation we can say things like

Starting at 12 o'clock we first reach 3 o'clock and then 6 o'clock (A′)

and

Starting at 12 o'clock we first reach 6 o'clock and then 3 o'clock (B′)

Now (A′) is true and (B′) is false, which makes the distinction we want.

So we think of our three-place relation as "starting at x and reading clockwise we encounter y first and then z"

This is a simple illustration of a fairly common move in metaphysics. It happens every now and then that there is a binary relation that you are trying vainly to make sense of and things start to clarify only once you realise that the relation holds not between the two things you were thinking of but between those two and an extra one lurking in the background that you had been overlooking.

A more mundane example is "No student has the same teacher for every subject"

Higher-order again

Notice that you are not allowed to **bind** predicate letters. It is in virtue of this restriction that this logic is sometimes called **first-order** Logic. As we explained in section 3.1.1 if you attempt to bind predicate letters you are engaging in what is sometimes called **second-order logic** and the angels will weep for you. It is the work of the Devil.

For the moment we are going to concentrate on just *reading* expressions of predicate calculus, so that we feel happy having them on our sofas and don't panic. And in getting used to them we'll get a feeling for the difference between those that are valid and those that aren't.

1. $(\forall x)(F(x) \vee \neg F(x))$

   This is always going to be true, whatever property F is. Every x is either F or it isn't. The formula is *valid*.

2. $(\forall x)(F(x)) \vee (\forall x)(\neg F(x))$

   This isn't always going to be true. It says (as it were) that everything is a frog or everything is not a frog; the formula is not valid. However it is satisfiable: take F to be a property that is true of everything, or a property that is true of nothing.

3. $(\exists x)(F(x) \vee \neg F(x))$

   This is always going to be true, whatever property F is, as long as there *is* something.

4. This next expression, too, is always going to be true—as long as there *is* something.

   $(\exists x)(F(x)) \vee (\exists x)(\neg F(x))$

   We adopt as a logical principle the proposition that the universe is not empty. That is to say we take these last two expressions to be logically true.

5. These two formulæ are logically equivalent:

   $$(\exists x)F(x) \qquad\qquad \neg(\forall x)\neg F(x)$$

   The only way it can fail to be the case that everything is a non-frog is if there is a frog! (The universe is not empty, after all)

   Similarly:

6. These two formulæ are logically equivalent:

$$(\forall x)F(x) \qquad\qquad \neg(\exists x)\neg F(x)$$

If there are no non-frogs then everything is a frog. These last two identities correspond to the de Morgan laws that we saw earlier.

7. These two formulæ are logically equivalent:

$$(\exists x)(\forall y)(F(y) \to F(x)) \qquad\qquad (\exists x)((\exists y)F(y)) \to F(x))$$

EXERCISE 32 *Consider the two formulæ*

$$(\forall x)(\exists y)(L(x,y)) \text{ and } (\exists y)(\forall x)(L(x,y)).$$

does either imply the other?

If we read '$L(x,y)$' as 'x loves y' then what do these sentences say in ordinary English?

fair test.

EXERCISE 33 *In each formula circle the principal connective. (This requires more care than you might think! Pay close attention to the brackets)*

In each of the following pairs of formulæ, determine whether the two formulæ in the pair are (i) logically equivalent or are (ii) negations of each other or (iii) neither. The last two are quite hard.

$(\exists x)(F(x))$;	$\neg\forall x\neg F(x)$
$(\forall x)(\forall y)F(x,y)$;	$(\forall y)(\forall x)F(x,y)$
$(\exists x)(F(x) \vee G(x))$;	$\neg(\forall x)(\neg F(x) \vee \neg G(x))$
$(\forall x)(\exists y)(F(x,y))$;	$(\exists y)(\forall x)(F(x,y))$
$(\exists x)(F(x)) \to A$;	$(\forall x)(F(x) \to A)$
$(\exists x)(F(x) \to A)$;	$(\forall x)(F(x)) \to A$

(In the last two formulæ 'x' is not free in A)

Bibliography

[1] J.L. Borges Labyrinths [publ'n data??]

[2] J.L. Borges *Siete Noches*

[3] Carnap, R. Meaning and Necessity. University of Chicago Press 1947

[4] Carroll, Lewis. Through the Looking Glass and what Alice found there.

[5] Church, A. The calculi of λ-conversion. Princeton 1941

[6] John Corcoran. Aristotle's Demonstrative Logic. History and Philosophy of Logic, **30** (February 2009), pp 1-20.

[7] Donald Kalish and Richard Montague: Logic: techniques of formal reasoning. Harcourt Brace 1964

[8] Quine, Mathematical Logic

[9] Quine, Ontological relativity and Other Essays
Selected Logic papers.

[10] Tversky and Kahnemann "Extensional versus Intuitive reasoning" Psychological Review **90** 293-315.

[11] Tymoczko and Henle. Sweet Reason: A Field Guide to Modern Logic. Birkhäuser 2000 ISBN