An Introduction to Formal Methods for Philosophy Students

Thomas Forster

February 20, 2021

Contents

Preface

The text that follows is worked up from material from which I lectured a first-year course in *Philosophical Skills* at the University of East Anglia and a second year course in Logic at the University of Canterbury at Christchurch (New Zealand). Both these courses are intended for philosophy students. The early chapters are intended for first-year students, while chapters further on in the book are intended for second and third-year students. There are various audiences for first-year logic courses: in many places there is a service course in logic for students doing psychology or economics or law or possibly other things. At UEA the first-year cohort consisted entirely of philosophy students, and that made it possible to include logical material with a flavour that is (one hopes!) useful to philosophy students in their subsequent studies, even if perhaps not to students in psychology, economics or law.

So what should one cover? I am on record as arguing that possible world semantics has no useful applications in philosophy, so the reader may fairly ask why I include it in this text. One answer is that people expect to find it; a better answer is that even tho' modal logic is of little or no use in philosophy, elementary modal logic is a good source of syntactic *drill*. There is good evidence that compelling students to acquire formal skills sharpens up their wits generally: studies suggest that secondary school pupils who are subjected to more mathematics than a control group write better essays. Think of doing modal logic as doing sudoku or crosswords. Sudoku is no more applicable (and for that matter no less applicable) than modal logic, and shares with it the useful feature that it keeps the lobes in trim. The chapter on modal logic is placed after the material (even the advanced material) on first-order logic, so that it can be omitted by those first-years who want nothing beyond the basics, but much of it is flagged as being elementary enough to be shown to them. There have been first-year courses in logic that include possible world semantics for constructive propositional logic . . . at Queen Mary, for example.

Chapters 1-3 cover the material one would expect to find in a first year logic course for philosophy majors. Chapters 4-6 cover second year material intended for more determined and interested students. There is inevitably material that falls between these two stools: stuff that—altho' elementary—is not basic enough to be compulsory for first years (and which would in any case distract and confuse weaker students who have no natural inclination to the subject) but which nevertheless in some sense belong with it. Predicate modifiers, natural deduction for predicate calculus...

When i have materials closely related to each other i have tried to keep them adjacent in the text. This has resulted in various bits of advanced material being insinuated into early chapters. The advantage (for enthusiasts) is a linear presentation of material; the disadvantage (for the rest of us) is that sometimes one has to skip ahead.

Thanks to (among others) Ted Harding, Aldo Antonelli, Ed Mares, Jeremy Seligman and of course my students, including Matt Grice . . .

Stuff to fit in

$\neg\neg$ distributes over $\wedge$ but not over $\vee$.

They stood so still that she quite forgot they were alive, and she was just looking round to see if the word 'TWEEDLE' was written at the back of each collar, when she was startled by a voice coming from the one marked 'DUM.'

"If you think we're wax-works," he said, "you ought to pay, you know. Wax-works weren't made to be looked at for nothing, Nohow!"

"Contrariwise," added the one marked 'DEE,' "if you think we're alive, you ought to speak."

$(A \vee B) \wedge (A \to C) \wedge (B \to D). \to (C \vee D)$

Never trust a man who, when left alone in a room with a tea-cosy, doesn't try it on.

$\mathtt{man}()$; $\mathtt{should\text{-}be\text{-}trusted}()$; $\mathtt{puts\text{-}on}$, $\mathtt{tea\text{-}cosy}()$; $\mathtt{is\text{-}in}(\ ,)$; $\mathtt{room}()$

$(\forall x)(\forall t)(\forall r)[\mathtt{man}(x) \wedge \mathtt{room}(r) \wedge \mathtt{tea\text{-}cosy}(t) \wedge \mathtt{in}(m, r) \wedge (\forall o)(\mathtt{man}(o) \wedge \mathtt{in}(o, r) \to o = x) \to (\mathtt{should\text{-}be\text{-}trusted}(x) \to \mathtt{puts\text{-}on}(x, t))]$

Man was never intended to understand things he meddled with.

Pratchett 'Pyramids' Page 361

Is this the contrapositive of "Man was never meant to meddle with things he didn't understand"? [Thanks to Tom Körner] Or it is the converse?

Remember to take your contrapositive every morning!

We talk about contrapositives only in the context of propositional logic. This example forces you to think about contrapositives in the context of FOL.

How many arguments does a function have? The refractive index of a material is the ratio of the angle of incidence to the angle of refraction, and interestingly (Thomas Hariot) it doesn't depend on the angle. However it does depend on the material the other side of the interface, so strictly one shouldn't talk of the refractive index of glass, but of the glass-air [or whatever] interface. But actually once one know the refractive indices of of the interfaces medium-A–to–vacuum and of medium-B–to–vacuum one can compute the refractive index of the interface medium-A–to–medium-B So the refractive-index function can be thought of as having only one argument after all.

If you think that logic is the study of inference that is truth-preserving-in-virtue of syntax then you will certainly be of the view that higher-order logic is not logic—there's no syntactic criterion for validity!

Notice that in the formal language for chemistry the occurence of 'N' inside 'Ni' or 'Na' cannot be seen … 'Ni' is a single symbol, as is 'Na'.

We must tie all the second-order logic stuff together.

"Have nothing in your houses that you do not know to be useful, or believe to be beautiful"

There are two desiderata for things you might want to have in your house: *known to be useful* and *believed to be beautiful*. Is Morris saying that in order to be admitted to your house a thing should tick *both* boxes? Or is it enough for it to tick merely one?

"She's sweeter than all the girls and
i've met quite a few.
Nobody in all the world can
do what she can do"

is wrong, but

"She's sweeter than all the *other* girls and
i've met quite a few.
Nobody *else* in all the world can
do what she can do"

doesn't scan!

Noöne [else] can play Schumann like Richter.

Where do we explain the use of the word "classical" as in "classical logic"?

Using function symbols express "Kim and Alex have the same father but different mothers".

We can't capture *all* structure:
France is a monarchy in the form of a republic
Britain is a republic in the form of a monarchy.

Jeremy Seligman says: you need only *one* formalisation of an argument that makes it look valid!

In cases like

Girls never share boyfriends

We are all different

Every foot is different

Need binary not unary: Degree is too low.

The point is not that $(\forall x)(\mathtt{foot}(x) \to \mathtt{different}(x))$ isn't *deep enough*, the point is that it's *wrong*.

Something to do with plural subjects? Probably not.

"wrong person in the wrong place at the wrong time".

Like "earlier than" should be ternary not binary. It's not the person who is wrong, nor the place, nor the time. It's the combination, the *triple*, of person, place and time.

Another case where surface grammar misleads.

"Some accountants are comedians, but comedians are never accountants"

Re: concealment... the type-theoretic analysis of the paradoxes is a *concealment* analysis: the trouble that you get into is the trouble that lies in wait for people who try to conceal the types. Clearest with Richard's paradox.

U/G logic conceals evaluation. But that's OK!

Somewhere we need to make the point that all the classical propositional connectives can be defined in terms of `NAND`. (or, for that matter, in terms of `NOR`). One can thereby set up the reader for the observation that the intuitionistic connectives are not mutually definable.

There is some mathematical material (induction, lambda-calc) assumed in later chapters that is not explained in earlier chapters. Can't make it *completely* self-contained ...

A section on synonymy: *theories* not languages. Take a novel in one language. translate it into another language and back.

If we had some bacon we could have some bacon and eggs ... if we had the eggs.

Formal languages have *only* literal meaning

There are other uses of 'or' ...

> *Experiments have shown that, at the pressure of the lower mantle, iron(II) oxide is converted into iron metal and iron(III) oxide—which means that large bodies such as the earth can self-oxidise their mantle, whereas smaller ones cannot (or do so to a lesser extent)*

Redox state of early magmas
Bruno Scaillet and Fabrice Gaillard, nature 1/12/2011 **180** pp 48–9.

Tested on animals

'connective' not defined yet

The neat way to deal with formulæ not having outside parentheses is to regard a (binary) connective ?? as (...) ?? (...)

I think this is what lies behind the tendency for people to write $(\forall x)(A(x) \wedge B(x))$' for "All A are B"

Conditionals often suggest their converses. The person who choses to say 'If A then B' rather than 'If C then B' might have chosen that option because A is a more likely cause of B than C is. This suggests the converse: If B held, then it was because A.

The lasagne example is not of this kind! "If you want to eat there's lasagne in the oven".

Need to have some live examples of sequent proofs

question-begging?

'refute'

We need a section on **Fallacies**
In particular fallacy of affirming the consequent.
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!

Talk about the problem of inferring individual obligation from collective obligation?

Chapter 1

Introduction

1.1 What is Logic?

Three logicians walk into a bar.

The bartender asks "Do you all want drinks?"

First logician says "I don't know."

Second logician says "I don't know either."

Third logician says "Yes!"

Allude back to this when talking about concealment. There is no way of analysing this without talking about dated deduction-tokens.

"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.

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 *A box is full of hats. All but three are red, all but three are blue, all but three are brown, all but three are white. How many hats are there in the box?*

EXERCISE 2 *The main safe at the bank is secured with three locks, A, B and C. Any two of the three system managers can cooperate to open it. How many keys must each manager have?*

EXERCISE 3 *A storekeeper has nine bags of cement, all but one of which weigh precisely 50kg, and the odd one out is light. He has a balance which he can use to compare weights. How can he identify the rogue bag in only three weighings? Can he still do it if he doesn't know if the rogue bag is light?*

EXERCISE 4 *There were five delegates, A, B, C, D and E at a recent summit.*

B and C spoke English, but changed (when D joined them) to Spanish, this being the only language they all had in common;
The only language A, B and E had in common was French;
The only language C and E had in common was Italian;
Three delegates could speak Portugese;
The most common language was Spanish;
One delegate spoke all five languages, one spoke only four, one spoke only three, one spoke only two and the last one spoke only one.

Which languages did each delegate speak?

EXERCISE 5 *People from Bingo always lie and people from Bongo always tell the truth.*

If you meet three people from these two places there is a single question you can ask all three of them and deduce from the answers who comes from where. What might it be?

If you meet two people, one from each of the two places (but you don't know which is which) there is a single question you can ask either one of them (you are allowed to ask only one of them!) and the answer will tell you which is which. What is it?

EXERCISE 6

Brothers and sisters have I none;
This man's father is my father's son

To whom is the speaker referring?

EXERCISE 7 *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?

EXERCISE 8 *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?

EXERCISE 9 *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?

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.

EXERCISE 10 *You are given a large number of lengths of fuse. The only thing 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 just about to go out. What other lengths can you measure?*

EXERCISE 11 *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:

EXERCISE 12 *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?

Chapter 2

Introduction to Logic

If you start doing Philosophy it's because 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 book 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. "this course"?

There is an important contrast with **Rhetoric** here. With rhetoric anything goes that works. 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])

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). Talk here about concealment?

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

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

fuller study of rhetoric.

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. . .

> 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 slightly less than a hundred 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—although it's only a part of Logic that is like that—it's that part of logic that will take up our time.

2.1 Statements, Commands, Questions, Performatives

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 with **questions**. or performatives. A statement

Say something about performatives Well, performatives have truth-values too, but that's not really what they are for

[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!

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.)

The idea of **evaluation** comes in useful when explaining the difference between the things in this section heading. 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—or perhaps to a search for 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.

Does this belong in intension/extension? see p **??**

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 2.5.5 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'). 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'.

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.

Delimiters for premisses. Do we want this to illustrate the identity rule or $\wedge$-elim?

Every argument has **premisses** and a **conclusion** (only one conclusion). We write premisses above the line, and the conclusion below the line, thus

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

Of course they may be more than one premise, and we can write them on the same line,

$$\frac{\text{Premise-1} \quad \text{Premise-2}}{\text{Conclusion}}$$

Or on more than one line:

$$\begin{array}{c} \text{Premise 1} \\ \text{Premise 2} \\ \underline{\text{Premise 3}} \\ \text{Conclusion} \end{array}$$

We tend not to mix the styles.

Here are two arguments, each with two premisses:

$$\frac{\text{It is monday} \quad \text{It is raining}}{\text{It is raining}}$$

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

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]

'and' and 'or' are examples of **connectives**, aften written with special symbols, as in the table below, which shows a couple more connectives.

DEFINITION 1

$\wedge$, &	*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*".*

These things are called **connectives** because they connect statements ("join statements together" would be better). (Confusingly, '$\neg$' is a connective even though it doesn't connect two things: it is a **unary** connective.)

EXERCISE 13

Let P *abbreviate "I bought a lottery ticket" and* Q *"I won the jackpot".*

To what natural English sentences do the following formulæ correspond?

$$\neg P; \;\; P \vee Q; \;\; P \wedge Q; \;\; P \rightarrow Q; \;\; \neg P \rightarrow \neg Q; \;\; \neg P \vee (P \wedge Q).$$

2.1.1 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 2.5.5.

There are plenty of non-truth-functional connectives to: “implies”, “because”—both of those binary, but there are unary non-truth-functional connectives too: “obviously”, “possibly”

We will later make the connection: truth-functional = extensional

2.1.2 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
At some point mention boolean satisfiability and 3-sat

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:

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	0	0
0	1	1
1	1	0
1	0	1

A	`NAND`	B
0	1	0
0	1	1
1	1	0
1	0	1

all of which are **binary** connectives (connect two statements) and

$\neg$	A
1	0
0	1

$\neg$ is a **unary** connective.

The connectives `NAND`, `NOR` and `XOR` are sometimes used, but altho’ you will see them in electronics you will never see them in the philosophical literature. The ternary (three-place) connective

$$\text{if } p \text{ then } q \text{ else } r$$

is used in Computer Science but we won't use it here. In Philosophy we tend to make do with $\wedge$, $\vee$ and $\neg$ and one more, an arrow '$\to$' for "if ... then" which we have just seen and which i shall explain soon.

We have just seen the five *binary* connectives $\wedge$, $\vee$, `XOR`, `NAND` and `NOR`. There are four one-place ("unary') truth-functional connectives. The only one we are interested in is negation, but there are three others. There is the one-place connective that always returns the `true` and one that always returns the `false`. Then there is the connective that just gives you back what you gave it: one might call it the identity connective. We don't have standard notations for these three connectives, since we never use them. In the truth-tables that follow I write them here with one, two and three question marks respectively.

?	A
T	T
T	F

??	A
F	T
F	F

???	A
T	T
F	F

How many binary truth-functional connectives are there?

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, the truth-functional connectives.** 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 a very fruitful restriction, and it 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, and *not* miss out on desirable food objects—nor 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 then you need to know it, ditto a food object or a potential mate. You will want modes of reasoning to be available to you that will deliver any and every truth that can be squeezed out of the data 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 evade the dinosaur and get the forbidden fruit; 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 do more than merely preserve truth. We will want to study such modes of inference (in later chapters below) but the above considerations do tell us that it is very sensible to start with truth-preservation!

2.2 The Language of Propositional Logic

Let's now 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.

Truth-tables are a convenient way of representing/tabulating all the valuations a formula can have. Each row of a truth-table for a formula encapsulates [the extension of] a valuation for the propositional letters in that formula. 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.

EXERCISE 14 *Can you see why it's* 2^n*?*

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
⋮	1		0		1
⋮	1		1		0
⋮	1		0		0
⋮	⋮		⋮		⋮
⋮	⋮		⋮		⋮

EXERCISE 15 *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. "*`if` *p* `then` *q* `else` *r" requires that q be true when p is (never mind what happens to r) and that r is true when p is false (never mind about q). (How many other rows will the truth-table have?)*

might want to make a point about lazy evaluation here. See section 3.14

2.2.1 Truth-tables for compound expressions

We haven't seen any co
expressions yet. Put in a
ercise here

Some illustrations needed here

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

DEFINITION 2 *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***;*
$\to$ *is a* **conditional** *and its immediate subformulæ are its* **antecedent** *and its* **consequent***;*
$\longleftrightarrow$ *is a* **biconditional***.*

collate with material on p 31.

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**;
$A \to B$ is a conditional, where A is the **antecedent** and B is the **consequent.**

EXERCISE 16 *What are the principal connectives and the immediate subformulæ of the formulæ below?*

$P \vee \neg P$
$\neg(A \vee \neg(A \wedge B))$
$(A \vee B) \wedge (\neg A \vee \neg B)$
$A \vee (B \wedge (C \vee D));$
$\neg(P \vee Q)$
$P \to (P \vee Q)$
$P \to (Q \vee P)$
$(P \to Q) \vee (Q \to P)$
$(P \to Q) \to \neg(Q \to P)$
$P \to \bot$
$P \to (P \wedge Q)$
$P \to (Q \to P)$
$(P \longleftrightarrow Q) \wedge (P \vee Q)$
$(P \longleftrightarrow Q) \longleftrightarrow (Q \longleftrightarrow P)$
$A \to [(A \to C) \to ((B \to C) \to C)]$
$B \to [(A \to C) \to ((B \to C) \to C)]$
$(A \vee B) \to [(A \to C) \to ((B \to C) \to C)].$

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

$A \vee (B \wedge C)$:

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 '$\vee$'[4]:

[4] If you think "but we knew what to put under the '$\vee$' in the bottom four rows as soon as we knew A was 1 in those four rows", you are not being an annoying smart-aleck... well, you are, but you are onto something. We will pursue that thought later, in section 3.14, but not here.

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.

Need another example here...?

These worked exercises 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**. Other communities (mathematicians, computer scientists …) use instead the word **recursive**.

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.

2.2.2 Logical equivalence

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

EXERCISE 17 *Write out truth-tables for the first five formulæ in exercise 16.*

EXERCISE 18 *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.

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

Mix these up a bit…

DEFINITION 3

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**.

Commutativity. *$A \wedge B$ is logically equivalent to $B \wedge A$; $A \vee B$ is equivalent to $B \vee A$: we say $\wedge$ and $\vee$ are* **commutative**.

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** $\wedge$.

Associativity means you can leave out brackets; idempotence means you can remove duplicates and commutativity means it doesn't matter which way round you write things. Readers will be familiar with these phenomena (even if not the terminology) from school arithmetic: $+$ and $\times$ are—both of them—associative and commutative: $x + y = y + x$, $x \times (y \times z) = (x \times y) \times z$ and so on ... and we are quite used to leaving out brackets. Also $+$ distributes over $\times$: $x \times (y + z) = x \times y \ + \ x \times z$. $\wedge$ and $\vee$ parallel $+$ and $\times$ in various ways—echoing these features of $+$ and $\times$ we've just seen, but $\wedge$ and $\vee$ are both idempotent, whereas $+$ and $\times$ are not!

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 to broach the subject.

We write '$\rightarrow$' for the connective that we use to formalise inference. It will obey the rule

"from P and $P \rightarrow Q$ infer Q".

or

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

This rule is called **modus ponens**. Q is the **conclusion**; P is the **minor premiss** and $P \rightarrow Q$ is the **major premiss**. (Thus the **conclusion** of a token of *modus ponens* is the **consequent** of its **major premiss**. The **minor premiss** is the antecedent of the **major premiss**. I writing this out not because this information will one day save your life[5] but merely so that you can check that you have correctly grasped how these gadgets fit together.) I know I haven't given you a truth-table for '$\rightarrow$' yet. All in good time! There is some other gadgetry we have to get out of the way first.

We haven't explained type-token yet

2.2.3 Non truth functional connectives

Causation and necessity are not truth-functional. Consider

1. Labour lost the 2010 election *because* unemployment was rising throughout 2009;
2. *Necessarily* Man is a rational animal.

The truth-value of (1) depends on more than the truth values of "unemployment was rising throughout 2009" and "Labour lost the 2010 election"; similarly 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.

[5] You never know what information might save your life: knowing the wind-speed af an african swallow?

Counterfactuals

Say something about counterfactuals

Can we say anything intelligent about the difference?

2.3 Intension and Extension

somewhere in this section put the aperçu that one should always legislate the intension not the extension, or at least that one should legislate the spec not the implementation. Cycle lights, glorifying terrorism, talking on a mobile while driving. A similar point about dfn of Life: Most of the literature about exobiology seems to identify life only extensionally, recognises only one implementation of the ideas: carbon based life in water.

The intension-extension distinction is a device of mediæval philosophy which was re-imported into the analytic tradition by Frege starting in the late nineteenth century and later Church (see [**?**] p 2) and Carnap [6]) in the middle of the last century, probably under the influence of Brentano. However, in its passage from the mediævals to the moderns it has undergone some changes and it might be felt that the modern distinction shares little more than a name with the mediæval idea.

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.

This slang turns up nowadays in the connection with the idea of evaluation. In recent times there has been increasingly the idea that intensions are the sort of things one *evaluates* and that the things to which they evaluate are extensions. One reason why it is useful is captured by an *aperçu* of Quine's ([36] 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, properties-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.

perhaps collate this with material on p 19.

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. In contrast 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? [6]

The trouble is that although 'don't-know' is a 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. This mistake—of thinking that your uncertainty is a property of the proposition rather than a manifestation of the fact that you are a ignorant worm[7]—is a manifestation of the *mind-projection fallacy*.

> 'Professor Cox, is there anyone out there?"
>
> "The answer is, we don't know"

He's wrong! He should've said "We don't know the answer"!

This merits a slightly more detailed discussion. The point is that he's not answering the question but rather talking metalanguage

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. [**we really need some illustrations at this point**] To argue for a third truth-value you have to be sure that none of the likely cases can plausibly 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—they don't correspond to anything out in the world. See section 8.1 The difference between the `true` and the `false` is uncontroversial but it's not clear when two propositions are the same proposition. (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 (only a tiny bit in section 10.4) 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

"equivocating"?

[6] 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`.

[7] "We are all worms, but I do believe that I am a glow-worm"—W. S. Churchill.

satisfactory identity criteria for propositions then fallacy of equivocation would be less of a danger, but we haven't! So what we want to do in logic—at least to start with—is study relations between *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.

Or we could say that we are interested only in their evaluation behaviour, loose tho' that formulation is.

DEFINITION 4 *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 (a relation between event-types) by means of constant conjunction between the corresponding event tokens[8] 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.

"coercion"?

EXERCISE 19 *Match them up:*

q if p	$p \to q$
p unless q	$\neg p \to q$
q only if p	$q \to p$
p despite q	$p \wedge q$

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.

[8] Haven't yet defined this word

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.

2.3.1 If–then

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, as we saw on page 24) 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:

Collate this with p 24.

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

Modus tollens is the rule:

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

Affirming the consequent and inferring the antecedent:

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

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**.

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 attempt to salvage if . . . then). Any candidate must at least obey *modus ponens*:

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

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 (it doesn't even try) but it is surprisingly useful, and we can go a long way merely with an extensional conditional. Understanding the intentional conditional 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 on page 29.) It has spawned a vast and inconclusive literature, and we will have to get at least some way into it. See chapters 5 and 6.

Once we've got it sorted out ...

A	$\rightarrow$	B
F	T	F
F	T	T
T	F	F
T	T	T

or, in 0/1 notation:

A	$\rightarrow$	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 at other times with a double shaft: $\Rightarrow$.

Going for the material conditional means we don't have to worry ourselves sick about whether or not $A \rightarrow (B \rightarrow A)$ 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 $A \rightarrow (B \rightarrow A)$ is a good principle of inference).

Preferably leave out the '$\bot$'s...? At some point we have to talk about how to fill in a truth-table for a formula with a '$\bot$' in it.

EXERCISE 20 *In the following table*

(1) $A \wedge A$	A
(2) $\neg(A \wedge \neg B)$	$A \rightarrow B$
(3) $A \rightarrow A$	$\top$
(4) $A \rightarrow B$	$\neg A \vee B$
(5) $A \rightarrow \bot$	$\neg A$
(6) $\top \rightarrow A$	A
(7) $\bot \rightarrow A$	$\top$
(8) $A \rightarrow \top$	$\top$
(9) $A \rightarrow B$	$\neg B \rightarrow \neg A$
(10) $A \rightarrow \neg A$	$\neg A$
(11) $\neg A \rightarrow A$	A
(12) $A \vee A$	A
(13) $\neg(A \vee B)$	$(\neg A) \wedge (\neg B)$

(14) $A \vee B$	$\neg((\neg A) \wedge (\neg B))$
(15) $\neg(A \wedge B)$	$(\neg A) \vee (\neg B)$
(16) $A \wedge B$	$\neg((\neg A) \vee (\neg B))$
(17) $A \to \bot$	$\neg A$
(18) $\bot \to A$	$\top$
(19) $\bot \vee A$	A
(20) $A \vee (B \vee C)$	$(A \vee B) \vee C$
(21) $A \wedge (A \vee B)$	A
(22) $A \vee (A \wedge B)$	A
(23) $(A \to B) \to B$	$A \vee B$
(24) $A \vee (B \vee C)$	$(A \vee B) \vee C$
(25) $A \wedge (B \wedge C)$	$(A \wedge B) \wedge C$
(26) $A \vee (B \wedge C)$	$(A \vee B) \wedge (A \vee C)$
(27) $(A \wedge B) \vee ((\neg A) \wedge C)$	$(A \to B) \wedge ((\neg A) \to C)$
(28) $A \to (B \to C)$	$B \to (A \to C)$
(29) $B \to (A \to C)$	$(A \wedge B) \to C$

we find that, in each line, the two formulæ in it are logically equivalent. In each case write out a truth-table to prove it.

Perhaps some of these could be put in bundles of three. Flag the **de Morgan laws**.

2.3.2 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 20) is truth-preserving because of its form.

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

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

$$\frac{\mathrm{A} \quad \mathrm{B}}{\mathrm{B}}$$

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.

2.3.3 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 by all means try constructing a truth-table for it!

Where were we? Ah! - *type-token...*

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 copyright in Diana's letters to him but not the letters themselves. (Or is it the other way round?)

- 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.[9] '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!

- What is the difference between "unreadable" and "illegible"? A book (type) is unreadable if it so badly written that one cannot force oneself to read it. A book (token) is illegible if it is so defaced or damaged that one cannot decypher the (tokens of) words on its pages.

- 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 collectively. However it is the number of *tokens* [of itself] that the type appears to be trying to maximise.

- *First diner*:

[9] Is the word 'heterological' heterological?

> "Isn't it a bit cheeky of them to put "vegetable*s* 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.

- 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*.

2.3.4 Copies

"specimen" instead of "token"

Buddhas

It is told that the Buddha could perform miracles. But—like Jesus—he felt they were vulgar and ostentatious, and they displeased him. But that didn't stop him from performing them himself when forced into a corner. In [4] 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.[10]

Apparently he routinely made copies of himself 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.

[10] 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.

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.

Sets

ordered pair

If A is a set with three members and B is a set with four members, how many ordered pairs can you make whose first component is in A and whose second component is in B?

W-e-e-ll ... you pick up a member of A and you pair it with a member of B ... that leaves two things in A so you can do it again .... The answer must be three!

Wrong! Once you have picked up a member of A and put it into an ordered pair—it's still there!

One would tend not to use the word *token* in this connection. One would be more likely to use a word like *copy*. One makes lots of copies of the members of A. Just as the Buddha made lots of copies of himself rather than lots of *tokens* of himself. I suppose you could say that the various tokens of a type are copies of each other.

It is possible to do a lot of rigorous analysis of this distinction, and a lot of refinements suggest themselves. However, in the culture into which you are moving the distinction is a piece of background slang useful for keeping your thoughts on an even keel, rather than something central you have to get absolutely straight. In particular we will need it when making sense of ideas like *disjoint union* which we will meet in chapter 7.

2.4 Tautology and Validity

2.4.1 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 5 *A* **valid** *argument (type) is one such that any argument of that form (any token of it) with true premisses has a true conclusion.*

See appendix 11.1.2.

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

DEFINITION 6 *A* **sound** *argument (token) is a token of a valid argument-type all of whose premisses are true.*

The final example on page 38 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 depends 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. Appeal to argument by analogy suggests that we recognise—at some level—that the structure of an argument is important in discovering whether or not it is a good one. It's not just the truth of the conclusion that makes the argument good[11].

Explain this

connect this with FTPM; material in philrave.tex

This was beautifully parodied in the following example (due, I think, to Dr. Johnson—the same who kicked the stone) of the young man who desired to have carnal knowledge of his paternal grandmother[12] and responded to his father's entirely reasonable objections with: "You, sir, did lie with my mother: why should I not therefore lie with yours?"

Need more illustrations before we ask them to do these exercises

EXERCISE 21 *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!)*

Identify each of the first six arguments as modus ponens, modus tollens or as an instance of the fallacy of affirming the consequent.

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

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

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

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

[11] This is one reason why the material conditional is so puzzling!

[12] Don't ask me, I don't know why either!

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

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

7. *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.
   Alfred receives good marks.

8. *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.
   Herbert should think twice.

9. *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.

10. *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.
    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 [24]. Long out of print, but you can sometimes find second-hand copies. If you find one, buy it. (Alfred is probably Alfred Tarski: `https://en.wikipedia.org/wiki/Alfred_Tarski`; Herbert is probably Herb Enderton: `https://en.wikipedia.org/wiki/Herbert_Enderton`, but we will now never know, because Kalish and Montague are now both dead.)

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.

Imply and infer

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**. We'd better have this up in lights:

DEFINITION 7
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. [13]

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

DEFINITION 8 *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:

[13] 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*.

> 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.

EXERCISE 22 *Which of the arguments in exercises 21 and 8 are valid?*

When do we talk about fallacy of equivocation?

2.4.2 $\bigwedge$ 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 26) 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. The curly brackets '{' and '}' you may remember from school. They are not mere punctuation, but two components of a piece of notation: $\{A, B, C\}$ is the set that contains the three things A, B and C. So $\bigvee$ is an operation that takes a set of propositions and `or`s them together

This motivates more complex notations like

$$\bigvee_{i \in I} A_i \tag{2.1}$$

...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![14] 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 tell yourself that '$A \vee B \vee C$' is a disjunction, that its principal connective is '$\vee$' and that its immediate subformulæ are 'A', 'B' and 'C'.

---

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

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 $A \wedge B$ on the same principle.

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 $\bigvee(X \cup Y)$[15] 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

$$(\bigvee X) \vee (\bigvee \emptyset) \;=\; (\bigvee X) \tag{2.2}$$

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 to evaluate to `true` then equation (2.2) would compel '$\bigvee X$' to evaluate to `true` whatever X was! 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 '$\bigwedge \emptyset$'.

2.4.3 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 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:

[15] Here we use the symbol '$\cup$' for the first time

$(A$	$\longleftrightarrow$	$B)$	$\longleftrightarrow$	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 ...
somewhere in connection with DNF mention P=NP...?
Define 'scope'

DEFINITION 9
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.[16]

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

EXERCISE 23 *Recall the formula "*`if` *p* `then` *q* `else` *r" from exercise 15. Put it into CNF and also into DNF.*

EXERCISE 24 *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

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

(vi) $(p \vee q)$
(vii) $(p \wedge q)$

THEOREM 10 *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 "import" $\neg$, so that the '$\neg$' sign appears only attached to propositional letters. How? We saw earlier (exercise 20) 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" or—as we say—*import* them.

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

$$A \vee (B \wedge C) \longleftrightarrow (A \vee B) \wedge (A \vee C) \tag{2.3}$$

means that $A \vee (B \wedge C)$ and $(A \vee B) \wedge (A \vee C)$ are logically equivalent, 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 to put the formula into CNF

or

$$A \wedge (B \vee C) \longleftrightarrow (A \wedge B) \vee (A \wedge C) \tag{2.4}$$

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). ■

Here are some examples:

1.
$$(p \vee q) \rightarrow r$$
convert the '$\rightarrow$':
$$\neg(p \vee q) \vee r$$
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.
$$p \rightarrow (q \wedge r)$$
convert the '$\rightarrow$':
$$\neg p \vee (q \wedge r)$$
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.

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

convert the '→'

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

Drop the brackets because '∨' 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 '∧' over the '∨'. (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 $\bot$...

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

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

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

finally dropping backets because '∨' is associative ...

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

Note that in CNF (DNF) there is no requirement that every conjunct (disjunct) has to contain every letter.

In DNF inconsistencies vanish: the empty disjunction is the `false`; in CNF tautologies vanish: the empty conjunction is the `true`. (Recall what we were saying on page 41 about the empty conjunction and the empty disjunction.)

Finally, by using CNF and DNF we can show that any truth-functional connective whatever can be expressed in terms of ∧, ∨ and ¬. Any formula is equivalent to the disjunction of the rows (of the truth-table) in which it comes out true. We illustrated this earlier 2.4.3 with the expression $A \longleftrightarrow (B \longleftrightarrow C)$.

Alleles and Normal Forms

CNF and DNF have ramifications outside logic. Plenty of human features have a genetic component, in the sense that—for example—for a given feature (perfect pitch, blue eyes, left-handedness) a particular combination of particular alleles at particular loci might give you a ≥50% chance of having the feature. Let us for the moment oversimplify by supposing that each locus has precisely two alleles. This simplification enables us to associate with each locus a single propositional letter. Then having-a-≥50% chance of having that feature turns out to be captured by a propositional formula over the alphabet consisting of these propositional letters. This propositional formula naturally has a CNF and a DNF. The DNF is the disjunction of all the sufficient conditions for having-a-≥50% chance of having that feature; the CNF is the conjunction of all the sufficient conditions for having-a-≥50% chance of having that feature.

Realistically each locus cannot be relied upon to have two alleles, so we don't have a single propositional letter for each locus, but rather a propositional letter for each allele at any given locus. Then we have to add axioms to express the obvious background assumption that each locus is occupied by precisely one allele . . . which is as much as to say that the propositional letters are not independent. It provides us wiht a natural example of a propositional theory.

2.5 Further Useful Logical Gadgetry

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

More detail needed here

2.5.1 The Analytic-Synthetic Distinction

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

Analytic = necessary = *a priori*

and

Synthetic = contingent = *a posteriori*.

Hume's wall indeed. Not everybody is convinced: Kant thought there were assertions that were synthetic but *a priori*; Kripke claims there are necessary truths that are *a posteriori*, and Quine famously claimed that the analytic-synthetic distinction at least (if not the others) is spurious.

Of these three distinctions (if there really are three, not one!) the one that most concerns us here is the analytic-synthetic distinction. The cast of philosophical pantomime includes the analytic truth "All bachelors are unmarried".[17] The idea is that you can see that this allegation is true merely by analysing it—hence *analytic*.

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

The analytic/synthetic distinction seems to be connected with the intension/extension distinction—see Carnap, [6]:

> Facts about intensions are analytic and
> Facts about extensions are synthetic;

specifically

> 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)

monadic?

explain the equivocation on 'featherless biped'

2.5.2 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.

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

Say something about unfortunate overloading of 'neccesary'

Say something about "overloading"

2.5.3 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’.

People complain that they don’t want their food to be full of E-numbers. What they mean is that they don’t want it to be full of the things denoted by the E-numbers.[18] [Alphabet soup!]

> “Put cream on the banana cake.”
>
> “There isn’t any cream!”
>
> “Then put ‘cream’ on the shopping list!”

Haddocks’ Eyes

As so often the standard example is from [8].

> [. . .] 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 or strings of words—‘The agèd, agèd man’.. Even so, it does seem clear that the White Knight has got it wrong. At the very least if the name of the song really *is* is “The agèd agèd man” (as he says) then clearly Alice was right to say that was what the song was called. Granted, it might have more names than just that one—‘Ways and means’ for example—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

[18]Mind you E-300 is Vitamin C and there’s nothing wrong with *that*!

should leave the single quotation marks off the title, or—failing that (as Martin Gardner points out in [18])—burst into song. These infelicities must be deliberate (Carroll does not make elementary mistakes like that), and one wonders whether or not the White Knight realises he is getting it wrong . . . is he an old fool and nothing more? Or is he a paid-up party to a conspiracy to make the reader's reading experience a nightmare? The Alice books are one long nightmare, and perhaps not just for Alice.

Some good advice

Q: Why should you never fall in love with a tennis player?
A: Because 'love' means 'nothing' to them.

'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[19] 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'.

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

The Deaf Judge

JUDGE (*to PRISONER*): Do you have anything to say before I pass sentence?

PRISONER: Nothing

JUDGE (*to COUNSEL* : Did your Client say anything?

COUNSEL: 'Nothing' my Lord.

JUDGE: Funny, I could have sworn I saw his lips move...

Fun on a train

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 tactfully 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 that was one they hadn't heard 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.

If we write numbers in various bases (Hex, binary, octal ...) the numbers stay the same, but we change the numerals we associate with each number. Thus the numerals '`XI`', 'B', '11', '13' '1011' all represent the same number.

EXERCISE 25 *What is that number, and under which systems do those numerals represent it?*

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 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.[20]

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 [33]. However it does introduce subtleties 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.

[20] But it's obvious anyway that bus numbers are not numbers but rather strings. Otherwise how could we have a bus with a "number" like '7A'?

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 [21]. 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.

2.5.4 Language-metalanguage distinction

We distinguish between a world and the language used to describe it. In the full rich complexity of real life the language we use to describe the world is of course part of the world, but there are plenty of restricted settings in which a clear distinction can be drawn. The language we use when we do chemistry is not part of the subject matter of chemistry: in chemistry we study chemical elements and their compounds, not language.

However there are also settings in which the object of study is itself a language, and in those circumstances there are *two* languages in play. Naturally we need terminology for this situation. The language that is the object of study[22] is called the *object language.* The language that we use for describing the object language is the *metalanguage.* Thus, when the subject we are investigating is a language, the object language corresponds to the chemical elements and their compounds while the metalanguage corresponds to the language we use to describe those elements and compounds.

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

The language-metalanguage distinction is important for rhetoric. Any debate will be 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

[21] Wikipædia confirms this: apparently he was an accomplished pianist.

[22] Not the *subject* of study? Confusing, I know!

Say something about this

to a 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.)

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 *own* 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.)

2.5.5 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, whereas (B) will.

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 [46] starting on p. 247.

2.5.6 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. And of course the person who says A is claiming knowledge of A!

This is probably correct—the principle of charity usually is! but if A really does hold then A-or-B really does hold too, and that is what concerns us here. The subtleties of why someone might assert A-or-B do not concern us here.

2.5.7 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 get them 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.

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 [45] (from which this comes) and also the Wikipædia article.

2.5.8 Overinterpretation

My brother-in-law once heard someone on the bus say "My mood swings keep changing." He—like you or I on hearing the story—knew at once that what the speaker was trying to say was that they suffer from mood swings!

Reinterpreting silly utterances like this so that they make sense is something that we are *incredibly* good at. And by 'incredibly good' I mean that this is one of the things we can do *vastly* better that computers do (in contrast to the things like multiplying 100-digit numbers, which computers can do very much better than we can). In fact we are *so* good at it that nobody has yet quite worked out how we do it, though there is a vast literature on it, falling under the heading of what people in linguistics call "pragmatics". Interesting though that literature is I am mentioning it here only to draw your attention to the fact that learning to do this sort of thing *better* is precisely what we are *not* going to do. I want you to recognise this skill, and know when you are using it, in order not to use it *at all!*

Why on earth might we *not* want to use it?? Well, one of the differences between the use of symbols in formal languages (like in logic) and the use of symbols in everyday language is that in formal settings we have to use symbols rigidly and we suffer for it if we don't. If you give your computer an instruction with a grammatical error in it the operating system will reject it: "Go away and try again." One of the reasons why we design mathematical language (and programming languages) in this po-faced fault-intolerant way is that that is the easiest way to do it. Difficult though it is to switch off the error-correcting pattern-matching software that we have in our heads, it is much more difficult still to discover how it works and thereby emulate it on a machine—which is what we would have to do if we were to have a mathematical or programming language that is fault-tolerant and yet completely unambiguous. In fact this enterprise is generally regarded as so difficult as to be not worth even attempting. There may even be some deep philosophical reason why it is impossible even in principle: I don't know.

Switching off our fault-tolerant pattern-matching is difficult for a variety of reasons. Since it comes naturally to us, and we expend no effort in doing it, it requires a fair amount of self-awareness even to realise that we *are* doing it. Another reason is that one feels that to refrain from sympathetically reinterpreting what we find being said to us or displayed to us is unwelcoming, insensitive, autistic and somehow not fully human. Be that as it may, you have to switch all this stuff off all the same. Tough!

So we all need some help in realising that we do it. I've collected in the appendix to this chapter a few examples that have come my way. I'm hoping that you might find them instructive.

2.5.9 Affirming the consequent

Years ago I was teaching elementary Logic to a class of first-year law students, and I showed them this syllogism:

> "If George is guilty he'll be reluctant to answer questions; George is reluctant to answer questions. Therefore George is guilty."

[23] Then I asked them: Is this argument valid? A lot of them said 'yes'.

We all know that an obvious reason—the first reason that comes to mind—why someone might be reluctant to answer questions is that they might have something

[23] What is this doing here? It should be in a section on fallacies. Perhaps that section should be here, OK, but either way some material needs to be moved around.

to hide. And that something might be their guilt. So if they are reluctant to answer questions you become suspicious at once. Things are definitely not looking good for George. Is he guilty? Yeah—string him up!

But what has this got to do with the question my first-years were actually being asked? Nothing whatever. They were given a premiss of the form $P \to Q$, and another premiss Q. Can one deduce P from this? Clearly not. Thinking that you can is the fallacy of *affirming the consequent* (which we first saw on page 31).

There are various subtle reasons for us to commit this fallacy, and we haven't got space to discuss them here. The question before the students in this case was not: do the premisses (in conjunction with background information) give evidence for the conclusion? The question was whether or not the inference from the premisses to the conclusion is logically valid. And that it clearly isn't. The mistake my students were making was in misreading the question, and specifically in misreading it as a question to which their usual fault-tolerant pattern-matching software would give them a swift answer.

Chapter 3

Proof Systems for Propositional Logic

3.1 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 use truth tables to *detect* them; In this chapter we explore a method for *generating* them.

3.2 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 C tells us how to **use** the information wrapped up in a formula whose principal connective is 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 Prawitz, and possibly to Gentzen.) references?

The rules tell us how to exploit the information contained in a formula.

(Some of these rules come in two parts.)

Introduction Rules	Elimination Rules
$\vee$-int: $\dfrac{A}{A \vee B}$; $\dfrac{B}{A \vee B}$;	$\boxed{\vee\text{-elim ???}}$
$\wedge$-int: $\dfrac{A \quad B}{A \wedge B}$;	$\wedge$-elim: $\dfrac{A \wedge B}{A}$; $\dfrac{A \wedge B}{B}$
$\boxed{\to\text{-int ???}}$	$\to$-elim: $\dfrac{A \quad A \to B}{B}$

'elim' is an abbreviation for 'elimination'; it does not allude to any religion.

You will notice the division into two columns. You will also notice the two *lacunæ*: for the moment there is no $\vee$-use rule and no $\to$-int rule.

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

Exercise 26

1. *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]

   $A,\ A \to B$;
   $A,\ A \to (B \to C)$;
   $A,\ A \to (B \to C),\ B$;
   $A,\ B,\ (A \wedge B) \to C$;
   $A,\ (A \vee B) \to C$;
   $A \wedge B,\ A \to C$;
   $A \wedge B,\ A \to C,\ B \to D$;
   $A \to (B \to C),\ A \to B,\ B \to C$;
   $A,\ A \to (B \to C),\ A \to B$;
   $A,\ \neg A$.

2. *Deduce C from $(A \vee B) \to C$ and A;*
   Deduce B from $(A \to B) \to A$ and $A \to B$;
   Deduce R from P, $P \to (Q \to R)$ and $P \to Q$;

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. (And of course you are allowed to use an assumption as often as you like. If it is a sunny tuesday you might use $\wedge$-elimination to infer that it is sunny so you can go for a walk in the botanics, but that doesn't relieve you of the obligation of inferring that it is tuesday and that you need to go to your 11 o'clock lecture.)

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

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. How so? Well, if you want to infer both A and B from $A \wedge B$ and you want to write '$A \wedge B$' only once, you will find yourself writing '$\frac{A \wedge B}{A \quad B}$' and then building proofs downward from the token of the 'A' on the lower line and also from the 'B' on the lower line. They might rejoin later on. Hence the plaiting.

Now we can introduce a new rule, the *ex falso sequitur quodlibet.*

Ex falso sequitur quodlibet; $\frac{\perp}{A}$

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

better explain at some point why it is truth-preserving

Is this the right place for double neg? aka classical contradiction.

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

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 $B \to \perp$, so the inference

$$\frac{A \qquad \neg A}{\perp} \tag{3.1}$$

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

The rule of $\to$-introduction

The time has now come to make friends with the rule of $\to$-introduction. Recalling what introduction rules do, you can se that the $\to$-introduction rule will be a rule that tells you how to prove things of the form $A \to B$. Well how, in real life, do you prove "if A then B"? Well, you assume A and deduce B from it. What could be simpler!? Let's have an illustration. We already know how to deduce $A \vee C$ from A (we use $\vee$-introduction) so we should be able to prove $A \to (A \vee C)$.

$$\frac{A}{A \vee C}\ \vee\text{-int} \tag{3.2}$$

So we just put '$A \to (A \vee C)$' on the end ...?

$$\cfrac{\cfrac{A}{A \vee C}\ \vee\text{-int}}{A \to (A \vee C)} \tag{3.3}$$

That's pretty obviously the right thing to do, but for one thing. The last proof has $A \to (A \vee C)$ as its last line (which is good) but it has A as a live premiss. We assumed A in order to deduce $A \vee C$, but although the truth of $A \vee C$ relied on the truth of A, the truth of $A \to (A \vee C)$ does not rely on the truth of A. (It's a tautology, after all.) We need to record this fact somehow. The point is that, in going from a deduction-of-$A \vee C$-from-A to a proof-of-$A \to (A \vee C)$, we have somehow *used up* the assumption A. We record the fact that it has been used up by putting square brackets round it, and putting a pointer from where the assumption A was made to the line where it was used up.

$$\dfrac{\dfrac{[A]^1}{A \vee C}\ \vee\text{-int}}{A \to (A \vee C)}\ \to\text{-int (1)} \tag{3.4}$$

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 rule of $\vee$-elimination

"they will either contradict the Koran, in which case they are heresy, or they will agree with it, so they are superfluous."

Do the sudoku on page **??**.

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.

Need detailed explanation of $\vee$-elim here

3.2.1 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". The person who says "A, therefore B" is not only asserting B but is also asserting A. The person who says "If A then B" is **not** asserting A! Despite this, the *relationship-between-A-and-B* to which our attention is being drawn is the same in the two cases: *that's* not where the difference lies. 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] see below, on dialetheism

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, 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, some of which we might feell are mistaken) then one would be liable to be labelled as autistic.

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]

The reader might feel that I have made absurdly heavy weather of this business of hypothetical reasoning. Not so: an inability to cope with hypothetical reasoning can be found even among people with lots of letters after their name. Se appendix 11.2.1.

The Identity Rule

Finally we need the identity rule:

$$\frac{A \; B \; C \ldots}{A} \tag{3.5}$$

(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 17.) One could do something like

[2] See [12]. 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 neccesarily 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.

[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".

$$\cfrac{\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A \wedge B}\ \wedge\text{-int}}{A}\ \wedge\text{-elim}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{3.6}$$

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

$$\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A}\ \text{identity rule}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{3.7}$$

If we take seriously the observation above concerning the rule of →-introduction—namely that you are not required to cancel *every* occurrence of an assumption—then you conclude that you are at liberty to 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 3.8 below. To my taste, it seems less bizarre to discard assumptions than it is to cancel assumptions that aren't there, so I prefer 3.7 to 3.8. It's a matter of taste.

$$\cfrac{\cfrac{[A]^1}{B \to A}\ \to\text{-int}}{A \to (B \to A)}\ \to\text{-int (1)} \tag{3.8}$$

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. (You will find them in section 9.2 if you last that long). 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.

define 'counterfactual'

3.2.2 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 $\vee$-elimination

The rule of $\vee$-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 knowing 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 $\vee$-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 $A \vee B$. This will come in useful on page 71.

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

$$\frac{\begin{array}{ccccc} [A_1]^1 & [A_2]^1 & \ldots & [A_n]^1 & \\ \vdots & \vdots & & \vdots & \\ C & C & & C & A_1 \vee A_2 \vee \ldots A_n \end{array}}{C} \vee\text{-elim (1)} \tag{3.9}$$

where we can cancel more than one assumption. That is to say we have a list $A_1 \ldots A_n$ of assumptions, and the rule accepts as input a list of proofs of C: a proof of C from A_1, a proof of C from A_2, and so on up to A_n. It also accepts the disjunction $A_1 \vee \ldots A_n$ of the assumptions $A_1 \ldots A_n$ 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. Some of the challenges in the exercise which follows require it.

EXERCISE 27
Deduce $P \to R$ from $P \to (Q \to R)$ and $P \to Q$;
Deduce $(A \to B) \to B$ from A;
Deduce C from A and $((A \to B) \to B) \to C$;
Deduce $\neg P$ from $\neg(Q \to P)$;
Deduce A from $B \vee C$, $B \to A$ and $C \to A$;
Deduce $\neg A$ from $\neg(A \vee B)$;
Deduce Q from P and $\neg P \vee Q$;
Deduce Q from $\neg(Q \to P)$.

3.2.3 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. 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*. 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.

Let's try an example. Suppose 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 have underlined. (1) in the box tells us to **assume** the antecedent (which is $(A \to B) \to A$), at which point the consequent (which is $(A \to B) \to B$) becomes our new goal. So we have traded the old goal $((A \to B) \to A) \;\to\; ((A \to B) \to B)$ for the new goal $(A \to B) \to B$ and generated the new assumption $(A \to B) \to A$. 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 $A \to B$. Now $A \to B$ might be an assumption. If it isn't, it becomes a new goal. As it happens, $A \to B$ is an assumption, because we had the goal $(A \to B) \to B$ and this—by rule-of-thumb-1) (in the box)—generates the assumption $A \to B$ and the goal B.

Do some very simple illustrations of compound proofs 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 16.) 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 $A \wedge B$.

3.2.4 The Small Print

This section contains some warnings that might save you from tripping yourself up . . .

Look behind you!

You can cancel an assumption only if it appears in the branch above you! You might care to study the following defective proof.

$$\dfrac{\dfrac{\dfrac{[A]^2 \qquad [A \to (B \vee C)]^3}{B \vee C}\,\to\text{-elim} \qquad \dfrac{\dfrac{[B]^1}{A \to B}\,\to\text{-int (2)}}{(A \to B) \vee (A \to C)}\,\vee\text{-int} \qquad \dfrac{\dfrac{[C]^1}{A \to C}\,\to\text{-int (2)}}{(A \to B) \vee (A \to C)}\,\vee\text{-int}}{(A \to B) \vee (A \to C)}\,\vee\text{-elim (1)}}{A \to (B \vee C). \to .(A \to B) \vee (A \to C)}\,\to\text{-int (3)} \tag{3.10}$$

An attempt is made to cancel—in the two branches in the middle and on the right—the 'A' in the leftmost of the three branches. (Look for the '$\to$-int (2)' at the top of the two branches.) This is not possible! Interestingly no proof of this formula can be given that does not use the rule of classical contradiction. You will see this formula again in exercise 6.4.2

to be continued . . .

Ellipsis

There is a temptation to ellipsis with $\vee$-elimination:

One of my students wrote

$$\frac{A \to C \qquad B \to C \qquad A \vee B}{C} \vee\text{-elim} \tag{3.11}$$

I can see what she meant! It was

$$\frac{\dfrac{[A]^1 \qquad A \to C}{C} \to\text{-elim} \qquad \dfrac{[B]^1 \qquad B \to C}{C} \to\text{-elim} \qquad A \vee B}{C} \vee\text{-elim (1)} \tag{3.12}$$

The two rules of thumb don't *always* work

The two rules of thumb are the bits of attack-advice in the box on page 64.

It isn't *invariably* 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 $A \vee \neg A$. (A 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

$$\frac{A}{A \vee \neg A} \vee\text{-int} \tag{3.13}$$

or the attempt

$$\frac{\neg A}{A \vee \neg A} \vee\text{-int} \tag{3.14}$$

and clearly neither of these is going to turn into a proof of $A \vee \neg A$, since we are not going to get a proof of A (nor a proof of $\neg A$). It turns out you have to use the rule of double negation: assume $\neg(A \vee \neg A)$ 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, Where do we explain classical contradiction?

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.

P=NP?

3.2.5 Some Exercises

EXERCISE 28 *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. $A \to ((((A \to B) \to B) \to C) \to C)$;
8. $(P \vee Q) \to (((P \to R) \wedge (Q \to S)) \to (R \vee S))$;
9. $(P \wedge Q) \to (((P \to R) \vee (Q \to S)) \to (R \vee S))$;
10. $\neg(A \vee B) \to (\neg A \wedge \neg B)$;
11. $A \vee \neg A$; *(*)*
12. $\neg(A \wedge B) \to (\neg A \vee \neg B)$; *(hard!) (*)*
13. $(A \wedge (B \vee C)) \to ((A \wedge B) \vee (A \wedge C))$;
14. $(((A \wedge B) \vee (A \wedge C)) \to (A \wedge (B \vee C))$;
15. $(A \vee (B \wedge C)) \to ((A \vee B) \wedge (A \vee C))$;
16. $((A \vee B) \wedge (A \vee C)) \to (A \vee (B \wedge C))$; *hard!*
17. $A \to [(A \to C) \to ((B \to C) \to C)]$; *(for this and the next you will need the identity rule)*;
18. $B \to [(A \to C) \to ((B \to C) \to C)]$; *then put these last two together to obtain a proof of*

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

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

20. $((B \vee (B \to A)) \to A) \to A$;

21. $(A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge B) \vee (\neg A \wedge \neg B)$. *(Hard! For enthusiasts only) (*)*

You should be able to do the first seven 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. We discuss it below, chapter 5.

Make sure they are roughly in increasing order of difficulty

Enthusiasts can also attempt the first two parts of exercise 61 on p. 152: they are like the exercises here but harder.

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 29

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 30

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

EXERCISE 31 *Annotate the following proofs, indicating which rules are used where and which premisses are being cancelled when.*

$$\cfrac{\cfrac{\cfrac{P \qquad P \to Q}{Q}}{(P \to Q) \to Q}}{P \to ((P \to Q) \to Q)} \tag{3.15}$$

$$\cfrac{\cfrac{\cfrac{P \wedge Q}{Q}}{P \vee Q}}{(P \wedge Q) \to (P \vee Q)} \tag{3.16}$$

$$\dfrac{\dfrac{\dfrac{P \qquad \neg P}{\bot}}{Q}}{P \to Q} \tag{3.17}$$

$$\dfrac{\dfrac{P \vee Q \qquad \dfrac{P \qquad P \to R}{R} \qquad \dfrac{Q \qquad Q \to R}{R}}{R}}{(P \vee Q) \to R} \tag{3.18}$$

$$\dfrac{\dfrac{\dfrac{A \qquad B}{A \wedge B}}{B \to (A \wedge B)}}{A \to (B \to (A \wedge B))} \tag{3.19}$$

$$\dfrac{\dfrac{\dfrac{(A \to B) \to B \qquad A \to B}{B}}{((A \to B) \to B) \to B}}{(A \to B) \to (((A \to B) \to B) \to B)} \tag{3.20}$$

A First Look at Three-valued Logic

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:

$\to$	1	2	3
1	1	2	3
2	1	1	3
3	1	1	1

(Notice that the two truth-tables you get if (i) strip out 3 or (ii) strip out 2 both look like the two-valued truth-table for $\to$. They have to, if you think of it. The only room for manœuvre comes with relations between 2 and 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.

EXERCISE 32 *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?

(i) $P \vee \neg P$*?*

(ii) $(\neg\neg P) \vee \neg P$*?*

(iii) $\neg\neg(P \vee \neg P)$*?*

(iv) $(\neg P \vee Q) \to (P \to Q)$*?*

3.3 Soundness of the Natural Deduction Rules

This section can be skipped by first year students.

The rules of natural deduction are **sound**: every formula we can prove using natural deduction is a tautology. The rules preserve truth: if you reason using these rules from true premisses your conclusions will be true as well. Whatever our logical machinery (and it might be deliberately over-simplified, as it is when we start off with propositional logic) we want to be sure that the rules that we decide on for reasoning with that machinery are sound in this sense.

Completeness is a feature complementary to soundness. Not only are the rules sound, but they exhaust the possible modes of truth-preserving reasoning (in this language) in the sense that any truth-preserving inference can be captured by reasoning according to these formulations. We say the rules are **complete**. We prove this in section 3.8. It is impossible to overstate the significance of this fact. There is a finite system of rules of inference which captures **all** truth-preserving reasoning expressible in this syntax. The power of this simplification is incalculable and has impressed generations of logicians. There is a tradition in modern logic that holds that a body of principles of reasoning that cannot be finitely codified is simply not part of Logic at all. Not everybody believes this, but it is a widely held view.

Double negation or classical negation?

In the case of propositional logic we have truth-tables, which enable us to decide quite quickly when a formula is valid (or when a principle of reasoning is truth-preserving otherwise-known-as sound). This is so convenient that one tends to forget that there is actually a method of *generating* all the valid principles (and all the tautologies—otherwise known as valid formulæ) over and above a method of recognising them when they pop up. In fact there are several ways of doing this, and we will see some of them, and we will prove that they do this: that is, that they are complete.

The rules are sound in that they preserve truth: in any token of the rule if the premisses are true then the conclusions are true too. For the rules like $\wedge$-introduction, $\vee$-introduction, $\wedge$-elimination, $\to$-elimination ... it's obvious what is meant: for any valuation v if the stuff above the line is true according to v then so is the stuff below the line.

What I am planning to convince you is that any complex proof made up by composing lots of tokens of $\wedge$-int, $\to$-elim and so on has the property that any valuation making all the premisses true also makes the conclusion true. That is to say, we claim that all complex proofs are **truth-preserving**. Notice that this has as a special case the fact that any complex proof with no premisses has a conclusion that is logically valid. Every valuation making all the premisses true will make the conclusion true. Now since there are no premisses, every valuation makes all the premisses true, so every valuation makes the conclusion true. So the conclusion is valid!

see also p 126

However this way of thinking about matters doesn't enable us to make sense of $\to$-introduction and $\vee$-elimination. To give a proper description of what is going on we need to think of the individual (atomic) introduction and elimination rules as gadgets for making new complex proofs out of old (slightly less complex) proofs.

That is to say you think of the rule of $\wedge$-introduction as a way of taking a complex proof $\mathcal{D}_1$ of A and a complex proof $\mathcal{D}_2$ of B and giving a complex proof $\mathcal{D}_3$ of $A \wedge B$. We are trying to show that all complex deductions are truth-preserving.

The fact that $\wedge$-introduction is truth-preserving in the sense of the previous paragraph now assures us that it has the new property that:

If

- $\mathcal{D}_1$ is a truth-preserving deduction of A (that is to say, any valuation making the premisses of $\mathcal{D}_1$ true makes A true); and
- $\mathcal{D}_2$ is a truth-preserving deduction of B (that is to say, any valuation making the premisses of $\mathcal{D}_2$ true makes B true);

Then the deduction $\mathcal{D}_3$:

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ A\end{array}\qquad\begin{array}{c}\mathcal{D}_2\\ \vdots\\ B\end{array}}{A \wedge B}\ \wedge\text{-int} \tag{3.21}$$

too, is truth-preserving in the sense that any valuation making the premisses of $\mathcal{D}_3$ true—and they are just (the premisses of $\mathcal{D}_1$) $\cup$ (premisses of $\mathcal{D}_2$)—makes $A \wedge B$ true too.

This sounds like a much more complicated way of thinking of $\wedge$-introduction as truth-preserving than the way we started out with, but we need this way of seeing things when we come to consider the rules that involve cancelling assumptions, namely

$\rightarrow$-introduction and $\vee$-elimination. Let us now consider these two.

$\rightarrow$-introduction

Suppose we have a deduction $\mathcal{D}$ of B from $A, C_1 \ldots C_n$, and that $\mathcal{D}$ is truth-preserving. That is to say, any valuation making all of $A, C_1 \ldots C_n$ true will also make B true. Now consider the deduction $\mathcal{D}'$ (of $A \rightarrow B$ from $C_1 \ldots C_n$) that is given us by an application of $\rightarrow$-introduction. We want this to be truth-preserving as well, that is to say, we want any valuation making $C_1 \ldots C_n$ true to make $A \rightarrow B$ true too. Rephrase this

Let's check this. Let v be a valuation making $C_1 \ldots C_n$ true. Then either

> (i) it makes A true in which case—because $\mathcal{D}$ was truth-preserving—it makes B true as well and thereby makes $A \rightarrow B$ true.

Or

> (ii) it makes A false. Any valuation making A false makes $A \rightarrow B$ true.

Remember: you don't have to cancel all occurrences of the premiss. (see page 60.)

$\vee$-elimination

We can tell a similar story about $\vee$-elimination. Suppose we have (i) a truth-preserving deduction $\mathcal{D}_1$ of C from A (strictly: from A and a bag of extra assumptions like the $C_1 \ldots C_n$ of the previous paragraph) and (ii) a truth-preserving deduction $\mathcal{D}_2$ of C from B (and extra assumptions). That is to say that any valuation making A (and the extra assumptions) true makes C true, and any valuation making B (and the extra assumptions) true makes C true. Now, any valuation making $A \vee B$ (and the extra assumptions) true will make one of A and B true. So the new proof

$$\dfrac{\begin{matrix}[A]^1 \\ \vdots \\ \mathcal{D}_1 \\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[B]^1 \\ \vdots \\ \mathcal{D}_2 \\ \vdots \\ C\end{matrix} \qquad A \vee B}{C}\ \vee\text{-elim (1)} \tag{3.22}$$

—that we make from $\mathcal{D}_1$ and $\mathcal{D}_2$ by applying $\vee$-elim to them—is truth-preserving as well. picture here

In excruciating detail: let v be a valuation that makes $A \vee B$ (and the extra assumptions) true. Since v makes $A \vee B$ true, it must either (i) make A true, in which case we conclude that C must be true because of $\mathcal{D}_1$; or (ii) make B true, in which case we conclude that C must be true because of $\mathcal{D}_2$. Either way it makes C true.

3.4 Harmony and Conservativeness

3.4.1 Conservativeness

Recall the discussion on page 62 about the need for the identity rule, and the horrendous proof of K that we would otherwise have, that uses the rules for $\wedge$.

Notice that the only proof of Peirce's Law that we can find uses rules for a connective ($\neg$, or $\perp$ if you prefer) that does not appear in the formula being proved. (Miniexercise: find a proof of Peirce's law). This rule is the rule of double negation of course. No-one is suggesting that this is illicit: it's a perfectly legal proof; however it does violate an æsthetic. (As does the proof of K on page 62 that uses the rules for $\wedge$ instead of the identity rule). The æsthetic is *conservativeness*: every formula should have a proof that uses only rules for connectives that appear in the formula. Quite what the metaphysical force of this æsthetic is is a surprisingly deep question. It is certainly felt that one of the points in favour of the logic without the rule of double negation (which we will see more of below) is that it respects this æsthetic.

The point of exercise 32 part 3 was to establish that there can be no proof of Peirce's law using just the rules for '$\to$'.

Put the curly Ds as markers to vertical brackets

Look at section 3.7

3.4.2 Harmony

A further side to this æsthetic is the thought that, for each connective, the introduction and elimination rules should complement each other nicely. What might this mean, exactly? Well, the introduction rule for a connective £ tells us how to parcel up information in a way represented by the formula A £ B, and the corresponding elimination ("use"!) rule tells us how to exploit the information wrapped up in A£B. We certainly don't want to set up our rules in such a way that we can somehow extract more information from A£B than was put into it in the first place. This would probably violate more than a mere æsthetic, in that it could result in inconsistency. But we also want to ensure that all the information that was put into it (by the introduction rules) can be extracted from it later (by the use rules). If our rules complement each other neatly in this way then something nice will happen. If we bundle information into A£B and then immediately extract it, we might as well have done nothing at all. Consider

$$\dfrac{\dfrac{\begin{array}{c}\mathcal{D}_1\\ \vdots\\ A\end{array} \qquad \begin{array}{c}\mathcal{D}_2\\ \vdots\\ B\end{array}}{A \wedge B}\ \wedge\text{-int}}{B}\ \wedge\text{-elim} \tag{3.23}$$

where we wrap up information and put it inside $A \wedge B$ and then immediately unwrap it. We can clearly simplify this to:

$$\begin{array}{c}\mathcal{D}_2\\ \vdots\\ B\end{array} \tag{3.24}$$

This works because the conclusion $A \wedge B$ that we infer from the premisses A and B is the strongest possible conclusion we can infer from A and B and the premiss $A \wedge B$ from which we infer A and B is the *weakest* possible premiss which will give us both those conclusions. If we are given the $\wedge$-elimination rule, what must the introduction rule be? From $A \wedge B$ we can get both A and B, so we must have had to put them in in the first place when we were trying to prove $A \wedge B$ by $\wedge$-introduction. Similarly we can infer what the $\wedge$-elimination rule must be once we know the introduction rule.

The same goes for $\vee$ and $\rightarrow$. Given that the way to prove $A \rightarrow B$ is to assume A and deduce B from it, the way to use $A \rightarrow B$ must be to use it in conjunction with A to deduce B; given that the way to use $A \rightarrow B$ is to use it in conjunction with A to infer B it must be that the way to prove $A \rightarrow B$ is to assume A and deduce B from it. That is why it's all right to simplify

$$\dfrac{\dfrac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \rightarrow B}\ \rightarrow\text{-int} \qquad A}{B}\ \rightarrow\text{-elim} \tag{3.25}$$

to

$$\begin{matrix}A\\ \vdots \\ B\end{matrix} \tag{3.26}$$

And, given that the way to prove $A \vee B$ is to prove one of A and B, the way to use $A \vee B$ must be to find something that follows from A and that also—separately—follows from B; given that the way to use $A \vee B$ is to find something that follows from A and that also—separately and independently—follows from B, it must be that the way to prove $A \vee B$ is prove one of A and B. That is why we can simplify

$$\dfrac{\begin{matrix}[A_1]^1\\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[A_2]^1\\ \vdots \\ C\end{matrix} \qquad \dfrac{A_1}{A_1 \vee A_2}\ \vee\text{-int}}{C}\ \vee\text{-elim (1)} \tag{3.27}$$

to

$$\begin{matrix}A_1\\ \vdots \\ C\end{matrix} \tag{3.28}$$

DEFINITION 11

We say a pair of introduction-plus-elimination rules for a connective £ *is* **harmonious** *if* *(i) A£B is the strongest thing we can infer from the premisses for £-introduction, and*

(ii) A£B is the weakest thing that (with the other premisses to the £-elimination rule, if any[5]) implies the conclusion of the £-elimination rule.

Plonk and tonk

What we have shown above is that the rules for $\to$, $\wedge$ and $\vee$ are harmonious.

3.4.3 Maximal Formulæ

... [*for enthusiasts only*!]

The first occurrence of '$A \to B$' in proof 3.25 page 73 above is a bit odd. It's the output of a $\to$-introduction and at the same time the (major) premiss of an $\to$-elimination. (We say such a formula is *maximal*.). That feature invites the simplification that we showed there. Presumably this can always be done? Something very similar happens with the occurrence of '$A_1 \vee A_2$' in proof 3.27 p. 73. One might think so, but the situation is complex and not entirely satisfactory. One way into this is to try the following exercise:

EXERCISE 33

Deduce a contradiction from the two assumptions $p \to \neg p$ and $\neg p \to p$.

(These assumptions are of course really $p \to (p \to \bot)$ and $(p \to \bot) \to p$).

Try to avoid having a maximal formula in your proof. see [42].

3.5 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:

$$\frac{\begin{array}{c} A \to (B \to C) \\ \vdots \end{array}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{3.29}$$

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 ... try doing this on paper or on a board and you'll see what i mean. 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

$$\frac{\dfrac{\vdots}{A \to C} \to\text{-int}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{3.30}$$

[5] Do not forget that the elimination rule for £ might have premisses in addition to $A£B$: $\to$-elimination and $\vee$-elimination do, for example.

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 '$\vdash$'.[6] Then stage two (which was formula 3.29) can be described by the formula

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

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 3.30, 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.

Capital Greek letters denote sets of formulæ and lower-case Greek letters denote formulæ. A **sequent** is an expression $\Gamma \vdash \psi$ where Γ is a set of formulæ and ψ is a formula. $\Gamma \vdash \psi$ says that there is a deduction of ψ 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.

plonk and tonk are clearly not harmonious!

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

$$\begin{array}{c}
\underline{A \to (B \to C), (A \to B), A \vdash C} \\
\underline{A \to (B \to C), (A \to B) \vdash (A \to C)} \\
\underline{A \to (B \to C) \vdash (A \to B) \to (A \to C)} \\
\vdash (A \to (B \to C)) \to ((A \to B) \to (A \to C))
\end{array}$$

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

and this means we are using a rule

$$\frac{\Gamma, A \ \vdash \ B}{\Gamma \ \vdash \ A \to B} \to \mathrm{R} \tag{3.31}$$

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

$$\vee L: \ \frac{\Gamma, \psi \vdash \Delta \quad \Gamma', \phi \vdash \Delta'}{\Gamma \cup \Gamma', \underline{\psi \vee \phi} \vdash \Delta \cup \Delta'} \qquad \vee \text{ R: } \frac{\Gamma \vdash \Delta, \phi}{\Gamma \vdash \Delta, \underline{\psi \vee \phi}}$$

$$\vee \text{ R: } \frac{\Gamma \vdash \Delta, \psi}{\Gamma \vdash \Delta, \underline{\psi \vee \phi}}$$

$$\wedge L: \ \frac{\Gamma, \psi, \phi \vdash \Delta}{\Gamma, \underline{\psi \wedge \phi} \vdash \Delta} \qquad \wedge R: \ \frac{\Gamma \vdash \Delta, \psi \quad \Gamma' \vdash \Delta', \phi}{\Gamma \cup \Gamma' \vdash \Delta \cup \Delta', \underline{\psi \wedge \phi}}$$

$$\neg L: \ \frac{\Gamma \vdash \Delta, \psi}{\Gamma, \underline{\neg\psi} \vdash \Delta} \qquad \neg R: \ \frac{\Gamma, \psi \vdash \Delta}{\Gamma \vdash \Delta, \underline{\neg\psi}}$$

This may be a bit compsci-ish for our audience

$$\to L: \ \frac{\Gamma \vdash \Delta, \phi \quad \Gamma', \psi \vdash \Delta'}{\Gamma \cup \Gamma', \underline{\phi \to \psi} \vdash \Delta \cup \Delta'} \qquad \to R: \ \frac{\Gamma, \psi \vdash \Delta, \phi}{\Gamma \vdash \Delta, \underline{\psi \to \phi}}$$

$$\text{Contraction-L:} \frac{\Gamma, \psi, \psi \vdash \Delta}{\Gamma, \underline{\psi} \vdash \Delta}; \qquad \text{Contraction-R: } \frac{\Gamma \vdash \Delta, \psi, \psi}{\Gamma \vdash \Delta, \underline{\psi}};$$

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

$$\text{Cut:} \qquad \frac{\Gamma \vdash \Delta, \underline{\psi} \quad \Gamma', \underline{\psi} \vdash \Delta'}{\Gamma \cup \Gamma' \vdash \Delta \cup \Delta'}$$

In this box I have followed the universal custom of writing 'Γ, ψ' for '$\Gamma \cup \{\psi\}$; 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. This might sound odd, but it starts to look natural quite early, and you will get used to it easily.

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 $p \to \bot$. (see appendix 11.2.3, page 217)

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

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

I shall explain soon (section 3.5.3) 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.

Here we are already allowing multiple formulæ on the right. Naughty!

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.

So we can't really think of a sequent as saying that there is a proof-of-something-on-the-right that uses premisses on the left, however nice that sounds, but by keeping that thought in mind one keeps up the good habit of thinking of sequents as *meta*formulæ, 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 here, is that sequent proofs with more than formula on the right correspond to natural deduction proofs using the rule of double negation.

multisets
Display this properly

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**.

EXERCISE 34 *Find sequent proofs for the formulæ in exercise 28 (page 66). 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.

When (if ever) do we talk about confluence of these rules?

3.5.1 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:

$$\frac{A, B \vdash C}{A \wedge B \vdash C} \wedge \mathrm{L} \tag{3.32}$$

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.

$$\frac{\Gamma \vdash \Delta, A \qquad \Gamma, B \vdash \Delta}{\Gamma, A \rightarrow B \vdash \Delta} \rightarrow \mathrm{L} \tag{3.33}$$

Assume the two sequents above the line. We want to use them to show that there is a derivation of something in Δ from $\phi \to \psi$ 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 ϕ. But we have $\phi \to \psi$, so we now have ψ. But then the second sequent above the line tells us that we can infer something in Δ.

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.

Need exercises here

Explain "effective"

(i) If $\mathcal{D}$ does not use A or B then it is a witness to the truth of $\Gamma, A \wedge B \vdash \Delta$;

(ii) If it uses either A or B (or both) then we can append[7] one (or two) applications of $\wedge$-elimination to it to obtain a new proof that is a witness to the truth of $\Gamma, A \wedge B \vdash \Delta$

'witness'

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

finish this off, with a picture

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 $\Gamma \vdash \phi$ arose as a way of saying that there was a proof of ϕ 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:

3.5.2 The rule of cut

It's not hard to check that—in the formula '*cut*' below—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,

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

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

[*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 ("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 $\vdash \Phi$ then any formula appearing anywhere in the proof is a subformula of Φ. 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 $\phi \vdash \psi$. 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 ψ, or you can use the rule-on-the-left for the chief connective of ϕ. 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: at all events the number of possibilities is finite.) If you are allowed cut then the task of proving $\phi \vdash \psi$ can spawn the two tasks of proving the two sequents

$$\phi \vdash \psi, \theta \quad \text{and} \quad \theta, \phi \vdash \psi$$

and θ could be anything at all! This means that the task of finding a proof of $\phi \vdash \psi$ 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 eventually 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 want to avoid infinite searches and eschew the rule of cut then if we are to be sure we are not missing out on some of the fun we will 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[8] 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 property. Therefore, if we manage to show that every sequent provable in the sequent version of T has a cut-free proof then

[8] 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. No but seriously... the empty sequent is not a naked turnstile but a turnstile flanked by two copies of the empty list of formulæ.

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.

3.5.3 Two tips

3.5.3.1 Keep a copy!!

One thing 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

$$\Gamma \vdash \phi \to \psi \tag{(1)}$$

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

$$\Gamma, \phi \vdash \psi \tag{2}$$

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

$$\Gamma, \phi \vdash \psi, \phi \to \psi \tag{3}$$

The point is that if you do a $\to$-R to sequent (3) you get sequent (1). Thus you get the same result as if you had done a $\to$-R to sequent (2). Sometimes *keeping a copy* of the eigenformula in this way is the only way of finding a proof.

Insert discussion of $\neg$-R rule here

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 35 *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:

$$\cfrac{\cfrac{\cfrac{\Gamma \vdash A, B}{\Gamma \vdash A \vee B, B}\ \vee\mathrm{R}}{\Gamma \vdash A \vee B, A \vee B}\ \vee\mathrm{R}}{\Gamma \vdash A \vee B}\ \text{contraction-R} \tag{3.34}$$

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

3.5.3.2 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 for logics other than classical logic, for which simple tautology-checking of this kind is not available. See section 3.11, p. 96.

Wrong! You don't need to keep a copy! Wrong illustration

3.5.4 Exercises

You can now attempt to find sequent proofs for all the formulæ in exercise 28 page 66. At this stage you can also attempt exercise 38 on page 84.

"classical"?

Exercise 36 *Find proofs of the following sequents*

$A \vee B \vdash \neg A \to B$

$\neg A \to B \vdash A \vee B$ *(defines $\vee$ in terms of $\neg$ and $\to$)*

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

$\neg(A \to \neg B) \vdash A \wedge B$ *(defines $\wedge$ in terms of $\neg$ and $\to$)*

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

$\neg(A \wedge \neg B) \vdash A \to B$ *(defines $\to$ in terms of $\neg$ and $\wedge$)*

We usually treat seq calculus as arising from ND but in fact the proofs that sequent calculus reasons about could be any proofs at all—even Hilbert-style proofs as below.

Insert Chapter break here …?

Exercise 37 *Find proofs of the following sequents*

$A \vee B \vdash \neg A \to B$

$\neg A \to B \vdash A \vee B$ *(defines $\vee$ in terms of $\neg$ and $\to$)*

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

$\neg(A \to \neg B) \vdash A \wedge B$ *(defines $\wedge$ in terms of $\neg$ and $\to$)*

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

$\neg(A \wedge \neg B) \vdash A \to B$ *(defines $\to$ in terms of $\neg$ and $\wedge$)*

If you are a first-year who is not interested in pursuing Logic any further you can skip the rest of this chapter and go straight to chapter 4. However, even students who do plan to refuse this particular jump should attempt exercise **??**.

Is this the right place for these exercises?

3.6 Hilbert-style Proofs

In this style of proof we have only three axioms

K: $A \to (B \to A)$

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

T: $(\neg A \to B) \to ((\neg A \to \neg B) \to A)$

and the rules of *modus ponens* and substitution. 'K' and 'S' are standard names for the first two axioms. There is a good reason for this, which we will see in chapter 7. The third axiom does not have a similarly standard name.

Notice that only two connectives appear here: $\to$ and $\neg$. How are we supposed to prove things about $\wedge$ and $\vee$ and so on? The answer is that we define the other connectives in terms of $\to$ and $\neg$, somewhat as we did on page 43—except that there we defined our connectives in terms of a different set of primitives.

Here is an example of a proof in this system:

1. $A \to ((A \to A) \to A)$	Instance of K
2. $(A \to ((A \to A) \to A)) \to$ $((A \to (A \to A)) \to (A \to A))$	Instance of S
3. $(A \to (A \to A)) \to (A \to A)$	Modus Ponens (1) and (2)
4. $A \to (A \to A)$	Instance of K:
5. $A \to A$	Modus Ponens (3) and (4)

I thought I would give you an illustration of a proof before giving you a formal definition. Here is the definition:

DEFINITION 12 *A Hilbert-style proof is a list of formulæ wherein every formula is either an axiom or is obtained from earlier formulæ in the list by* modus ponens *or substitution.*

Some comments...

1. We can do without the rule of substitution, simply by propagating the substitutions we need back to the axioms in the proof and ruling that a substitution instance of an axiom is an axiom.

2. We can generalise this notion to allow assumptions as well as axioms. That way we have—as well as the concept of an outright (Hilbert)-proof—the concept of a *Hilbert-proof of a formula from a list of assumptions.*

3. An initial segment of a Hilbert-style proof is another Hilbert-style proof—of the last formula in the list.

4. Hilbert-style proofs suffer from not having the subformula property, as the boxed proof aboveshows.

EXERCISE 38 *You have probably already found natural deduction proofs for K and S. If you have not done so, do it now. Find also a natural deduction proof of T, the third axiom. (You will need the rule of double negation).*

EXERCISE 39 *Go back to Zarg (exercise 32 p. 68) and—using the truth-table for $\neg$ that you decided that the Zarglings use—check that the Zarglings do not believe axiom T to be a tautology.*

I will spare you the chore of testing whether or not the Zarglings believe S to be a tautology. One reason is that it would involve writing out a truth-table with a dispiritingly large number of rows. How many rows exactly?

EXERCISE 40 *[For enthusiasts only]*

Find Hilbert-style proofs of the following tautologies
(a) $B \to \neg\neg B$.
(b) $\neg A \to (A \to B)$.
(c) $A \to (\neg B \to \neg(A \to B))$.
(d) $(A \to B) \to ((\neg A \to B) \to B)$.

reconstruct this exercise

Notice how easy it is to prove that the Hilbert-style proof system is sound! After all, every substitution-instance of a tautology is a tautology, and if $A \to B$ and A are tautologies, so is B.

3.6.1 The Deduction Theorem

In this Hilbert-style proof system the only rules of inference are *modus ponens* and substitution. Establishing that $A \to A$ is a theorem—as we did above—is quite hard work in this system. If we had a derived rule that said that if we have a Hilbert-style proof of A using a premiss B then we have a Hilbert-style proof of $A \to B$ then as a special case we would know that there was a Hilbert-proof of $A \to A$.

To justify a derived rule that says that if we have a Hilbert-proof of A from B then there is a Hilbert-proof of $A \to B$ we will have to show how to transform a proof of B with an assumption A in it into a proof of $A \to B$. Let the Hilbert-proof of B be the list whose ith member is B_i. The first thing we do is replace every B_i by $A \to B_i$ to obtain a new list of formulæ. This list isn't a proof, but it is the beginnings of one.

Suppose B_k had been obtained from B_i and B_j by *modus ponens* with B_i as major premiss, so B_i was $B_j \to B_k$. This process of whacking '$A \to$' on the front of every formula in the list turns these into $A \to (B_j \to B_k)$ and $A \to B_j$. Now altho' we could obtain B_k from B_j and $B_j \to B_k$ by *modus ponens* we clearly can't obtain $A \to B_k$ from $A \to B_j$ and $A \to (B_j \to B_k)$ quite so straightforwardly. However we can construct a little Hilbert-style proof of $A \to B_k$ from $A \to B_j$ and $A \to (B_j \to B_k)$ using S. When revising you might like to try covering up the next few formulæ and working it out yourself.

1. $(A \to (B_j \to B_k)) \to ((A \to B_j) \to (A \to B_k))$	S
2. $A \to (B_j \to B_k)$	
3. $(A \to B_j) \to (A \to B_k)$	*modus ponens* (1), (2)
4. $A \to B_j$	
5. $A \to B_k$	*modus ponens* (3), (4)

Lines (2) and (4) I haven't labelled. Where did they come from? Well, what we have just seen is an explanation of how to get $A \to B_k$ from $A \to (B_j \to B_k)$ and $A \to B_j$ given that we can get B_k from B_j and $B_j \to B_k$. What the box shows us is how to rewrite any **one** application of *modus ponens*. What we have to do to prove the deduction theorem is to do this trick to every occurrence of *modus ponens*.

This needs massive expansion

If we apply this process to:

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

we obtain

1. $(A \to B) \to$
   $(((A \to B) \to (A \to B)) \to (A \to B))$ Instance of K
2. $((A \to B) \to (((A \to B) \to (A \to B)) \to (A \to B))) \to$
   $(((A \to B) \to ((A \to B) \to (A \to B))) \to ((A \to B) \to (A \to B)))$ Instance of S
3. $((A \to B) \to ((A \to B) \to (A \to B))) \to ((A \to B) \to (A \to B))$
   Modus Ponens (1) and (2)
4. $(A \to B) \to ((A \to B) \to (A \to B))$ Instance of K:
5. $(A \to B) \to (A \to B)$ Modus Ponens (3) and (4)
6. $((A \to B) \to (A \to B)) \to$
   $(((A \to B) \to A) \to ((A \to B) \to B)$ Instance of S
7. $((A \to B) \to A) \to ((A \to B) \to B)$ Modus ponens (6), (5)
8. A Assumption
9. $A \to ((A \to B) \to A)$ Instance of K.
10. $(A \to B) \to A$ Modus ponens (9), (8).
11. $(A \to B) \to B$ modus ponens (10), (7).

(Of course the annotations at the beginning and end of the lines are not part of the proof but are part of a commentary on it. That's the language-metalanguage distinction again.)

Revise this: it isn't correct

THEOREM 13 *If* $\Gamma, A \vdash B$ *then* $\Gamma \vdash A \to B$

3.7 Interpolation

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 those three. There is a very strong intuition of *irrelevance* at work here. It's strong, but it's so natural that you probably didn't notice that you had it. The time has now come to discuss it. But we need a bit more gadgetry first.

More to do here

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 was 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''"

Cite the text

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 20 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 20 that

$p \rightarrow \top$ is logically equivalent to $\top$
$\top \rightarrow p$ is logically equivalent to p
$\bot \rightarrow p$ is logically equivalent to $\top$
$p \rightarrow \bot$ 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:

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

1. In the case where p is true this becomes

   $$(\top \vee A) \to (B \wedge C)$$

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

   $$(p \vee A) \to (B \wedge C)$$

   becomes

   $$\top \to (B \wedge C)$$

   which is just

   $$B \wedge C.$$

   So we get

   $$p \wedge (B \wedge C).$$

2. In the case where p is false this becomes

   $$(\bot \vee A) \to (B \wedge C)$$

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

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

   and

   $$\neg p \wedge (A \to (B \wedge C))$$

So $(p \vee A) \to (B \wedge C)$ is equivalent to

$$(p \wedge (B \wedge C)) \vee (\neg p \wedge (A \to (B \wedge C)))$$

	$\overset{[\![p]\!] = \top}{\leftsquigarrow}$		$\overset{[\![p]\!] = \bot}{\rightsquigarrow}$	
$B \vee C$		$p \wedge (B \vee C)$		$\bot$
B		$(p \wedge B) \vee (\neg p \wedge C)$		C
$B \wedge C$		$p \longleftrightarrow (B \wedge C)$		$\neg B \vee \neg C$
$A \vee B$		$p \to (A \vee B)$		$\top$
$B \wedge C$		$(p \vee A) \to (B \wedge C)$		$A \to (B \wedge C)$

This illustrates what one might call

THEOREM 14 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 $(A_1 \wedge p) \vee (A_2 \wedge \neg p)$.

Evidently

$$A \longleftrightarrow (A \wedge (p \vee \neg p))$$

which we can distribute to

$$A \longleftrightarrow (A \wedge p) \vee (A \vee \neg p)$$

but we can simplify $A \wedge p$ to $A' \wedge p$ where A' is the result of substituting $\top$ for p in A; A'' similarly, obtaining

$$A \longleftrightarrow (A' \wedge p) \vee (A'' \vee \neg p)$$

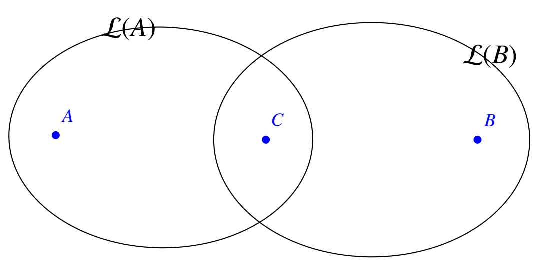
■

DEFINITION 15 *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, since $A \to B$ is a tautology 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 $\perp$ is the disjunction of the empty set of formulæ (see section 2.4.2 and therefore both '$\top$' and '$\perp$' belong to the language over the empty alphabet—which is to say to $\mathcal{L}(A) \cap \mathcal{L}(B)$. We established that either $A \to \perp$ is a tautology (in which case A is the negation of a tautology) or $\top \to B$ is a tautology (in which case B is a tautology). But, since both $A \to \top$ and $\perp \to B$ are always tautologies (as we saw in exercise 20) we can tell ourselves that what we have established is that there is some formula C in the common vocabulary such that both $A \to C$ and $C \to B$ are tautologies. C must be either '$\top$' or '$\perp$'.

supply here the result of splitting Carroll's example "over"



Here we use the symbol '∩' for the first time... do we need to explain it?

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

THEOREM 16 *(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: if v is a valuation making A' true then it needn't mention 'p' at all, so we can extend it to a valuation $v' =: v \cup \{\langle p, \mathtt{true}\rangle\}$ that makes 'p' true. So v' is a valuation making $(p \wedge A')$ true, so it must make $(p \wedge B')$ true. So v must have made A' true. (A'' and B'' *mutatis mutandis*.)

Next 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 every 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 41 *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"?

3.8 Completeness of Propositional Logic

This section is not recommended for first-years.

[this section uses notions from comp th that we haven't introduced]

How about the proof i used for the CS MPhil?

We have to use one or the other

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.

3.8.1 Completeness

Completeness is harder than soundness. When we say that the system of rules of natural deduction is complete we mean that it provides proofs of every tautology.

A **row** is a conjunction of atomics and negatomics in which every propositional letter appears precisely once. There is an obvious correlation between rows and valuations.

For A a propositional formula let A^* be the disjunction of all the rows that make A come out true.

We write '$\vdash \phi$' for "there is a natural deduction proof of ϕ".

LEMMA 17 *For all propositional formulæ A, there is a natural deduction proof of $A \longleftrightarrow A^*$.*

Proof:

By structural induction on formulæ. The base case concerns individual propositional letters and $\bot$, the false. If A is a propositional letter or $\bot$ then A^* is just A. Clearly $\vdash A \longleftrightarrow A^*$.

There is an induction step for each of $\wedge$, $\vee$ and $\to$.

This section needs to be cut drastically

$\wedge$ $(A \wedge B)^*$ is the disjunction of those rows common to A^* and B^*, and is therefore interdeducible with $A^* \wedge B^*$. By induction hypothesis A is interdeducible with A^* and B is interdeducible with B^* so $A^* \wedge B^*$ (which we have just seen is interdeducible with $(A \wedge B)^*$) is interdeducible with $A \wedge B$.

$\vee$ $(A \vee B)^*$ is the disjunction of those rows appearing in the truth-table for $A \vee B$, and is therefore interdeducible with $A^* \vee B^*$. By induction hypothesis A is interdeducible with A^* and B is interdeducible with B^* so $A^* \vee B^*$ (which we have just seen is interdeducible with $(A \vee B)^*$) is interdeducible with $A \vee B$.

$\to$ $(A \to B)^*$ is of course the disjunction of all rows that make A false or B true. We prove the two directions separately.

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

Let r be one of the rows of $(A \to B)^*$.

(i) If r is a row that makes B true, then it is a disjunct of B^* so $\vdash r \to B^*$ whence $\vdash r \to B$ by induction hypothesis. So definitely $\vdash r \to (A \to B)$.

(ii) If r is a row that makes A false, then it is inconsistent with every row that makes A true, so it is inconsistent with their disjunction—which is A^*. A and A^* are interdeducible by induction hypothesis, so $\vdash r \to (A \to \bot)$. But $\vdash (A \to \bot) \to (A \to B)$, so $\vdash r \to (A \to B)$.

Either way, if r is a row of $(A \to B)$, $\vdash r \to (A \to B)$. $(A \to B)^*$ is the disjunction of all the rows of $A \to B$ so, by $\vee$-elimination, $\vdash (A \to B)^* \to (A \to B)$.

$\vdash (A \to B) \to (A \to B)^*$.

Assume $A \to B$ and $\neg(A \to B)^*$. We will deduce the false. $\neg(A \to B)^*$ denies every row in B^*, so refutes B^* and therefore refutes B (by induction hypothesis). $\neg B$ gives $\neg A$ by *modus tollens*. Now by induction bypothesis on A we can refute every disjunct in A^* (every row that makes A true). But our denial of $(A \to B)^*$ refuted every row that made A false. So we have refuted *all* rows! Recall that we can prove the disjunction of all the rows. $(A \vee \neg A)^*$ is provable. This gives us the contradiction we seek. Then we use the rule of classical negation to deduce $(A \to B)^*$. We now use $\to$-introduction to obtain a proof of $(A \to B) \to (A \to B)^*$.

■

We can now prove

THEOREM 18 *Every truth-table tautology has a natural deduction proof*

Proof: Suppose that A is a truth-table tautology. Observe that, if $a_1 \ldots a_n$ are the propositional letters that appear in A, then we can prove the disjunction of the 2^n rows to be had from $a_1 \ldots a_n$. We do this by induction on n. Since A is a truth-table tautology this disjunction is in fact A^*. Lemma 17 tells us that there is a natural deduction proof of $A \longleftrightarrow A^*$ so we conclude that there is a natural deduction proof of A. ■

If you are still less than 100% happy about this, attempt the following exercise:

EXERCISE 42

1. *Find a natural deduction proof of* $A \vee \neg A$ *(case* $n = 1$*);*
2. *Find a natural deduction proof of* $(A \wedge B) \vee (\neg A \wedge B) \vee (A \wedge \neg B) \vee (\neg A \wedge \neg B)$ *(case* $n = 2$*);*
3. *Explain the induction step.*

Admittedly this seems excessively laborious but the result is important even if the proof isn't. Important too, is the experience of discovering that soundness proofs are easy and completeness proofs are hard(er)!

3.8.2 Completeness using Sequents

We can show easily that if ϕ is a truth-table tautology then the sequent $\vdash \phi$ has a proof using our sequent rules.

This is probably the correct place to discuss **confluence**.

3.9 What is a Completeness Theorem?

The completeness+soundness result we have just seen for the rules of natural deduction and the concept of a propositional tautology connects two sets. One set is defined by a semantical property (being satisfied by all valuations) and the other is defined by a syntactic property (being generated by a set of rules). Indeed the property of being generated by a set of rules is equivalent to being what is called in the literature a recursively enumerable ("r.e.") set or (more illuminatingly) a **semidecidable** set. We say a set X is semidecidable if there is a procedure $\mathcal{P}$ that will authenticate its members (so whenever a candidate for membership is in fact a member this will be confirmed in finite time). Notice that this does not require that the method $\mathcal{P}$ will reject any unsuitable candidate in finite time. If there is a method that will additionally reject any unsuitable candidate in finite time then the complement of X, too, is semidecidable and we say X is **decidable** ("recursive" is the old terminology).

So typically a completeness theorem is an assertion about two sets X and Y where X is a set defined semantically (as it might be, the set of tautologies) and Y is a semidecidable set defined by a syntactic criterion (as it might be the set of strings that have natural deduction proofs) and says that $X = Y$.

You may have felt tempted to say that the completeness theorem for propositional logic was no big deal. So we have this set of tautologies . . . well, cook up some rules that generate them all. What's the problem? The problem is that there might be no such set of rules. We will see later that there are Logics which cannot be captured by a set of rules in this way: every set of rules either generates things it shouldn't or fails to generate some things it should. (Trakhtenbrot's theorem; second-order logic)

"structural induction"?!

There is a completeness theorem for predicate logic, as we shall see. There is also a completeness theorem for constructive logic, but that is beyond the scope of this book.

Do we actually see these things?

THEOREM 19 *The completeness theorem for propositional logic.*

The following are equivalent:

(1) ϕ is provable by natural deduction.

(2) ϕ is provable from the three axioms K, S and T.

(3) ϕ is truth-table valid.

Proof: We will prove that (3) → (2) → (1) → (3).

(2) → (1).

First we show that all our axioms follow by natural deduction–by inspection. Then we use induction: if there are natural deduction proofs of A and $A \to B$, there is a natural deduction proof of B!

(1) → (3).

To show that everything proved by natural deduction is truth-table valid we need only note that, for each rule, if the hypotheses are true (under a given valuation), then the conclusion is too. By induction on composition of rules this is true for molecular proofs as well. If we have a molecular proof with *no* hypotheses, then vacuously they are all true (under a given valuation), so the conclusion likewise is true (under a given valuation). But the given valuation was arbitrary, so the conclusion is true under all valuations.

"constructive" doesn't appear until p 145.

(3) → (2).

(This proof is due to Mendelson [30] and Kalmár [25].) Now to show that all tautologies follow from our three axioms.

At this point we must invoke exercise 40, since we need the answers to complete the proof of this theorem. It enjoins us to prove the following:

(a) $B \to \neg\neg B$.
(b) $\neg A \to (A \to B)$.
(c) $A \to (\neg B \to \neg(A \to B))$.
(d) $(A \to B) \to ((\neg A \to B) \to B)$.

If we think of a propositional formula in connection with a truth-table for it, it is natural to say things like: $p \longleftrightarrow q$ is true as long as p and q are both true or both false, and false otherwise. Thus truth-tables for formulæ should suggest to us deduction relations like

$$A, B \vdash A \longleftrightarrow B,$$

$$\neg A, \neg B \vdash A \longleftrightarrow B,$$

and similarly

$$A, \neg B \vdash \neg(A \longleftrightarrow B).$$

To be precise, we can show:

Let A be a molecular wff containing propositional letters $p_1 \ldots p_n$, and let f be a map from $\{k \in \mathbb{N} : 1 \leq k \leq n\}$ to $\{\texttt{true}, \texttt{false}\}$. If A is satisfied in the row of the truth-table where p_i is assigned truth-value $f(i)$, then

$$P_1 \ldots P_n \vdash A,$$

where P_i is p_i if $f(i) = \texttt{true}$ and $\neg p_i$ if $f(i) = \texttt{false}$. If A is not satisfied in that row, then

$$P_1 \ldots P_n \vdash \neg A,$$

and we prove this by a straightforward induction on the rectype of formulæ.

First use of 'arbitrary' in this sense

We have only two primitive connectives, $\neg$ and $\to$, so two cases.

$\neg$

Let A be $\neg B$. If B takes the value $\texttt{true}$ in the row $P_1 \ldots P_n$, then, by the induction hypothesis, $P_1 \ldots P_n \vdash B$. Then, since $\vdash p \to \neg\neg p$ (this is exercise 40 p. 84), we have $P_1 \ldots P_n \vdash \neg\neg B$, which is to say $P_1 \ldots P_n \vdash \neg A$, as desired. If B takes the value $\texttt{false}$ in the row $P_1 \ldots P_n$, then, by the induction hypothesis, $P_1 \ldots P_n \vdash \neg B$. But $\neg B$ is A, so $P_1 \ldots P_n \vdash A$.

$\rightarrow$

Let A be $B \rightarrow C$.

Case (1): B takes the value `false` in row $P_1 \ldots P_n$. If B takes the value `false` in row $P_1 \ldots P_n$, then A takes value `true` and we want $P_1 \ldots P_n \vdash A$. By the induction hypothesis we have $P_1 \ldots P_n \vdash \neg B$. Since $\vdash \neg p \rightarrow (p \rightarrow q)$ (this is exercise 40(b)), we have $P_1 \ldots P_n \vdash B \rightarrow C$, which is $P_1 \ldots P_n \vdash A$.

Case (2): C takes the value `true` in row $P_1 \ldots P_n$. Since C takes the value T in row $P_1 \ldots P_n$, A takes value `true`, and we want $P_1 \ldots P_n \vdash A$. By the induction hypothesis we have $P_1 \ldots P_n \vdash C$, and so, by K, $P_1 \ldots P_n \vdash B \rightarrow C$, which is to say $P_1 \ldots P_n \vdash A$.

Case (3): B takes value `true` and C takes value `false` in row $P_1 \ldots P_n$. A therefore takes value `false` in this row, and we want $P_1 \ldots P_n \vdash \neg A$. By the induction hypothesis we have $P_1 \ldots P_n \vdash B$ and $P_1 \ldots P_n \vdash \neg C$. But $p \rightarrow (\neg q \rightarrow \neg(p \rightarrow q))$ is a theorem (this is exercise 40(c)) so we have $P_1 \ldots P_n \vdash \neg(B \rightarrow C)$, which is $P_1 \ldots P_n \vdash \neg A$.

Suppose now that A is a formula that is truth-table valid and that it has propositional letters $p_1 \ldots p_n$. Then, for example, both $P_1 \ldots P_{n-1}, p_n \vdash A$ and $P_1 \ldots P_{n-1}, \neg p_n \vdash A$, where the capital letters indicate an arbitrary choice of $\neg$ or null prefix as before. So, by the deduction theorem, both p_n and $\neg p_n \vdash (P_1 \wedge P_2 \ldots \wedge P_{n-1}) \rightarrow A$ and we can certainly show that $(p \rightarrow q) \rightarrow (\neg p \rightarrow q) \rightarrow q$ is a theorem (this is exercise 40(d)), so we have $P_1 \ldots P_{n-1} \vdash A$, and we have peeled off one hypothesis. Clearly this process can be repeated as often as desired to obtain $\vdash A$. ■

There is another assertion—equivalent to theorem 19—which, too, is known as the completeness theorem. Sometimes it is a more useful formulation.

‘rectype’ not explained

COROLLARY 20 *ϕ is consistent (not refutable from the axioms) iff there is a valuation satisfying it.*

first occ of ‘iff’

Proof: $\nvdash \neg\phi$ (i.e., ϕ is consistent) iff $\neg\phi$ is not tautologous. This is turn is the same as ϕ being satisfiable. ■

3.10 Compactness

We close this chapter with an observation which—altho’ apparently banal—actually has considerable repercussions. Suppose there is a deduction of a formula ϕ from a set Γ of formulæ. In principle Γ could of course be an infinite set (there are infinitely many formulæ after all) but any deduction of ϕ from Γ is a finite object and can make use of only finitely many of the formulæ in Γ. This tells us that

THEOREM 21 *If θ follows from Γ then it follows from a finite subset of Γ*

This is actually pretty obvious. So obvious in fact that one might not think it was worth pointing out. However, it depends sensitively on some features one might take for granted and therefore not notice. If we spice up our language into something more expressive in a manner that does not preserve those nice features we might find that it isn't true any more. For example it won't work if our formulæ can be infinitely long or if our proofs are allowed to be infinitely long.

Here is a realistic illustration. Since the infinite sequence 0, 1, 2, 3, ... exhausts the natural numbers, it seems entirely reasonable to adopt a rule of inference:

$$\frac{F(0),\ F(1),\ F(2),\ \ldots}{(\forall n)(F(n))}$$

... where there are of course infinitely many things on the top line. This is called the ω-**rule**[9]. There are infinitely many premisses. However it is clear that the conclusion does not follow from any finite subset of the premisses, so we would not normally be licenced to infer the conclusion. Thus the ω-rule is strong: it enables us to prove things we would not otherwise be able to prove.

3.10.1 Why "compactness"?

The word 'compactness' comes from nineteenth century topologists' attempts to capture the difference between plane figures of finite extent (for example, the circle of radius 1 centred at the origin) and plane figures of infinite extent (for example the left half-plane)—and to do this without talking about any numerical quantity such as *area*. The clever idea is to imagine an attempt to cover your chosen shape with circular disks. A set of disks that covers the figure in question is a *covering* of the figure. It's clearly going to take infinitely many disks to cover the half-plane. A plane figure F that is finite (perhaps 'bounded' is a better word) in the sense we are trying to capture has the feature that whenever we have a set O of disks that cover F then there is a finite subset $O' \subseteq O$ of disks that also covers F. Such a figure is said to be **compact**.

'iff' not defined yet

The connection between these two ideas (compactness in topology, and the finite nature of logic) was made by Tarski.

3.11 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.

[9] 'ω' (pronounced 'omega') is the last letter of the Greek alphabet. The capital form is 'Ω'.

In particular we will be studying in later chapters both predicate calculus and constructive logic (where we reason without using the laws of excluded middle and double negation). In neither of these cases are truth-tables sensitive enough for us to be able to use them *on their own* for checking the validity of inferences.

3.12 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 43 *We saw in exercise 20 page 32 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.*[10]

3.13 Formal Semantics for Propositional Logic

This section not recommended for first years

The key to not getting lost in this enterprise is to bear in mind that the expressions of propositional logic are built up from atomic formulæ (letters) whose meaning is not reserved: they can be anything: *Herbert takes the flat* or *Herbert's wife insists on keeping Fido*, but the symbols in the logical vocabulary—'$\wedge$', '$\vee$' and so on—emphatically *are* reserved.

A valuation [for propositional language] is a function that assigns truth-values (not meanings!) to the primitive letters of that language. We will use the letter 'v' to range over valuations. Now we define a satisfaction relation `sat` between valuations and complex expressions.

DEFINITION 22

A complex expression ϕ *might be a propositional letter and—if it is—then* $\mathtt{sat}(v, \phi)$ *is just* $v(\phi)$, *the result of applying* v *to* ϕ;

If ϕ *is the conjunction of* ψ_1 *and* ψ_2 *then* $\mathtt{sat}(v, \phi)$ *is* $\mathtt{sat}(v, \psi_1) \wedge \mathtt{sat}(v, \psi_2)$;

If ϕ *is the disjunction of* ψ_1 *and* ψ_2 *then* $\mathtt{sat}(v, \phi)$ *is* $\mathtt{sat}(v, \psi_1) \vee \mathtt{sat}(v, \psi_2)$;

If ϕ *is the conditional whose antecedent is* ψ_1 *and whose consequent is* ψ_2 *then* $\mathtt{sat}(v, \phi)$ *is* $\mathtt{sat}(v, \psi_1) \to \mathtt{sat}(v, \psi_2)$;

If ϕ *is the negation of* ψ_1 *then* $\mathtt{sat}(v, \phi)$ *is* $\neg\mathtt{sat}(v, \psi_1)$;

If ϕ *is the biconditional whose two immediate subformulæ are* ψ_1 *and* ψ_2 *then* $\mathtt{sat}(v, \phi)$ *is* $\mathtt{sat}(v, \psi_1) \longleftrightarrow \mathtt{sat}(v, \psi_2)$.

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

Notice that here i am using the letters 'ϕ' and 'ψ_1' and 'ψ_2' as variables that range over formulæ, as in the form of words "If ϕ is the conjunction of ψ_1 and ψ_2 then …". They are not abbreviations of formulæ. There is a temptation to write things like

"If ϕ is $\psi_1 \wedge \psi_2$ then $\mathtt{sat}(v, \phi)$ is $\mathtt{sat}(v, \psi_1) \wedge \mathtt{sat}(v, \psi_2)$"

or perhaps

$$\mathtt{sat}(v, \psi_1 \wedge \psi_2) \text{ is } \mathtt{sat}(v, \psi_1) \wedge \mathtt{sat}(v, \psi_2) \tag{3.35}$$

Now although our fault-tolerant pattern matching enables us to see immediately what is intended, the pattern matching does, indeed, need to be fault-tolerant. (In fact it corrects the fault so quickly that we tend not to notice the processing that is going on.)

In an expression like '$\mathtt{sat}(v, \phi)$' the 'ϕ' has to be a name of a formula, as we noted above, not an abbreviation for a formula. But then how are we to make sense of

$$\mathtt{sat}(v, \psi_1 \wedge \psi_2) \tag{3.36}$$

The string '$\psi_1 \wedge \psi_2$' has to be the name of a formula. Now you don't have to be The Brain of Britain to work out that it has got to be the name of whatever formula it is that we get by putting a '$\wedge$' between the two formulæ named by 'ψ_1' and 'ψ_2'—and this is what your fault-tolerant pattern-matching wetware (supplied by Brain-Of-Britain) will tell you. But we started off by making a fuss about the fact that names have no internal structure, and now we suddenly find ourselves wanting names to have internal structure after all!

In fact there is a way of making sense of this, and that is to use the cunning device of *corner quotes* to create an environment wherein compounds of names of formulæ (composed with connectives) name composites (composed by means of those same connectives) of the formulæ named..

So 3.35 would be OK if we write it as

$$\mathtt{sat}(v, \ulcorner \psi_1 \wedge \psi_2 \urcorner) \text{ is } \mathtt{sat}(v, \psi_1) \wedge \mathtt{sat}(v, \psi_2) \tag{3.37}$$

Corner quotes were first developed in [33]. See pp 33-37. An alternative way of proceding that does not make use of corner quotes is instead to use an entirely new suite of symbols—as it might be '**and**' and '**or**' and so on, and setting up links between them and the connectives '$\wedge$' and so on in the object language so that—for example

$$\psi_1 \text{ and } \psi_2 \tag{A}$$

is the conjunction of ψ_1 and ψ_2. In other words, we want 'ψ_1 **and** ψ_2' to name the conjunction of ψ_1 with ψ_2. The only drawback to this is the need to conjure up an entire suite of symbols, all related suggestively to the connectives they are supposed to name… in the way that '**and**' names the symbol '$\wedge$'. Here one runs up against the fact that any symbols that are suitably suggestive will also be laden with associations from their other uses, and these associations may not be helpful. Suppose we were to use an ampersand instead of '**and**'; then the fact that it is elsewhere used instead of '$\wedge$' might cause the reader to assume it is a *synonym* for '$\wedge$' rather than a name of it. There is no easy way through.

3.14 Eager and Lazy Evaluation

Recall "recursive" from p. 26

The recursive definition of `sat` in the previous section gives us a way of determining what truth-value a formula receives under a valuation. Start with what the valuation does to the propositional letters (the leaves of the parse tree) and work up the tree. Traditionally the formal logic that grew up in the 20th century took no interest in how things like $\mathtt{sat}(\phi, v)$ were actually *calculated*. The recursive definition tells us uniquely what the answer must be, but it doesn't tell us uniquely how to calculate it.

The way of calculating $\mathtt{sat}(\phi, v)$ that we have just seen (start with what the valuation does to the propositional letters—the leaves of the parse tree—and work up the tree) is called **Eager evaluation** also known as **Strict evaluation**. But there are other ways of calculating that will give the same answer. One of them is the beguilingly named **Lazy evaluation** which we will now describe.

A nice example of formalisation

Consider the project of filling out a truth-table for the formula $A \wedge (B \vee (C \wedge D))$. One can observe immediately that any valuation (row of the truth-table) that makes 'A' false will make the whole formula false:

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

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

Now, in the remaining cases we can observe that any valuation that makes 'B' true will make the whole formula true.:

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

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

In the remaining cases any valuation that makes ‘C’ false will make the whole formula false.

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

A	$\wedge$	$(B$	$\vee$	$(C$	$\wedge$	$D))$
0	0	0		0		0
0	0	0		0		1
0	0	0		1		0
0	0	0		1		1
0	0	1		0		0
0	0	1		0		1
0	0	1		1		0
0	0	1		1		1
1	0	0	0	0	0	0
1	0	0	0	0	0	1
1	0	0	0	1	0	0^*
1	1	0	1	1	1	1^*
1	1	1	1	0		0
1	1	1	1	0		1
1	1	1	1	1		0
1	1	1	1	1		1

The starred ‘0^*’ and ‘1^*’ are the only cases where we actually have to look at the truth-value of D.

These illustrations concern evaluation in languages whose expressions evaluate to truth-values. The idea originally arose in connection with languages whose expressions evaluate to numbers or other data objects.

$$\texttt{if}\ x \geq 0\ \texttt{then}\ f(x)\ \texttt{else}\ g(x).$$

This expression evaluates to [the value of] $f(x)$ or to [the value of] $g(x)$, depending on what x evaluates to. There is no point in calculating both $f(x)$ and $g(x)$ when you are clearly going to need only one of them! So you evaluate lazily: first you evaluate x

to see whether it is above or below 0 and then you evaluate whichever of $f(x)$ and $g(x)$ that it turns out that you need.

Notice also in this connection that i might not even have to evaluate x *completely* in order to know which way to jump. If x is presented to me as a double-precision decimal number i have 12 decimal places to evaluate, but i will know already after evaluating the first of them whether x is positive or negative.

We noted above that you will get the one and the same truth-value for any given proposition under any given evaluation whether you evaluate eagerly or lazily. This makes it possible to think of [this part of] logic *statically*: it's safe to ignore or *hide* the process of evaluation. We are not interested in the *process* since we always reach the same *result*.

we seem to have covered it already in section 3.14

This is the point behind the joke about the three logicians who go into a bar.

> "Do you all want drinks?" asks the barman. "I don't know" says the first logician; "I don't know" says the second logician; "Yes!" says the third logician.

The point is that if the first logician *didn't* want a drink (s)he could evaluate "Logician 1 wants a drink $\wedge$ logician 2 wants a drink $\wedge$ logician 3 wants a drink" *lazily* and answer 'No'. (S)he doesn't, and therefore answers 'I don't know'... from which the other two logicians (who know (s)he is evaluating lazily) can infer that (s)he *does* want a drink.

Chapter 4

Predicate (first-order) Logic

4.1 Towards First-Order Logic

We saw earlier (in section 3.7) the following puzzle 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. (For some reason they always have beef?!)
>
> 1. If Barry takes salt, then either Cole or Lang takes only *one* of the two condiments, salt and mustard (and *vice versa*). If he takes mustard then either Dix takes neither condiment or Mill takes both (and *vice versa*).
> 2. If Cole takes salt, then either Barry takes only *one* condiment, or Mill takes neither (and *vice versa*). If he takes mustard then either Dix or Lang takes both (and *vice versa*).
> 3. If Dix takes salt, then either Barry takes neither condiment or Cole takes both (and *vice versa*). If he takes mustard then either Lang or Mill takes neither (and *vice versa*).
> 4. If Lang takes salt, then either Barry or Dix takes only *one* condiment (and *vice versa*). If he takes mustard then either Cole or Mill takes neither (and *vice versa*).
> 5. If Mill takes salt, then either Barry or Lang takes both condiments (and *vice versa*). If he takes mustard then either Cole or Dix takes only one (and *vice versa*).

The task Carroll sets us is to ascertain whether or not these conditions can in fact be met. I do not know the answer, and finding 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, nor have I any intention of doing it—I do have a life, after all) but it's a good idea to at least think (a bit) about some of the preparatory work that would be involved.

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 this internal structure is not visible to the logic. The logic regards 'ds' as a single propositional letter, so we do not yet have a logic that can see the structure we want: this first step is not enough, and we have to do a bit more if we are to make 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 to each other by the (meaning of the) relation symbol.

We can apply this to Lewis Carroll's problem on page 86 by having, for each condiment, a one-place predicate (of diners) of being a consumer of that condiment, and constant symbols 'd', 'l', 'm', 'b' and 'c' for Dix, Lang, Mill, Barry and Cole, respectively. I am going to write them in lower case because we keep upper case letters for predicates—relation symbols.

'$S(b)$': Barry takes salt	'$U(b)$': Barry takes mustard
'$S(m)$': Mill takes salt	'$U(m)$': Mill takes mustard
'$S(c)$': Cole takes salt	'$U(c)$': Cole takes mustard
'$S(l)$': Lang takes salt	'$U(l)$': Lang takes mustard
'$S(d)$': Dix takes salt	'$U(d)$': Dix takes mustard

or the other way round—having, for each diner, a one-place predicate of being consumed by that diner, and a constant symbol 's' for salt, and another 'u' for mustard. (We've already used 'm' for Mill.)

'$B(s)$': Barry takes salt	'$B(u)$': Barry takes mustard
'$M(s)$': Mill takes salt	'$M(u)$': Mill takes mustard
'$C(s)$': Cole takes salt	'$C(u)$': Cole takes mustard
'$L(s)$': Lang takes salt	'$L(u)$': Lang takes mustard
'$D(s)$': Dix takes salt	'$D(u)$': Dix takes mustard

But perhaps the most natural 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, and 's' for salt and 'u' for mustard. 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—in all three approaches—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 parentheses: we don't write '$m\,T\,s$'. However identity is a special case and 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 38, 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', for takes-the-flat, to be applied to Herbert? No-one else is conjectured to take the flat after all, so you'd have no other use for that predicate letter ... so perhaps not. 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 44 *Formalise the following, using a lexicon of your choice*
Romeo loves Juliet; Juliet loves Romeo.

Balbus loves Julia. Julia does not love Balbus. What a pity.[1]

Fido sits on the sofa; Herbert sits on the chair.

Fido sits on Herbert.

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

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

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

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.]

4.1.1 The Syntax of First-order Logic

There is really an abuse of notation here: we should use quasi-quotes ...

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.

[1] I found this in a Latin primer: *Balbus amat Juliam; Julia non amat Balbum* ....

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. **Unary** predicates have one argument, **binary** predicates have two; n-**ary** have n. '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**; elsewhere such strings are 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.

Similarly arity of functions. [say a bit more about this]

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

Quantifiers

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;
$(\forall x)(\exists y)L(x, y)$	Everybody loves someone

To save space we might write this second as

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

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. Beware! This does not mean that constants are reserved words! The constant 'a' can denote anything the user wants it to denote, it doesn't wander around like the denotation of a variable such as 'x'. Confusingly that's not to say that there are no reserved constants; there are plenty in formalised mathematics, the numerals 0, 1 ... for starters.

should probably Say something about this

...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 are we trying 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.

Say something about *father* as a one-place predicate is really defined in terms of a two-place predicate *father-of* and since it satisfies a uniqueness condition there is naturally a function *father-of*.

4.1.2 Warning: Scope ambiguities

Perhaps this could be moved into a section called something like 'subtleties of evaluation'

> "All that glisters is not gold"

is not

> $(\forall x)(\text{glisters}(x) \to \neg\text{gold}(x))$

and

> "All is not lost"

is not

> $(\forall x)(\neg\text{lost}(x))$

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.

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 $(\forall x)(\exists y)(x \text{ loves } y)$ and $(\exists y)(\forall x)(x \text{ loves } y)$, 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!

4.1.3 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.

Along the same lines one can observe that in the formal languages of logic there is no *tense* or *aspect* or *mood*.

4.2 Some exercises to get you started

You might also like to think if any of these arguments are valid.

EXERCISE 45
Render the following fragments of English into predicate calculus, using a lexicon of your choice.

This first bunch involve monadic predicates only and no nested quantifiers.

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. *A wise man walks on his feet;*
   an unwise man on his hands.
   Therefore no man walks on both.

10. *No fossil can be crossed in love;*
    an oyster may be crossed in love.
    Therefore oysters are not fossils.

11. *All who are anxious to learn work hard;*
    some of these students work hard.
    Therefore some of these students are anxious to learn.

12. *His songs never last an hour;*
    A song that lasts an hour is tedious;
    Therefore his songs are never tedious.

13. *Some lessons are difficult;*
    what is difficult needs attention;
    Therefore some lessons need attention.

14. *All humans are mammals;*
    all mammals are warm blooded;
    Therefore all humans are warm-blooded.

15. *Warmth relieves pain;*
    nothing that does not relieve pain is useful in toothache;
    Therefore warmth is useful in toothache.

16. *Louis is the King of France;*
    all Kings of France are bald.
    Therefore Louis is bald.

EXERCISE 46 *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.*

1. *Anyone who has forgiven at least one person is a saint.*
2. *Nobody in the logic class is cleverer than everybody in the history class.*
3. *Everyone likes Mary—except Mary herself.*
4. *Jane saw a bear, and Roger saw one too.*
5. *Jane saw a bear and Roger saw it too.*
6. *Some students are not taught by every teacher;*
7. *No student has the same teacher for every subject.*
8. *Everybody loves my baby, but my baby loves nobody but me.*

EXERCISE 47 *These involve nested quantifiers and dyadic predicates*
Match up the formulæ on the left with their English equivalents on the right. Duplicates page 113?

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

EXERCISE 48 *Render the following pieces of English into Predicate calculus, using a lexicon of your choice.*

1. *Everyone who loves is loved;*
2. *Every horse is an animal so every head of a horse is the head of an animal.*
3. *Everyone loves a lover;*
4. *The enemy of an enemy is a friend*
5. *The friend of an enemy is an enemy*
6. *Any friend of George's is a friend of mine*
7. *Jack and Jill have at least two friends in common*
8. *Two people who love the same person do not love each other.*
9. *None but the brave deserve the fair.*
10. *If there is anyone in the residences with measles then anyone who has a friend in the residences will need a measles jab.*

11. *No two people are separated by more than six steps of aquaintanceship.*

This next batch involves nested quantifiers and dyadic predicates and equality.

EXERCISE 49 *Render the following pieces of English into Predicate calculus, using a lexicon of your choice.*

1. *There are two islands in New Zealand;*
2. *There are* three[2] *islands in New Zealand;*
3. *tf knows (at least) two pop stars;*

   (You must resist the temptation to express this as a relation between tf and a plural object consisting of two pop stars coalesced into a kind of plural object like Jeff Goldblum and the Fly[3]*. You will need to use '=', the symbol for equality.)*
4. *If there is to be a jackpot winner it will be me.*

   This is much trickier than it looks. If you are having difficulty with it, here is a hint . . . . What is the top-level connective?
5. *You are loved only if you yourself love someone [other than yourself!];*
6. *At least two Nobel prizewinners have changed Professor Körner's nappies.*
7. *God will destroy the city unless there are (at least) two righteous men in it;*
8. *There is at most one king of France;*
9. *I know no more than two pop stars;*
10. *There is precisely one king of France;*
11. *I know three FRS's and one of them is bald;*
12. *Brothers and sisters have I none; this man's father is my father's son.*
13. ** Anyone who is between a rock and a hard place is also between a hard place and a rock.*

EXERCISE 50 *Using 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) y;*
$F(x,y)$:	*x and y are friends;*
$R(x)$:	*x lives in the residences;*
$M(x)$:	*x has measles;*

[2] The third is Stewart Island

[3] See [16]

*Turn the following into English. (**normal** English: something you can imagine yourself saying—no xs and ys.)*

$(\forall x)(F(\texttt{Kim}, x) \to F(\texttt{Alex}, x))$

$(\forall x)(\exists y)(F(x, y) \wedge M(y) \wedge Z(y))$

$(\forall x)(F(\texttt{Kim}, x) \to Z(x))$

$(\forall x)((Z(x) \wedge M(x)) \to F(\texttt{Kim}, x))$

$(\forall x)(Z(x) \to (\exists y)(F(x, y) \wedge M(y)))$

$(\forall x)(S(x) \to (\exists yz)(T(y, x, z)))$

$(\exists x)(S(x) \wedge (\forall z)(\neg A(x, z)))$

$(\exists x)(C(x) \wedge (\forall z)(\neg A(z, x)))$

$(\exists x)(L(x) \wedge (\forall yz)(\neg T(x, y, z)))$

$(\forall x_1 x_2)[(\forall z)(A(x_1, z) \longleftrightarrow A(x_2, z)) \to x_1 = x_2]$

$(\forall x_1 x_2)[(\forall z)(A(z, x_1) \longleftrightarrow A(z, x_2)) \to x_1 = x_2]$

$(\forall y)(\exists x, z)(T(x, y, z)) \to (\forall u, v)(\neg T(y, u, v))$

$(\forall xy)(x \neq y \to (\exists z)(F(z, x) \longleftrightarrow \neg F(z, y)))$

EXERCISE 51 *Look up 'monophyletic'. Using only the auxilliary relation "is descended from" give a definition in first-order logic of what it is for a monadic predicate of lifeforms to be monophyletic.*

F is monophyletic iff both $(\forall xy)((F(x) \wedge F(y) \to (\exists z)(F(z) \wedge D(z, x) \wedge D(z, y)))$ and $(\forall x)(\forall y)(F(x) \to (D(x, y) \to F(y)))$ hold
Should probably also accept $(\exists x)(\forall y)(D(x, y) \longleftrightarrow F(y))$

EXERCISE 52 *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? Duplicates page 111?

I think of this exercise as a kind of touchstone for the first-year student of logic. It would be a bit much to ask a first-year student (who, after all, might not be going on to do second-year logic) to give a formal proof of the implication or to exhibit a countermodel to demonstrate the independence, but exercise 52 is a fair test.

Need some formalisation exercises using function symbols

4.3 Comparatives and Superlatives

Perhaps here some chat about comparatives and superlatives. The superlatives will prepare us for Russell's theory of descriptions. And also prepare us for some chat about predicate modifiers.

In the old grammar books I had at school we were taught that adjectives had three forms: *simple* ("cool") *comparative* ("cooler") and *superlative* ("coolest").

We can define superlatives in terms of comparatives because if you are the coolest thing around then no-one is cooler than you. Or does it mean that you are cooler than everyone [else]? Can you be first-equal? cooler-than is obviously transitive. Is it irreflexive?

haven't defined 'transitive' yet

4.4 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 the contrapositive of the converse)? 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 be cautious and say that we can't properly use expressions like "the King of France" *at all* 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 one might feel.

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 that *do* contain 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 in these 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.

Is this the first place where we talk about translations?

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.**

4.5 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—'p' 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.

4.5.1 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*.

Review comparative and superlative

They could be represented in higher order logic by two predicate modifiers. The 'ER' (comparative) modifier takes a one-place predicate letter and returns a two-place predicate letter. The 'EST' (superlative) operator takes a one-place predicate letter and returns another one-place predicate letter.

Another predicate modifier is *too*.

> No woman can be too thin or too rich.

Not a good example

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.

4.6 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:

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

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 and we will see it in section 4.7 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.

Armed with this new language we can characterise some important properties:

DEFINITION 23 *A relation R is*

symmetrical *if* $(\forall x)(\forall y)(R(x, y) \longleftrightarrow R(y, x))$
transitive *if* $(\forall x)(\forall y)(\forall z)((R(x, y) \wedge R(y, z)) \to R(x, z))$
reflexive *if* $(\forall x)(R(x, x))$
irreflexive *if* $(\forall x)(\neg R(x, x))$
extensional *if* $(\forall x)(\forall y)(x = y \longleftrightarrow (\forall z)(R(x, z) \longleftrightarrow R(y, z))$.

Finally a relation that is transitive, reflexive and symmetrical is an **equivalence relation**.

(People aften say 'symmetric' instead of 'symmetrical'.)

The binary relation "full sibling of" is symmetrical, and so is the binary relation "half-sibling of". However, "full sibling of" is transitive whereas "half-sibling of" is not; "full sibling of" is an equivalence relation (always assuming you are your own full sibling)

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 (because two women with the same children are the same woman) but its converse isn't (because 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 is bigger than y; y is bigger than z. Therefore x is bigger than z. (S)

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

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

were a logical truth. However (T) is not a logical truth. (S) is truth-preserving all right, but not 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

$$(\forall x \forall y \forall z)(x = y \wedge y = z \;\to\; x = z)$$

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.

These properties of relations are in any case useful in general philosophy but they are useful in particular in connection with possible world semantics to be seen in chapter 6.

Observe that **transitive**, **symmetrical**, **irreflexive**, **extensional** etc are second-order properties, being properties of properties/relations.

Be careful! We want to say that they are first-order but that wellfounded is second-order. Find the correct thing to say.

Wellfounded relations

ontological dependence

Not all relations are unary or binary

It may be worth making the point that not all relations are unary or binary relations. Topical mundane examples are the three-place relation "student s is lectured by lecturer l for course c", or the four-place relation "Course c is lectured by lecturer l in room r at time-slot t".

There is also a natural three-place relation of *betweenness* that relates points on a line, but that doesn't concern us much as philosophers. Yet another example (again not of particular philosophical interest but cutely everyday) is the three-place relation of "later than" between times 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 an (apparently) binary relation that you are trying vainly to make sense of, but 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.

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 4.1.1 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 around and don't panic when they come and sit next to us on the sofa. 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. So it is valid.

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, in exercise 20.

7. These two formulæ are logically equivalent:

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

(hint: what is the principal connective of each of the formulæ?)

EXERCISE 53 *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)

Wouldn't it be nice to do without variables, since once they're bound it doesn't matter which they are? It would—and there is a way of doing it, called *Predicate Functor Logic*. Quine [38] and Tarski-Givant [44] wrote about it, and we will see glimpses in chapter 7. Unfortunately it seems that the human brain (most of them anyway—certainly mine and probably yours) are not configured to process the kind of syntax one is forced into if one doesn't have variables. As far as I know all natural languages have pronouns rather than the contrivances required by variable-free syntax.

4.7 Natural Deduction Rules for First-Order Logic

To the natural deduction rules for propositional calculus we add rules for introducing and eliminating the quantifiers:

Rules for ∃

$$\frac{A(t)}{(\exists x)(A(x)}\ \exists\text{-int} \qquad\qquad \frac{\begin{array}{c}[A(t)]^{(1)}\\ \vdots\\ C\end{array} \qquad (\exists x)(A(x))}{C}\ \exists\text{-elim}(1) \tag{4.1}$$

Notice the similarity between ∨-elimination and ∃-elimination.

Rules for ∀

$$\frac{\begin{array}{c}\vdots\\ A(t)\end{array}}{(\forall x)(A(x))}\ \forall\text{-int} \qquad\qquad\qquad \frac{(\forall x)(A(x))}{A(t)}\ \forall\text{-elim}$$

To prove that everything has property A, reason as follows. Let x be an object about which we know nothing, reason about it for a bit and deduce that x has A; remark that no assumptions were made about x; Conclusion: *all* xs must therefore have property A. But it is important that x should be an object about which we know nothing, otherwise we won't have proved that every x has A, merely that A holds of all those x's that satisfy the conditions x satisfied and which we exploited in proving that x had A. The rule of $\forall$-introduction therefore has the important side condition: 't' **not free in the premisses**. The idea is that if we have proved that A holds of an object x *selected arbitrarily*, then we have actually proved that it holds for *all* x.

The rule of $\forall$-introduction is often called **Universal Generalisation** or **UG** for short. It is a common strategy and deserves a short snappy name. It even deserves a joke.[4]

explain 'arbitrary'

THEOREM 24 *Every government is unjust.*

Proof: Let G be an arbitrary government. Since G is arbitrary, it is certainly unjust. Hence, by universal generalization, every government is unjust. ■

This is of course a fallacy of equivocation.

In the propositional calculus case a theorem was a formula with a proof that had no undischarged assumptions. We have to tweak this definition slightly in this new situation of natural deduction rules for first-order logic. We have to allow undischarged assumptions like $t = t$: it's hard to see how else we are going to prove obvious logical truths like $(\forall x)(\forall y)(x = y \to y = x)$. (The fact that symmetry of equality is a *logical* truth is worth noting. This is because equality is part of the *logical* vocabulary ...)

Must say a lot more about equality being part of the logical vocabulary

However, we will not develop this further but will procede immediately to a sequent treatment.

more exercises here?

4.8 Sequent Rules for First-Order Logic

$\forall$ left:

$$\frac{F(t), \Gamma \vdash \Delta}{(\forall x)(F(x)), \Gamma \vdash \Delta} \qquad \forall - \text{left}$$

where 't' is an arbitrary term

Do they know what a *term* is?

(If Δ follows from Γ plus the news that Trevor has property F then it will certainly follow from Γ plus the news that *everybody* has property F.)

eigenvariable

$\forall$ right:

$$\frac{\Gamma \vdash \Delta, F(t)}{\Gamma \vdash \Delta, (\forall x)(F(x))} \qquad \forall - \text{right}$$

[4] Thanks to Aldo Antonelli.

where 't' is a term not free in the lower sequent. We explain $\forall$-R by saying: if we have a deduction of $F(a)$ from Γ then if we replace every occurrence of 'a' in Γ by 'b' we have a proof of $F(b)$ from the modified version $[b/a]\Gamma$. If there are no occurrences of 'a' to replace then $[b/a]\Gamma$ is just Γ and we have a proof of $F(b)$ from the original Γ. But that means that we have proved $(\forall x)(F(x))$ from Γ.

Surely 't' has to be a constant not an arbitrary closed term

$\exists$ **left:**

$$\frac{F(a), \Gamma \vdash \Delta}{(\exists x)(F(x)), \Gamma \vdash \Delta} \qquad \exists - \text{left}$$

where 'a' is a variable not free in the lower sequent.

$\exists$ **right:**

$$\frac{\Gamma \vdash \Delta, F(t)}{\Gamma \vdash \Delta, (\exists x)(F(x))} \qquad \exists - \text{right}$$

where 't' is an arbitrary term.

(Notice that in $\forall$-L and $\exists$-R the thing that becomes the bound variable (the eigenvariable) is an arbitrary term whereas with the other two rules it has to be a variable)

We will of course have to allow sequents like $\vdash x = x$ as initial sequents.

You might like to think about what the subformula property would be for first-order logic. What must the relation "subformula-of" be if it is to be the case that every proof of a formula ϕ using only these new rules is to contain only subformulæ of ϕ?

4.8.1 Repeat a warning

Now is probably as a good a places as any to remind oneself that the sequent rules for the quantifiers—like the sequent rules for the propositional connectives—always work at top level only. One of my students attempted to infer

$$(\forall x)(A \vee F(x)) \vdash A \vee (\forall x)F(x)$$

from

$$(\forall x)(A \vee F(x)) \vdash A \vee F(x)$$

by means of $\forall$-R. Check that you understand why this is wrong.

4.8.2 Some more exercises

EXERCISE 54 *Find proofs of the following sequents:*
$\neg\forall x\phi(x) \vdash \exists x\neg\phi(x)$*;*
$\neg\exists x\phi(x) \vdash \forall x\neg\phi(x)$*;*
$\phi \wedge \exists x\psi(x) \vdash \exists x(\phi \wedge \psi(x))$*;*
$\phi \vee \forall x\psi(x) \vdash \forall x(\phi \vee \psi(x))$,
$\phi \to \exists x\psi(x) \vdash \exists x(\phi \to \psi(x))$,
$\phi \to \forall x\psi(x) \vdash \forall x(\phi \to \psi(x))$,
$\exists x\phi(x) \to \psi \vdash \forall x(\phi(x) \to \psi)$

$\forall x\phi(x) \to \psi \vdash \exists x(\phi(x) \to \psi)$,
$\exists x\phi(x) \vee \exists x\psi(x) \vdash \exists x(\phi(x) \vee \psi(x))$,
$\forall x\phi(x) \wedge \forall x\psi(x) \vdash \forall x(\phi(x) \wedge \psi(x))$,

In this exercise ϕ and ψ are formulæ in which 'x' is not free, while $\phi(x)$ and $\psi(x)$ are formulæ in which 'x' may be free.

EXERCISE 55 *Prove the following sequents. The first one is really quite easy. (It is Russell's paradox of the set of all sets that are not members of themselves, and it's related to Grelling's paradox that we saw on p. 34) (See section 10.4.) The third sequent underlines the fact that you do not need a biconditional in the definition of 'symmetric'.*

1. $\vdash \neg(\exists x)(\forall y)(P(y,x) \longleftrightarrow \neg(P(y,y)))$

2. $\vdash [(\exists x)(\forall y)(P(y,x) \longleftrightarrow (P(y,y) \to p))] \to p$

3. $\forall x\forall y(R(x,y) \to R(y,x)) \vdash \forall x\forall y(R(x,y) \longleftrightarrow R(y,x))$;

4. $\vdash \neg(\exists x)(\forall y)(P(y,x) \longleftrightarrow (\forall z)(P(z,y) \to \neg P(y,z)))$

   This formula concerns the modified paradox of Russell concerning the set of those sets that are not members of any member of themselves.

   It is noticeably harder, and is recommended mainly for enthusiasts. You will certainly need to "keep a copy"! You will find it much easier to find a proof that uses cut. Altho' there is certainly a proof that never has more than one formula on the right you might wish to start off without attempting to respect this constraint.

5. *Find a proof of the following sequent:*

$$(\forall x)[P(x) \to P(f(x))] \vdash (\forall x)[P(x) \to P(f(f(x)))]$$

   For this you will definitely need to keep a copy. (On the left, as it happens)

We need some natural deduction exercises for FOL

6. *Find natural deduction and sequent proofs of*

$$(\exists x)A(x), (\forall x)(A(x) \to B(f(x))) \vdash (\exists x)B(x).$$

   [This formula doesn't need you to keep a copy but it does use function symbols. Need more exercises on function symbols.]

4.9 Equality and Substitution

Frege gave a definition of equality, using higher-order logic. Equality is a deeply deeply problematic notion in all branches of philosophy, so it was really quite brave of Frege to even attempt to define it. His definition of equality says that it is the intersection of all reflexive relations. Recall from definition 23 that a binary relation R is reflexive if

$R(w, w)$ holds for all w: (That's what the '$(\forall w)(R(w, w))$' is doing in the formula 4.9 below.) So Frege's definition is

$$x = y \quad \text{iff} \quad (\forall R)[(\forall w)(R(w, w)) \to R(x, y)] \tag{4.2}$$

The first thing to notice is that this definition is second-order! You can tell that by the '$(\forall R)$' and the fact that the 'R' is obviously a predicate letter because of the '$R(w, w)$'.

Notice that this definition is not circular (despite what you might have expected from the appearance of the word 'reflexive') since the *definiendum* does not appear in the *definiens*.

4.9.1 Substitution

Consider the binary relation "every property that holds of x holds also of y and *vice versa*". This is clearly reflexive! If x and y are equal then they stand in this relation (because two things that are equal stand in every reflexive relation, by definition) so they have the same properties. This justifies the rule of substitution. (If you have good French have a look at [13]).

$$\frac{A(t) \qquad t = x}{A(x)} \text{ subst} \tag{4.3}$$

In the rule of substitution you are not obliged to replace *every* occurrence of 't' by 'x'. (This might remind you of the discussion on page 60 where we consider cancelling premisses.)

The following example is a perfectly legitimate use of the rule of substitution, where we replace only the *first* occurrence of 't' by 'x'. In fact this is how we prove that equality is a symmetrical relation!

$$\frac{t = t \qquad t = x}{x = t} \text{ subst} \tag{4.4}$$

Given that, the rule of substitution could more accurately be represented by

$$\frac{A[t/x] \qquad t = x}{A} \text{ subst} \tag{4.5}$$

. . . the idea being that A is some formula or other—possibly with free occurrences of 'x' in it—and $A[t/x]$ is the result of replacing all free occurrences of 'x' in A by 't'. This is a bit pedantic, and on the whole our uses of substitution will look more like 4.3 than 4.5.

However we will definitely be using the $A[t/x]$ notation in what follows, so be prepared. Sometimes the $[t/x]$ is written the other side, as

$$[t/x]A. \tag{subst}$$

This notation is intended to suggest that $[t/x]$ is a function from formulæ to formulæ that is being applied to the formula A.

One thing that may cause you some confusion is that sometimes a formula with a free variable in it will be written in the style "$A(x)$" making the variable explicit. Sometimes it isn't made explicit. When you see the formula *subst* above it's a reasonable bet that the variable 'x' is free in A, or at least could be: after all, there wouldn't be much point in substituting 't' for 'x' if 'x' weren't free, now would it?!

4.9.2 Leibniz's law

"The identity of indiscernibles". This is a principle of second-order logic:

$$(\forall xy)((\forall R)(R(x) \longleftrightarrow R(y)) \to x = y) \tag{4.6}$$

The converse to 4.6 is obviously true so we can take this as a claim about the nature of equality: $x = y$ iff $(\forall R)(R(x) \longleftrightarrow R(y))$.

It's not 100% clear how one would infer that x and y are identical in Frege's sense merely from the news that they have the same *monadic* properties: Frege's definition talks about reflexive relations, which of course are *binary*. The claim that 4.6 characterises equality (by which we mean that if we replace '=' in 4.6 by any other binary relation symbol the result is no longer true) is potentially contentious. It is known as **Leibniz's Law**.

4.10 Prenex Normal Form

There is a generalisation of CNF and DNF to first-order logic: it's called **Prenex normal form**. The definition is simplicity itself. A formula is in Prenex normal form if it is of the form

$$(Qv_1)(Qv_2)\cdots(Qv_n)(....)$$

where the Qs are quantifiers, and the dots at the end indicate a purely propositional formula: one that contains no quantifiers, and is in conjunctive normal form. All quantifiers have been "pulled to the front".

EXERCISE 56 *Which of the following formulæ are in Prenex normal form?* Insert some formulae here!!

THEOREM 25 *Every formula is logically equivalent to one in PNF.*

To prove this we need to be able to "*pull all quantifiers to the front*". What does this piece of italicised slang mean? Let's illustrate:

$$(\forall x)F(x) \wedge (\forall y)G(y)$$

is clearly equivalent to

$$(\forall x)(\forall y)(F(x) \wedge G(y))$$

(If everything is green and everything is a frog then everything is both green and a frog, and *vice versa*).

In exercise 54 the point in each case is that in the formula being deduced the scope of the quantifier is larger: it has been "pulled to the front". If we keep on doing this to a formula we end up with something that is in PNF.]

. . . and explain to your flatmates what this has to do with theorem 25.

Explain why PNF is important—why normal form theorems are important in general. It imposes a linear order on the complexity of formulæ.

4.11 Soundness again

At this point we should have a section analogous to section 3.3 where we prove the soundness of natural deduction for propositional logic and section 3.5.1 where we prove the soundness of sequent calculus for propositional logic.

Work to be done here

You will discover that it's nowhere near as easy to test predicate calculus formulæ for validity as it is to test propositional formulæ: there is no easy analogue of truth-tables for predicate calculus. Nevertheless there is a way of generating all the truth-preserving principles of reasoning that are expressible with this syntax, and we will be seeing them, and I hope to prove them complete.

well, will you or won't you?

You must get used to the idea that all notions of logical validity, or of sound inference, can be reduced to a finite set of rules in the way that propositional logic can and predicate calculus can. Given that—as we noted on p 37—the validity of an argument depends entirely on its syntactic form, perhaps we should not be surprised to find that there are finite mechanical methods for recognising valid arguments. However this holds good only for arguments of a particularly simple kind. If we allow variables to range over predicate letters then things start to go wrong. Opinion is divided on how important is this idea of completeness. If we have something that looks like a set of principles of reasoning but we discover that it cannot be generated by a finite set of rules, does that mean it isn't part of logic?

see page 69

In contrast to soundness, *completeness* is hard. See section 4.13.

Mention here other notions of validity: true in all finite models: true in all infinite models

4.12 Hilbert-style Systems for First-order Logic

At this point there should be a section analogous to section 3.6. However I think we can safely omit it.

4.13 Semantics for First-order Logic

This section is not recommended for first-years.

4.13.1 stuff to fit in

We arrived at the formulæ of first-order logic by a process of codifying what was logically essential in some scenario or other. Semantics is the reverse process: picking up a formula of LPC and considering what situations could have given rise to it by the kind of codification that we have seen in earlier exercises such as 45.

A valid formula is one that is true in all models. We'd better be clear what this means! So let's define what a model is, and what it is for a formula to be true in a model.

[Signatures, structures, carrier set. Then we can explain again the difference between a first-order theory and a higher-order theory.]

Now we have to give a rigorous explanation of what it is for a formula to be true in a structure.

Say something about how the $'$ operation on variables has no semantics.

Also, the cute facts about words beginning with 'sn' or with 'gl' will not be captured by the kind of semantics we are about to do.

Semantics is the process of allocating extensions to intensions.

Semantics is useful for independence results. Peirce's law independent of K and S. Mind you, to do that we have to un-reserve $\to$.

4.13.2 Syntactic types for the various pieces of syntax

It may be worth making the point that you can actually tell, on being a formula in some language (plus the information that it is a well-formed formula) what "part of speech" (to use the old terminology) each piece of syntax is at least if you have some minimal extra information—such as being able to recognise the quantifiers and the left and right parentheses, perhaps. Any symbol that immediately follows a quantifier must be a variable. Any symbol that immediately precedes a variable must be (if not a quantifier) then either a function symbol or a predicate symbol.

Thus each piece of syntax has a *type*, which reflects its rôle in the assembly of formulæ. For example we say that constant symbols (and variables) have type `ind`, since (in an interpretation) they point to *individuals*—inhabitants of the carrier[5] set. Similarly propositional constants are of type `bool`, since in any interpretation they point to a truth-value.

Other pieces of syntax have more complicated ("molecular") types:

One-place predicate symbols are of type `ind -> bool`
One-place function symbols are of type `ind -> ind`
Quantifiers are of type `(ind -> bool) -> bool`
Determiners are of type `(ind -> bool) -> ((ind -> bool) -> bool)`
Naturally we also have:
n-place predicate symbols are of type `ind`^{n} `-> bool`
n-place function symbols are of type `ind`^{n} `-> ind`

Marry this up with the material on page 131

Enlarge on this a bit

For example in

$$(\forall x)(f(x) = 1)$$

you can tell that 'x' is a variable since one occurrence of it is preceded by a quantifier. The symbol '1' must be a constant symbol since it sits immediate to the right of n occurrence of '='. 'f' must now be either a function symbol or a predicate symbol

[5] "Carrier set" not yet defined

because it is applied to a variable. Now the result of applying it to the variable sits to the left of an occurrence of '=' and therefore must be of type `ind`. This tells us that 'f' is a function symbol (has type `ind -> ind`) rather than type `ind -> bool`.

Aquiring the habit of performing this kind of calculation in your head makes formulæ much more digestible and less scary, since it tells you how to read them. Indeed this "part-of-speech" system of typing for the various pieces of syntax is a useful way of thinking when we come to semantics ... to which we now turn.

4.14 Truth and Satisfaction

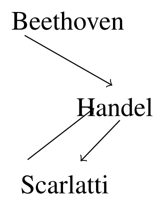
In this section we develop the ideas of truth and validity (which we first saw in the case of propositional logic) in the rather more complex setting of predicate logic.

We are going to say what it is for a formula to be **true** in a structure. We will achieve this by doing something rather more general. What we will give is—for each language $\mathcal{L}$—a definition of what it is *for a formula of* $\mathcal{L}$ *to be true in a structure*. Semantics is a relation not so much between an expression and a structure as between a *language* and a structure. [Slogan: semantics for an expression cannot be done in isolation.]

We know what expressions are, so what is a structure? It's a set with knobs on. You needn't be alarmed here by the sudden appearance of the word 'set'. You don't need to know any fancy set theory to understand what is going on. The set in question is called the *carrier set*, or *domain*. One custom in mathematics is to denote structures with characters in uppercase $\mathfrak{FRAKTUR}$ font, typically with an '$\mathfrak{M}$'.

The obvious examples of structures arise in mathematics and can be misleading, and in any case are not really suitable for our expository purposes here. We can start off with the idea that a structure is a set-with-knobs on. Here is a simple example that cannot mislead anyone.

The carrier set is the set {Beethoven, Handel, Domenico Scarlatti} and the knobs are (well, *is* rather than *are* because there is only one knob in this case) the binary relation *is-the-favourite-composer-of*. We would obtain a different structure by adding a second relation: *is-older-than* perhaps.



If we are to make sense of the idea of an expression being true in a structure then the structure must have things in it to match the various gadgets in the language to which the expression belongs. If the expression contains a two-place relation symbol 'loves' then the structure must have a binary relation on it to correspond. This information is laid down in the **signature**. The signature of the structure in the composers example above has one binary relation symbol and three constant symbols; the signature of set theory is equality plus one binary predicate; the signature of the language of first-order Peano arithmetic has slots for one unary function symbol, one nullary function symbol (or constant) and equality.

Let's have some illustrations, at least situations where the idea of a signature is useful.

- Cricket and baseball resemble each other in a way that cricket and tennis do not. One might say that cricket and baseball have the same signature. Well, more or less! They can be described by giving different values to the same set of parameters.

- It has been said that a French farce is a play with four characters, two doors and one bed. This *aperçu* is best expressed by using the concept of signature.

- Perhaps when you were little you bought mail-order kitsets that you assembled into things. When your mail-order kitset arrives, somewhere buried in the polystyrene chips you have a piece of paper (the "manifest") that tells you how many objects you have of each kind, but it does not tell you what to do with them. Loosely, the manifest is the *signature* in this example. Instructions on what you do with the objects come with the *axioms* (instructions for assembly).

The modern version is IKEA

- Recipes correspond to theories: lists of ingredients to signatures.

Structure	**Signature**	**Axioms**
French Farce	4 chars, 2 doors 1 bed	Plot
Dish	Ingredients	Recipe
Kitset	list of contents	Instructions for assembly
Cricket/baseball	innings, catches, etc	Rules
Tennis/table tennis		

A telephone directory is a structure with a total order, and a name for every element. It has no other structure. As the old joke has it: *lots of characters but no plot.* A rather linear narrative.

4.14.1 Definition of "truth-in-a-structure"

First we need to decide what our structure is to be. Suppose it is $\mathfrak{M}$, with carrier set M. Next we need the concept of an **interpretation**. This is a function assigning to each predicate letter, function letter and constant in the language a subset of M^n, or a function $M^k \to M$, or element of M *mutatis mutandis.* That is to say, to each syntactic device in the language, the interpretation assigns a component of $\mathfrak{M}$ of the appropriate arity. For this to be possible it is necessary that the structure and our language have the same signature.

Now we have to get straight the difference between the rôles played by the various different kinds of lexical items ... what one might call the different parts of speech.

The first difference is between the logical vocabulary and the non-logical vocabulary. The non-logical vocabulary is open to various kinds of stipulation, whereas the nonlogical vocabulary isn't. You might object that the symbol '$\wedge$' acquires its meaning (logical conjunction) by stipulation, which of course it does. However, in this game, that particular kind of stipulation happened at the beginning of time, and it is not for us to choose to make '$\wedge$' mean anything else. Similarly '$\forall$', '$\exists$' and even '$=$'.

We are free to stipulate the meanings of the predicate symbols, the function letters, the constants and the variables. There are two ways in which we stipulate these, and it is helpful at this point to draw a contrast with the propositional case. The only nonlogical vocabulary in the propositional case is the propositional letters, and we stipulate them by means of **valuation**s. The meaning of the items of the logical vocabulary emerges from the way in which the semantics of complex expressions (complexed from the propositional letters) emerges from the semantics of those propositional letters. (I don't want to say *meaning* of the propositional letters co's in this case I mean *truth-value* rather than anything intensional such as meaning.)

In the propositional case we stipulate (by means of valuations) the semantics of the propositional letters, so that then the compositional rules for the connectives (the bits of logical vocabulary) tell us the truth value of the complex formulæ.

In the first-order case there are two kinds of non-logical vocabulary, and they are controlled by stipulation in two different ways. The two kinds are (i) constant symbols, predicate letters and function symbols and (ii) individual variables. Now this difference between the two kinds of non-logical vocabulary doesn't corrrespond neatly to a type distinction such as we have just seen. Individual variables will be given a semantics differently from constants, even tho' both these gadgets are of type `ind`. Constants are of a different syntactic type from predicate symbols, but their meanings are stipulated in the same way... in a way which we must now explain.

reserved word

There is a difference in Compsci-speak between *configure time* and *run time*. If these expressions have any meaning for you, savour it. If not, don't worry.

It's at configure-time that we decide on the semantics of the constant symbols, predicate letters and function symbols. That's when we decide what the structure is to be. Suppose we fasten on a structure $\mathfrak{M}$. This act determines what the universe (what the totality of things over which our variables range) is going to be: it's going to be the carrier set M of $\mathfrak{M}$. The interpretation now tells us which element of M corresponds to which constant symbol, which subset of M corresponds to which one-place predicate letter, which subset of $M \times M$ corresponds to which two-place predicate letter, which subset of M^n corresponds to which n-place predicate letter, which function $M^n \to M$ corresponds to which n-place function letter. Thus an interpretation is a function that sends

This is the first mention of cartesian product

(individual) constant symbols	to	elements of M;
n-ary predicate symbols	to	subsets of M^n;
n-adic function symbols	to	functions $M^n \to M$;
(propositional) constant symbols	to	`true` or `false`.

An interpretation matches up the signature-in-the-language with the signature-in-the-structure.

Marry this up with the material on page 127

I shall use the calligraphic letter '$\mathcal{I}$' to vary over interpretations.

We have now equipped the language with an interpretation so we know what the symbols mean, but not what the values of the variables are. In other words, settling on an interpretation has enabled us to reach the position from which we started when doing propositional logic.

When we did semantics for propositional logic I encouraged you to think of valuations (rows in the truth-table) as *states*, as *scenarios*. The idea will be useful here too. The interpretation tells us what the constant symbols point to, but it can't tell us what the variables point to. To understand how variables point to things we need a notion of state very redolent of the notion of state suggested by propositional valuations.

It's rather like the position we are in when contemplating a computer program but not yet running it. When we run it we have a concept of instantaneous state of the program: these states (snapshots) are allocations of values to the program variables. Let us formalise a concept of state.

An **assignment function** is a function that takes a variable and returns a member of M. (If we want to make it clear that we are talking about assignment functions under an interpretation $\mathcal{I}$ we will speak of $\mathcal{I}$-assignment functions.) What do we do with assignment functions? Reflect that we don't really want to say of a formula with a free variable in it that it is *true* or that it is *false*—not straightforwardly anyway. But we do want to say something of that nature that takes the assignment functions into account. The gadget we need is a relation-between-assignment-functions-and-formulæ of *satisfaction*. We want to be able to say that an assignment function **satisfies** a formula. Thus we say—for example—that an assignment function f satisfies the formula '$R(x)$' as long as that element of the domain M to which f sends the variable 'x' belongs to the subset of M to which the interpretation $\mathcal{I}$ sent the letter 'R'.

Got that last bit? Read it again just to be sure.

The satisfaction relation between formulæ and assignment functions is defined "compositionally"—by recursion on the subformula relation. Clearly an assignment function will satisfy a conjunction iff it satisfies both conjuncts and will satisfy a disjunction iff it satisfies at least one of the disjuncts. The recursion steps for the quantifiers are fiddly, and depend on whether or not we insist that the assignment functions be total.

The recursions for the quantifiers

This is annoying and fiddly, because what one has to do depends very sensitively on whether one's assignment functions are total or partial.

- If our assignment functions are total, one says that
  - f satisfies $(\exists x)(\phi(x))$ as long as f satisfies $\phi(x)$;
  - f satisfies $(\forall x)(\phi(x))$ as long as $\phi(x)$ is satisfied by every f' which differs from f only for input 'x'.
- If our assignments functions are allowed to be partial one says that
  - f satisfies $(\exists x)(\phi(x))$ as long as *either* (i) f is defined on 'x' and satisfies $\phi(x)$, *or* (ii) f is not defined on 'x' but has an extension $f' \supset f$ which satisfies $\phi(x)$;
  - f satisfies $(\forall x)(\phi(x))$ as long as $\phi(x)$ is satisfied by every f' which differs from f only for input 'x'.

No such disclaimer allowed

I cannot put my hand on my heart and swear that I have got these right. Fiddly details like these one tends to disregard!! If you ever find that you absolutely have to be on top of this detail contact me and we'll go over it together.

Now we are in a position to define what it is for an expression ϕ to be **true-according-to-the-interpretation** $\mathcal{I}$. Again, we have two ways to proceed: (i) with total assignment functions (ii) with partial assignment functions.

1. If our assignment functions have to be total then ϕ is **true-according-to-the-interpretation** $\mathcal{I}$ iff every $\mathcal{I}$–assignment-function satisfies ϕ.

2. If our assignment functions may be partial then ϕ is **true-according-to-the-interpretation** $\mathcal{I}$ iff the empty assignment function satisfies ϕ.

4.14.2 An illustration

Let's illustrate with an example

$$(\exists x)(F(x) \wedge \neg(x = a))$$

The '$\exists$', the '$\wedge$' and the '$=$' are part of the logical vocabulary. No stipulation. The 'F' is a part of the nonlogical vocabulary and when we stipulate our domain M we also stipulate which subset of M is to be the extension of 'F'. 'a' too is part of the nonlogical vocabulary and when we stipulate our domain M we also stipulate which element of M is to be the extension/denotation of 'a'. The variable 'x' is a different matter. A valuation f satisfies '$F(x) \wedge \neg(x = a)$' if the member of M to which it sends the variable 'x' is something other than the denotation (according to the interpretation $\mathcal{I}$) of the constant symbol 'a' and moreover is a member of that subset of M to which the interpretation $\mathcal{I}$ has sent the predicate letter 'F'. So is $(\exists x)(F(x) \wedge \neg(x = a))$ true? It is true as long as every valuation satisfies it. (Or if the empty valuation satisfies it—if our valuations are allowed to be partial)

In search of another example, we can return to Handel, Scarlatti and Beethoven. The language $\mathcal{L}$ has the signature of one binary relation "is the favourite composer of". The structure into which we are going to interpret $\mathcal{L}$ has carrier set

{Beethoven, Handel, Domenico Scarlatti}

and our interpretation sends "is the favourite composer of" to the set (the relation-in-extension)

{⟨Handel, Beethoven⟩, ⟨Handel, Domenico Scarlatti⟩, ⟨Domenico Scarlatti, Handel⟩}.

H I A T U S — Novels, plays here?

DEFINITION 26
A **theory** *is a set of formulæ closed under deduction.*
We say T **decides** *ψ if $T \vdash \psi$ or $T \vdash \neg\psi$.*
Let us extend our use of the '$\mathcal{L}$' notation to write '$\mathcal{L}(T)$' for the language to which T belongs.[6]
A theory T is **complete** *if T decides every closed ϕ in $\mathcal{L}(T)$.*
A **Logic** *is a theory closed under uniform substitution.*

Theories can arise in two ways: syntactically and semantically.

Syntactically a theory can arise as the set of deductive consequences of a set of assumptions (in this context always called *axioms* that grab our attention somehow. An example is set theory: we discover these entities which we call *sets* and we think for a bit and come up with some principles that they might obey.

[6] For sticklers:

$$\mathcal{L}(T) =: \bigcup_{s \in T} \mathcal{L}(s)$$

where $\mathcal{L}(s)$ is as defined in the second part of definition 15 on page 89.

A typical semantic way for a theory to arise is as the set of things true in a given structure $\mathfrak{M}$. Such a theory is denoted by '$Th(\mathfrak{M})$'. Thus $Th(\mathfrak{M}) = \{\phi : \mathfrak{M} \models \phi\}$. Theories that arise in this way, as the set of things true in a particular structure, are of course complete—simply because of excluded middle.

A related typical way in which a theory can arise is as *the set of all sentences true in a given class of structures.*

Surprisingly some theories that arise in this second way can be complete too: DLO is the theory of dense linear orders. It is expressed in a language $\mathcal{L}$(DLO) with equality and one two-place predicate $<$. Its axioms say that $<$ is transitive and irreflexive, and that between any two things there is a third, and that there is no first or last element.

EXERCISE 57 *Write out the axioms of DLO. Can there be a finite model of DLO?*[7]

It's not hard to show that this theory is complete, using a famous construction of Huntington. However we cannot do that until we have corollary 30.

A famous example of an *in*complete theory is the theory known as *Peano Arithmetic*. Its incompleteness was proved by Gödel.

We need one more technicality: the concept of a **countable language**. A first-order language with a finite lexicon has infinitely many expressions in it, but the set of those expressions is said to be *countable*: that is to say we can count the expressions using the numbers 1, 2, 3, 4 . . . which are sometimes called the *counting numbers* and sometimes called the *natural numbers*. (If you were a mathematics or computer science student I would drag you kicking and screaming through a proof of the fact that the set of finite strings you can form from a finite alphabet can be counted.). The set of natural numbers is usually written with a capital 'N' in a fancy font, for example $\mathbb{N}$. There is some small print to do with the fact that we might have an infinite supply of variables . . . . After all, there is no limit on the length of expressions so there is no limit on the number of variables that we might use, so we want to be sure we will never run out. The best way to do this is to have infinitely many variables to start with. We can achieve this while still having a finite alphabet by saying that our variables will be not 'x', 'y' . . . but 'x', 'x'', 'x''' . . . the idea being that you can always make another variable by plonking a '$'$' on the right of a variable. (Notice that the systematic relation that holds between a variable and the new variable obtained from it by whacking it on the right with a '$'$' has no semantics: the semantics that we have cannot see through into the typographical structure of the variables.)

THEOREM 27 *Every theory in a countable language can be extended to a complete theory.*

Proof: Suppose T is a theory in a language $\mathcal{L}(T)$ which is countable. Then we count the formulæ in $\mathcal{L}(T)$ as $\phi_1, \phi_2 \ldots$ and define a sequence of theories T_i as follows.

$T_0 = T$ and thereafter

T_{i+1} is to be T_i if T_i decides ϕ_i and is $T_i \cup \{\phi_i\}$ otherwise.

[7] Some of my students said there is if you drop axiom 6, or something like that. There's a belief there that the object persists through changes done to it, like Theseus' ship. Sets aren't like that. Must find something useful to say about this. . .

The theory $T_\infty =: \bigcup\{T_i : i \in \mathbb{N}\}$ is now a complete theory. ■

There is no suggestion that this can be done effectively

Set-theoretic notation. . . ?

4.14.3 Completeness

∈-terms

For any theory T we can always add constants to $\mathcal{L}(T)$ to denote witnesses to $\exists^n$ sentences in T.

$\exists^n$ sentence? 'witness' to existential quantifiers not explained yet

Suppose $T \vdash (\exists x)(F(x))$. There is nothing to stop us adding to $\mathcal{L}(T)$ a new constant symbol 'a' and adding to T an axiom $F(a)$. Clearly the new theory will be consistent if T was. Why is this? Suppose it weren't, then we would have a deduction of $\perp$ from $F(a)$. But T also proves $(\exists x)(F(x))$, so we can do a $\exists$-elimination to have a proof of $\perp$ in T. But T was consistent.

Notice that nothing about the letter 'a' that we are using as this constant tells us that a is a thing which is F. We could have written the constant 'a_F' or something suggestive like that. Strictly it shouldn't matter: variables and constant symbols do not have any internal structure that is visible to the language[8], and the 'F' subscript provides a kind of spy-window available to anyone *mentioning* the language, but not to anyone merely *using* it. The possibility of writing out novel constants in suggestive ways like this will be useful later.

EXERCISE 58

(1) Find a proof of the sequent $\vdash (\exists x)(\forall y)(F(y) \to F(x))$*;*
(2) Find a natural deduction proof of $(\exists x)(\forall y)(F(y) \to F(x))$*;*
(3) Find a proof of the sequent $\vdash (\exists x)(F(x) \to (\forall y)(F(y)))$*;*
(4) Find a natural deduction proof of $(\exists x)(F(x) \to (\forall y)(F(y)))$.

The first item tells us that for any F with one free variable we can invent a constant whose job it is to denote an object which has property F as long as anything does. If there is indeed a thing which has F then this constant can denote one of them, and as long as it does we are all right. If there isn't such a thing then it doesn't matter what the constant denotes. There is a similar argument for the formula in parts 3 and 4. The appeal to the law of excluded middle in this patter should alert you to the possibility that this result is not constructively correct. (So you should expect to find that you have to use the rule of double negation in parts 2 and 4 and will have two formulæ on the right at some point in the proof of parts 1 and 3.)

Do they know "constructively correct"?

This constant is often written $(\epsilon x)F(x)$. Since it points to something that has F as long as there is something that has F, we can see that

$$(\exists x)(F(x)) \quad \text{and} \quad F((\epsilon x)F(x))$$

are logically equivalent. So we have two rules

$$\frac{(\exists x)(F(x))}{F((\epsilon x)F(x))} \quad \text{and} \quad \frac{F((\epsilon x)F(x))}{(\exists x)(F(x))}$$

[8] Look again at formula 11.2 on page 218 and the discussion on page 104.

The right-hand rule is just a special case of ∃-introduction but the left-hand rule is new, and we call it ϵ-introduction. In effect it does the work of ∃-elimination, because in any proof of a conclusion ϕ using ∃-elimination with an assumption $(\exists x)F(x)$ we can replace the constant (as it might be) 'a' in the assumption $F(a)$ being discharged by the ϵ term '$(\epsilon x)F(x)$' to obtain a new proof of ϕ, thus:

$$\dfrac{\begin{array}{c}[A(t)]^{(1)}\\ \vdots \\ C\end{array} \qquad (\exists x)(A(x))}{C}\ \exists\text{-elim}(1) \tag{4.7}$$

with

$$\begin{array}{c}\dfrac{(\exists x)(A(x))}{A((\epsilon x)(A(x)))}\ \epsilon\text{-int} \\ \vdots \\ C\end{array} \tag{4.8}$$

... where, in the dotted part of the second proof, 't' has been replaced by '$(\epsilon x)(A(x))$'

Notice that this gives us an equivalence between a formula that definitely belongs to predicate calculus (co's it has a quantifier in it) and something that appears not to[9]. Hilbert was very struck by this fact, and thought he had stumbled on an important breakthrough: a way of reducing predicate logic to propositional logic. Sadly he hadn't, but the ϵ-terms are useful gadgets all the same, as we are about to see.

Explain 'binder'?

Observe that the failure of Exercise 58 constructively is not a fact purely about the constructive existential quantifier, but also the constructive conditional. Check that constructively there isn't even a proof of $\neg(\forall x)\neg(\forall y)(F(y) \to F(x))$; nor is there a proof of $(\exists x)(\forall y)\neg(F(y) \wedge \neg F(x))$—so it isn't the constructive conditional either!

THEOREM 28 *Every consistent theory in a countable language has a model.*

Proof:

Let T_1 be a consistent theory in a countable language $\mathcal{L}(T_1)$.

We do the following things

1. Add axioms to T_1 to obtain a complete extension;
2. Add ϵ terms to the language.

Notice that when we add ϵ-terms to the language we add new formulæ: if '$(\epsilon x)F(x))$' is a new ϵ-term we have just added then '$G((\epsilon x)F(x)))$' is a new formula, and T_1 doesn't tell us whether it is to be true or to be false. That is to say $\mathcal{L}(T_1)$ doesn't contain '$(\epsilon x)F(x)$' or '$G((\epsilon x)F(x)))$'. Let $\mathcal{L}(T_2)$ be the language obtained by adding to $\mathcal{L}(T_1)$ the expressions like '$(\epsilon x)F(x)$' and '$G((\epsilon x)F(x)))$'.

We extend T_1 to a new theory in $\mathcal{L}(T_2)$ that decides all these new formulæ we have added. This gives us a new theory, which we will—of course—call T_2. Repeat and

[9]The 'ϵ' is not a quantifier, but it is a *binder*: something that binds variables. '∃' and '∀' are binders of course, and so is 'λ' which we will meet in chapter 7.

take the union of all the theories T_i we obtain in this way: call it T_∞. (Easy to see that all the T_i are consistent—we prove this by induction).

It's worth thinking about what sort of formulæ we generate. We added terms like $(\epsilon x)(F(x))$ to the language of T_1. Notice that if H is a two-place predicate in $\mathcal{L}(T)$ then we will find ourselves inventing the term $(\epsilon y)H(y, (\epsilon x)F(x))$ which is a term of—one might say—*depth* 2. And there will be terms of depth 3, 4 and so on as we persist with this process. All atomic questions about ϵ terms of depth n are answered in T_{n+1}.

T_∞ is a theory in a language $\mathcal{L}_\infty$, and it will be complete. The model $\mathfrak{M}$ for T_∞ will be the structure whose carrier set is the set of ϵ terms we have generated *en route*. All questions about relations between the terms in the domain are answered by T_∞. Does this make $\mathfrak{M}$ into a model of T? We will establish the following:

LEMMA 29 $\mathfrak{M} \models \phi(t_1, \ldots t_n)$ *iff* $T_\infty \vdash \phi(t_1, \ldots t_n)$

Proof: We do this by induction on the logical complexity of ϕ. When ϕ is atomic this is achieved by stipulation. The induction step for propositional connectives is straightforward. (Tho' for one direction of the '$\vee$' case we need to exploit the fact that T_∞ is complete, so that if it proves $A \vee B$ then it proves A or it proves B.)

The remaining step is the induction step for the quantifiers. They are dual, so we need consider only $\forall$. We consider only the hard direction ($L \rightarrow R$).

Suppose $\mathfrak{M} \models (\forall x)\phi(x, t_1, \ldots t_n)$. Then $\mathfrak{M} \models \phi(t_0, t_1, \ldots t_n)$ for all terms t_0. In particular it must satisfy it even when $t_0 = (\epsilon x)(\neg\phi(x, t_1, \ldots t_n))$, which is to say

$$\mathfrak{M} \models \phi((\epsilon x)(\neg\phi(x, t_1, \ldots t_n)), t_1, \ldots t_n)$$

So, by induction hypothesis we must have

$$T_\infty \vdash \phi((\epsilon x)(\neg\phi(x, t_1, \ldots t_n)), t_1, \ldots t_n)$$

whence of course $T_\infty \vdash (\forall x)\phi(x, t_1, \ldots t_n)$.

■

This completes the proof of theorem 28. Observe the essential rôle played by the ϵ terms.

This is a result of fundamental importance. Any theory that is not actually self-contradictory is a description of *something*. It's important that this holds only for first-order logic. It does not work for second-order logic, and this fact is often overlooked. (If you want a discussion of this, look at appendix 11.3.2). A touching faith in the power of the completeness theorem is what lies behind the widespread error of reifying possibilities into possible worlds. See [17].

Notice that this proof gives us something slightly more than I have claimed. If the consistent theory T we started with was a theory in a countable language then the model we obtain by the above method is also countable. It's worth recording this fact:

COROLLARY 30 *Every consistent theory in a countable language has a* **countable** *model.*

Is this the point at which to start making a fuss about nonconstructive proof?

4.15 Interpolation

There is a precise analogue in predicate calculus of the interpolation lemma for propositional logic of section 3.7.

THEOREM 31 *The Interpolation Lemma*
If $A \to B$ is a valid formula of first-order logic then there is a formula C containing only predicate letters that appear in **both** *A and B such that $A \to C$ and $C \to B$ are both valid formulæ of first-order logic.*

A proof of this fact is beyond the scope of this course. The proof relies on the **subformula property** mentioned earlier. The disjoint-vocabulary case is intuitively obvious, but it's not at all clear how to do the induction.

Close attention to the details of the proof of the completeness theorem will enable us to prove it and get bounds on the complexity of the interpolating formula. These bounds are not very good!

The interpolation lemma is probably the most appealing of the consequences of the completeness theorem, since we have very strong intuitions about irrelevant information. Hume's famous dictum that one cannot derive an "ought" from an "is" certainly arises from this intuition. The same intuition is at work in the hostility to the *ex falso sequitur quodlibet* that arises from time to time: if there has to be a connection in meaning between the premisses and the conclusion, then an empty premiss—having no meaning—can presumably never imply anything.

4.16 Compactness

Recall section 3.10 at this point.

4.17 Skolemisation

EXERCISE 59 *Using either natural deduction or sequent calculus, deduce*
$(\forall x_1)(\exists y_1)(\forall x_2)(\exists y_2)(R(x_1, y_1) \wedge R(x_2, y_2) \wedge (x_1 = x_2 \to y_1 = y_2))$
from
$\forall x \exists y R(x, y)$

Schütte's proof

Skolemised theories and synonymy.

4.18 What is a Proof?

"No entity without identity" said Quine. The point he is making is that you can't claim to know what your entities (widgets, wombats …) are until you have a way of telling whether two given widgets, wombats … are the same widget, wombat … or not. One of the difficulties with proof theory is that although our notions of proof allow us to tell whether two proofs are the same or not, they generally seem to be too fine.

Consider the sequent

$$P(a), P(b) \vdash (\exists x)P(x).$$

Given our concept of sequent proof it has two proofs depending on whether we instantiate 'x' to 'a' or to 'b'. But do we really want to distinguish between these two proofs? Aren't they really the same proof? This looks like a problem: if one had the correct formal concept of proof, one feels, it would not be making spurious distinctions like this. A correct formalisation would respect the folk-intuitions that the prescientific notion comes with. Not all of them, admittedly, but some at least, and surely this one. Arriving at the most parsimonious way of thinking about phenomena is part of what good old conceptual analysis is for.

But does it matter? There are people who say that it doesn't. Ken Manders is a philosopher of Mathematics at Pittsburgh who says that *all* formalisations of prescientific concepts result in spurious extra detail in this way, and that it's just a fact of life. His favoured examples are knots and computable functions. He thinks this is inevitable: this is the kind of thing that does just happen if you mathematise properly. Typically there won't be just one right way of thinking about any mathematical entity. The error of thinking that there is always precisely one he lays at Frege's door. Work to do here

This makes me think: might we not be able to avoid this overdetermination by having an operationalist view of mathematical entities? Operationalism is usually a dirty word in Philosophy of Science and Ken says this is because it results in very impoverished theoretical entities (He mentions Bridgeman in this connection).

So why might it be less problematic in Mathematics? Anything to do with the idea that Mathematics has no subject matter? If you are a scientific realist then operationalism is clearly a bad idea because it won't capture the full throbbing reality of the entities you are looking at. But in *Mathematics*...? If it is true that anything done with sufficient rigour is part of Mathematics then we might be all right. Of course the idea that Mathematics has no subject matter is just, in new dress, the old idea that all of Mathematics is *a priori* and has no empirical content. Better still, it might even be the correct expression of that insight. Reference?

EXERCISE 60 *If S and T are theories, $S \cap T$ is the set of those formulæ that are theorems of both S and T.*

Show that, if S and T are both finitely axiomatisable, so is $S \cap T$.

4.19 Relational Algebra

Some syllogisms seem to invite formalisation without variables:

"All A are B, all B are C, so all A are C"

asks to be formalised as:

$$A \subseteq B, B \subseteq C, \text{ so } A \subseteq C;$$

Similarly

"All A are B, some A are C, so some B are C"

asks to be formalised as

$$A \subseteq B, A \cap C \neq \emptyset, \text{ so } B \cap C \neq \emptyset.$$

We are tempted to try to get rid of the "intrusive" variable 'x' in '$(\forall x)(A(x) \to B(x))$', '$(\exists x)(A(x) \wedge C(x))$' and so on, and to express this all using only the Venn diagram stuff we learnt at school.

We recycle the predicate letters as letters that point to sets. Predicate letters point to predicates, which can be thought of as either properties-in-intension or properties-in-extension—it doesn't matter. The sets that the recycled letters point to are definitely extensional things!

Fit this in somewhere: take a formula of the Σ^1 fragment of second-order logic. Delete the existential quantifiers. The result is a formula in 1st order logic with function letters. If it is refutable then so was the Σ^1 formula we started with. So there is a refutation procedure for the Σ^1 fragment of second-order logic.
Similarly there is a refutation procedure for the set of formulæ true in all finite structures.

All this is fine. In fact all the earlier exercises 45 [and possibly others—check!] that use only monadic logic can be captured in this way. The reader might like to attempt a few, just to get the hang.

We can do even more, and quite neatly, but we need cartesian product. (This is where it turns up first)

We encountered ordered pairs on page 36. $X \times Y$ is the set of ordered pairs whose first components are members of X and whose second components are members of Y. A binary relation-in-extension on a set X is simply a subset of $X \times X$. We write 'X^2' instead of '$X \times X$'.

Thus armed we can give cute characterisations of the properties of *transitive*, *reflexive*, *symmetrical* that we saw first in definition 23:

DEFINITION 32 *If $R \subseteq D^2$ then*

R is symmetrical if $R = R^{-1}$;
R is transitive if $R^2 \subseteq R$;
R is reflexive if $\mathbb{1}_D \subseteq R$;
R is irreflexive if $\mathbb{1}_D \cap R = \emptyset$.

What is $\mathbb{1}_D$? Well, given that you know what a reflexive relation is you can deduce what $\mathbb{1}_D$ must be . . . it's the identity relation on members of D: $\{\langle x, x\rangle : x \in D\}$.

The first line of definition 32 involves an 'R^2'. Do we ever naturally have R^3, R^4 and so on? After all, if R is transitive we will have $R^3 \subseteq R$, $R^4 \subseteq R$ and so on. Observe that we can infer all these from our definition of transitivity because

$$R \subseteq S \to R \circ T \subseteq S \circ T$$

Composition of two reflexive relations is reflexive;
Intersection of two reflexive relations is reflexive[10];
Intersection of two transitive relations is transitive;

All these facts are expressible in first-order logic and are in fact theorems of first-order logic.

[10] connect this with Frege's dfn of equality

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

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

A proof would be quite a mouthful.

Look at exercise 48 parts 4 and following:

The friend of my friend is my friend;	$F \circ F \subseteq F$;
The enemy of an enemy is a friend;	$E \circ E \subseteq F$;
The enemy of a friend is an enemy;	$E \circ F \subseteq E$;
The friend of an enemy is an enemy.	$F \circ E \subseteq E$.

The way to prove these equations is by appeal to extensionality.

It's a nice fragment of second-order logic

Chapter 5

Constructive and Classical truth

stuff to fit in

Include my constructive treatment of Sorites? Supervaluationism?

There is one great advantage of constructive existence proofs: there is a technique for detecting mistakes in (purported) constructive existence proofs that isn't available for detecting mistakes in (purported) *non*constructive existence proofs. The constructive existence proof of a wombat tells you where to find the wombat; if you go there and find no wombat, you know you've made a mistake. A nonconstructive existence proof doesn't provide that sort of check. The existence proof of The Great Watchmaker is nonconstructive, so it doesn't tell us where to find the Great Watchmaker. (There is Mock Turtle soup so there must be a Mock Turtle!) Thus Paley cannot be refuted merely by saying "I went to where you told me, and i found nothing!". He may be refuted by other means of course, but . . . indeed there are still people hoping to find him.

It's very tempting to think that because two proofs give you different information, they must therefore be proofs of different propositions. This reminds me of the way in which the deduction theorem for a logic internalises information about that logic. Constructive logic internalises some things that classical logic does not. See also Kleene's theorem about finite axiomatisability.

To the classically-minded, when someone asserts $\neg\neg p$ but doesn't assert p—thereby revealing that they think there is a difference—it seems that that difference is that they are not making-an-assertion-about-the-subject-matter-of-p [making an assertion *at the same level* as p so to speak] so much as an assertion about the assertion p itself—about [for example] the prospects for establishing or refuting it.

There are times when you *don't* care—in the slightest—that your existence proof is nonconstructive. Here is a real-life example. The City of Wellington reconfigured its one-way system in the city centre, to make motorway access and egress easier. Some one-way streets became one-way in the opposite direction (sense!). In these circumstances one can be entirely confident that, one day, some Wellingtonian crossing one of these streets on autopilot will look for oncoming traffic in the wrong direction, see none, step out and get mown down.

burble *ensemble properties*.

If you are kind of constructivist who says that because this existence proof (of a pedestrian casualty) is not constructive then we don't have to accept its conclusion, then you are misunderstanding something! The fact that the proof is nonconstructive doesn't give us any grounds to hope that no-one will be mown down.

There is an issue in here of more than merely logical concern. Public policy affects the environment within which individual people make their own decisions. If some of the decisions made by those individuals result in blameworthy acts, where does the blame lie? If morality is to have any meaning then people must be blamed for the blameworthy acts they perform, but should they cop *all* the blame that is going? If some of the blameworthy acts being performed would not have been performed if public policy had taken a different tack, presumably some of the blame being dished out should lie with the people who control public policy. But how much?

There are some features of this situation that may be worth discussing. The actual real-life chain of reasoning doesn't stop with "$(\exists x)(x$ is going to get killed)" but rather "We shouldn't reconfigure". Constructively we seem to have established "$\neg\neg(\exists x)(x$ is going to get killed)" What is the conclusion we are trying to draw? It will be one of

> "We should not reconfigure"
>
> and
>
> "It is not the case that we should reconfigure"

and these two are distinct, even classically. Constructively (tho' not classically) we can distinguish between negative and non-negative expressions. And any negative assertion that follows constructively from p also follows from $\neg\neg p$. So, since we have

> "$(\exists x)(x$ is going to get killed)" implies "It is not the case that we should reconfigure"

and we know "$\neg\neg(\exists x)(x$ is going to get killed)", and that the conclusion is negative, we can infer it from the doubly negated

> "$\neg\neg(\exists x)(x$ is going to get killed)"

which we *have* established, we can safely conclude that it is not the case that we should configure.

Nonconstructive proof is widespread in real life.

Every artefact has a purpose. It is often possible to determine that something is an artefact without knowing what that purpose is. When you've done that you have a nonconstructive existence proof.

When you find in the library a body with a knife in its heart you know there is a murderer but you have no idea who it is. One could say that—seen from outside—the logical structure of a murder mystery is the process of replacing a nonconstructive proof of the existence of a murderer by a constructive existence proof.

end of digression

Suppose you are the hero in a mediæval romance. You have to rescue the Princess from captivity and maltreatment at the hands of the Evil Dragon. To do this you will of course need a Magic Sword to cut off the head of the Evil Dragon, and a Magic Key to open the door of the dungeon, co's it had a Spell put on it by the Evil Dragon so, if you are to open it, only a Magic Key will do. How are you to procede? You can cheer yourself up with the thought: "Things aren't as bad as they might look. After all, God In His Infinite Mercy would not have given me this task if it were impossible, if there were no Magic Key or no Magic Sword it would be impossible. This being the case there must be a Magic Key and there must be a Magic Sword!

A better example is ϵ-terms

You can say this to yourself—and you'd be right: there must indeed be a Magic Sword and a Magic Key. However this is not a great deal of use to you. It doesn't begin to solve the problem, since what you want is not an existence theorem for Magic Keys and Magic Swords—what you actually want is to *find* the gadgets and have them in your little hot hand. And the chain of reasoning you have just put yourself through, sound tho' it undeniably is, tells you nothing about where to find the damned things. It's reassuring up to a point, in that this inference-from-authorial-omniscience constitutes a sort of prophecy that the Magic Key and Magic Sword will turn up eventually, but it doesn't put them in your wee sweaty hand. (We will return to Prophecy in section 5.3).

The problem I am trying to highlight[1] is one that arises most naturally in Mathematics, and it is in Mathematics that the clearest examples are to be found. (See the discussion in appendix 11.4.) The mediæval romance is the best I can do in the way of a non-mathematical example. The phenomenon goes by the name of **nonconstructive existence theorem.** We have a proof that there is a whatever-it-is, but the proof that there is one does not reveal where the whatever-it-is is to be found. Further, this is an example of a situation where a nonconstructive existence theorem is of very little use, which of course is why we worry about having a nonconstructive existence proof.

In order not to find ourselves in the predicament of the hero of the mediæval romance who has proved the existence of the sword and the key but does not know where to find them we could consider restricting the principles of reasoning we use to those principles which, whenever they prove that $(\exists x)(\text{Sword}(x))$, also prove $\text{Sword}(a)$ for some a. The thinking behind this suggestion is that the Hero's energies (and perhaps his wits) are limited, and there is therefore no point in having clever inferences that supply him with information that he cannot use and which will only distract him.

The principles of reasoning it is suggested we should restrict ourselves to are said to be **constructive** and proofs constructed in accordance with them are also said to be constructive. We have to be able to *exhibit* the things whose existence we think we have proved. In fact, one of my students said that principles of reasoning that were well-behaved in this way should be called "exhibitionist" and that the philosophy of mathematics that insisted on them should be called "exhibitionism".

(A reluctance to infer $\forall x F(x)$ from $\neg\exists x \neg F(x)$ may be what is behind the reluctance a lot of people have in concluding that vacuous universal quantification always gives you the `true`. $\bigcap \emptyset = V$. Trivially everything belongs to all members of the empty set.

One reason why people are reluctant to accept "All A are B" as a good inference when A is empty is that when A is empty there is no experimental support for "All A are B"

[1] I nearly said *showcase* there…

Clearly there cannot be a member of the empty set to which you do not belong (that's a $\neg\exists\neg$) so you belong to all of them.)

BAD JOIN

You are sitting in your shed, your workshop, at your bench, building a widget. Now it happens that at some point you have to ascertain whether or not p holds. If it does, your construction of a widget procedes along one path; if it doesn't, it proceeds along a different path. The law of excluded middle tells you that one of the two will hold, so you're OK either way: you succeed in constructing you widget. You have an existence proof: there is (or will be) a widget. You don't know which of the two (or more—there may be more than one fork in the road) widgets you get, but you know you will get one.

So you at least get your existence proof. However, your proof is not constructive. Notice the significance of the law of excluded middle here. The foregoing means that if excluded middle didn't hold then you wouldn't *have* an existence proof. You come to a fork in the road but not only do you not know which branch you will be taking you don't even know that there is a way through it. So, in particular ('*a fortiori*"!) you don't get a *nonconstructive* existence proof. The point is not that if you didn't have excluded middle your proof of the existence theorem would have been constructive; the point is that if you didn't have excluded middle you wouldn't have an existence theorem at all. But this does at least mean that if you eschew excluded middle then you have closed off *one* way, at least, in which nonconstructive existence theorems can happen.

There is a standard example of a conconstructive proof that features the law of excluded middle in precisely this way. Suppose you are given the challenge of finding two irrational numbers α and β auch that α^β is rational. It is in fact the case that both e and $log_e(2)$ are transcendental but this is not easy to prove. Is there an easier way in? Well, one thing every schoolchild knows is that $\sqrt{2}$ is irrational, so how about taking both α and β to be $\sqrt{2}$? This will work if $\sqrt{2}^{\sqrt{2}}$ is rational. Is it? As it happens, it isn't (but that, too, is hard to prove). If it isn't, then we take α to be $\sqrt{2}^{\sqrt{2}}$ (which we now believe to be irrational—had it been rational we would have taken the first horn) and take β to be $\sqrt{2}$.

α^β is now

$$(\sqrt{2}^{\sqrt{2}})^{\sqrt{2}} = \sqrt{2}^{\sqrt{2}\cdot\sqrt{2}} = \sqrt{2}^{2} = 2$$

which is rational.[2]

What does this prove? It certainly doesn't straightforwardly show that the law of excluded middle is *false*; it does show that there are situations where you don't want to reason with it. There is a difference between proving that there is a widget, and actually getting your hands on the widget. Sometimes it matters, and if you happen to be in the kind of pickle where it matters, then you want to be careful about reasoning with excluded middle. But if it doesn't matter, then you can happily use excluded middle.

[2]We can actually exhibit such a pair, and using only elementary methods, at the cost of a little bit more work. $log_2(3)$ is obviously irrational: $2^p \neq 3^q$ for any naturals p, q. $log_{\sqrt{2}}(3)$ is also irrational, being $2 \cdot log_2(3)$. Clearly $(\sqrt{2})^{(log_{\sqrt{2}}(3))} = 3$.

"But what" i can hear you asking "is special about excluded middle? What happens if the fork in the road is $p \vee q$ where we know that one of them must hold (so we know that the process continues) but we don't know which? Don't we have the same problem?" A very good point. This is why we have to design constructive logic so that whenever you prove $p \vee q$ it is beco's you have proved (or can prove) one of the two, p and q.

BAD JOIN

Excluded middle isn't the only thing that can give us nonconstructive proofs. Let's go back to our mediæval-romance illustration. What principle of reasoning have we used that conflicts with exhibitionism? Well, we started off by supposing that there was no key and no sword, and found that this contradicted the known fact there is a happy ending. So our assumption must have been wrong. It isn't true that there is no key and no sword. That is to say

$\neg$(There is no key) and $\neg$(There is no sword) (*)

And from this we wished to infer

There is a key and a sword (**)

Now our proof of (*) can't violate exhibitionism—not literally at least—co's (*) isn't of the form $(\exists x)\ldots$. But our proof of (**) definitely can—and it does. And since the only thing we did to our proof of (*) (which was exhibitionist) to obtain the proof of (**) (which is not exhibitionist) is to apply of the law of double negation then clearly that application of the law of double negation was the fatal step.

(And this isn't a rerun of the problem with *reductio ad absurdum* that we saw in section 3.2.1!!!)

Since we can sometimes find ourselves in situations where a nonconstructive proof is no use to us, we want to distinguish between constructive and nonconstructive proofs of, say

$(\exists x)(x$ is a Magic Sword). (MS)

Typically (tho' not invariably) a nonconstructive proof of MS will take the form of an assumption that there are no Magic Swords followed by a deduction of a contradiction from it. Such a proof can be divided into two parts:

1. a first half—*not* using excluded middle or double negation—in which we derive a contradiction from $\neg(\exists x)(x$ is a Magic Sword), and thereby prove $\neg\neg(\exists x)(x$ is a Magic Sword); and

2. a second part in which we use the law of double negation to infer $(\exists x)(x$ is a Magic Sword).

This certainly throws the spotlight on the law of double negation. Let's intermit briefly to think about it. One thing to notice is that we can give a natural deduction proof of *triple negation*: $\neg\neg\neg p \to \neg p$ without using the rule of double negation. Indeed we can prove $(((p \to q) \to q) \to q) \to (p \to q)$ just using the rules for $\to$. (This was part 6 of exercise 28 on page 66.)

Classically we acknowledge nonconstructive proof (in that we think the second part of the proof is legitimate) and we believe that $(\exists x)(x$ is a Magic Sword) and $\neg\neg(\exists x)(x$ is a Magic Sword) are the same proposition—and we can do this even while recognising the important difference between constructive proof and nonconstructive proof. Is there anything to be said for a contrasting viewpoint in which we acknowledge only constructive proof and we believe that $(\exists x)(x$ is a Magic Sword) and $\neg\neg(\exists x)(x$ is a Magic Sword) are *different* propositions? That is to say we renounce step (2) not because it gives us misleading information about a true conclusion (namely that it tells us that there is a Magic Sword without telling us where to find it) but rather because it tells us something that is simply not true!

The first thing to say here is that our desire to distinguish between constructive and nonconstructive proof absolutely does not commit us to this second position. It would be an error to think that because we wish to eschew certain kinds of proof it therefore follows either that the proofs are not good proofs or that the things whose proofs are eschewed are not true, or have not been proved. This error has parallels elsewhere. Here are five I can think of, and no doubt the reader can think of more.

- Philosophers of Science are—rightly—concerned that the endeavour to understand science done by earlier people in the West should not be seen merely as part of a process whose culminating point is us. They warn us against doing 'whig history'. One strategy for doing this is to pretend that there is no such thing as progress in the sciences.

- People who study sociology of science are concerned with how scientific theories propagate though communities. For them, questions of the content and truth of those theories are a distraction, and one strategy for not being distracted is to pretend that the theories simply do not have content.

- A strategy for not worrying about the ills to which flesh is heir is to deny the reality of matter.

- The law of rape protects girls under the age of consent from the sexual attentions of men. It protects them whether they are 4, or even 15 and sexually mature. People who are concerned to protect adolescent girls will not wish any debate on how to do it to be sidetracked into a discussion of precisely how much worse a rape of a 4-year old is that of a 15-year old. One way of forstalling such a discussion is to deny that between these two crimes is there any difference to be discussed.

- Psychotherapists have to help their clients in their (the clients') difficulties in personal relations. The pyschotherapist has no way of telling whether or not the client's version of events is true, but they have to help anyway. *Therefore the truth (or otherwise) of the client's story cannot be a consideration.* In these circumstances it is easy to slip into the position that there is no such thing as truth.

The fact that the inference from considerations like MS to exhibitionism is fallacious doesn't mean that exhibitionism is mistaken. (If you wish to pursue this look at [14], [27] and [28].)

Even if constructivism is a mistake there might nevertheless be something to be said for exploring some of the consequences of adopting it: plenty of truths have been inferred from falsehoods (see `www.dpmms.cam.ac.uk/~tf/kannitverstan.html`).

5.1 The Radical Translation Problem with Classical and Constructive Logic

Radical Translation is the problem confronted by the field anthropologist observing members of an exotic tribe going about their everyday business, doing things, making utterances and expressing agreement or disagreement. All this is going on in a language the field anthropologist has no dictionary for, and no interpreter. The problem is: how does the anthropologist translate utterances of the exotic tribe's language into his-or-her own language? There is a procedural problem of course: ("how do you set about it?") but there is also a more philosophical problem: what are the criteria for success or failure? Should the anthropologist be willing to ascribe deviant notions of *truth* or deviant notions of *inference* to the tribe if that makes the translation go more smoothly? Might the anthropologist ever be forced to the conclusion that the people of the alien tribe do not believe the law of non-contradiction, for example?

Quine wrote extensively about this problem of radical translation (it all starts in [37]), and his general drift is that the anthropologist would never (or hardly ever!) be forced into concluding that the tribe has a deviant notion of truth or a deviant logic; there would always be enough slop in the system for one to be able to reinterpret one's way out of such an impasse. The catchphrase associated with this view was "the indeterminacy of translation".

The general view nowadays seems to be that Quine was wrong in at least some of what he wrote about this, if not all of it. However he did at least do us a favour by making us think about what the criteria for correct *versus* incorrect translations might be. Constructive and classical logic might be a good case study because we have quite a lot of data to work on. How are classical logicians and constructive logicians to make sense of what the other is saying?

Edit below here

Do constructivists have a different concept of proposition from the rather operational concept held by classical logicians? For that matter do paraconsistentists have a different concept of proposition..?

Is it that the two parties have different propositional attitudes that they are calling by the same name? Or do they have the same attitudes to two different propositions for which they are using the same description? Can they agree on a description of their disagreement?

Readers do not know what a propositional attitude is; this is the first appearance of this idiom

This touches a very delicate area in philosophy, and one on which there is very little satisfactory literature. How can one give a coherent acount of the incoherent? (cf Prior on the Liar paradox ?reference)

Say something about how there is nothing like the negative interpretation for paraconsistent logics.

The classical logician probably regards the intuitionist's insistence on putting double negations in front of propositions that haven't been proved constructively as a manœuvre that imports into the language some considerations that properly belong to pragmatics. He would say "The constructivist and I agree that there is a Magic Sword, but our reasons for being sure there is one don't actually give us a recipe for finding it. Why not just leave it at that? The logic is surely the last thing to mutilate!" This point (that *Logic is the last thing you tweak in order to accommodate data*) is one that Quine was fond of making.

The classical principles of reasoning preserve truth. What do the principles of constructive reasoning preserve? The answer you will give seems to depend on whether you are a classical or a constructive mathematician/logician/philosopher. From the classical point of view the answer seems to be that they preserve the property of having-a-proof-that-respects-exhibitionism. And from the constructive point of view? Some constructivists think that constructive reasoning preserves truth, and some would say that it preserves something rather-like-truth-but-not-exactly.

Leaving this second flavour of constructivist out of the debate for the moment one can ask: given that classical and constructive logicians agree that the purpose of reasoning is to preserve truth, is the disagreement between them a disagreement about

(i) which things are true? or

(ii) the nature of truth? or

(iii) which rules preserve truth?

If (ii) then does this disagreement arise from a different view of what propositions are?

For the classical logician a proposition is something that in each setting evaluates to a truth-value determined by that setting. You hold it up to the light and you see `true` or `false`.[3]

I suspect that the disagreement is rather over the idea that propositions are characterised by their propensity to evaluate to truth-values.

What is a proposition, constructively?

see [41] and [11].

5.2 Classical Reasoning from a Constructive Point of View

Let's approach this radical translation problem from the point of view of the constructive logician. Quine somewhere alludes to a *principle of charity*: there is a default assumption that what the foreigner is saying not only can be made sense of but can probably be made sense of in such a way that it comes out true.

[3] As it happens there are only two truth-values in this picture but the number of truth-values is not, I think, the point at issue. Indeed, constructivists even agree that (in some sense) there are no more than two truth values: the assumption that *no two of the three propositions A, B and C agree on their truth value* leads to a contradiction. That is to say there is a constructive proof of the sequent

$$\neg(A \longleftrightarrow B),\ \neg(B \longleftrightarrow C),\ \neg(A \longleftrightarrow C)\ \vdash$$

Do not waste time trying to find it—it is very big!

The considerations that led us to consider constructive logic lead us to expect that if A is a classical tautology then $\neg\neg A$ should be constructively correct. This is straightforwardly true in the propositional case, and was proved by Glivenko many years ago ([19] and [20].) Let's announce this fact as a theorem.

THEOREM 33 *If there is a classical proof of a formula Φ of propositional logic then there is a constructive proof of $\neg\neg\Phi$.*

Proof:

To do this properly we have to have a Hilbert-style axiomatisation (one whose sole rule of inference is *modus ponens*) that does not exploit any definitions of connectives in terms of other connectives. (We retain the definition of $\neg$ in terms of $\perp$). The obvious thing to do is replace every rule of inference by an axiom taking the form of a conditional whose antecedent is the premiss and whose consequent is the conclusion. This gives us immediately the following axioms:

$A \to A \vee B$	(from $\vee$-introduction)
$A \to B \vee A$	(from $\vee$-introduction)
$A \wedge B \to A$	(from $\wedge$-elimination)
$A \wedge B \to B$	(from $\wedge$-elimination)

If we have more than one premiss in the rule then one gets the following

$A \to (B \to (A \wedge B))$	(from $\wedge$-introduction)
$A \to (B \to (B \wedge A))$	(from $\wedge$-introduction)
$A \to ((A \to B) \to B)$	(from $\to$-elimination)
$(A \to B) \to (A \to B)$	(from $\to$-elimination)

The rule of double negation can be captured easily:

$$\neg\neg A \to A$$

The two "action at a distance" rules require a bit more thought. First $\to$-introduction: Suppose we have a Hilbert proof

$$\begin{array}{c} A \\ \vdots \\ B \end{array}$$

We want to obtain from this a Hilbert proof of

$$\begin{array}{c} \vdots \\ A \to B \end{array}$$

To do this we exploit the deduction theorem from section 3.6.1. For this it is sufficient to have K and S as axioms.

$\vee$-elimination is a bit harder. Suppose we have two proofs of C, one from A and the other from B:

$$\begin{array}{cc} A & B \\ \vdots & \vdots \\ C & C \end{array}$$

and we have $A \vee B$. How are we to obtain a proof of C?

Well, the two proofs of C will give us proofs of $A \to C$ and of $B \to C$ by means of the deduction theorem (theorem 13) . So all we need now is an axiom that says

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

Now, to complete the proof of Glivenko's theorem, suppose we have a Hilbert-style proof $\mathcal{D}$ of Φ:

$$\begin{array}{c} \vdots \\ \Phi \end{array}$$

Suppose we simply prefix every formula in the list with $\neg\neg$. What does that give us? The result—let us call it $\mathcal{D}^*$—isn't a Hilbert-style proof of $\neg\neg\Phi$ but we are very nearly there. It is a string of formulæ wherein every formula is either the double negation of a (substitution instance of an) axiom or the double negation of a theorem. There are two key facts that we now need:

1. The double negation of each of the new axioms is constructively provable;
2. There is a Hilbert-style proof (not using double negation!) of $\neg\neg B$ from $\neg\neg A$ and $\neg\neg(A \to B)$.

So, to obtain our proof of $\neg\neg\Phi$ from our proof $\mathcal{D}$ of A we first decorate $\mathcal{D}$ with double negations to obtain $\mathcal{D}^*$ as above. We next replace every occurrence of a doubly negated axiom in $\mathcal{D}^*$ with a prefix containing a proof of that doubly negated axiom that does not use the rule of double negation. Next, wherever we have an entry $\neg\neg B$ in the list that is preceded by $\neg\neg(A \to B)$ and $\neg\neg A$ we insert the missing lines from the Hilbert-style proof of $\neg\neg B$ from $\neg\neg(A \to B)$ and $\neg\neg A$.

The result is a Hilbert-style proof of $\neg\neg\Phi$. ■

EXERCISE 61

Provide, **without** *using the rule of double negation,*

a natural deduction proof of $\neg\neg((\neg A \to B) \to ((\neg A \to \neg B) \to A))$*;*

a natural deduction proof of $\neg\neg B$ *from* $\neg\neg A$ *and* $\neg\neg(\neg\neg A \to \neg\neg B)$.

Provide proofs of the following sequents, respecting the one-formula-on-the-right constraint.

$\vdash \neg\neg((\neg A \to B) \to ((\neg A \to \neg B) \to A))$*;*

$\neg\neg A,\ \neg\neg(\neg\neg A \to \neg\neg B) \vdash \neg\neg B$.

It's much harder to prove theorem 33 by reasoning about natural deduction proofs or sequent proofs instead of Hilbert proofs, though it can be done. The reader may have been wondering why we ever used Hilbert-style proofs in the first place, since they do not have the subformula property and are so hard to find. The reason is that they are much better than natural deduction proofs when it comes to proving results like this.

Theorem 33 doesn't work for predicate calculus because

$$\neg\neg(\neg(\forall x)(F(x)) \to (\exists x)(\neg F(x))) \tag{5.1}$$

is classically valid but is not constructively provable. Something like theorem 33 is true, but the situation is more complicated. In the propositional case, the constructive logician who hears the classical logician assert A can interpret it as $\neg\neg A$. If there are quantifiers lurking then the constructive logician not only has to whack '$\neg\neg$' on the front of A but has to do something to the inside of A, and it's not immediately obvious what that might be. Working out quite what has to be done to the inside of A was one of the many major contributions to Logic of Gödel [21].

There should be an exercise to find a countermodel for it

I think the analogue of Glivenko works for the $\exists$ fragment but not if you allow $\forall$.

5.2.1 Interpretations, specifically the Negative Interpretation

(If you are to do any philosophy you will need in any case to think a bit about the explanatory power of interpretations. It's behind a lot of reductionist strategies in the sciences. The negative interpretation is a nice simple example to start on.)

The way the constructive logician narrates this situation is something like the following. Here *grokking* is a propositional attitude whose precise nature is known at any rate to the constructive logician but possibly not to anyone else. The constructive logician muses:[4]

Readers do not know what a propositional attitude is

> "The classical logician reckons he can grok $A \vee B$ whenever he groks A or groks B but he also says that when he groks $A \vee B$ it doesn't follow from that—according to him—that he groks either of them. How different from me! When *I* grok $A \vee B$ it certainly follows that I grok at least one of them. Since—when he says that he groks $A \vee B$—he does at least say that in those circumstances he cannot grok either $\neg A$ or $\neg B$, it might be that what he really means is that he groks something like $\neg(\neg A \wedge \neg B)$, since he can at least grok *that* without grokking A or grokking B. Accordingly henceforth, whenever I hear him assert $A \vee B$, I shall mentally translate this into $\neg(\neg A \wedge \neg B)$. At least for the moment."

Or again:

> "When the classical logician says that he groks $(\exists x)W(x)$ it doesn't follow from that—according to him—that there is anything which he groks to be W, though he certainly groks $(\exists x)W(x)$ whenever there is an a such that he groks $W(a)$. How different from me! When *I* grok $(\exists x)W(x)$ there

[4] For you SciFi buffs: Robert Heinlein: Stranger in a Strange Land.

> most certainly is an x which I grok to be W. Since—when he says that he groks $(\exists x)W(x)$—it is entirely possible that there is no x which he groks to be W—it must be that what he really means is that he groks something like $\neg(\forall x)(\neg W(x))$ since he can at least grok *that* even without there being anything which he groks to be W. Accordingly henceforth whenever I hear him assert $(\exists x)W(x)$ I shall mentally translate this into $\neg(\forall x)(\neg W(x))$—at least until anybody comes up with a better idea."

and again:

> "Given what the classical logician says about the conditional and truth preservation, it seems to me that when (s)he claims to grok $A \rightarrow B$ all one can be certain of it that it cannot be the case that A is true and B is false. After all, (s)he claims to have a proof of $\neg\neg A \rightarrow A$! Accordingly henceforth whenever I hear them assert $A \rightarrow B$ I shall mentally translate this into $\neg(A \wedge \neg B)$. That covers the $\neg\neg A \rightarrow A$ case nicely, because it cannot be the case that $\neg\neg A$ is true but that A is false *and* it captures perfectly what the buggers say they mean."

Let us summarise the clauses in the translation here. ϕ^* is what the constructive logician takes the classical logician to be saying when they say ϕ.

"negatomic"?

DEFINITION 34 *We define ϕ^* by recursion on the subformula relation:*
ϕ^* is $\neg\neg\phi$ when ϕ is atomic; ϕ^* is ϕ when ϕ is negatomic;

$(\neg\phi)^*$ *is* $\neg(\phi^*)$;
$(\phi \vee \psi)^*$ *is* $\neg(\neg\phi^* \wedge \neg\psi^*)$;
$(\phi \wedge \psi)^*$ *is* $(\phi^* \wedge \psi^*)$;
$(\phi \rightarrow \psi)^*$ *is* $\neg(\phi^* \wedge \neg\psi^*)$;
$((\forall x)\phi(x))^*$ *is* $(\forall x)(\phi(x)^*)$;
$((\exists x)\phi(x))^*$ *is* $\neg(\forall x)(\neg\phi(x)^*)$.

What drives the constructivists' choices of readings of the classical logicians' utterances? How did they know to interpret $A \vee B$ as $\neg(\neg A \wedge \neg B)$? Why do they not just throw up their hands? Because of the principle of charity from p. 150: this interpretative ruse enables the constructivist to pretend, whenever the classical logician is uttering something that (s)he believes to be a classical tautology, that what is being uttered is something that the constructivist believes to be constructively correct. Isn't that a feature one would desire for a translation from my language into yours, that it should send things that look good in my world to things that look good in yours...? (One wouldn't want to go so far as to say that it enables the constructivist to actually *understand* the classicist, but it does enable him to construe what he hears as both sensible and true.)

The claim is that if ϕ is a classical tautology then ϕ^* is constructively provable. In fact we will prove something rather more fine-grained. For this we need the notion of a stable formula.

DEFINITION 35 *A formula ϕ is* **stable** *if $\neg\neg\phi \to \phi$ is constructively correct.*

This is an important notion because if we add the law of double negation to constructive propositional logic we get classical propositional logic; nothing more is needed.

We will need the following

LEMMA 36 *Formulæ built up from negated and doubly-negated atomics solely by $\neg$, $\wedge$ and $\forall$ are stable.*

Proof: We do this by induction on quantifiers and connectives.

For the base case we have to establish that $\neg\neg A \to A$ holds if a is a negatomic or a doubly negated atomic formula. This is easy. The induction steps require a bit more work.

$\neg$:

For the case of $\neg$ we need merely the fact that triple negation is the same as single negation. In fact we can do something slightly prettier.[5]

$$\begin{prooftree}
\AxiomC{$[p]^2$}
\AxiomC{$[p \to q]^1$}
\RightLabel{$\to$-elim}
\BinaryInfC{q}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$(p \to q) \to q$}
\AxiomC{$[((p \to q) \to q) \to q]^3$}
\RightLabel{$\to$-elim}
\BinaryInfC{q}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$p \to q$}
\RightLabel{$\to$-int (3)}
\UnaryInfC{$(((p \to q) \to q) \to q) \to (p \to q)$}
\end{prooftree} \tag{5.2}$$

...noting that $\neg p$ is just $p \to \bot$.

$\wedge$:

We want to deduce $(p \wedge q)$ from $\neg\neg(p \wedge q)$ given that we can deduce p from $\neg\neg p$ and that we can deduce q from $\neg\neg q$. The following is a derivation of $\neg\neg p$ from $\neg\neg(p \wedge q)$:

$$\begin{prooftree}
\AxiomC{$[p \wedge q]^1$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{p}
\AxiomC{$[\neg p]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$\neg(p \wedge q)$}
\AxiomC{$\neg\neg(p \wedge q)$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$\neg\neg p$}
\end{prooftree} \tag{5.3}$$

and the following is a derivation of $\neg\neg q$ from $\neg\neg(p \wedge q)$:

$$\begin{prooftree}
\AxiomC{$[p \wedge q]^1$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{q}
\AxiomC{$[\neg q]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$\neg(p \wedge q)$}
\AxiomC{$\neg\neg(p \wedge q)$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$\neg\neg q$}
\end{prooftree} \tag{5.4}$$

[5] This was part 6 of exercise 28 on page 66.

But both p and q are stable by induction hypothesis, so we can deduce both p and q and thence $p \wedge q$.

$\forall$:

First we show $\neg\neg\forall \to \forall\neg\neg$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[(\forall x)\phi(x)]^1}{\phi(a)}\ \forall\text{ elim} \quad [\neg\phi(a)]^2}{\bot}\ \to\text{-elim}}{\neg(\forall x)\phi(x)}\ \to\text{-int }(1) \quad [\neg\neg(\forall x)\phi(x)]^{(3)}}{\dfrac{\bot}{\neg\neg\phi(a)}\ \to\text{-int }(2)}\ \to\text{-elim}}{(\forall x)\neg\neg\phi(x)}\ \forall\text{-int}}{\neg\neg(\forall x)\phi(x) \ \to\ (\forall x)\neg\neg\phi(x))}\ \to\text{-int }(3) \tag{5.5}$$

So $\neg\neg\forall x\phi$ implies $\forall x\neg\neg\phi$. But $\neg\neg\phi \to \phi$ by induction hypothesis, whence $\forall x\phi$.

■

So in particular everything in the range of the negative interpretation is stable. Also, ϕ and ϕ^* are classically equivalent. So the negative interpretation will send every formula in the language to a stable formula classically equivalent to it.

Lemma 37 *If ϕ is classically valid then ϕ^* is constructively correct.*

Proof: We do this by showing how to recursively transform a classical proof of ϕ into a constructive proof of ϕ^*.

There is no problem with the three connectives $\neg$, $\wedge$ or $\forall$ of course. We deal with the others as follows.

$\vee$-introduction

$$\dfrac{\dfrac{\dfrac{[\neg p^* \wedge \neg q^*]^1}{\neg p^*}\ \wedge\text{-elim} \quad p^*}{\bot}\ \to\text{-elim}}{\neg(\neg p^* \wedge \neg q^*)}\ \to\text{-int }(1) \qquad \dfrac{\dfrac{\dfrac{[\neg p^* \wedge \neg q^*]^1}{\neg q^*}\ \wedge\text{-elim} \quad q^*}{\bot}\ \to\text{-elim}}{\neg(\neg p^* \wedge \neg q^*)}\ \to\text{-int }(1) \tag{5.6}$$

are derivations of $(p \vee q)^*$ from p^* and from q^* respectively.

$\vee$-elimination

We will have to show that whenever there is (i) a deduction of r^* from p^* and (ii) a deduction of r^* from q^*, and (iii) we are allowed $(p \vee q)^*$ as a premiss, then there is a constructive derivation of r^*.

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{array}{c}[p^*]^1\\ \vdots \\ r^*\end{array} \qquad [\neg r^*]^3}{\bot}\,\to\text{-elim}}{\neg p^*}\,\to\text{-int (1)} \qquad \dfrac{\dfrac{\begin{array}{c}[q^*]^2\\ \vdots \\ r^*\end{array} \qquad [\neg r^*]^3}{\bot}\,\to\text{-elim}}{\neg q^*}\,\to\text{-int (2)}}{\neg p^* \wedge \neg q^*}\,\wedge\text{-int} \qquad \neg(\neg p^* \wedge \neg q^*)}{\dfrac{\bot}{\neg\neg r^*}\,\to\text{-int (3)}}\,\to\text{-elim} \tag{5.7}$$

... and we infer r^* because r^* is stable.

→-introduction

Given a constructive derivation $\dfrac{\underline{p^*}}{\begin{array}{c}\vdots\\ q^*\end{array}}$ we can build the following

$$\dfrac{\dfrac{\dfrac{[p^* \wedge \neg q^*]^1}{p^*}\,\wedge\text{-elim}}{\begin{array}{c}\vdots\\ q^*\end{array}} \qquad \dfrac{[p^* \wedge \neg q^*]^1}{\neg q^*}\,\wedge\text{-elim}}{\dfrac{\bot}{\neg(p^* \wedge \neg q^*)}\,\to\text{-int (1)}}\,\to\text{-elim} \tag{5.8}$$

which is of course a proof of $(p \to q)^*$.

→-elimination

The following is a deduction of q^* from $(p \to q)^*$ and p^*:

$$\dfrac{\dfrac{p^* \qquad [\neg q^*]^{1)}}{p^* \wedge \neg q^*}\,\wedge\text{-int} \qquad \neg(p^* \wedge \neg q^*)}{\dfrac{\bot}{\neg\neg q^*}\,\to\text{-int (2)}}\,\to\text{-elim} \tag{5.9}$$

... q^* is stable so we can infer q^*.

∃-introduction

Constructively $\exists$ implies $\neg\forall\neg$ so this is immediate.

∃-elimination

We use this where we have a classical derivation

$$\frac{\dfrac{\phi(x)}{\vdots}}{p}$$

and have been given $\exists y\phi(y)$.
By induction hypothesis this means we have a constructive derivation

$$\frac{\dfrac{\phi^*(x)}{\vdots}}{p^*}.$$

Instead of $\exists y\phi(y)$ we have $\neg(\forall y)\neg\phi^*(y)$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\phi^*(a)]^2}{\vdots} \quad\quad [\neg p^*]^1}{\perp}\ \text{→-elim}}{\neg\phi^*(a)}\ \text{→-int (2)}}{(\forall y)\neg\phi^*(y)}\ \forall\text{-int} \quad\quad \neg(\forall y)\neg\phi^*(y)}{\dfrac{\perp}{\neg\neg p^{*(1)}}\ \text{→-int (1)}}\ \text{→-elim} \tag{5.10}$$

and p^* follows from $\neg\neg p^*$ because p^* is stable.

We want double negation not classical negation here. Sort this out

The Classical Rules

In a classical proof we will be allowed various extra tricks, such as being able to assume $p \vee \neg p$ whenever we like. So we are allowed to assume $(p \vee \neg p)^*$ whenever we like. But this is $\neg(\neg p^* \wedge \neg\neg p^*)$ which is of course a constructive theorem.

The starred version of the rule of double negation tells us we can infer p^* from $\neg\neg p^*$. By lemma 36 every formula built up from $\forall$, $\wedge$ and $\neg$ is stable. But, for any formula p whatever, p^* is such a formula.

There are other rules we could add—instead of excluded middle or double negation—to constructive logic to get classical logic, and similar arguments will work for them.

Substitutivity of Equality

To ensure that substitutivity of equality holds under the stars we want to prove

$$(\forall xy)(\neg\neg\phi(x) \to \neg\neg(x = y) \to \neg\neg\phi(y))$$

This we accomplish as follows:

$$\dfrac{\dfrac{\dfrac{\dfrac{[\neg\phi(y)]^1 \qquad [x=y]^2}{\neg\phi(x)}\text{ subst} \qquad \neg\neg\phi(x)}{\bot}\to\text{-elim}}{\neg(x=y)}\to\text{-int (2)} \qquad \neg\neg(x=y)}{\dfrac{\bot}{\neg\neg\phi(y)}\to\text{-int (1)}}\to\text{-elim} \tag{5.11}$$

which is a proof of $\neg\neg\phi(y)$ from $\neg\neg\phi(x)$ and $\neg\neg(x = y)$.
This completes the proof of lemma 37 ■

5.3 Prophecy

What does this * interpretation tell the constructive logician? Let us consider a simple case where $\phi(x)$ and $\phi(x)^*$ are the same, and the classical logician has a proof of $(\exists x)(\phi(x))$. Then the constructive logician acknowledges that there is a proof of $\neg(\forall x)(\neg\phi(x))$. What is (s)he to make of this? There isn't officially a proof of $(\exists x)(\phi(x))$, but they can at least conclude that there can never be a proof of $\neg(\exists x)(\phi(x))$. This makes a good exercise!

EXERCISE 62 *Using the natural deduction rules derive a contradiction from the two assumptions* $\neg(\forall x)(\neg\phi(x))$ *and* $\neg(\exists x)(\phi(x))$.

If there can never be a proof of $\neg(\exists x)(\phi(x))$ then the assumption that there is an x which is ϕ cannot lead to contradiction. In contrast the assumption that there isn't one *will* lead to contradiction. So would your money be on the proposition that you will find an x such that ϕ or on the proposition that you won't? It's a no-brainer. This is why people say that, to the constructive logician, nonconstructive existence theorems have something of the character of prophecy.

Chapter 6

Possible World Semantics

Split the quantifiers off. Need two sections

This should really be called "Multiple Model Semantics" but the current terminology is entrenched.

How is the classical logician supposed to react when the constructive logician does something obviously absurd like deny the law of excluded middle? (S)he will react in the way we all react when confronted with apparently sensible people saying obviously absurd things: we conclude that *they must mean something else*.

Possible world semantics is a way of providing the classical logician with something sensible that the constructive logician might mean when they come out with absurdities like excluded-middle-denial. It's pretty clear that constructive logicians don't actually mean the things that classical logicians construe them as meaning in their (the classicists') attempt to make sense of their (the constructivists') denial of excluded middle. But that doesn't mean that the exercise is useless. It's such a good story that it doesn't matter *where* it comes from.

DEFINITION 38 *A* **possible world model** $\mathfrak{M}$ *has several components:*

- *There is a collection of* **worlds** *with a binary relation* $\leq$ *between them; If* $W_1 \leq W_2$ *we say* W_1 *can* **see** W_2.
- *There is also a binary relation between worlds and formulæ, written '*$W \models \phi$*';*
- *Finally there is a* **designated** *(or 'actual' or 'root') world* W_0^M.

We stipulate the following connections between the ingredients:

1. $W \models \bot$ *never holds. We write this as* $W \not\models \bot$.
2. $W \models A \wedge B$ *iff* $W \models A$ *and* $W \models B$*;*
3. $W \models A \vee B$ *iff* $W \models A$ *or* $W \models B$*;*
4. $W \models A \to B$ *iff every* $W' \geq W$ *that* $\models A$ *also* $\models B$*;*
5. $W \models \neg A$ *iff there is no* $W' \geq W$ *such that* $W' \models A$*;*

At some point we have to explain how to use this st… $\Diamond$ and $\Box$

(We will deal with the quantifiers later)

We stipulate further that **for atomic formulæ** *ϕ, if $W \models \phi$ and $W \leq W'$, then $W' \models \phi$. (The idea is that if $W \leq W'$, then W' in some sense contains more information than W.) We will call this phenomenon* **persistence**.

Then we say

$$\mathfrak{M} \models A \ \ \text{if} \ \ W_0^M \models A$$

5 is a special case of 4: $\neg A$ is just $A \to \bot$, and no world believes $\bot$!

The relation which we here write with a '$\leq$' is the **accessibility** relation between worlds. We assume for the moment that it is **transitive** and **reflexive**. Just for the record we note that '$A \leq B$' will sometimes be written as '$B \geq A$'.

[Chat about quantifier alternation. There is a case for writing out the definitions in a formal language, on the grounds that the quantifier alternation (which bothers a lot of people) can be made clearer by use of a formal language. The advantage of not using a formal language is that it makes the language-metalanguage distinction clearer.]

If one takes these worlds too seriously then one will find that the $\models$ relation between worlds and propositions is epistemically problematic. For example W believes $\neg p$ iff no world beyond W believes p. This being so, how can anyone in W come to know $\neg p$? They would have to visit all worlds $\geq W$! So this possible worlds talk is not part of an *epistemic* story! This being the case, one should perhaps beware of the danger of taking the "world W believes ϕ" slang too literally. Even if W believes $\neg\phi$ then in some sense it doesn't know that it believes $\neg\phi$... unless of course W includes among its inhabitants all the worlds $\geq W$. But that makes for a scenario far too complicated for us to entertain in a book like this. And it is arguable that it is a scenario of which no coherent account can be given. See [17].

The possible worlds semantics is almost certainly not part of a constructivist account of truth or meaning at all. (Remember: we encountered it as the classical logicians' way of making sense of constructive logic!) If it were, the fact that it is epistemically problematic would start to matter.

The relation $\leq$ between worlds is transitive. A model $\mathfrak{M}$ believes ϕ (or not, as the case may be) iff the designated world W_0 of $\mathfrak{M}$ believes ϕ (or not). When cooking up W_0 to believe ϕ (or not) the recursions require us only to look at worlds $\geq W_0$. This has the effect that the designated world of $\mathfrak{M}$ is $\leq$ all other worlds in $\mathfrak{M}$. This is why we sometimes call it the 'root' world. This use of the word 'root' suggests that the worlds beyond W_0 are organised into a tree: so if W_1 and W_2 are two worlds that cannot see each other then there is no world they can both see. However we are emphatically *not* making this assumption.

[bad join: this para uses 'R' for the accessibility relation, so it should probably be put in after we have generalised, and also introduced $\Diamond$ and $\Box$.]

It's probably a good idea to think a bit about how this gadgetry is a generalisation of the semantics for propositional logic that we saw in section 3.13, or—to put it the other way round—how the semantics there is a degenerate case of what we have here.

Worlds here correspond to *valuations* (or rows of a truth-table) there. In section 3.13 each valuation went on its merry way without reference to any other valuation: if you wanted to know whether a valuation v made a formula ϕ true you had to look at subformulæ of ϕ but you didn't have to look at what any other valuation did to ϕ or to any of its subformulæ. If you think about this a bit you will realise that if you have a possible world model where the the accessibility relation R is the identity relation [so that the only valuation that a valuation v ever has to consult is v itself] then the semantics you get will be the same as in section 3.13. Another thing that happens if R is the identity is that $\Diamond p$, p and $\Box p$ turn out to be the same. duplicates section 6.2.2

6.0.1 Quantifiers

Definition 38 didn't have rules for the quantifiers. We'd better have them now.

Definition 38 continued

6 : $W \models (\exists x)A(x)$ iff there is an x in W such that $W \models A(x)$;

7 : $W \models (\forall x)A(x)$ iff for all $W' \geq W$ and all x in W', $W' \models A(x)$.

In the first instance the only thing the worlds have to do is believe (or not believe) atomic propositions: the rules in definition 38 for the connectives don't compel us to think of the worlds as having inhabitants. In contrast the rules for the quantifiers *do* assume that worlds have inhabitants. In the propositional case we stipulate which atomic propositions each world believes, and the rest of the semantics is done by the recursion. When we add quantifiers we stipulate which atomic formulaæ a world believes of which of its inhabitants.

I think we generally take it that our worlds are never empty: every world has at least one inhabitant. However there is no global assumption that all worlds have the same inhabitants. Objects may pop in and out of existence as we turn our gaze from one world to another. However we do take the identity relation between inhabitants across possible worlds as a given.

6.1 Language and Metalanguage again

It is very important to distinguish between the stuff that appears to the left of a '$\models$' sign and that which appears to the right of it. The stuff to the right of the '$\models$' sign belongs to the *object language* and the stuff to the left of the '$\models$' sign belongs to the *metalanguage*. So that we do not lose track of where we are I am going to write '$\rightarrow$' for *if–then* in the metalanguage and '&' for *and* in the metalanguage instead of '$\wedge$'. And I shall use square brackets instead of round brackets in the metalanguage.

If you do not keep this distinction clear in your mind you will end up making one of the two mistakes below (tho' you are unlikely to make both.)

Remember what the aim of the Possible World exercise was. It was to give people who believe in classical logic a way of making sense of the thinking of people who believe in constructive logic. That means that it's perfectly OK to use classical logic in reasoning with/manipulating stuff to the left of a '$\models$' sign.

For example here is a manœuvre that is perfectly legitimate:

If

$$\neg[W \models A \to B]$$

then it is not the case that

$$(\forall W' \geq W)(W' \models A \;\to\; W' \models B)$$

So, in particular,

$$(\exists W' \geq W)(W' \models A \;\&\; \neg(W' \models B))$$

The inference drawn here from $\neg\forall$ to $\exists\neg$ is perfectly all right in the classical metalanguage, even though it's not allowed in the constructive object language.

In contrast it is *not* all right to think that—for example—$W \models \neg A \vee \neg B$ is the same as $W \models \neg(A \wedge B)$ (on the grounds that $\neg A \vee \neg B$ is the same as $\neg(A \wedge B)$). One way of warding off the temptation to do it is to remind ourselves—again—that the aim of the Possible World exercise was to give people who believe in classical logic a way of making sense of the mental life of people who believe in constructive logic. That means that it is not OK to use classical logic in reasoning with/manipulating stuff to the right of a '$\models$' sign.

Another way of warding off the same temptation is to think of the stuff after the '$\models$' sign as stuff that goes on in a fiction. You, the reader of a fiction, know things about the characters in the fiction that they do not know about each other. Just because something is true doesn't mean they know it!! (This is what the literary people call **Dramatic Irony**.)[1]

(This reflection brings with it the thought that reading "$W \models \neg\neg A$" as "W believes not not A" is perhaps not the happiest piece of slang. After all, in circumstances where $W \models \neg\neg A$ there is no suggestion that the fact-that-no-world-$\geq$-W-believes-A is encoded in W in any way at all.)

Could say more about this

Another mistake is to think that we are obliged to use constructive logic in the metalanguage which we are using to discuss constructive logic—to the left of the '$\models$' sign. It is probably the same mistake made by people who think that hypothetical reasoning—and specifically *reductio ad absurdum* is incoherent. If you are a philosophy student you might find this an interesting topic in its own right. I suspect it's a widespread error. It may be the same mistake as the mistake of supposing that you have to convert to Christianity to understand what is going on in the heads of Christians. Christians of some stripes would no doubt agree with the assertion that there are bits of it you can't understand until you convert, but I think that is just a mind-game.

Doesn't this duplicate earlier stuff? section **??**?

[1]Appreciation of the difference between something being true and your interlocutor knowing it is something that autists can have trouble with. Some animals that have "a theory of other minds" (in that they know that their conspecifics might know something) too can have difficulty with this distinction. Humans seem to be able to cope with it from the age of about three.

We could make it easier for the nervous to discern the difference between the places where it's all right to use classical reasoning (the metalanguage) and the object language (where it isn't) by using different fonts or different alphabets. One could write "For all W" instead of $(\forall W)\ldots$". That would certainly be a useful way of making the point, but once the point has been made, persisting with it looks a bit obsessional: in general people seem to prefer overloading to disambiguation.

6.1.1 A Possibly Helpful Illustration

Let us illustrate with the following variants on the theme of "there is a Magic Sword." All these variants are classically equivalent. The subtle distinctions that the possible worlds semantics enable us to make are very pleasing.

$\neg\forall x \neg MS(x)$

$\neg\neg\exists x MS(x)$

$\exists x \neg\neg MS(x)$

$\exists x MS(x)$

The first two are constructively equivalent as well.

To explain the differences we need the difference between **histories** and **futures.**

- A *future* (from the point of view of a world W) is any world $W' \geq W$.
- A *history* is a string of worlds—an unbounded trajectory through the available futures.

No gaps between worlds...?

$\neg\forall x \neg MS(x)$ and $\neg\neg\exists x MS(x)$ say that every future can see a future in which there is a Magic Sword, even though there might be histories that avoid Magic Swords altogether: *Magic Swords are a permanent possibility: you should never give up hope of finding one.*

How can this be, that every future can see a future in which there is a magic sword but there is a history that contains no magic sword–ever? It could happen like this: each world has precisely two immediate children. If it is a world with a magic sword then those two worlds also have magic swords in them. If it is a world without a magic sword then one of its two children continues swordless, and the other one acquires a sword. We stipulate that the root world contains no magic sword. That way every world can see a world that has a magic sword, and yet there is a history that has no magic swords.

$\exists x \neg\neg MS(x)$ says that every history contains a Magic Sword and moreover the thing which is destined to be a Magic Sword is already here. Perhaps it's still a lump of silver at the moment but it will be a Magic Sword one day.

6.2 Some Useful Short Cuts

6.2.1 Double negation

The first one that comes to mind is $W \models \neg\neg\phi$. This is the same as $(\forall W' \geq W)(\exists W'' \geq W')(W'' \models \phi)$. "Every world that W can see can see a world that believes ϕ". Let's thrash this out by hand.

By clause 5 of definition 38

$$W \models \neg(\neg\phi)$$

iff

$$(\forall W' \geq W)\neg[W' \models \neg\phi] \tag{6.1}$$

Now

$W' \models \neg\phi$ iff $(\forall W'' \geq W')\neg[W'' \models \phi]$ by clause 5 of definition 38 so

$\neg[W' \models \neg\phi]$ is the same as $\neg(\forall W'' \geq W')\neg[W'' \models \phi]$ which is

$$(\exists W'' \geq W')(W'' \models \phi).$$

Substituting this last formula for for '$W' \models \neg\phi$' in (6.1) we obtain

$$(\forall W' \geq W)(\exists W'' \geq W')(W'' \models \phi)$$

6.2.2 If there is only one world then the logic is classical

If $\mathfrak{M}$ contains only one world—W, say—then $\mathfrak{M}$ believes classical logic. Let me illustrate this in two ways:

1. Suppose $\mathfrak{M} \models \neg\neg A$. Then $W \models \neg\neg A$, since W is the root world of $\mathfrak{M}$. If $W \models \neg\neg A$, then for every world $W' \geq W$ there is $W'' \geq W$ that believes A. So in particular there is a world $\geq W$ that believes A. But the only world $\geq W$ is W itself. So $W \models A$. So every world $\geq W$ that believes $\neg\neg A$ also believes A. So $W \models \neg\neg A \to A$.

2. W either believes A or it doesn't. If it believes A then it certainly believes $A \vee \neg A$, so suppose W does not believe A. Then W can see no world that believes A. So $W \models \neg A$ and thus $W \models (A \vee \neg A)$. So W believes the law of excluded middle.

   It looks as if we have used the law of excluded middle to prove the law of excluded middle ("W either believes A or it doesn't"). But we have used in it the *metalanguage*—and the metalanguage is classical.

In these circumstances the logic of quantifiers is classical too.
Suppose $W \models \neg\forall x \neg F(x)$. This is

$$(\forall W' \geq W)(\neg[W' \models \forall x \neg F(x)])$$

But, since there is only one $W' \geq W$, and that W' is W itself, this becomes

$$\neg[W \models \forall x \neg F(x)]$$

which becomes

$$\neg[(\forall W' \geq W)(\forall x \in W')(W' \models \neg F(x))]$$

Now, again, since there is only one $W' \geq W$, and that W' is W itself, this becomes

$$\neg[(\forall x \in W)(W \models \neg F(x))]$$

which becomes

$$(\exists x \in W)(\neg[W \models \neg F(x)])$$

and

$$(\exists x \in W)(\exists W' \geq W)\neg[W' \models \neg F(x)])$$

giving

$$(\exists x \in W)\neg[W \models \neg F(x)]$$

If $\neg[W \models \neg F(x)]$ it must be that W can see a world that believes $F(x)$. But the only world it can see is itself, so we infer

$$(\exists x \in W)(W \models F(x))$$

which is to say

$$W \models (\exists x)(F(x))$$

as desired. ■

One can perform a similar calculation to reduce $\neg\exists\neg$ to $\forall$, but one can obtain this second equivalence as a corollary of the first.

$\neg\exists\neg$ is equivalent to $\neg\neg\forall\neg\neg$ by the forgoing, but now we can cancel the two '$\neg\neg$'s to obtain $\forall$.

The same arguments can be used even in models with more than one world, as long as the worlds in question can see only themselves.

6.3 Persistence

[

stuff to fit in

Persistence enables us to connect possible world semantics with many-valued logic. Each truth-value corresponds to an upper set in the quasiorder, in the sense that $[[\phi]] = \{W : W \models \phi\}$. Upper sets in quasiorders form a Heyting Algebra, so that the truth-values $[[A]]$ are members of a Heyting algebra, and $[A]] = \top$ iff $\mathfrak{M} \models A$.

For the other direction any Heyting valuation can be turned into a possible-world model by appealing to the representation theorem for distributive posets: every Heyting algebra is isomorphic to a subset of a power set algebra. The truth-value $[[A]]$ then corresponds to a set $\mathcal{A}$, which we can take to be a set of worlds. We then rule that $W \models A$ iff $W \in \mathcal{A}$.

Supply missing details; calculate $[[A \to B]$ and $[[A \vee B]]$ and $[[A \wedge B]]$.

(1) $[[A \to B]] = \{W : W \models A \to B\}$
(2) $\{W : (\forall W' \geq W)((W' \models A) \to (W' \models B)\}$
(3) $\{W : (W \models A) \to (W \models B)\}$
(4) $\{W : (W \not\models A) \vee (W \models B)\}$
(5) $\{W : W \not\models A\} \cup \{W : W \models B)\}$
(6) $\{W : W \not\models A\} \cup [[B]]$
(7) $\{W : W \models A\} \to [[B]]$
(8) $[[A]] \to [[B]]$

We need persistence to ensure $\models p \to \neg\neg p$!

For atomic formulæ ϕ we know that if $W \models \phi$ then $W' \models \phi$ for all $W' \geq W$. We achieved this by stipulation, and it echoes our original motivation. Even though $\neg\neg(\exists x)(x$ is a Magic Sword) is emphatically not to be the same as $(\exists x)(x$ is a Magic Sword), it certainly is inconsistent with $\neg(\exists x)(x$ is a Magic Sword) and so it can be taken as *prophecy* that a Magic Sword will turn up one day. The idea of worlds as states of knowledge where we learn more as time elapses sits very well with this. By interpreting $\neg\neg(\exists x)(x$ is a Magic Sword) as "Every future can see a future that contains a Magic Sword" possible world semantics captures the a way in which $\neg\neg(\exists x)(x$ is a Magic Sword) can be incompatible with the nonexistence of Magic Swords while nevertheless not telling us how to find a Magic Sword.

We will say ϕ is **persistent** if whenever $W \models \phi$ then $(\forall W' \geq W)(W' \models \phi)$

We want to prove that all formulæ are persistent.

THEOREM 39 *All formulæ are persistent.*

Proof:

We have taken care of the atomic case. Now for the induction on quantifiers and connectives.

$\neg$ $W \models \neg\phi$ iff $(\forall W' \geq W)\neg(W' \models \phi)$. Therefore if $W \models \neg\phi$ then $(\forall W' \geq \phi)\neg[W' \models \phi]$, and, by transitivity of $\geq$, $(\forall W'' \geq W')\neg[W'' \models \phi]$. But then $\neg[W' \models \neg\phi]$.

$\bigvee$ Suppose ϕ and ψ are both persistent. If $W \models \psi \vee \phi$ then either $W \models \phi$ or $W \models \psi$. By persistence of ϕ and ψ, every world $\geq$ satisfies ϕ (or ψ, whichever it was) and will therefore satisfy $\psi \vee \phi$.

$\bigwedge$ Suppose ϕ and ψ are both persistent. If $W \models \psi \wedge \phi$ then $W \models \phi$ and $W \models \psi$. By persistence of ϕ and ψ, every world $\geq$ satisfies ϕ and every world $\geq$ satisfies ψ and will therefore satisfy $\psi \wedge \phi$.

$\exists$ Suppose $W \models (\exists x)\phi(x)$, and ϕ is persistent. Then there is an x in W which W believes to be ϕ. Suppose $W' \geq W$. As long as x is in W' then $W' \models \phi(x)$ by persistence of ϕ and so $W' \models (\exists x)(\phi(x))$.

$\forall$ Suppose $W \models (\forall x)\phi(x)$, and ϕ is persistent. That is to say, for all $W' \geq W$ and all x, $W' \models \phi(x)$. But if this holds for all $W' \geq W$, then it certainly holds for all $W' \geq$ any given $W'' \geq W$. So $W'' \models (\forall x)(\phi(x))$.

$\to$ Finally suppose $W \models (A \to B)$, and $W' \geq W$. We want $W' \models (A \to B)$. That is to say we want every world beyond W' that believes A to also believe B. We do know that every world beyond W that believes A also believes B, and every world beyond W' is a world beyond W, and therefore believes B if it believes A. So W' believes $A \to B$.

That takes care of all the cases in the induction. ■

It's worth noting that we have made heavy use of the assumption that $\leq$ is transitive. Later we will consider other more general settings where this assumption is not made. In those more general settings we will use symbols other than '$\leq$' to denote the accessibility relation (since the use of that symbol inevitably connotes transitivity) and we will drop the assumption of persistence for atomic formulæ. However such generality is beyond the scope of this book.

Now we can use persistence to show that this possible world semantics always makes $A \to \neg\neg A$ comes out true. Suppose $W \models$. Then every world $\geq W$ also believes A. No world can believe A and $\neg A$ at the same time. ($W \models \neg A$ only if none of the worlds $\geq W$ believe A; one of the worlds $\geq W$ is W itself.) So none of them believe $\neg A$; so $W \models \neg\neg A$.

This is a small step in the direction of a completeness theorem for the possible world semantics.

THEOREM 40 *Every propositional formula with a natural deduction proof using only the constructive rules is true in all possible world models.*

Proof:

In the propositional case at least we prove by induction on proofs ...

Let $\mathfrak{M}$ be a possible world model. We prove by induction on proofs $\mathcal{D}$ in $\mathcal{L}(\mathfrak{M})$ that, for all $W \in \mathfrak{M}$, if $W \models$ every premiss in $\mathcal{D}$, then $W \models$ the conclusion of $\mathcal{D}$.

It's pretty straightforward. Consider the case of a proof of $A \to B$. The last step is a $\to$-introduction, The induction hypothesis will be that every world that believes A (and the other premisses) also believes B. Now let W be a world that believes all the other premisses. Then certainly (by persistence) every $W' \geq W$ also believes all the other premisses, so any such W' that believes A also believes B. But that is to say that any world that believes all the other premisses believes $A \to B$. ■

I don't know how to prove the other direction!

A later version of this material is to be found in partiiilogic2017

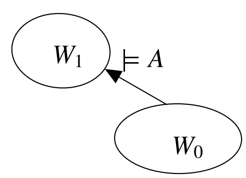
6.4 Independence Proofs Using Possible world semantics

6.4.1 Some Worked Examples

Challenge 6.4.1.1: Find a countermodel for $A \vee \neg A$

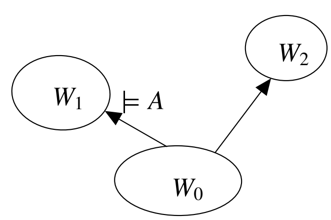
The first thing to notice is that this formula is a classical (truth-table) tautology. Because of subsection 6.2.2 this means that any countermodel for it must contain more than one world.

The root world W_0 must not believe A and it must not believe $\neg A$. If it cannot see a world that believes A then it will believe $\neg A$, so we will have to arrange for it to see a world that believes A. One will do, so let there be W_1 such that $W_1 \models A$.



Challenge 6.4.1.2: Find a countermodel for $\neg\neg A \vee \neg A$

The root world W_0 must not believe $\neg\neg A$ and it must not believe $\neg A$. If it cannot see a world that believes A then it will believe $\neg A$, so we will have to arrange for it to see a world that believes A. One will do, so let there be W_1 such that $W_1 \models A$. It must also not believe $\neg\neg A$. It will believe $\neg\neg A$ as long as every world it can see can see a world that believes A. So there had better be a world it can see that cannot see any world that believes A. This cannot be W_1 because $W_1 \models A$, and it cannot be W_0 itself, since $W_0 \leq W_1$. So there must be a third world W_2 which does not believe A.



Challenge 6.4.1.3: Find a model that satisfies $(A \to B) \to B$ **but does not satisfy** $A \vee B$

The root world W_0 must not believe $A \vee B$, so it must believe neither A nor B. However it has to believe $(A \to B) \to B$, so every world that it can see that believes $A \to B$ must also believe B. One of the worlds it can see is itself, and it doesn't believe B, so it had better not believe $A \to B$. That means it has to see a world that believes A but does not believe B. That must be a different world (call it W_1). So we can recycle the model from Challenge 6.4.1.2.

Challenge 6.4.1.4: Find a countermodel for $((A \to B) \to A) \to A$

You may recall from exercise 32 on page 68 that on Planet Zarg this formula is believed to be false[2]. There we had a three-valued truth table. Here we are going to use possible worlds. As before, with $A \vee \neg A$, the formula is a truth-table tautology and so we will need more than one world.

Recall that a model $\mathfrak{M}$ satisfies a formula ψ iff the root world of $\mathfrak{M}$ believes ψ: that is what it is for a model to satisfy ψ. Definition!

As usual I shall write 'W_0' for the root world; and will also write '$W \models \psi$' to mean that the world W believes ψ; and $\neg[W \models \psi]$ to mean that W does not believe ψ.

So we know that $\neg[W_0 \models ((A \to B) \to A) \to A]$.

Now the definition of $W \models X \to Y$ is (by definition 38)

$$(\forall W' \geq W)[W' \models X \ \to \ W' \models Y] \tag{6.1}$$

So since

$$\neg[W_0 \models ((A \to B) \to A) \to A]$$

we know that there must be a $W' \geq W_0$ which believes $(A \to B) \to A$ but does not believe A. (In symbols: $(\exists W' \geq W_0)[W' \models ((A \to B) \to A) \,\&\, \neg(W' \models A)]$.) Remember too that in the metalanguage we are allowed to exploit the equivalence of $\neg\forall$ with $\exists\neg$. Now every world can see itself, so might this W' happen to be W_0 itself? No harm in trying...

So, on the assumption that this W' that we need is W_0 itself, we have:

1. $W_0 \models (A \to B) \to A$; and
2. $\neg[W_0 \models A]$.

This is quite informative. Fact (1) tells us that every $W' \geq W_0$ that believes $A \to B$ also believes A. Now one of those W' is W_0 itself (Every world can see itself: remember that $\geq$ is reflexive). Put this together with fact (2) which says that W_0 does not believe A, and we know at once that W_0 cannot believe $A \to B$. How can we arrange for W_0 not to believe $A \to B$? Recall the definition 38 above of $W \models A \to B$. We have to ensure that there is a $W' \geq W_0$ that believes A but does not believe B. This W' cannot be W_0 because W_0 does not believe A. So there must be a *new* world (we always knew there would be!) visible from W_0 that believes A but does not believe B. (In symbols this is $(\exists W' \geq W_0)[W' \models A \quad \& \quad \neg(W' \models B)]$.)

So our countermodel contains two worlds W_0 and W', with $W_0 \leq W'$. $W' \models A$ but $\neg[W_0 \models A]$, and $\neg[W' \models B]$.

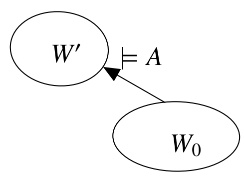
[2] I have just corrected this from "You may recall from exercise 32 on page 68 that this formula is believed to be false on Planet Zarg"—which is not the same!

Let's check that this really works. We want

$$\neg[W_0 \models ((A \to B) \to A) \to A]$$

We have to ensure that at least one of the worlds beyond W_0 satisfies $(A \to B) \to A$ but does not satisfy A. W_0 doesn't satisfy A so it will suffice to check that it does satisfy $(A \to B) \to A$.

So we have to check (i) that if W_0 satisfies $(A \to B)$ then it also satisfies A and we have to check (ii) that if W' satisfies $(A \to B)$ then it also satisfies A. W' satisfies A so (ii) is taken care of. For (i) we have to check that W_0 does not satisfy $A \to B$. For this we need a world $\geq W_0$ that believes A but does not believe B and W' is such a world. This is actually the same model as we used in Challenge 6.4.1.1.



Challenge 6.4.1.5: Find a model that satisfies $(A \to B) \to B$ but does not satisfy $(B \to A) \to A$

We must have

$$W_0 \models (A \to B) \to B \qquad (1)$$

and

$$\neg[W_0 \models (B \to A) \to A] \qquad (2)$$

By (2) we must have $W_1 \geq W_0$ such that

$$W_1 \models B \to A \qquad (3)$$

but

$$\neg[W_1 \models A] \qquad (4)$$

We can now show

$$\neg[W_1 \models A \to B] \qquad (5)$$

If (5) were false then $W_1 \models B$ would follow from (1) and then $W_1 \models A$ would follow from (3). (5) now tells us that there is $W_2 \geq W_1$ such that

$$W_2 \models A \qquad (6)$$

and

$$\neg[W_2 \models B] \tag{7}$$

From (7) and persistence we infer

$$\neg[W_1 \models B] \tag{8}$$

and

$$\neg[W_0 \models B] \tag{9}$$

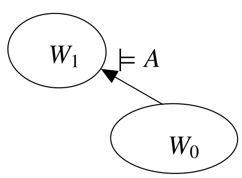
Also, (4) tells us

$$\neg[W_0 \models A]. \tag{10}$$

So far we have nothing to tell us that $W_0 \neq W_1$. So perhaps we can get away with having only two worlds W_0 and W_1 with $W_1 \models A$ and W_0 believing nothing.

W_0 believes $(A \to B) \to B$ vacuously: it cannot see a world that believes $A \to B$ so—vacuously—every world that it can see that believes $A \to B$ also believes B. However, every world that it can see believes $(B \to A)$ but it does not believe A itself. That is to say, it can see a world that does not believe A so it can see a world that believes $B \to A$ but does not believe A so it does not believe $(B \to A) \to A$.

Thus we have the by-now familiar picture:



6.4.2 Exercises

EXERCISE 63 *Return to Planet Zarg!*[3]

The truth-tables for Zarg-style connectives are on p 68.

1. *Write out a truth-table for* $((p \to q) \to q) \to (p \vee q)$.

   (Before you start, ask yourself how many rows this truth-table will have).

2. *Identify a row in which the formula does not take truth-value* 1.

3. *Find a sequent proof for* $((p \to q) \to q) \to (p \vee q)$.

EXERCISE 64

Find a model that satisfies $(p \to q) \to q$ *but does not satisfy* $p \vee q$.

It turns out that Zarg-truth-value 1 means "true in W_0 and in W_1"; Zarg-truth-value 2 means "true in W_1", and Zarg-truth-value 3 means "true in neither"—where W_0 and W_1 are the two worlds in the countermodel we found for Peirce's law. (Challenge 6.4.1.5)

Where does this *aperçu* long?

EXERCISE 65 *Find a model that satisfies* $p \to q$ *but not* $\neg p \vee q$.

EXERCISE 66 *Find a model that doesn't satisfy* $p \vee \neg p$. *How many worlds has it got? Does it satisfy* $\neg p \vee \neg\neg p$? *If it does, find one that doesn't satisfy* $\neg p \vee \neg\neg p$.

EXERCISE 67

1. *Find a model that satisfies* $A \to (B \vee C)$ *but doesn't satisfy* $(A \to B) \vee (A \to C))$[4].
2. *Find a model that satisfies* $(A \to B) \wedge (C \to D)$ *but doesn't satisfy* $(A \to D) \vee (C \to B))$[5].
3. *Find a model that satisfies* $\neg(A \wedge B)$ *but does not satisfy* $\neg A \vee \neg B$
4. *Find a model that satisfies* $(A \to B) \to B)$ *and* $(B \to A) \to A$ *but does not satisfy* $A \vee B$.

   Check that in the three-valued Zarg world $((A \to B) \to B) \wedge ((B \to A) \to A)$ *always has the same truth-table as* $A \vee B$.

EXERCISE 68 *Find countermodels for:*

1. $(A \to B) \vee (B \to A)$;
2. $(\exists x)(\forall y)(F(y) \to F(x))$ *(which is the formula in exercise 58 part 1 on page 135).*

EXERCISE 69 *Consider the model in which there are two worlds,* W_0 *and* W_1, *with* $W_0 \leq W_1$. W_0 *contains various things, all of which it believes to be frogs;* W_1 *contains everything in* W_0 *plus various additional things, none of which it believes to be frogs. Which of the following assertions does this model believe?*

1. $(\forall x)(F(x))$;
2. $(\exists x)(\neg F(x))$;
3. $\neg\exists x \neg F(x)$;
4. $\neg\neg(\exists x)(\neg F(x))$.

[3] Beware: Zarg is a planet not a possible world!

[4] You saw a fallacious attempt to prove this inference on page 64.

[5] This is a celebrated illustration of how $\to$ does not capture 'if-then'. Match the antecedent to "If Jones is in Aberdeen then Jones is in Scotland and if Jones is in Delhi then Jones is in India".

Find countermodels for
$(\exists x)(\forall y)(F(y) \to F(x))$
$\neg(\forall x)\neg(\forall y)(F(y) \to F(x))$
$(\exists x)(\forall y)\neg(F(y) \land \neg F(x))$

At some point demonstrate that K does not hold for the strict implication of S5, for example. Failure of deduction theorem.

6.5 Modal Logic

Do we really want a section of modal logic???

Some modal logics do not obey 3.6.1, the deduction theorem. A logic obeys the deduction theorem iff it proves both K and S. [We should really prove this at some point] K is rejected by the modal logicians who (following C. I. Lewis) have the effrontery to call it a *paradox of material implication*.

A word is in order about this, since K is valid constructively. The key point is that in constructive logic we assume persistence of atomic formulæ. Persistence of atomics implies persistence generally,

Persistence for complex formulæ enforces the constructive theory of the conditional: First we show that this enforces K:

We wish to establish that every world visible from the root world W_0 that believes A also believes $B \to A$. So ignore all worlds that do not believe A. So let W be any world that believes A. Then (by persistence) any world that it can see believes A, so certainly any world that it can see that believes B also believes A. So definitely $W \models B \to A$. But W was an arbitrary world visible from W_0 that believes A. So every world visible from W_0 that believes A also believes $B \to A$, which is to say that $W_0 \models A \to (B \to A)$.

Persistence in a possible world model also implies that S holds, even if the accessibility relation is not transitive (as long as it is at least reflexive). Let us write '$\mapsto$' for the accessibility relation, so we don't find ourselves assuming that it is transitive.

We want the following for all W.
if $W \models A \to (B \to C)$ then $W \models (A \to B) \to (A \to C)$
So, for all W, we want
For all $W' \mapsto W$, if $W' \models A \to B$ then $W' \models A \to C$.

Suppose $W' \mapsto W$. We have $W' \models A \to (B \to C)$ by persistence. Suppose further that $W' \models A \to B$. We desire $W' \models A \to C$. Suppose it doesn't. Then there is $W'' \mapsto W'$ s.t. $W'' \models A$ but $W'' \not\models C$. By persistence we have
$W'' \models A \to (B \to C)$
and
$W'' \models A \to B$
and (by assumption)
$W'' \models A$.

On the (surely safe. . . ?) assumption that the beliefs of any one world are deductively closed we can infer
$W'' \models C$,
contradicting assumption.

Is it safe to assume that $\{\phi : W \models \phi\}$ is deductively closed? For it to be closed under *modus ponens* we need the acessibility relation to be reflexive. It's certainly closed under $\wedge$-int and *ex falso*

Clearly persistence enforces constructive logic but not excluded middle: the usual countermodel does the trick.

Presumably persistence for complex formula doesn't follow from persistence from atomics in the absence of transitivity.

However, if we do not assume persistence, then we can cook up countermodels for $A \to (B \to A)$, even if the accessibility relation is transitive and reflexive.

We need two worlds: the designated world W_0 (which believes both A and B) and W_1 that believes only B. These two worlds can see themselves, and W_0 can see W_1; there are no other worlds. Persistence fails because W_0 believes A but it can see W_1 which has "forgotten" A. $W_0 \not\models A \to (B \to A)$ because it can see a world (itself) that believes A but does not believe $B \to A$. $W_0 \not\models B \to A$ because it can see the world W_1 that does not believe $B \to A$. $W_1 \not\models B \to A$ because it can see a world (itself) that believes B but does not believe A.

Chapter 7

Curry-Howard

This chapter is particularly recommended for anyone who is thinking of going on to do linguistics. It's actually less alarming than most first-years will think, and it may well be worth having a bash at.

Must introduce the slang expression "propositions-as-types"

The Curry-Howard trick is to exploit the possibility of using the letters 'A', 'B' *etc.* to be dummies not just for propositions but for sets. This means reading the symbols '$\to$', '$\wedge$', '$\vee$' etc. as symbols for operations on sets as well as on formulæ. The ambiguity we will see in the use of '$A \to B$' is quite different from the ambiguity arising from the two uses of the word 'tank'. *Those* two uses are completely unrelated. In contrast the two uses of the arrow in '$A \to B$' have a deep and meaningful relationship. The result is a kind of cosmic pun. Here is the simplest case.

Altho' we use it as a formula in propositional logic, the expression '$A \to B$' is used by various mathematical communities to denote the set of all functions from A to B. To understand this usage you don't really need to have decided whether your functions are to be functions-in-intension or functions-in-extension; either will do. The ideas in play here work quite well at an informal level: the best way to think of a function from A to B is purely *operationally*, as a thing such that when you give it a member of A it gives you back a member of B. You don't know anything else about it and you don't need to.

This duplicate points made on page 179

[before we decorate formulæ we need some discussion of the cosmic pun: how K lives inside every $A \to (B \to A)$. The various manifestations of K depend on A and B of course, but they are all things that output constant functions in a way that invites us to give them all the same name 'K'. I find myself reaching for the wonderful conceit in *The Hitch-hiker's Guide to the Galaxy* that every galactic civilisation has a drink called gin-and-tonix. They're all different of course, but every civilisation has one. Whatever the sets A and B are, the drinks cabinet $A \to (B \to A)$ has a bottle in it called 'K'[1].]

[1] "It is a curious fact, and one to which no one knows quite how much importance to attach, that something like 85% of all known worlds in the Galaxy, be they primitive or highly advanced, have invented a drink called jynnan tonnyx, or gee-N'N-T'N-ix, or jinond-o-nicks, or any one of a thousand or more variations on the same phonetic theme. The drinks themselves are not the same, and vary between the Sivolvian 'chinanto/mnigs' which is ordinary water served at slightly above room temperature, and the Gagrakackan

7.1 Decorating Formulæ

7.1.1 The rule of →-elimination

Consider the rule of →-elimination

$$\frac{A \qquad A \to B}{B} \to\text{-elim} \tag{7.1}$$

If we are to think of A and B as sets then this will say something like "If I have an A (abbreviation of "if i have a member of the set A") and an $A \to B$ then I have a B". So what might an $A \to B$ (a member of $A \to B$) be? Clearly $A \to B$ must be the set of those functions that give you a member of B when you feed them a member of A. Thus we can decorate 7.1 to obtain

$$\frac{a : A \qquad f : A \to B}{f(a) : B} \to\text{-elim} \tag{7.2}$$

which says something like: "If a is an A and f takes As to Bs then $f(a)$ is a B.[2] This gives us an alternative reading of the arrow: '$A \to B$' can now be read ambiguously as either the conditional "if A then B" (where A and B are propositions) or as a notation for the set of all functions that take members of A and give members of B as output (where A and B are sets).

These new letters preceding the colon sign are **decorations**. The idea of Curry-Howard is that we can decorate *entire proofs*—not just individual formulæ—in a uniform and informative manner.

We will deal with →-int later. For the moment we will look at the rules for ∧.

7.1.2 Rules for ∧

7.1.2.1 The rule of ∧-introduction

Consider the rule of ∧-introduction:

$$\frac{A \qquad B}{A \wedge B} \wedge\text{-int} \tag{7.3}$$

If I have an A and a B then I have a . . . ? thing that is both A and B? No. If I have one apple and I have one bratwurst then I don't have a thing that is both an apple and a bratwurst; what I do have is a sort of plural object that I suppose is a pair of an apple and a bratwurst—a packed lunch perhaps. (By the way I hope you are relaxed about having compound objects like this in your world. Better start your breathing exercises *now*.) The thing we want is called an **ordered pair**: $\langle a, b\rangle$ is the ordered pair of a and b. So the decorated version of 7.3 is

'tzjin-anthony-ks' which kill cows at a hundred paces; and in fact the one common factor between all of them, beyond the fact that the names sound the same, is that they were all invented and named before the worlds concerned made contact with any other worlds. "

The Restaurant at the End of the Universe, Ch. 24, p. 138 of the Pan paperback.

[2] So why not write this as '$a \in A$' if it means that a is a member of A? There are various reasons, some of them cultural, but certainly one is that here one tends to think of the denotations of the capital letters 'A' and 'B' and so on as predicates rather than sets.

$$\frac{a : A \qquad b : B}{\langle a, b\rangle : A \times B} \wedge\text{-int} \tag{7.4}$$

Say something about how we use $\times$ here ...

What is the ordered pair of a and b? It might be a kind of funny plural object, like the object consisting of all the people in this room, but it's safest to be entirely *operationalist* [3] about it: all you know about ordered pairs is that there is a way of putting them together and a way of undoing the putting-together, so you can recover the components. Asking for any further information about what they *are* is not cool: they are what they do. Be doo be doo. That's operationalism for you. Computer Scientists often describe this in terms of *security*: you are given information about ordered pairs only on a *need-to-know* basis ... and all you need to know is how to pair things up and how to extract components. So that's all you are told.

This duplicate points made on page 177

7.1.2.2 The rule of ∧-elimination

If you can do them up, you can undo them: if I have a pair-of-an-A-and-a-B then I can have an A and I can have a B.

$$\frac{\langle a, b\rangle : A \wedge B}{a : A} \qquad\qquad \frac{\langle a, b\rangle : A \wedge B}{b : B}$$

$A \times B$ is the set $\{\langle a, b\rangle : a \in A \wedge b \in B\}$ of pairs whose first components are in A and whose second components are in B. $A \times B$ is the **Cartesian product** of A and B.

(Never forget that it's $A \times B$ not $A \cap B$ that we want. A thing in $A \cap B$ is a thing that is both an A and a B: it's not a pair of things one of which is an A and the other a B; remember the apples and bratwürste above.)

We need to explain `fst` and `snd`...

$$\frac{x : A \wedge B}{\mathtt{fst}(a) : A} \qquad\qquad \frac{x : A \wedge B}{\mathtt{snd}(x) : B}$$

7.1.3 Rules for ∨

To make sense of the rules for ∨ we need a different gadget.

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

If I have a thing that is an A, then I certainly have a thing that is either an A or a B—namely the thing I started with. And in fact I know which of A and B it is—it's an A. Similarly If I have a thing that is a B, then I certainly have a thing that is either an A or a B—namely the thing I started with. And in fact I know which of A and B it is—it's a B.

Just as we have cartesian product to correspond with ∧, we have **disjoint union** to correspond with ∨. This is not like the ordinary union you may remember from school maths. You can't tell by looking at a member of $A \cup B$ whether it got in there by being a member of A or by being a member of B. After all, if $A \cup B$ is $\{1, 2, 3\}$ it could have

[3] Have a look at chapter 1

been that A was $\{1, 2\}$ and B was $\{2, 3\}$, or the other way round. Or it might have been that A was $\{2\}$ and B was $\{1, 3\}$. Or they could both have been $\{1, 2, 3\}$! We can't tell. However, with disjoint union you *can* tell.

To make sense of disjoint union we need to rekindle the idea of a *copy* from section 2.3.4. The disjoint union $A \sqcup B$ of A and B is obtained by making copies of everything in A and marking them with wee flecks of *pink* paint and making copies of everything in B and marking them with wee flecks of *blue* paint, then putting them all in a set. We can put this slightly more formally, now that we have the concept of an ordered pair: $A \sqcup B$ is

$$(A \times \{\texttt{pink}\}) \quad \cup \quad (B \times \{\texttt{blue}\}),$$

where `pink` and `blue` are two arbitrary labels.

(Check that you are happy with the notation: $A \times \{\texttt{pink}\}$ is the set of all ordered pairs whose first component is in A and whose second component is in $\{\texttt{pink}\}$ which is the singleton of[4] `pink`, which is to say whose second component *is* `pink`. Do not ever confuse any object x with the set $\{x\}$—the set whose sole member is x! We can think of such an ordered pair as an object from A labelled with a pink fleck.)

Notice that, since pink $\neq$ blue, $A \sqcup B \neq B \sqcup A$ ...in contrast to $A \cup B = B \cup A$!

$\vee$-introduction now says:

$$\frac{a : A}{\langle a, \texttt{pink}\rangle : A \sqcup B} \qquad\qquad \frac{b : B}{\langle b, \texttt{blue}\rangle : A \sqcup B}$$

$\vee$-elimination is an action-at-a-distance rule (like $\rightarrow$-introduction) and to treat it properly we need to think about:

7.2 Propagating Decorations

The first rule of decorating is to decorate each assumption with a variable, a thing with no internal syntactic structure: a single symbol.[5] This is an easy thing to remember, and it helps guide the beginner in understanding the rest of the gadgetry. Pin it to the wall:

Decorate each assumption with a variable!

Two assumptions that are discharged at the same place should be given the same decoration, otherwise they should be given different decorsations.

How are you to decorate formulæ that are not assumptions? You can work that out by checking what rules they are the outputs of. We will discover through some examples what extra gadgetry we need to sensibly extend decorations beyond assumptions to the rest of a proof.

[4]The singleton of x is the set whose sole member is x.

[5]You may be wondering what you should do if you want to introduce the same assumption twice. Do you use the same variable? The answer is that if you want to discharge two assumptions with a single application of a rule then the two assumptions must be decorated with the same variable.

7.2.1 Rules for $\wedge$

7.2.1.1 The rule of $\wedge$-elimination

$$\frac{A \wedge B}{B} \wedge\text{-elim} \tag{7.5}$$

We decorate the premiss with a variable:

$$\frac{x : A \wedge B}{B} \wedge\text{-elim} \tag{7.6}$$

...but how do we decorate the conclusion? Well, x must be an ordered pair of something in A with something in B. What we want is the second component of x, which will be a thing in B as desired. So we need a gadget that, when we give it an ordered pair, gives us its second component. Let's write this '`snd`'.

$$\frac{x : A \wedge B}{\mathtt{snd}(x) : B}$$

By the same token we will need a gadget '`fst`' which gives the first component of an ordered pair so we can decorate[6]

$$\frac{A \wedge B}{A} \wedge\text{-elim} \tag{7.7}$$

to obtain

$$\frac{x : A \wedge B}{\mathtt{fst}(x) : A}$$

7.2.1.2 The rule of $\wedge$-introduction

Actually we can put these proofs together and whack an $\wedge$-introduction on the end:

$$\frac{\dfrac{x : A \wedge B}{\mathtt{snd}(x) : B} \quad \dfrac{x : A \wedge B}{\mathtt{fst}(x) : A}}{\langle \mathtt{snd}(x), \mathtt{fst}(x) \rangle : B \wedge A}$$

7.2.2 Rules for $\to$

7.2.2.1 The rule of $\to$-introduction

Here is a simple proof using $\to$-introduction.

$$\dfrac{\dfrac{[A \to B]^1 \qquad A}{B} \to\text{-elim}}{(A \to B) \to B} \to\text{-int (1)} \tag{7.8}$$

[6] Agreed: it's shorter to write 'x_1' and 'x_2' than it is to write 'fst(x)' and 'snd(x)' but this special use of subscripts would prevent us using 'x_1' and 'x_2' as variables on their own account. In any case I feel strongly that the fact that there is a function that extracts components from ordered pairs is an important one, and it should be brought out into the open and given its own name.

We decorate the two premisses with single letters (variables): say we use 'f' to decorate '$A \to B$', and 'x' to decorate 'A'. (This is sensible. 'f' is a letter traditionally used to point to functions, and clearly anything in $A \to B$ is going to be a function.) How are we going to decorate 'B'? Well, if x is in A and f is a function that takes things in A and gives things in B then the obvious thing in B that we get is going to be denoted by the decoration '$f(x)$':

$$\dfrac{\dfrac{f : [A \to B]^1 \quad x : A}{f(x) : B}}{??? : (A \to B) \to B}$$

So far so good. But how are we to decorate '$(A \to B) \to B$'? What can the '???' stand for? It must be a notation for a thing (a function) in $(A \to B) \to B$; that is to say, a notation for something that takes a thing in $A \to B$ and returns a thing in B. What might this function be? It is given f and gives back $f(x)$. So we need a notation for a function that, on being given f, returns $f(x)$. (Remember, we decorate all assumptions with variables, and we reach for this notation when we are discharging an assumption so it will always be a variable). We write this

$$\lambda f.f(x)$$

This notation points to the function which, when given f, returns $f(x)$. In general we need a notation for a function that, on being given x, gives back some possibly complex term t. We will write:

$$\lambda x.t$$

for this. Thus we have

$$\dfrac{\dfrac{f : [A \to B]^1 \qquad x : A}{f(x) : B}\ \to\text{-elim}}{\lambda f.f(x) : (A \to B) \to B}\ \to\text{-int (1)} \tag{7.9}$$

Thus, in general, an application of $\to$-introduction will gobble up the proof

$$\begin{array}{c} \underline{x : A} \\ \vdots \\ \overline{t : B} \end{array}$$

and emit the proof

$$\begin{array}{c} \underline{[x : A]} \\ \vdots \\ \overline{t : B} \\ \hline \lambda x.t : A \to B \end{array}$$

This notation—$\lambda x.t$—for a function that accepts x and returns t is incredibly simple and useful. Almost the only other thing you need to know about it is that if we apply the function $\lambda x.t$ to an input y then the output must be the result of substituting 'y' for all the occurrences of 'x' in t. In the literature this result is notated in several ways, for example $[y/x]t$ or $t[y/x]$.

Go over a proof of S at this point

7.2.3 Rules for $\vee$

We've discussed $\vee$-introduction but not $\vee$-elimination. It's very tricky and—at this stage at least—we don't really need to. It's something to come back to—perhaps!

EXERCISE 70 *Go back and look at the proofs that you wrote up in answer to exercise 26, and decorate those that do not use '$\vee$'.*

7.2.4 Remaining Rules

7.2.4.1 Identity Rule

Here is a very simple application of the identity rule.

See [39]:
Semantical Archæology.

$$\dfrac{\dfrac{\dfrac{A \quad B}{B}}{B \to A}}{A \to (B \to A)}$$

Can you think of a function from A to the set of all functions from B to A? If I give you a member a of A, what function from B to A does it suggest to you? Obviously the function that, when given b in B, gives you a.

This gives us the decoration

$$\dfrac{\dfrac{\dfrac{a : A \quad b : B}{b : B}}{\lambda b.a : B \to A}}{\lambda a.(\lambda b.a) : A \to (B \to A)}$$

The function $\lambda a.\lambda b.a$ has a name: K for Konstant. (See section 3.6.)

Show how do do this using the option of cancelling non-existent assumptions.

7.2.4.2 The *ex falso*

The *ex falso sequitur quodlibet* speaks of the propositional constant $\perp$. To correspond to this *propositional* constant we are going to need a *set* constant. The obvious candidate for a set corresponding to $\perp$ is the empty set. Now $\perp \to A$ is a propositional tautology. Can we find a function from the empty set to A which we can specify without knowing anything about A? Yes: the empty function! (You might want to check very carefully that the empty function ticks all the right boxes: is it really the case that whenever we give the empty function a member of the empty set to contemplate it gives us back one and only one answer? Well yes! It has never been known to fail to do this!! Look again at page 145.) That takes care of $\perp \to A$, the *ex falso*.

7.2.4.3 Double Negation

What are we to make of $A \to \perp$? Clearly there can be no function from A to the empty set unless A is empty itself. What happens to double negation under this analysis?

$$((A \to \perp) \to \perp) \to A$$

- If A is empty then $A \to \perp$ is the singleton of the empty function and is not empty. So $(A \to \perp) \to \perp$ is the set of functions from a nonempty set to the empty set and is therefore the empty set, so $((A \to \perp) \to \perp) \to A$ is the set of functions from the empty set to the empty set and is therefore the singleton of the empty function, so it is at any rate nonempty.

- However if A is nonempty then $A \to \perp$ is empty. So $(A \to \perp) \to \perp$ is the set of functions from the empty set to the empty set and is nonempty—being the singleton of the empty function—so $((A \to \perp) \to \perp) \to A$ is the set of functions from the singleton of the empty function to a nonempty set and is sort-of isomorphic to A. empty.

So $((A \to \perp) \to \perp) \to A$ is not reliably inhabited, in the sense that it's inhabited but not uniformly. This is in contrast to all the other truth-table tautologies we have considered. Every other truth-table tautology that we have looked at has a lambda term corresponding to it.

to be continued

A final word of warning: notice that we have not provided any λ-gadgetry for the quantifiers. This can in fact be done, but there is no spacetime here to do it properly.

7.3 Exercises

In the following exercises you will be invited to find λ terms to correspond to particular wffs—in the way that the λ term $\lambda a.\lambda b.a$ (aka 'K') corresponds to $A \to (B \to A)$ (also aka 'K'!) You will discover very rapidly that the way to find a λ-term for a formula is to find a proof of that formula: λ-terms encode proofs!

Exercise 71 *Find λ-terms for*
$(A \wedge B) \to A$*;*
$((A \to B) \wedge (C \to D)) \to ((A \wedge C) \to (B \wedge D))$*;*
$(A \to B) \to ((B \to C) \to (A \to C))$*;*
$((A \to B) \to A) \to ((A \to B) \to B)$*;*
$(A \to (B \to C)) \to (B \to (A \to C))$*;*
$(A \to (B \to C)) \to ((B \wedge A) \to C))$*;*
$((B \wedge A) \to C)) \to (A \to (B \to C))$*.*

Finding λ-terms in exercise 71 involves of course first finding natural deduction proofs of the formulæ concerned. A provable formula will always have more than one proof. (It won't always have more than one *sensible* proof!) For example the tautology $(A \to A) \to (A \to A)$ has these proofs (among others)

$$\dfrac{\dfrac{[A \to A]^1}{A \to A}\ \text{identity rule}}{(A \to A) \to (A \to A)}\ {\to}\text{-int (1)} \tag{7.10}$$

$$\dfrac{\dfrac{\dfrac{[A]^1 \qquad [A \to A]^2}{A}\ {\to}\text{-elim}}{A \to A}\ {\to}\text{-int (1)}}{(A \to A) \to (A \to A)}\ {\to}\text{-int (2)} \tag{7.11}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{[A]^1 \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim}}{A \to A}\ {\to}\text{-int (1)}}{(A \to A) \to (A \to A)}\ {\to}\text{-int (2)} \tag{7.12}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[A]^1 \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim}}{A \to A}\ {\to}\text{-int (1)}}{(A \to A) \to (A \to A)}\ {\to}\text{-int (2)} \tag{7.13}$$

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[A]^1 \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim} \qquad [A \to A]^2}{A}\ {\to}\text{-elim}}{A \to A}\ {\to}\text{-int (1)}}{(A \to A) \to (A \to A)}\ {\to}\text{-int (2)} \tag{7.14}$$

EXERCISE 72 *Decorate all these proofs with λ-terms. If you feel lost, you might like to look at the footnote[7] for a HINT.*

[7] Notice that in each proof of these proofs all the occurrences of '$A \to A$' are cancelled simultaneously.. Look at the footnote on page 180.

On successful completion of exercise 72 you will be in that happy frame of mind known to people who have just discovered **Church numerals**.

Then we will define `plus` ...

****

Things still to do in this chapter.

Is every λ term a decoration of a proof? No. There is an obvious way to run-in-reverse the process-of-decoration to obtain a proof, but it doesn't always work. Sometimes it fails, and when it fails it will be because the λ-term is UNTYPED!

Should we Say something about typing algorithms? There is a connection here with syntactic typing in the section on semantics

(i) Make more explicit the connection with constructive logic

(ii) Scott's cute example in [39]:

(iii) So far we've been inputting proofs and outputting λ-terms. It's now time to start doing it the other way round.

(iv) Church Numerals, fixed point combinators

(v) Explain α and β conversion.

(vi) Do something with

$$\lambda x.(\lambda w.w(\lambda z.x(\lambda y.z)))$$

We might find some nice things to say about

$$\bigwedge_C [(A \to C) \to ((B \to C) \to C)]$$

which is supposed to be $A \vee B$. After all,

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

and

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

are both provable and therefore

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

will be provable as well. (We saw this in exercise 28. Think about how both A and B imply $(\forall x)[(A \to F(x)) \to ((B \to F(x)) \to F(x))]$)

blah harmony section 3.4

$(\forall x)[(A \to F(x)) \to ((B \to F(x)) \to F(x))]$

We could have a rule of $\vee$-elimination that goes like this

$$\frac{A \vee B}{(A \to C) \to ((B \to C) \to C)}$$

... where C can be anything. How about a λ-term?

$\lambda fg.fa$ and $\lambda fg.fb$

(vii) Provide a similar analysis of $\bigwedge_C(A \to (B \to C)) \to C$.

We want to show that it follows from $\{A, B\}$. Use $\to$-elimination twice on $\{A, B, A \to (B \to C)\}$. To infer A instantiate 'C' to 'A' getting $(A \to (B \to A)) \to A$; to infer B instantiate 'C' to 'B' getting $(A \to (B \to B))toB$. How, one might ask, have we made any progress by explaining finitary con junction in terms of infinitary conjunction? The answer is that $\bigwedge$ isn't really conjunction (aka intersection)—remember that $A \wedge B$ is $A \times B$ rather than $A \cap B$.

K isn't really of type $A \to (B \to A)$, but of type $\bigwedge_A \bigwedge_B(A \to (B \to A))$.

A λ-term? '$\lambda f.fab$' should do the trick.

This shows how, by using $\bigwedge$, we can find λ-terms inside the pure lambda calculus and don't need pairing and unpairing.

Worth pointing out that if we think of $A \vee B$ and $A \wedge B$ as syntactic sugar for the complicated formulæ with $\bigwedge$ then the introduction and elimination rules are harmonious.

(viii) Say something about Paul Taylor's *aperçu* in [43] that Curry-Howard retrospectively makes sense of logicism.

7.4 Combinators and Hilbert Proofs

Work to do here

Take the proof on p 83 and decorate it with combinators:

K	$A \to ((A \to A) \to A)$
S	$(A \to ((A \to A) \to A)) \to ((A \to (A \to A)) \to (A \to A))$
SK	$(A \to (A \to A)) \to (A \to A)$
K	$A \to (A \to A)$
$(SK)K$	$A \to A$

'S' is the proper name for the λ term $\lambda adc.(ac)(dc)$

Chapter 8

How Not to Use Logic

Logic an exercise in concealment. Shouldn't conceal the wrong things. Talk about this in connection with hasty formalisation. A useful form of words: "must be held constant during the exercise". Use this in connection with *drilling down* in formalisation of English in LPC.

hasty formalisation → excessive concealment → fallacies of equivocation.

The key to using logic correctly is concealing precisely the right amount . . . showing precisely the right amount of leg.

If you want to apply Logic (or formal methods) you first have to decide what it is you want a formal theory for. I mean 'for' not 'of'. Why do you want this formal theory? This may be determined for you, of course. Then you have to identify those features that can safely be hidden.

For formalisation to work, two things must be in place. (i) the concepts you are trying to formally reason about must be well-defined, hard-edged; (ii) the concepts you are *not* going to reason about must be concepts that can be safely partitioned-off-from/invariant-over the things you *are* going to reason about.

I would go so far as to say that—to a first approximation—all the failures and infelicities of bad formalisations come from inattention to these considerations.

What does this mean, exactly? (i) means that if the things you are reasoning about are measurements of values of a parameter then you'd better be sure that if you measure some quantity a second time, and at a time when it hasn't changed, you'd better record the same value. (ii) means . . .

It may be of course that your instruments are imperfect, and you have no guarantee that recording a quantity for a second time will give the same reading as you got the first time. Such is life. One can't give up altogether of course, but what it does mean is that one shouldn't reason about the figures that one gets as if they are snapshots of the system, one has to take into account that are at one remove from the system of interest, and the fact that they are at one remove cannot be safely concealed. They are snapshots not of the system, but of the system+observer. The error made by "Fuzzy

Logic" is that it attempts to conceal this remove. It's a form of the mind-projection fallacy. see http://en.wikipedia.org/wiki/Mind_projection_fallacy. Identifying the world with your knowledge of it is really a form of solipsism, of extreme vanity.

If, after having made your initial assessment of what it is safe to conceal, and after having started work on that basis, you feel the urge to use a nonclassical logic, that almost certainly means that you were mistaken in your assessment of what it was safe to conceal. See appendix p 221.

It may be felt that this appeal to the legitimacy of classical logic is circular, and it can certainly seem that way. The best way to set out one's stall is to say that IF you start off with a presumption in favour of classical logic, your bad experiences with attempts at formalisation will never overturn it, for all such bad experiences arise from hasty formalisation and inappropriate concealment. You might have a *ab initio* objection to classical logic, but don't try to justify it by narratives of attempted formalisation.

The iteration test?

Some very special conditions have to be met before you can properly use **a logic** to formalise anything. Of course sometimes you can use a logic as a kind of syntactic sugar for the formalism that you are are constructing (but withholding from the public)

"syntactic sugar"?

So why do people do it? I think the answer is, in part, that there are a lot of syntax buffs around, who just *love* inventing new logics.

Propositional logic is a triumph of ellipsis. We can get away with writing 'p' instead of '$p_t^{ch,co}$' (which would mean that Chap ch in context co asserts p at time t) as long as we can hold all these other parameters constant. In settings where the other parameters cannot be held constant the ellipsis is not safe. Yet it is precisely this kind of ellipsis we have to perform if what we want is a *logic* rather than a first-order theory of deduction-tokens-in-context. Here is an example of how *not* to do it, taken from a standard text (names changed to preserve anonymity). Kevin (not his real name) and his friends have been having some fun in the chemistry lab, and they wrote:

ceteris paribus other things being equal

$$MgO + H_2 \to Mg + H_2O$$

(This would require extreme conditions but never mind[1]; this is at least standard chemical notation.)

Then assume we have some MgO and some H_2. They (Kevin and his friends) end up representing reactions by means of *logical* formulæ like

$$(MgO \wedge H_2) \to (Mg \wedge H_2O) \qquad \text{(K1)}$$

This is on the basis that if one represents "I have some MgO" by the propositional letter 'MgO' (and others similarly)[2] then the displayed formula does not at all represent the reaction it is supposed to represent. $p \to q$ does not say anything like "p and then q" (at which point p no longer!) but once one "has" Mg and H_2O as a result of the reaction

[1] I suppose it just *might* work if you roasted magnesia in a stream of hydrogen at a temperature above the melting point of magnesium metal. However I suspect not: Kevin probably not only knows no logic but no chemistry either.

[2] Do not be led astray by the fact that 'MgO' is three letters in English! It's only *one* in the propositional language we are setting up here!

allegedly captured by the displayed formula one no longer "has" any Mg or H_2O: it's been used up! In contrast, p and $p \to q$ are not in any sense "used up" by modus ponens. And nothing will be achieved by trying to capture the fact that the reagants are used up by writing something like repetition

$$(\mathrm{MgO} \wedge \mathrm{H}_2) \to ((\mathrm{Mg} \wedge \mathrm{H}_2\mathrm{O}) \wedge \neg\mathrm{MgO} \wedge \neg\, \mathrm{H}_2)$$

Consider what this would mean. It would mean that from the assumption $\mathrm{MgO} \wedge \mathrm{H}_2$ we would be able to infer $\neg\mathrm{MgO} \wedge \neg\, \mathrm{H}_2$, and this conclusion contradicts the assumption, so we would infer $\neg(\mathrm{MgO} \wedge \mathrm{H}_2)$, and that is clearly not what was intended. The problem—an important part of it at least—is that we have tried to get away without datestamping anything.

Now if we spice up the formalism we are using by means of datestamping, then it all becomes much more sensible. Rather than write 'MgO' to mean "Kevin has some magnesia" we write 'MgO(t)' to mean "at time t Kevin [or whoever it happens to be] has some magnesia"—and the other reagents similarly—then instead of (K1) we have

$$\mathrm{MgO}(t) \wedge \mathrm{H}_2(t) \to \mathrm{Mg}(t+1) \wedge \mathrm{H}_2\mathrm{O}(t+1) \tag{K2}$$

which is altogether more sensible. Notice that just as we left the datestamps out of the original formulation, here we have left out the name of the poor helot in the lab coat. That is perfectly OK, because the chemistry doesn't depend on the chemist.

In writing 'MgO(t)' we have taken the (possession-by-Kevin of) magnesia to be a predicate, and points-in-time as arguments. We could have written it the other way round: 't(MgO)' with time as the predicate and magnesia as the argument. That way it more obviously corresponds to "at time t there is some magnesia". Or we could make the lab technician explicit by writing something like '$K(\mathrm{MgO}, t)$' with a two-place predicate $K(,)$ which would mean something like "Kevin has some magnesia at time t". Indeed we could even have had a *three*-place predicate and a formulation like 'H(k,MgO,t)' to mean that "k has some magnesia at time t". All of these can be made to work.

The moral of all this is that if there are important features—such as datestamping—that your formalisation takes no account of, then you shouldn't be surprised if things go wrong.

To forestall the charge that I have tried to burn a straw man instead of a heretic, I should point out that this example (of how *not* to do it) comes from a textbook (which should be showing us how we *should* do it), to wit [censored][3]

Mind you, there might be a useful parallel one can draw here between logical-truth-plus-evaluation-strategies on the one hand and chemical-reactions-with-mechanism on the other. The arrow syntax that one remembers from school chemistry

[3]Most of us have at one time or another committed to paper *sottises* like this—if not worse—and nevertheless subsequently gone on to lead entirely blameless lives. The law providing that spent convictions should be overlooked is a good one, and it is there to protect others as well as me. Kevin has been granted name suppression.

$$NaOH + HCl \to NaCl + H_2O$$

hides mechanism.

8.1 Many-Valued Logic is an Error

Edit this section

[summary: to be eventually suppressed

Many–states does not correspond to many–truth-values–of–a–single–proposition. Sometimes (if the number of states is a power of 2) they might correspond to compound propositions. But never if the number of states is not a power of two. Why not? Why might there not be lots of truth-values? The point is that all the different stories in the different case have to cohere.

Nobody thinks that trichotomy is a reason for having a 3-valued logic for describing total order. Why are people attracted to this error? Probably lots of reasons. evaluation blah mind-projection fallacy]

In section 3.2.5 we met a device called *three-valued logic*. No suggestion was made there that the third truth-value has any *meaning*. It is purely a device to show that certain things do not follow from certain other things. However there is the obvious thought that perhaps the third truth-value really *might* mean something, and that the logic with the three values is a theory of *something*, something different from classical two-valued logic. Might there be any mileage in this idea? Might there, perhaps, have always been three truth-values, and the thought (often called *bivalence*) that there are only two a naïve delusion?

I don't see any way of refuting the idea that there really are more than two truth values (in fact I don't even know what such a refutation would look like) but what one can do is analyse all the lines of chat that suggest that there might be, and show how they stand up under examination.

What phenomena become intelligible, what puzzles disappear, if we think there are more than two truth-values?

Not to say that there cannot be particular restricted settings in which a pretence that there are more than two truth-values might not be a useful programming trick. But that's not metaphysics.

8.1.1 Many states means many truth-values . . . ?

One of the impulses towards many-valued logic is the thought that for any system with many states there is a special proposition about that system, such that the system being in any one state is simply that proposition having a particular truth value.

This is completely crazy, but there are special cases in which it can look plausible.

If a parameter can take one of two values, A and B, there is no harm in thinking of these two outcomes as two truth-values of a proposition. Indeed it's obvious what that proposition must be: "The value of the parameter is A". If $\neg p$ then the value isn't A so it must be B, co's the parameter is two-valued. The sensitive reader might feel that this is a bit of a cheat, a *virtus dormitiva*. It is, but it does at least work. What happens if the parameter can take more than two values?

Suppose my local supermarket has four brands of vegetarian sausage. It may be that there are two natural yes/no parameters that I can use to describe them, say `gluten-free` *vs* ¬`gluten-free` and `with-garlic` *vs* `without-garlic`, so I can characterise my purchase by giving truth values to two independent propositions (*my–sausage–has–garlic* and *my–sausage–is–gluten-free*) but equally it might be that there are no natural divisions of that nature. But even if there is no natural division—as in the case where there are four varieties unilluminatingly called A, B, C and D—I can still cook up two propositions p and q such that four possibilities A, B, C and D correspond to the four possible combinations $p \wedge q$, $p \wedge \neg q$, $\neg p \wedge q$ and $\neg p \wedge \neg q$.

EXERCISE 73 *What might p and q be? (Remember that A, B, C are mutually exclusive and jointly exhaustive. There is more than one correct answer!)*

[*strongly agree / agree / disagree /strongly disagree* does look like four truth-values rather than two propositions. agree/disagree and strong/mild doesn't seem to do it. Must discuss this. " "To what four situations do the four truth-value attributions of "Professor Copeland is with us today" correspond?" Our course even if there are four truth-values some props behave as if they have only two. The real global truth-value algebra must be something of which all these little algebras are quotients. It is very striking that there is no consideration in the literature of what that global truth-value algebra might be.]

The Blood Group Exercise

Consider the relation between humans "It is safe for x to receive a transfusion of blood from y." Ignoring the blood-borne diseases like HIV, CJD, Hep C and so on, we find that if x can safely receive a transfusion of blood from y, and y' belongs to the same blood group as y, then x can safely receive a transfusion of blood from y'. That is to say, the equivalence relation of having-the-same-blood-group is a congruence relation for the binary relation "x can safely receive a transfusion of blood from y". That way we can think of the relation "x can safely receive a transfusion of blood from y" as really a relation between the blood groups, and summarise it in the following matrix.

Are we going to explain congruence relations?

Columns are donors, rows are recipients.

	$O-$	$O+$	$B-$	$B+$	$A-$	$A+$	$AB-$	$AB+$
$O-$	×							
$O+$	×	×						
$B-$	×		×					
$B+$	×	×	×	×				
$A-$	×				×			
$A+$	×	×			×	×		
$AB-$	×		×		×		×	
$AB+$	×	×	×	×	×	×	×	×

So the table is telling you that O- people can donate blood to everyone, and so on.

Your blood group is a parameter that can take one of eight values. You would have to be very obtuse not to suspect that there are three switches (three propositions) each of which can take two values.

Perhaps enough has been said for the reader to appreciate that a system with 2^n states can be described by n propositions each taking two truth-values.

These analyses—and some others like them—are sensible, but they don't argue for lots of *truth-values*, rather for lots of *propositions*, and they don't work at all when the number of states is not a power of 2:

(i) We don't want to infer from the fact that there are n alleles at a locus that there is a proposition with n truth values.

(ii) The electronic lock on my building at work has three states: locked, unlocked and disabled. Nobody thinks there is a proposition which has three truth-values corresponding to these three states.

A jumble below here

By implication if it doesn't work for a number of states that isn't a power of 2 then it might perhaps not be a good idea even when the number of states is a power of 2. A cricket match has four possible outcomes: win, lose, draw and tie. But this situation doesn't seem to invite an analysis in terms of two propositions. Even less does it seem to invite an analysis using a single proposition and four truth-values, where—say—winning corresponds to true, losing to false, and the other two to two novel truth-values `draw` and `tie`, say.

If there really were four such truth-values, one would be able to ask (and would have to be able to answer) "What state of affairs is described by the proposition "Balbus loves Julia" having truth-value `tie`? Why? The point is that the number of truth-values is a fact about nature herself. No-one is suggesting that propositions about wombats are two-valued but propositions about dingbats are three-valued, propositions about widgets have four values but propositions about gadgets have five. One policy to rule them all. If we have one story that suggests there are three truth-values and they behave like this, and another story that says there are five truth-values and they behave like that, then we don't have two arguments against bivalence, we merely have two *potential* arguments against bivalence, and—sadly—they are mutually contradictory, so at least one of them must be wrong.

failure to evaluate

refer back to p 29.

It's an example of the mind projection fallacy.

It's a bad idea to think that the failure of evaluation is a third truth-value. For one thing, if you wait a bit longer, the thing you are looking at that hasn't yet evaluated might yet evaluate, which means you are hiding a (date) parameter that you shouldn't have hidden. That probably won't matter, but you might want to take seriously the time consumption of the evaluation. So that, if A and B are both undetermined at time t, then $A \wedge B$ is undetermined at time $t + 1$. (It takes you one clock tick to compute the conjunction.) For another (and more seriously) if we pretend that failure-to-evaluate-at-time-t is a truth-value we find that the intended semantics for this logic is no longer compositional, and we no longer have the theorem that eager and lazy evaluation give the same answer. The truth value of $A \to B$ is undetermined if the truth-values of A and B are undetermined, but the truth value of $A \to A$ is always 1. Thus an eager semantics will not detect the fact that the truth-value of $A \to A$ is 1 but a lazy semantics (or at least a suitably ingenious top-down semantics) will.

Should have a proof of this. See section 3.14.

On a related topic. We have to be very careful with the impulse to use three-valued logic in connection with Sorites. If we believe that logical equivalence of two expressions is having-the-same-truth-value-under-all-valuations then A and $A \wedge A$ will have the same truth-value-under-all-valuations. Further, we desire that the truth-value of $A \wedge B$ under any given valuation will depend solely on the truth-values of A and B under that valuation. (This is a point about *truth-functionality* not *compositionality*.) So if A and B have the same truth-value under a given valuation, $A \wedge B$ will have the same truth-value as $A \wedge A$, which is to say, the same as A. In other words **the conjunction operator acting on truth-values must be idempotent.** (Ditto disjunction, for that matter). But then that means that the ruse of giving all the assertions

Do they know 'idempotent'?

> "If a man with n hairs on his head is not bald, then
> neither is a man with $n - 1$ hairs on his head" `bald`(n)

the same truth value of 0.99999 will not have the desired effect. For then the conditional

> "If a man with 150000 hairs on his head is not bald, then neither is a man
> with 0 hairs on his head"

which must have a truth value at least as great as the conjunction of the truth values of all the 150000 expressions `bald`(n). But conjunction is idempotent and truth-functional, so this conjunction has truth-value 0.99999, *not* the truth-value $(0.99999)^{150000} = 0.223$ which we wanted, so we don't get the *leaking away of truth* that we wanted. This kind of thing works with *probabilities* or *credences* but not truth-values.

could say more about truth-functionality-plus-idempotence

8.2 Beware of the concept of *logically possible*

For the purposes of this discussion a (Chalmerian) *zombie* is a creature physiologically like us but without any mental life. I have heard it claimed (as part of a wider programme) was that zombies are logically possible but perhaps not metaphysically possible.

The only sane point of departure for a journey that uses the concept of *logical possibility* is that of *satisfiable* formula of first-order logic. ('Logically possible' is presumably a logical notion so one's first enquiries are to logicians—presumably!) An important but elementary point which we have been emphasising (see section 2.3.2 *et seq*) is that whether or not a proposition is logically possible depends only on its logical structure, and not on the meanings of any tokens of the non-logical vocabulary to be found in it. It's not logically possible that, say

All bachelors are unmarried and at least some bachelors are married.

because there is no way of interpreting those two predicates in such a way that the conjunction comes out true. (that is, if we agree to take 'unmarried' to be syntactic sugar for 'not-married', so that the displayed expression has the logical form '$(\forall x)(B(x) \to \neg M(x)) \land (\exists x)(B(x) \land M(x))$'). So is it logically possible that there are zombies? Yes, clearly. The physiological vocabulary and the mental vocabulary are disjoint so it's easy (using the interpolation lemma, thm 16) to find a structure which contains zombies (= things that according to that structure are zombies).

Now i don't think that was what is meant by people who think it's logically possible that there should be zombies. They want to "reserve" the denotations of the mental predicates and the physiological predicates, rather in way that the meaning of the symbol '=' is *reserved.* (Forgive me using the Compsci jargon word 'reserved': it happens to be exactly what i mean—even tho' it is the property of a different community!) Now reserving a word in this way is a very significant move. The only predicate letter that logicians ever reserve is '=', and when they do that they are aware that they are doing something that needs to be flagged. They speak of 'predicate calculus with equality' or 'predicate calculus without equality'. Nowhere in the logical literature is the possibility of reserving any other predicate letter ever considered. Nevertheless philosophers appear to be talking as if such reservations were routine. Predicate calculus with zombies(!) But never mind. Let us suppose one can reserve mental predicates in this way and see what happens. But if we really knew what the denotation of the mental predicates were—so that we could fix them—the question of whether or not there are interpretations of the predicate letters in "There are zombies" which make that sentence true would reduce to the question of whether or not there are, in fact, any zombies.

I think this is where philosophers reach for the modish expression 'rigid designator'

And i don't think that is what was meant either!

Chapter 9

Other Logics

Logic of questions. Logic of commands is the study of programming languages. A rich and beautiful topic. Here the distinction between different kinds of evaluations lazy vs strict really matters

Infinitary Logic? The quantifiers as infinitary conjunction and disjunction. Harmony tells us that $A \vee B$ can be seen as the conjunction of all $[(A \to C) \wedge (B \to C)] \to C$

Since '$A \vee B$' can be subst for 'C' we have an illfounded subformula relation.

Monadic second-order logic is OK. Possibly talk about branching-quantifier logics.

The logics we consider have arisen from thinking about modes of inference that preserve certain features of the propositions we are reasoning about. In the case of classical logic, the thing being preserved is truth. As it happens it doesn't make any difference whether it is truth-in-a-particular-interpretation we are preserving or truth-in-all-interpretatations. With constructive logic we restrict ourselves to rules that preserve the property of having a proof that respects the existence property. That *tightens* up the rules we are allowed to use. Sometimes it might make sense to *relax* them. When? If we think that the universe is not infinite then we might drop our guard to the extent of allowing ourselves the luxury of rules of inference that do not preserve the property of being true in *all* models but do at least preserve the property of being true in all *finite* models. After all, if the only models in which the conclusions we are now allowed to draw could fail would be infinite models then we have nothing to fear. As it happens, the extra principles we could safely allow ourselves to use on this assumption are neither easy to capture with rules nor particularly useful. But the thought opens the door to other possibilities.

Say something about this

We might be interested in principles of reasoning that—even if they don't work in *all* models, at least preserve truth-in-models-where-there-is-a-God, or truth-in-models-that-contain-humans (we aren't going to find ourselves in any that don't, are we!) or all truth-in-all-models-where-there-has-not-been-a-nuclear-war (not much point in planning for that contingency really, is there?—it's something you try to avoid, not something you prepare for.)

However most of these relaxations don't result in new *logics* as that word is generally understood...

9.1 Relevance Logic

K seems obvious: Clearly we can deduce A from A. Adding further information might enable us to deduce more things, but it cannot prevent us from deducing things we could deduce already. If we could deduce A from A, we can certainly deduce it from A and B. Thus no-one can argue against $A \wedge B \to A$. And—as long as we accept the equivalence of $A \to (B \to C)$ and $A \wedge B \to C$—we can deduce K. (There are connections here to the principle of *Independence of Irrelevant Alternatives* from economics.).

This thought—that if we can deduce A from A we can deduce it from $\{A, B\}$—is sometimes referred to as the *monotonicity* of deductive logic.

Digression on Monotonicity

It can do no harm to have this word explained. The word 'monotone' in mathematics refers to functions f which satisfy conditions like

$$x \leq y \to f(x) \leq f(y).$$

We say such a function is *monotone increasing* with respect to $\leq$. (If instead f satisfies $x \leq y \to f(x) \geq f(y)$ we say f is monotone *de*creasing with respect to $\leq$.) Of course, it may be (as it is in fact the case here) that the partial order in the antecedent of the condition is not the same partial order as in the consequent, so ideally we would need a more complex form of words along the lines of "f is monotone [increasing] with respect to $\leq$ and $\leq'$". However this ideal notation is never used, being sacrificed by ellipses to the form of words "f is monotone [increasing]".

We use it here because the function F that takes a set of assumptions A and returns the set $F(A)$ of its logical consequences is monotone with respect to set-inclusion :

$$A \subseteq B \to F(A) \subseteq F(B).$$

[We have in fact encountered this notion of monotonicity earlier under a different name: the phenomenon of *persistence* that we saw in section 6.3 tells us that in a possible world model for constructive logic the function $\lambda W.\{\Phi : W \models \Phi\}$ (that sends each world to the set of things it believes) is a function that is monotone with respect to the accessibility relation and logical strength.]

However, ordinary commonsense reasoning is not monotone in this way. In real life we might infer[1] A from B even if we are not deductively authorised to do so, as long as the evidence is suggestive enough—while reserving the right to change our minds later. There are circumstances in which I might risk inferring A from B but definitely not from $B \wedge C$. This can happen if A is true in most cases where B holds (so we are generally happy to risk inferring A from B) but not in the unlikely event of C.

The standard example is

$$\frac{\text{Tweety is a bird}}{\text{Tweety can fly.}}$$

[1]Notice that I am using the word 'infer' not the word 'deduce' here!

The sting in the tail is that Tweety is a penguin. I am writing this in NZ so actually Tweety was a kiwi but never mind. In those cases we infer q from p but not from $p \wedge r$. Defeasible reasoning (thought of as a function from sets-(of assumptions) to (sets-of) conclusions) is not monotone with respect to set-inclusion. Nor is it monotone with respect to temporal-order and set-inclusion. If i am allowed to retract beliefs then the set of things $K(t)$ that I hold true at time t is not a monotone functon of t: $t \leq t'$ does not imply $K(t) \subseteq K(t')$. After all, if at time t I am told that Tweety is a bird, then I may well hold-true-at-time-t that Tweety can fly. However, when I learn—at time $t+1$—that Tweety is in fact a kiwi I no longer hold true that Tweety can fly.

Blah a theory of inference-tokens not inference-types Blah

9.2 Resource Logics

Drop weakening and contraction.

p doesn't suddenly cease to be true just because I act on the assumption that p. Let's return to our example from page 58 ... where it is sunny and it's a tuesday. By $\wedge$-elimination I infer that it is sunny and consequently that it would be good to go for a walk in the botanics. However altho' i have *used* the assumption that it-is-sunny-and-it's-a-tuesday i definitely haven't *used-it-up*. It remains available to me, for me to infer from it also that it is tuesday—which will remind me that I have an 11 o'clock lecture to go to. No doubt it would be nice if i didn't have to go to 11 o'clock lectures on sunny tuesdays but logic gives me no help there.

So the idea that you can use each "assumption" precisely once means that the ps and qs that you are minding so carefully are not propositions, but something else: they must be dollar coins, or something.[2] When I make this point and say "Why call it a logic, not just a first-order theory of double-entry bookkeeping?", Ed Mares replies: "Because it has cut-elimination". What I should have replied then (but i am doing it belatedly now, because i didn't see it at the time) is that if it's not propositions we are manipulating then why is cut-elimination such a big deal? I suppose it means that this theory of double-entry bookkeeping has additive cancellation: you can borrow a resource and pay it back. Or perhaps it means that you can lay off your bets. That makes it important all right, but what does that have to do with ***logic***?

[2]One of these logics was started by the great French proof-theorist Jean-Yves Girard, and I remember the example he used in a talk I went to: "*Eeef* I have zee dollaire, I can buy zee packet of condoms, and *Eeef* I have zee dollaire, I can buy zee packet of fags; but *Eeef* I have zee dollaire I cannot buy *boace* zee packet of fags and zee packet of condoms!".

Chapter 10

Some Applications of Logic

The logical gadgetry we have seen so far can lead in two directions:

1. The gadgetry can be exposed to philosophical analysis. We can try to get straight things like the constructive concept of *proposition* or *truth*. See, for example, [27], [28] and [14].

2. We can use the logic tools to attack problems in philosophy. e.g. *Synonymy* is a concept from philosophy of language which logic has quite a lot to say about.

This chapter will have examples of both.

Exercise 21, part 10 is a minor example of the kind of thing we are after. This argument, valid tho' it is, isn't going to impress a firm believer. That's not to say that it is a completely pointless exercise. It can help to focus our minds on the issues.

H I A T U S marry these two paras together

In the two examples we are about to see Logic is useful to us because it gives us a formal proof of something which appears to be fairly clearly absurd. The dscipline of formalisation helps us to see which of the conditions-for-correct-use-of-logic have been violated, and this gives us feedback about the way we have been thinking about the phenomena.

10.1 Berkeley's Master Argument for Idealism

Berkeley's Master Argument [1] for Idealism combines immediate appeal and extreme murk, which makes it an ideal thing for logicians to practice their gadgets on. In this section I cover some of the modern work on it using those gadgets. My purpose here is pædogogical rather than investigatory: I want to show what can be done with the logical machinery we developed in the earlier chapters. I want to put the machinery through its paces, and Berkeley's Master argument is a stretch where the going is hard enough to test all the runners thoroughly: the murkiness of Berkeley's argument makes for lots of pitfalls in the application of logical gadgetry—and that of course suits my pædogogical purpose.

In what follows we will see natural deduction, modal operators, ϵ-terms, and the intension/extension distinction.

The purpose of *Berkeley's Master Argument* is to prove that everything exists in the mind. Berkeley cast it in the form of dialogue, as people did in those days. The two interlocutors are *Philonous* and *Hylas*. Philonous is Berkeley's mouthpiece, Hylas the stooge[1].

> HYLAS : What more easy than to conceive of a tree or house existing by itself, independent of, and unperceived by any mind whatsoever. I do at present time conceive them existing after this manner.
>
> PHILONOUS : How say you, Hylas, can you see a thing that is at the same time unseen?
>
> HYLAS : No, that were a contradiction.
>
> PHILONOUS : Is it not as great a contradiction to talk of *conceiving* a thing which is *unconceived*?
>
> HYLAS : It is
>
> PHILONOUS : This tree or house therefore, which you think of, is conceived by you?
>
> HYLAS : How should it be otherwise?
>
> PHILONOUS : And what is conceived is surely in the mind?
>
> HYLAS : Without question, that which is conceived exists in the mind.
>
> PHILONOUS How then came you to say, you conceived a house or a tree existing independent and out of all mind whatever?
>
> HYLAS That was I own an oversight . . .

Readers do not know what a proppositional attitude is

slooplessness

There is surely some simple point to be made by appealing to the difference between "intensional" and "extensional" attitudes. You can desire-a-sloop without there being a sloop. Don't we have to ask some awkward questions about which of these "conceive" is, intensional or extensional? Surely it is only if it is extensional that Philonous' trick ever gets started; and it is surely clear that Hylas reckons that the conceiving he is doing is *intensional*. Though this is probably intentional with a 't', as in Chisholm.

10.1.1 Priest on Berkeley

In [32] Graham Priest gives a very elegant formulation of Berkeley's Master Argument, and I will recapitulate it here.

Priest starts off by distinguishing, very properly, between **conceiving objects** and **conceiving propositions**. This answers our concerns in the previous section about equivocating between intensional and extensional kinds of conceiving. Accordingly

[1] I am endebted to my colleague Aneta Cubrinovska for pointing out to me that 'Hylas' comes from a Greek word meaning 'matter' and 'Philonous' means lover of 'nous' or (loosely) mind.

in his formalisation he will have *two* devices. One is a sentence operator T which is syntactically a modal operator and a the other is a predicate τ whose intended interpretation is that $\tau(x)$ iff x is conceived. It is presumably true that both these devices should be relativised—in the sense of having an extra argument place where we can put in a name of whoever is doing the conceiving, but since this is probably going to be plugged with some fixed constant we will ignore it. As a matter of record, no developments of this problem have depended on "conceivable" meaning conceivable not just by me but by anyone else, but one certainly should not exclude this possibility. Until further notice, then $T\phi$ means that the proposition ϕ is being conceived by me, and $\tau(x)$ means that I am contemplating the object x.

In addition to the usual devices of the ϵ-calculus from section 4.14.3 we will adopt the following schemes for these syntactic devices.

$$\frac{\phi \to \psi}{T(\phi) \to T(\psi)}$$

Priest calls this **affixing**. The other rule is one that tells us that if we conceive an object to be something then we conceive it.

$$\frac{T(\phi(x))}{\tau(x)}$$

Let us call it the **mixed rule**. We of course have the usual rule of ϵ-introduction for ϵ-terms (from page 136) namely

$$\frac{\exists y \Psi(y)}{\Psi(\epsilon x \Psi(x))}$$

Priest's formalisation procedes as follows. One particular ϵ-term we shall need a great deal is $\epsilon x.\neg\tau(x)$. Since it takes up a lot of space it will be abbreviated to 'c'. The only thing we know about this c is that $(\exists x \neg\tau(x)) \to \neg\tau(c)$. This indeed is a logical truth, so we can allow it to appear as an undischarged assumption in the natural deduction proof which we will now exhibit:

$$\dfrac{\dfrac{\dfrac{\dfrac{(\exists x\neg\tau(x)) \to \neg\tau(c)}{T(\exists x\neg\tau(x)) \to T(\neg\tau(c))}\text{ Affixing} \qquad T(\exists x\neg\tau(x))}{T(\neg\tau(c))}\to\text{-elim}}{\tau(c)}\text{ Mixed Rule} \qquad \dfrac{(\exists x\neg\tau(x)) \to \neg\tau(c) \qquad (\exists x\neg\tau(x))}{\neg\tau(c)}\to\text{-elim}}{\bot}\to\text{-elim} \tag{10.1}$$

Thus, by judicious use of the ϵ machinery, we have derived a contradiction from the two assumptions $(\exists x)\neg\tau(x)$ and $T(\exists x\neg\tau(x))$, which both sound like possible formulations of realism.

Do we accept this argument as legitimate? Or, for that matter, as an accurate formalisation of what Berkeley was trying to do? The answer to the first question depends on whether or not we think that the premiss $T(\neg\tau(c))$ really has an occurrence of '$\epsilon x.\neg\tau(x)$' in it. If '$\neg\tau(c)$' is merely syntactic sugar for '$(\exists x)\neg\tau(x)$' then $T(\neg\tau(c))$ has no occurrence of '$\epsilon x.\neg\tau(x)$' in it, so the use of the mixed rule is illegimate. If we think it does have an occurrence of '$\epsilon x.\neg\tau(x)$' in it then the use of the mixed rule is perfectly legitimate. But then we feel cheated. Surely it cannot be intended that we should be allowed to use the mixed rule when the argument to 'τ' in the conclusion is a dodgy term invented only as syntactic sugar for something? Perhaps we can get round this by insisting that the 'ϕ' in the mixed rule must be in primitive notation . . . ?

Let us follow this idea wherever it may lead and try accordingly for a proof that uses no ϵ machinery. I have also kept the proof constructive. This is in part because I want to use as little machinery as possible, but another consideration is that I suspect that the existence proof for a contemplation-of-an-arbitrary-object that Berkeley has in mind might be nonconstructive, and this possibility is something worth keeping an eye on. (We certainly don't want any badnesses in the proof to be attributable to peculiarly classical modes of reasoning if we can prevent it.)

$$
\begin{prooftree}
\AxiomC{$[p]^1$}
\AxiomC{$[\tau(x) \to \bot]^2$}
\RightLabel{identity rule}
\BinaryInfC{$\tau(x) \to \bot$}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$p \to (\tau(x) \to \bot)$}
\RightLabel{Affixing}
\UnaryInfC{$Tp \to T(\tau(x) \to \bot)$}
\AxiomC{Tp}
\RightLabel{$\to$-elim}
\BinaryInfC{$T(\tau(x) \to \bot)$}
\RightLabel{Mixed Rule}
\UnaryInfC{$\tau(x)$}
\AxiomC{$[\tau(x) \to \bot]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$(\tau(x) \to \bot) \to \bot$}
\RightLabel{$\forall$-int}
\UnaryInfC{$(\forall x)((\tau(x) \to \bot) \to \bot)$}
\end{prooftree}
\tag{10.2}
$$

The presence of the undischarged assumption Tp is admittedly an infelicity, but it's one we cannot hope to be rid of. Neither the affixing rule nor the mixed rule have anything of the form $T\phi$ as a conclusion. This means that if we want to draw conclusions of the form $T\phi$ then we have to have premisses of that form. So if we (i) interpret T as a *falsum* operator—so that Tp is always false, and (ii) interpret τ as a predicate with empty extension—so $\tau(x)$ is always false, then the rules are truth-preserving. So we cannot expect to be able to prove that even one thing is τ without some extra premisses.

"Side condition"?

Do we like this proof any better? It seems to capture Berkeley's *aperçu* just as well as the last one—10.1—did. But there still seems to be trouble with the mixed rule. The idea that one is contemplating x whenever one is entertaining a proposition about x seems entirely reasonable, but surely this is only because one is tacitly assuming that the term denoting the contemplated object is a constant not a variable. If we allow the argument to 'τ' in the conclusion to be a variable then one derives the absurd conclusion that Berkeley is trying to foist on us. The mixed rule surely invites the side condition that the argument to 'τ' in the conclusion must be a closed term in primitive notation. If we respect that restriction we then find that from the assumption Tp (which, as we have seen, we cannot avoid) we can infer $\neg\neg\tau(t)$ for any closed term t. This is an entirely congenial conclusion: if we have a name for something then we can graciously concede that it is not unconceived.

What about affixing?

The complications have all arisen in connection with the mixed rule, but the rule of affixing invites comment too. It does seem disquietingly strong. Suppose there is a proposition p such that we entertain p and[2] we entertain $\neg p$; Then, by affixing, we infer $Tp \to T\bot$, whence $T\bot$. But we certainly have $\bot \to q$ whence $T\bot \to Tq$ by affixing and then Tq. But q was arbitrary.

$$\dfrac{\dfrac{\dfrac{T(p\to\bot)}{Tp\to T\bot}\text{Affixing}\qquad Tp}{T\bot}{\to}\text{-elim}\qquad \dfrac{\dfrac{\dfrac{[\bot]^1}{q}\text{ex falso}}{\bot\to q}{\to}\text{-int }(1)}{T\bot\to Tq}\text{Affixing}}{Tq}{\to}\text{-elim} \tag{10.3}$$

The conclusion is that if one contemplates/entertains both p and $\neg p$ (and what fair-minded person entertaining p does not also entertain $\neg p$? one might ask) then every proposition is contemplated. This seems too strong. Let us not forget that whatever affixing-style rule we use in an effort to recuperate Berkeley's argument should be as weak as possible.

Priest, being a dialetheist, might deny that $\bot \to q$ holds for arbitrary q. Again, one might deny that $\neg p$ has the logical form of a conditional, and insist that therefore affixing cannot be used in the left branch of 10.3. If affixing is not hamstrung by some manœuvre such as these it does seem very strong: too strong to be what we want, perhaps.

Another possibility to ponder is that if $\tau(x)$ then there must be a ϕ with a free occurrence of 'x' in it such that $T\phi$...

10.2 Fitch's Knowability Paradox

See the Stanford Encyclopædia article at `http://plato.stanford.edu/entries/fitch-paradox/`

The idea is to show that there is no truth which is not known.

Clearly if p is a truth not known to you then this very fact itself (that p is a truth not known to you) is not a truth you can know, so we should expect trouble if we have a principle that says that all truths are knowable.

The apparatus we are about to use needs a couple of new principles. One is **neccessitation**, which says that if we have proved something outright then we have proved also that it is necessary [we should have seen this in the chapter on modal logic]. The reference, please
other is the principle that from p we can infer $\Diamond Kp$, that is to say: if p then it is possible for p to be known. We also need (tho' these are hardly as contentious) the two principles that (i) if $p \wedge q$ is known then so are p and q, and (ii) the equally uncontentious view that p follows from p-is-known. (We say K is *factive*.)

[2]Surely, for at least some notions of *entertaining a proposition* entertaining p and entertaining $\neg p$ are the same thing ...? Plausibly this holds for the kind of entertaining at play in Berkeley's Master argument.

$$
\begin{prooftree}
\AxiomC{$[K(p \wedge \neg Kp)]^1$}
\RightLabel{K-distrib}
\UnaryInfC{$Kp \wedge K\neg Kp$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{Kp}
\AxiomC{$[K(p \wedge \neg Kp)]^1$}
\RightLabel{K-distrib}
\UnaryInfC{$Kp \wedge K\neg Kp$}
\RightLabel{$\wedge$-elim}
\UnaryInfC{$K\neg Kp$}
\RightLabel{factivity}
\UnaryInfC{$\neg Kp$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$\neg(K(p \wedge \neg Kp))$}
\RightLabel{necessitation}
\UnaryInfC{$\Box\neg(K(p \wedge \neg Kp))$}
\AxiomC{p}
\AxiomC{$\neg Kp$}
\RightLabel{$\wedge$-int}
\BinaryInfC{$p \wedge \neg Kp$}
\RightLabel{possible knowledge}
\UnaryInfC{$\neg\Box\neg K(p \wedge \neg Kp)$}
\RightLabel{$\to$-elim}
\BinaryInfC{$\bot$}
\end{prooftree}
\tag{10.4}
$$

So we have derived a contradiction from p and $\neg Kp$.

Something has clearly gone wrong: for we *know* that there are unknown truths. So what has happened, and what was the point of this exercise?

The first thing to say is that we shouldn't think of the above as a failed attempt to use Logic to prove that every truth is known. That's not the kind of thing Logic does. The rôle of Logic here is to formalise our arguments and thereby focus our attention on features of those arguments that may be suspect. The enterprise throws up the following questions that one might like to ask:

(i) What does Kp mean? Does it mean: Everybody knows p? Someone knows p? Someone specific knows p? Most people know p? Were we justified in concealing the person who is doing the knowing? And what concept of knowledge is in play?

(ii) Are the two notions of necessity that are in play in the two branches of the proof one and the same notion? If they aren't we have a fallacy of equivocation.

(iii) Notice that we have used the classical concept of negation in assuming the duality of $\Box$ and $\Diamond$. Do we really have to use classical logic? After all, one normally expects any contradiction provable classically to be provably constructively.

(iv) One feature that should make you suspicious is that the contradiction concerns only the K operator, but we seem nevertheless to have had to use the modal operators to obtain the contradiction. The way the modal operators pop in, do their work and then disappear seems to violate the spirit of the interpolation lemma. Agreed, it may just be that interpolation doesn't hold for the logic with both K and $\Box$, but the matter needs looking into. Perhaps we could recast the paradox using, instead of the principle that *every truth is knowable*, rather the principle that *no truth is known to be false*—so that we just replace '$\Box$' by 'K' throughout.

10.3 Curry-Howard Unifies Two Riddles

Copy in aphrodisiac stuff from bradley.tex

Curry-Howard enables us to make a connection between two riddles familiar from the philosophical literature. The two riddles are Lewis Carroll's discussion "What the tortoise said to Achilles" in [7] and F.H. Bradley's infinite regress argument about predication.

10.3.1 What the Tortoise Said to Achilles

The Tortoise challenges Achilles to reach the end of a logical race-course that begins with a 'Hypothetical Proposition'. The race runs something like this: suppose that there are two formulæ A and B, and that we have proved A and $A \to B$; we want to infer B. Achilles is ready to race immediately to this conclusion, but the Tortoise objects that Achilles is being too hasty. The Tortoise professes unwillingness to obtain B from $\{A, A \to B\}$. He demands reassurance that this is legitimate, the sought reassurance being along the lines of a certificate that $(A \wedge (A \to B)) \to B$. This Achilles is happy to furnish, but the Tortoise now professes unwillingness to obtain B from $A \wedge (A \to B)$ and $(A \wedge (A \to B)) \to B$. He demands reassurance that this is legitimate, the sought reassurance being along the lines of a certificate that $(A \wedge (A \to B) \wedge (A \wedge (A \to B)) \to B) \to B$. This Achilles is happy to furnish, but ...

10.3.2 Bradley's regress

Bradley's riddle is to be found in the text

> Let us abstain from making the relation an attribute of the related, and let us make it more or less independent. "There is a relation C, in which A and B stand; and it appears with both of them." But here again we have made no progress. The relation C has been admitted different from A and B, and no longer is predicated of them. Something, however, seems to be said of this relation C; and said, again, of A and B. And this something is not to be the ascription of one to the other. If so, it would appear to be another relation, D, in which C, on one side, and, on the other side, A and B, stand. But such a makeshift leads at once to the infinite process. The new relation D can be predicated in no way of C, or of A and B; and hence we must have recourse to a fresh relation, E, which comes between D and whatever we had before. But this must lead to another, F; and so on, indefinitely.

F. H. Bradley: [5], p 27.[3]

Let me recast Bradley's argument in a form that is slightly more suitable for our purposes. We have a function f (unary, to keep things simple) and we are going to apply it to things. How are we to think of the result of applying f to an argument x? Presumably as f applied to x, so that we denote it '$f(x)$'. The regress is launched by the thought: should we not think of $f(x)$ as the result of applying the (binary) function `apply` to the pair of arguments f and x? And why stop there? Should we not be thinking of it as the result of applying the (binary) function `apply` to the pair of arguments `apply` and the pair f-and-x? And why stop there...!?

The thinker of the recurring thought "should we not be thinking of this object as `apply` applied to the two argument ...?" is of course the Tortoise in disguise. The Carroll regress is the the *proposition* version and the Bradley regress the *types* version

[3]Thanks to Paul Andrews for supplying the reference and the source code!

of some thing that can with the help of the Curry-Howard propositions-as-types insight be seen as *one* regress. The particular instance of the correspondence that concerns us is not $\vee$ or $\wedge$ but $\rightarrow$ and specifically the rule of $\rightarrow$-elimination.

So, when f is a function from A to B, are we to think of $f(x)$ as the result of applying f (which is of type $A \rightarrow B$) to x (which is of type A) so that we have the picture

$$\frac{A \qquad A \rightarrow B}{B} \rightarrow\text{-elim} \tag{10.5}$$

? Or are we to think of it as the result of applying `apply` (which is of type $((A \rightarrow B) \times A) \rightarrow B)$ to the pair $\langle f, a\rangle$ (which is of type $(A \rightarrow B) \times A)$ so that we have the picture

$$\frac{(A \rightarrow B) \times A \qquad ((A \rightarrow B) \times A) \rightarrow B}{B} \rightarrow\text{-elim} \tag{10.6}$$

? Or are we to think of it as the result of applying `apply` (which is of type $((((A \rightarrow B) \times A) \rightarrow B) \times ((A \rightarrow B) \times A)) \rightarrow B)$ to the pair $\langle \texttt{apply}, \langle f, a\rangle\rangle$ (which is of type $((A \rightarrow B) \times A)) \rightarrow B) \times ((A \rightarrow B) \times A))$ so that we have the picture

$$\frac{((A \rightarrow B) \times A)) \rightarrow B) \times ((A \rightarrow B) \times A) \qquad ((((A \rightarrow B) \times A) \rightarrow B) \times ((A \rightarrow B) \times A)) \rightarrow B}{B} \rightarrow\text{-elim} \tag{10.7}$$

Where will it all end?!

10.4 The Paradoxes

I have apologized to my readers—several times—(e.g. section 3.11)—for inflicting on them all this apparatus of proof systems despite the fact that it is usually easier to check for validity of a formula by inspecting a truth-table than it is to find a proof. Even in predicate calculus (where there is no straightforward analogue of truth-tables) it seems easier to check validity by inspection than by looking for a proof. In this section we are going to see some natural deduction and sequent rules for a subject matter in a setting where the proof-theoretical gadgetry is genuinely illuminating.

Let us consider Russell's paradox of the set of those sets that are not members of themselves. Let us use the notation '$\{x : \phi(x)\}$' for the set of all x that are ϕ. We write '$x \in y$' to mean that x is a member of y. Since $\{x : \phi(x)\}$ is the set of all things that are ϕ we want to have

$$\phi(a) \longleftrightarrow a \in \{x : \phi(x)\} \tag{10.8}$$

This gives us two natural deduction rules

$$\frac{\phi(a)}{a \in \{x : \phi(x)\}} \in\text{-int}; \qquad\qquad \frac{a \in \{x : \phi(x)\}}{\phi(a)} \in\text{-elim}$$

(Do not confuse these with the rules for ϵ-terms from page 136!! '$\in$' is not a typographical variant of 'ϵ'!)

Let us now use this to analyse Russell's paradox of the set of all sets that are not members of themselves. We can, as before, write '$\neg(x \in x)$' instead of '$x \in x \to \bot$', but to save yet more space we will instead write '$x \notin x$'.

The following is a proof that $\{x : x \in x \to \bot\}$ is not a member of itself.

$$\dfrac{\dfrac{\dfrac{[(\{x : x \notin x\} \in \{x : x \notin x\})]^1}{(\{x : x \notin x\} \in \{x : x \notin x\}) \to \bot}\ \in\text{-elim} \qquad [\{x : x \notin x\} \in \{x : x \notin x\}]^1}{\bot}\ \to\text{-elim}}{\{x : x \notin x\} \in \{x : x \notin x\}) \to \bot}\ \to\text{-int (1)} \tag{10.9}$$

Clearly space is going to be a problem, so let's abbreviate '$\{x : x \notin x\}$' to 'R' (for Russell).

$$\dfrac{\dfrac{\dfrac{[R \in R]^1}{(R \in R) \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{(R \in R) \to \bot}\ \to\text{-int (1)} \tag{10.10}$$

But we can extend this proof by one line to get a proof that $\{x : x \notin x\}$ *is* a member of itself after all!

$$\dfrac{\dfrac{\dfrac{\dfrac{[R \in R]^1}{(R \in R) \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{(R \in R) \to \bot}\ \to\text{-int (1)}}{R \in R}\ \in\text{-int} \tag{10.11}$$

. . . and put these two proofs together to obtain a proof of a contradiction

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[R \in R]^1}{(R \in R) \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{(R \in R) \to \bot}\ \to\text{-int (1)}}{R \in R}\ \in\text{-int} \qquad \dfrac{\dfrac{\dfrac{[R \in R]^1}{(R \in R) \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{(R \in R) \to \bot}\ \to\text{-int (1)}}{\bot}\ \to\text{-elim} \tag{10.12}$$

Must say something about how this ties in with the proof of the nonexistence of the Russell class in exercise 55.

Chapter 11

Appendices

11.1 Notes to Chapter one

11.1.1 The Material Conditional

Lots of students dislike the material conditional as an account of implication. The usual cause of this unease is that in some cases a material conditional $p \to q$ evaluates to `true` for what seem to them to be spurious and thoroughly unsatisfactory reasons, namely: that p is false, or that q is true. How can q follow from p merely because q happens to be true? The meaning of p might have no bearing on q whatever! Standard illustrations in the literature include

> If Julius Cæsar is Emperor then sea water is salt.

need a few more examples

These example seem odd because we feel that to decide whether or not p implies q we need to know a lot more than the truth-values of p and q.

This unease shows that we have forgotten that we were supposed to be examining a relation between *extensions*, and have carelessly returned to our original endeavour of trying to understand implication between *intensions*. $\wedge$ and $\vee$, too, are relations between intensions but they also make sense applied to extensions.[1] Now if p implies q, what does this tell us about what p and q evaluate to? Well, at the very least, it tells us that p cannot evaluate to `true` when q evaluates to `false`. That is to say that we require—at the very least— that the *extension* corresponding to a conditional should satisfy *modus ponens*.

How many extensions are there that satisfy *modus ponens*? For a connective C to satisfy *modus ponens* it suffices that in each of the two rows of the truth table for C where p is true, if $p\,C\,q$ is true in that row then q is true too.

[1] should say something here about how $\vee$ and $\wedge$ commute with evaluation but that conditionals don't . . . think along those lines

p	C	q
1	?	1
0	?	1
1	0	0
0	?	0

We cannot make $p\,C\,q$ true in the third row, because that would cause C to disobey *modus ponens*, but it doesn't matter what we put in the centre column in the three other rows. This leaves eight possibilities:

$$(1):\ \frac{p \quad q}{q} \qquad (2):\ \frac{p \quad p \longleftrightarrow q}{q} \qquad (3):\ \frac{p \quad \neg p}{q} \qquad (4):\ \frac{p \quad p \to q}{q}$$

p	C^1	q
1	1	1
1	0	0
0	1	1
0	0	0

p	C^2	q
1	1	1
1	0	0
0	0	1
0	1	0

p	C^3	q
1	0	1
1	0	0
0	1	1
0	1	0

p	C^4	q
1	1	1
1	0	0
0	1	1
0	1	0

$$(5):\ \frac{p \quad \bot}{q} \qquad (6):\ \frac{p \quad p \wedge q}{q} \qquad (7):\ \frac{p \quad \neg p \wedge q}{q} \qquad (8):\ \frac{p \quad \neg p \wedge \neg q}{q}$$

p	C^5	q
1	0	1
1	0	0
0	0	1
0	0	0

p	C^6	q
1	1	1
1	0	0
0	0	1
0	0	0

p	C^7	q
1	0	1
1	0	0
0	1	1
0	0	0

p	C^8	q
1	0	1
1	0	0
0	0	1
0	1	0

obtained by replacing the major premiss '$p \to q$' in the rule of *modus ponens* by each of the eight extensional binary connectives that satisfy the rule.

(1) will never tell us anything we didn't know before; we can never use (5) because its major premiss is never true; (6) is a poor substitute for the rule of $\wedge$-elimination; (3), (7) and (8) we will never be able to use if our premisses are consistent.

(2), (4) and (6) are the only sensible rules left. (2) is not what we are after because it is symmetrical in p and q whereas "if p then q" is not. The advantage of (4) is that you can use it whenever you can use (2) or (6). So it's more use!

We had better check that we do not get into trouble with this policy of adopting (4), and evaluating $p \to q$ to `true` unless there is a very good reason not to. Fortunately, in cases where the conditional is evaluated to `true` *merely* for spurious reasons, then no harm can be done by accepting that evaluation. For consider: if it is evaluated to `true` *merely* because p evaluates to `false`, then we are never going to be able to invoke it (as a major premiss at least), and if it is evaluated to `true` *merely* because q evaluates to `true`, then if we invoke it as a major premiss, the only thing we can conclude—namely q—is something we knew anyway.

This last paragraph is not intended to be a *justification* of our policy of using only the material conditional: it is merely intended to make it look less unnatural than it otherwise might. The astute reader who spotted that nothing was said there about conditionals as *minor* premisses should not complain. They may wish to ponder the reason for this omission.

11.1.2 The Definition of Valid Argument

There is a difficulty with definition 5, namely that there could be an argument type so vast that no token of it small enough to fit into the universe has true premisses and a false conclusion. Such an argument type would be waved on as valid when—really—it shouldn't. If you are a purist of a certain stamp you might lose sleep over it, feeling that altho' definition 5 captures the *ex*tension of 'valid' it doesn't capture its *in*tension. There isn't really much practical point to this worry: after all, *ex hypothesi* the universe isn't big enough to contain the scenario that you were worrying might be a counterexample to definition 5. As soon as the universe becomes big enough to contain a hitherto convincing counterexample the definition of valid changes to accommodate it.

An example could help

The alternative to the admittedly rather brutal strategy of definition 5 would be to give a definition in terms of *possible* argument-tokens. The trouble with this approach is that it tries to explain something apparently straightforward (namely *valid argument*) in terms of something metaphysically much more problematic—*possibilia*. And what is to be gained? Not much. At all events, if Tweedledum goes for definition 5 and Tweedledee goes for a definition involving possibilia, they will still never find an instance to disagree on. "If the Good Lord had meant us to worry, Mr Fawlty, he would have given us something to worry about!" Tweedledum and Tweedledee will instead have to fight over their rattle.

11.1.3 Valuation and Evaluation

see also subsection 8.1.

Needs to be drastically rewritten

Bog standard first-year propositional logic does not examine *evaluation* at all. We are aware in an abstract sort of way that if we know what a valuation does to propositional letters (atomic formulæ then we know what it does to molecular formulæ but we pay no attention to the mechanics of how these truth-values are actually calculated ...

This is a perfectly sensible thing to do because all the different ways of computing the truth-value give the same answer. [prove it by induction on composition of formulæ]. So of course it's safe to hide the calculation ("Hide anything that it is safe to hide")

explain: eager vs lazy

see section 3.14. This needs to be properly rewritten

Suppose we allow valuations that are merely *partial* functions.) What is the truth-value given by v to $A \vee B$ when A is not given a truth-value by v? Do we want to say that it, too, is undecided, on the grounds that indecision about a subformula is enough to contaminate the whole formula? Or do we rather say "Well, it's true if the truth-value of B is true and undecided otherwise".

If we are evaluating eagerly then one does not give a value to $A \vee B$ until one has given values both to A and to B. Thus, when evaluating $A \vee B$ eagerly, we cannot give it a value without having a value for A. So *even if we know that B has evaluated to true* we cannot give a value to $A \vee B$. On the other hand, if we are evaluating lazily, once we know that B is true we are allowed to infer that $A \vee B$ must have the value true. Our choice of policy concerning this question is crucial. If we rule that $A \vee B$ is to be true [under a valuation] as long as A is true under that valuation (never mind B) then eager evaluation and lazy evaluation no longer reliably give the same answer, so one cannot regard evaluation as something that takes place instantaneously in a black box somewhere: it is no longer something one can *hide* in the way we could when we were using complete ("total") valuations.

Reflecting on this can banish some of the puzzlement that people feel when they incautiously think that they can treat 'undecided' as a third truth-value.

One's intuitions about what $v(A \vee B)$ should be when one or both of $v(A)$ and $v(B)$ is intermediate seems to depend to a certain extent on whether our evaluation strategy is eager or lazy.

Suppose we kid ourselves that there should be three truth-values: true, false and undecided.

$$A \vee (B \wedge (C \vee (D \wedge \ldots))))$$

If we really want to get this straight we have to have a binary relation "undecided at time t"

H I A T U S

Computer science has been a wonderfully fertilising influence on modern philosophy. Not only has it brought new ideas to the subject, but it has breathed new life into old ones. A striking example is the way in which Computer Science's concern with evaluation and strategies (lazy, eager and so on) for evaluation has made the intension/extension distinction nowadays almost more familiar to computer scientists than to philosophers. Intensions evaluate to extensions. In the old, early-twentieth century logic, evaluation just happened, and the subject was concerned with that part of metaphysics that was unaffected by how evaluation was carried out. For example, the completeness theorem for propositional logic says that a formula is derivable iff it is true under all valuations: the internal dynamic of valuations is not analysed or even considered. Modern semantics for programming languages has a vast amount to say about the actual dynamics of evaluation *as a process*. The old static semantics gave a broad and fundamental picture, but was unsuited for the correct analysis of certain insights that happened to appear at that time. A good example of an insight whose proper unravelling was hampered by this lack of a dynamic perspective is Popper's idea of falsifiability. Let us examine a natural setting for the intuition that gave rise to it.

Do we inseert here stuff from lazy.tex Hereditary Hæmorrhagic telangiectasia?

As well as thinking a bit about the evaluation strategy one is to use (the order in which one is to attempt to evaluate subformulæ one can also take some thought about the order in which one asks the given valuation for truth-values of the atomics that go into the complex propositions we are going to evaluate.

Let us suppose that, in order to be confirmed as a widget, an object x has to pass a number of independent tests, all of similar cost. If investigator $\mathcal{I}$ wants to test whether a candidate x is a widget or not, $\mathcal{I}$ subjects it to these tests, all of which it has to pass. Which test does $\mathcal{I}$ run first? Obviously the one that is most likely to fail! It will of course be said that this is so that if x passes it then the theory T (that x is a widget) is more strongly confirmed than it would have been if it had passed an easy one. Indeed I have heard Popperians say precisely this.

It seems to me that although this may be true, it does not go to the heart of the insight vouchsafed to Popper. This traditional account concerns merely the theory that is being confirmed, and not any of $\mathcal{I}$'s other preoccupations. By taking into account a more comprehensive decription of $\mathcal{I}$ we can give a more satisfactory account of this intuition. Specifically it is helpful to bear in mind the cost to $\mathcal{I}$ of these tests. Suppose candidate x has to pass two tests T_1 and T_2 to be confirmed as a widget, and the costs-to-$\mathcal{I}$ of the two tests are similar. Suppose also that most candidates fail T_1 but most pass T_2. What is $\mathcal{I}$ to do? Obviously $\mathcal{I}$ can minimise his expected expenses of investigation by doing T_1 *first*. It is of course true that if x is indeed a widget then by the time it has passed both tests, $\mathcal{I}$ will have inevitably have incurred the costs of running both T_1 and T_2. But a policy of doing T_1 first rather than doing T_2 first will in the long run save $\mathcal{I}$ resources because of the cases where x is not a widget.

Notice that this point of view has something to say also about the situation dual to the one we have just considered, in which the investigator $\mathcal{I}$ has a number of tests and a candidate x can be shown to be a widget by passing even *one* of them. In this situation the dual analysis tells $\mathcal{I}$ that the best thing to do in order to minimise the expected cost of proving x to be a widget is to try first the test most likely to *succeed*. Although this is logically parallel ("dual") to the situation we have just considered, the traditional Popperian analysis has nothing to say about it at all. This is surely a warning sign.

This is not to say that Popper's insight is not important: it clearly is. The claim is rather that it has not been received properly. Properly understood it is not straightforwardly a piece of metaphysics concerning verification and support, but a superficially more mundane fact about strategies for minimising costs for agents in an uncertain world.

One is reminded of the story of the drunkard who has dropped his house keys and is looking for them under the street light. He didn't drop them exactly under the streetlight but unless they are dropped reasonably near the street light he has no hope of finding them anyway. So he looks for them under the street light.

11.2 Notes to Chapter 3

11.2.1 Hypothetical Reasoning: an Illustration

The late Lance Endersbee's [15] is an intruiging book about the origin of groundwater, and promotes the thesis that much groundwater—even in artesian basins—is plutonic in origin, and is *not* (as is popularly supposed) water that fell as rain and became trapped between impervious layers of rock. Endersbee is worth quoting *in extenso*...

> "A key feature of the [Queensland Government's] report is the use of radioactive isotope ratios to indicate age of the groundwater. The assessment of age of groundwater is based on the knowledge of these isotope ratios in

rainfall and surface waters, and the change of these ratios, over time, when the water is no longer exposed to the atmosphere. It is a technique that is used worldwide.

The logical steps in this method of assessment of age of groundwater are as follows:

(a) the isotope ratios in rainwater are known.

(b) It is assumed that the groundwater was originally derived from rainfall; and that the groundwater once had the isotope ratios of rainwater.

(c) the decay times of the two marker isotopes are known, and thus the change in these ratios over time provides an estimate of age since the groundwater was originally rainwater.

(d) the age of the groundwater is estimated from the difference of the isotope ratios of the groundwater and rainwater. In the case of the Great Artesian Basin, the estimate of age may be up to 2 million years.

Note the circular argument. The procedure is directed towards proving the assumption that the groundwater was originally rainfall, but that fact is not recognised. The procedure specifically excludes the possibility that the groundwater was never rainfall. Unfortunately, it is normal for groundwater hydrologists to be quite unaware of the assumptions involved. From their perspective, a date determined by nuclear physics must be right, and they thereby manage to prove that all groundwater is derived from surface rainfall."

[15] pp 51–2.

It is not straightforward to reconstruct the author's thought-processes, but it does seem to be a safe bet that he doesn't see the discussions he alludes to as establishing the proof of a *conditional* along the lines of

> "If the groundwater in the Great Artesian basin was ever rainwater, that was at least 2×10^6 years ago."

And this conditional is entirely consistent with his thesis that the groundwater was never rainfall.

This is a salutory lesson. Endersbee was no fool, and if even he can make errors of logic of this magnitude then we all of us need to be on our guard.

11.2.2 $\vee$-elimination and the *ex falso*

What happens with $\vee$-elimination if the set of proofs (and therefore also the set of assumptions) is empty? That would be a rule that accepted as input an empty list of proofs of C, and an empty disjunction of assumptions (recall from section 2.4.2 that the empty disjunction is the **false**). This is just the rule of *ex falso sequitur quodlibet.*

If you are a third-year pedant you might complain that all instances of $\vee$-elimination have two inputs, a disjunction and a list of proofs; *ex falso sequitur quodlibet* in contrast only has one: only the empty disjunction. So it clearly isn't a special case of $\vee$-elimination. However if you want to get A rather than B as the output of an instance of $\vee$-elimination with the empty disjunction as input then you need as your other input the empty list of proofs of A, rather than the empty list of proofs of B. So you are right, there *is* something fishy going on: the rule of *ex falso sequitur quodlibet* strictly has two inputs: (i) the empty disjunction and (ii) the empty list of proofs of A. It's a bit worrying that the empty list of proofs of A seems to be the same thing as the empty list of proofs of B. If you want to think of the *ex falso sequitur quodlibet* as a thing with only one input then, if you feed it the `false` and press the start button, you can't predict which proposition it will give you a proof of! It's a sort of nondeterministic engine. This may or may not matter, depending on how you conceptualise proofs. This is something that will be sorted out when we reconceptualise proof theory properly if we ever do. We will think about this a bit in section 4.18. For the moment just join the first and second years in not thinking about it at all.

"Will think"? We are already in an appendix!

Of course you can also think of the `false` as the *con*junction of *all* propositions (instead of as the disjunction of none of them). In that case you will believe the *ex falso* because it is a special case of $\wedge$-elimination.

11.2.3 Negation in Sequent Calculus

It is possible to continue thinking of $\neg P$ as $P \to \bot$. In a sequent calculus context this means we have to think of any sequent as having an unlimited supply of '$\bot$'s on the right. (Miniexercise: why is this all right?) Infinitely many? Not necessarily: it will suffice to have an indeterminate finite number of them. Or perhaps a magic pudding (see [26]) of '$\bot$'s: something that emits a $\bot$ whenever you ask it nicely. In these circumstances the $\neg$-R rule simply becomes a special case of $\to$-R. Considerations of this kind are an essential input into any discussion that aims to determine precisely what sort of data object the right-hand-side of a sequent is: `list`, `set`, `multiset`, `stream`, `magic-pudding` ...

Similarly we can think of the rule of $\neg$-L as a $\to$-L:

$$\frac{\Delta \vdash \Gamma, A \qquad \bot, \Delta \vdash \Gamma}{\Delta, \neg A \vdash \Gamma} \to L \tag{11.1}$$

We can think of the sequent $\bot, \Delta \vdash \Gamma$ as an initial sequent because of the cryptic '$\bot$' on the right, so there is a '$\bot$' on both sides.

11.2.4 What is the right way to conceptualise sequents?

From time to time people have felt that part of the the job of philosophy is to find the right way to think about—to *conceptualise*—certain phenomena... to *carve nature at the joints* to use Plato's imagery (one assumes he was no vegetarian).

In chapter 3 I equivocated as much as I decently could between the various ways of conceptualising sequents. Here are some of the issues[2]

- Are the two parts (left and right) of the sequent to be sets, or multisets, or lists?

  > If they are multisets or lists we need contraction;
  >
  > If they are lists we need exchange;
  >
  > If they are sets we have to specify the eigenformula. This suggests we should be using lists, so that the eigenformula is always the first (or always the last) formula.

To be continued

  Edmund writes:

  > the good answer is that it all comes down to the kinds of distinctions you're trying to make. The CS answer is that prop = type, proof = term and what you are doing is a kind of type theory. The things on either side are contexts.
  >
  > This gives you an answer in which you can make all the distinctions you want, encoding just about all the rules.
  >
  > Then you can "forget" bits of that structure. The key bit being a theorem that says the logical rules are properly supported on the quotient structure. Sets are OK from a certain perspective, but you'd have problems in detail co's you might not be able to lift a set-based proof back to an arbitrary term-based one.

- other issues?

11.3 Notes to Chapter 4

11.3.1 Subtleties in the definition of first-order language

The following formula looks like a first-order sentence that says there are at least n distinct things in the universe. (Remember the $\bigvee$ symbol from page 40.)

$$(\exists x_1 \ldots x_n)(\forall y)(\bigvee_{i \leq n} y = x_i) \tag{11.2}$$

But if you are the kind of pedant that does well in Logic you will notice that 11.2 isn't a formula of the first-order logic we have just seen because there are variables (the subscripts) ranging over variables! If you put in a concrete actual number for n then what you have is an *abbreviation* of a formula of our first-order language. Thus

$$(\exists x_1 \ldots x_3)(\forall y)(\bigvee_{i \leq 3} y = x_i) \tag{11.3}$$

is an abbreviation of

[2] And, yes, i do mean issues, not problems.

$$(\exists x_1 x_2 x_3)(\forall y)(y = x_1 \vee y = x_2 \vee y = x_3) \tag{11.4}$$

(Notice that formula 11.2 isn't actually *second*-order either, because the dodgy variables are not ranging over subsets of the domain.)

11.3.2 Failure of Completeness of Second-order Logic

Second-order arithmetic includes as one of its axioms the following:

Second-order induction axiom:

$$(\forall F)([F(0) \wedge (\forall n)(F(n) \to F(n+1))] \to \forall n)(F(n)))$$

And we can give a second-order definition of what it is for a natural number to be standard:

DEFINITION 41
standard(n) $\longleftrightarrow_{df}$ $(\forall F)([F(0) \wedge (\forall m)(F(m) \to F(m+1))] \to F(n))$

This axiom enables us to prove—in second-order arithmetic—that every natural number is standard: simply take '$F(n)$' to be 'standard(n)'.

Another thing we can prove by induction is the following:

if n is a natural number then, for any model $\mathfrak{M}$ of arithmetic, there is a unique embedding from $[0, 1, \ldots n]$ into an initial segment of $\mathfrak{M}$.

This isn't really *second*-order; it's much worse…

It requires a little bit of work to show that the converse is true, but it is.

One consequence of this is that second-order arithmetic is what they call **categorical**: it is a theory with only one model. We exploit this fact here. Add to the language of second-order arithmetic the constant symbol 'a', and infinitely many axioms $a \neq 0$, $a \neq 1$, $a \neq 2$, $a \neq 3$ …. This theory is now consistent, since no contradiction can be deduced from it in finitely many steps, but it has no models.

11.4 Notes to Chapter 5

If my favourite theorem-prover says $\vdash 0 = 1$ then i know it has made a mistake, even if i can't find it!

Another—more striking—example is Prior's Cretan: "Everything I say is false". It is clear that he must have said something else. For suppose that were the only thing he had said. Then we would have the liar paradox, since "Everything I say is false" is equivalent to "what I am now saying is false" if that is the only thing the speaker says. Yet we cannot determine what else he has said!

This is more striking, but it is a less satisfactory example, since it relies on self-reference, which is fraught with problems. Those problems have nothing to do with nonconstructivity, so it is best not to use an example that drags them in.

HOLE A very good example is the pigeonhole principle. If you have more pigeons then pigeonholes then it can't be the case that every pigeon has a pigeonhole all of its own: at least one hole must have at least two pigeons. But it doesn't tell you which hole, or which pigeons (even if all the pigeons and all the holes have names).

11.4.1 Church on Intension and Extension

> *"The foregoing discussion leaves it undetermined under what circumstances two functions shall be considered the same. The most immediate and, from some points of view, the best way to settle this question is to specify that two functions f and g are the same if they have the same range of arguments and, for every element a that belongs to this range, $f(a)$ is the same as $g(a)$. When this is done we shall say that we are dealing with functions in extension.*
>
> *It is possible, however, to allow two functions to be different on the ground that the rule of correspondence is different in meaning in the two cases although always yielding the same result when applied to any particular argument. When this is done we shall say that we are dealing with functions in intension. The notion of difference in meaning between two rules of correspondence is a vague one, but, in terms of some system of notation, it can be made exact in various ways. We shall not attempt to decide what is the true notion of difference in meaning but shall speak of functions in intension in any case where a more severe criterion of identity is adopted than for functions in extension. There is thus not one notion of function in intension, but many notions; involving various degrees of intensionality".*

Church [10]. p 2.

The intension-extension distinction has proved particularly useful in computer science—specifically in the theory of computable functions, since the distinction between a *program* and the *graph* of a function corresponds neatly to the difference between a function-in-intension and a function-in-extension. Computer Science provides us with perhaps the best-motivated modern illustration. A piece of code that needs to call another function can do it in either of two ways. If the function being called is going to be called often, on a restricted range of arguments, and is hard to compute, then the obvious thing to do is compute the set of values in advance and store them in a look-up table in line in the code. On the other hand if the function to be called is not going to be called very often, and the set of arguments on which it is to be called cannot be determined in advance, and if there is an easy algorithm available to compute it, then the obvious strategy is to write code for that algorithm and call it when needed. In the first case the embedded subordinate function is represented as a function-in-extension, and in the second case as a function-in-intension.

The concept of algorithm seems to be more intensional than function-in-extension but not as intensional as function-in-intension. Different programs can instantiate the same algorithm, and there can be more than one algorithm for computing a function-in-extension. Not clear what the identity criteria for algorithms are. Indeed it has been argued that there can be no satisfactory concept of algorithm (see [2]). This is particularly unfortunate because of the weight the concept of algorithm is made to bear in some philosophies of mind (or some parodies of philosophy-of-mind ["strong AI"] such as are to be found in [31]).[3]

11.5 Notes to Chapter 8

This doesn't mean that it's a mistake to add modal operators, but it does mean that you want them to sit on top of a classical logic.

[3]Perhaps that is why is is made to carry that weight! If your sights are set not on devising a true philosophical theory, but are set merely on cobbling together a philosophical theory that will be hard to refute then a good strategy is to have as a keystone concept one that is so vague that any attack on the theory can be repelled by a fallacy of equivocation. The unclarity in the key concept ensures that the target presented to aspiring refuters is a fuzzy one, so that no refutation is ever conclusive. This is why squids have ink.

Chapter 12

Answers to some Exercises

Exercises from Chapter 2

Exercise 14

Each time you add a new propositional letter you double the number of possible combinations.

Exercise 21, part 10

Here are the basic propositions and the letters we are going to abbreviate them to.

God exists	E
God is omnipotent	P
God is omniscient	O
God is benevolent	B
God can prevent Evil	D
God knows that Evil exists	K
God prevents Evil	J
Evil exists	V

If God exists then He is omnipotent.	$E \to P$	(1)
If God exists then He is omniscient.	$E \to O$	(2)
If God exists then He is benevolent.	$E \to B$	(3)
If God can prevent Evil then—if He knows that Evil exists—then He is not benevolent if He does not prevent it.	$D \to (K \to (\neg J \to \neg B))$	(4)
If God is omnipotent, He can prevent Evil.	$P \to D$	(5)

If God is omniscient then He knows that Evil exists if it does indeed exist.	$O \to (V \to K)$	(6)
Evil does not exist if God prevents it.	$J \to \neg V$	(7)
Evil exists.	V	(8)

We want to persuade ourselves that God does not exist. Well, suppose he does. Let's deduce a contradiction

Assume E. Then (1), (2) and (3) give us

$$P \qquad (9),$$

$$O \qquad (10)$$

and

$$B \qquad (11)$$

Now that we know O, (7) tells us that

$$V \to K \qquad (12)$$

But we know V (that was (8)) so we know

$$K \qquad (13)$$

We know P, so (5) tells us that

$$D \qquad (14)$$

We can feed D into (4) and infer

$$K \to (\neg J \to \neg B) \qquad (15)$$

But we know K (that was line 13) so we get

$$\neg J \to \neg B \qquad (16)$$

(8) and (7) together tell us $\neg J$, so we get $\neg B$. But we got B at line 11.

Exercises from Chapter 3

Exercise 28, part 16

$((A \vee B) \wedge (A \vee C)) \to (A \vee (B \wedge C))$; *hard!*

Here is a proof:

$$
\dfrac{
\dfrac{
\dfrac{
\dfrac{\dfrac{E \qquad E \to P}{P}\,\to\text{-elim} \qquad P \to D}{D}\,\to\text{-elim} \qquad D \to (K \to (B \to J))
}{K \to (B \to J)}\,\to\text{-elim}
\qquad
\dfrac{\dfrac{\dfrac{E \qquad E \to O}{O}\,\to\text{-elim} \qquad O \to (V \to K)}{V \to K}\,\to\text{-elim} \qquad V}{K}\,\to\text{-elim}
}{B \to J}\,\to\text{-elim}
\qquad
\dfrac{E \qquad E \to B}{B}\,\to\text{-elim}
}{J}\,\to\text{-elim}
\qquad
\dfrac{V \qquad V \to \neg J}{\neg J}\,\to\text{-elim}
}{\bot}\,\to\text{-elim}
\tag{12.1}
$$

$$\dfrac{\dfrac{\dfrac{\dfrac{[(A\vee B)\wedge(A\vee C)]^1}{A\vee B}\,\wedge\text{-elim}\qquad \dfrac{[A]^2}{A\vee(B\wedge C)}\,\vee\text{-int}\qquad \dfrac{\dfrac{[B]^2\qquad [C]^3}{B\wedge C}\,\wedge\text{-int}}{A\vee(B\wedge C)}\,\vee\text{-int}}{A\vee(B\wedge C)}\,\vee\text{-elim (2)}\qquad \dfrac{[A]^3}{A\vee(B\wedge C)}\,\vee\text{-int}\qquad \dfrac{[(A\vee B)\wedge(A\vee C)]^1}{A\vee C}\,\wedge\text{-elim}}{A\vee(B\wedge C)}\,\vee\text{-elim (3)}}{((A\vee B)\wedge(A\vee C))\to(A\vee(B\wedge C))}\,\to\text{-int (1)} \tag{12.2}$$

Why is this exercise hard? The point is that in this proof the two conjuncts in the antecedent—namely $A \vee B$ and $A \vee C$—play *differing* rôles in the proof, despite the fact that their two contributions to the truth of the consequent seems to be the same. This last fact means that one naturally starts by looking for a proof wherein these two conjuncts are symmetrically placed. Sadly there is no such proof. Instead we have *two distinct* proofs, where each is obtained from the other by permuting the two conjuncts. Here is the other one:

$$\dfrac{\dfrac{\dfrac{\dfrac{[(A\vee B)\wedge(A\vee C)]^1}{A\vee C}\,\wedge\text{-elim}\qquad \dfrac{[A]^2}{A\vee(C\wedge B)}\,\vee\text{-int}\qquad \dfrac{\dfrac{[C]^2\qquad [B]^3}{C\wedge B}\,\wedge\text{-int}}{A\vee(B\wedge C)}\,\vee\text{-int}}{A\vee(C\wedge B)}\,\vee\text{-elim (2)}\qquad \dfrac{[A]^3}{A\vee(B\wedge C)}\,\vee\text{-int}\qquad \dfrac{[(A\vee B)\wedge(A\vee C)]^1}{A\vee B}\,\wedge\text{-elim}}{A\vee(C\wedge B)}\,\vee\text{-elim (3)}}{((A\vee B)\wedge(A\vee C))\to(A\vee(B\wedge C))}\,\to\text{-int (1)} \tag{12.3}$$

work on this

I think that the sequent calculus proof *is* symmetrical. Supply it here.

Exercise 39

Go back to Zarg (exercise 32 p. 68) and—using the truth-table for $\neg$ that you decided that the Zarglings use—check that the Zarglings do not believe axiom T to be a tautology.

T fails when $[[A]] = 3$ and $[[B]] = 1$.
The truth-table for S has $3^3 = 27$ rows.

Exercise 41

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

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

The interpolant can contain only the letters 'B' and 'C'. It's pretty clear that the antecedent implies $B \vee C$ so might that be what we are looking for? Yes!

Exercises from Chapter 4

Exercise 45

In my model answers I have tended to use bits of English text in `verbatim` font (as is the habit in certain computer science cultures) for predicate letters, rather than use the single letters that are more customary in most logical cultures. I have done this merely in order to make the notation more suggestive: there is no cultural significance to it. And in any case, further down in the list of model answers I have reverted to the philosophico-logical standard practice of using single capital Roman letters.

This first bunch involve monadic predicates only and no nested quantifiers.

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

   Lexicon:
   Unary predicate letters: `good-boy`( ); `deserves-favour`( )
   Constant symbol: `George`

   Formalisation

   $$\begin{array}{l}(\forall x)(\texttt{good-boy}(x) \to \texttt{deserves-favour}(x));\\ \underline{\texttt{good-boy(George)};}\\ \texttt{deserves} - \texttt{favour}(\texttt{George})\end{array}$$

   You might prefer to have two unary predicate letters `good`( ) and `boy`( ), in which case you would have

   $$\begin{array}{l}(\forall x)((\texttt{good}(x) \wedge \texttt{boy}(x)) \to \texttt{deserves-favour}(x));\\ \underline{\texttt{good}(\texttt{George}) \wedge \texttt{boy}(\texttt{George}));}\\ \texttt{deserves} - \texttt{favour}(\texttt{George}).\end{array}$$

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

   Lexicon:
   Unary predicate letters: `eats-grass`( ), `Cow`( );
   Constant symbol: `Daisy`.

   Formalisation

   $$\begin{array}{l}(\forall x)(\texttt{Cow}(x) \to \texttt{eats} - \texttt{grass}(x))\\ \underline{\texttt{eats} -} \texttt{grass}(\texttt{Daisy});\\ (\texttt{Daisy}).\end{array}$$

3. Socrates is a man; all men are mortal. Therefore Socrates is mortal.

   Lexicon:
   Unary predicate letters: `man`( ), `mortal`( ),
   Constant symbol: `Socrates`.

   Formalisation

$$\begin{array}{l} \texttt{man}(\texttt{Socrates}); \\ \underline{(\forall x)(\texttt{man}(x) \to \texttt{mortal}(x));} \\ \texttt{mortal}(\texttt{Socrates}). \end{array}$$

4. Daisy is a cow; all cows eat grass. Therefore Daisy eats grass.

   Lexicon:
   Unary predicate letters: `eats-grass`( ), `cow`( );
   Constant symbol: `Daisy`.

   Formalisation

$$\begin{array}{l} \texttt{cow}(\texttt{Daisy}); \\ \underline{(\forall x)(\texttt{cow}(x) \to \texttt{eats} - \texttt{grass}(x));} \\ \texttt{eats} - \texttt{grass}(\texttt{Daisy}). \end{array}$$

5. Daisy is a cow; all cows are mad. Therefore Daisy is mad.

   Lexicon:
   Unary predicate letters: `mad`( ), `cow`( );
   Constant symbol: `Daisy`.

   Formalisation

$$\begin{array}{l} \texttt{cow}(\texttt{Daisy}); \\ \underline{(\forall x)(\texttt{cow}(x) \to \texttt{mad}(x));} \\ \texttt{mad}(\texttt{Daisy}). \end{array}$$

6. No thieves are honest; some dishonest people are found out. Therefore some thieves are found out.

   Lexicon:
   Unary predicate letters: `thief`( ), `honest`( ), `found-out`( ).

   Formalisation

$$\begin{array}{l} (\forall x)(\texttt{thief}(x) \to \neg(\texttt{honest}(x)); \\ \underline{(\exists x)(\neg\texttt{honest}(x) \land \texttt{found-out}(x));} \\ (\exists x)(\texttt{thief}(x) \land \texttt{found} - \texttt{out}(x)). \end{array}$$

7. No muffins are wholesome; all puffy food is unwholesome. Therefore all muffins are puffy.

   Lexicon:
   Unary predicate letters: `muffin`( ), `wholesome`( ), `puffy`( ).

   Formalisation

$$\begin{array}{l} \neg(\exists x)(\texttt{muffin}(x) \land \texttt{wholesome}(x)); \\ \underline{(\forall x)(\texttt{puffy}(x) \to \neg(\texttt{wholesome}(x)));} \\ (\forall x)(\texttt{muffin}(x) \to \texttt{puffy}(x)). \end{array}$$

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.

   Lexicon:
   Unary predicate letters: `peacock`( ), `can-sing`( ), `proud-of-tail`( ).

   Formalisation

$$\frac{\begin{array}{l}(\forall x)(\texttt{proud-of-tail}(x) \to \texttt{peacock}(x));\\ (\exists x)(\texttt{proud-of-tail}(x) \wedge \neg(\texttt{can-sing}(x)));\end{array}}{(\exists x)(\texttt{peacock}(x) \wedge \neg(\texttt{can} - \texttt{sing}(x))).}$$

9. A wise man walks on his feet; an unwise man on his hands. Therefore no man walks on both.

   Formalisation

$$\frac{\begin{array}{l}(\forall x)(\texttt{wise}(x) \to \texttt{walks-on-feet}(x));\\ (\forall x)(\neg(\texttt{wise}(x)) \to \texttt{walks-on-hands}(x));\end{array}}{(\forall x)(\neg(\texttt{walks} - \texttt{on} - \texttt{feet}(x) \wedge \texttt{walks} - \texttt{on} - \texttt{hands}(x))).}$$

   You might want to try to capture the fact that `walks-on-feet`( ) and `walks-on-hands`( ) share some structure, and accordingly have a two-place relation `walks-on`(,).
   Then i think you will also want binary predicate letters `feet-of`(,) and `hands-of`(,) so you would end up with

   $(\forall x)(\texttt{wise}(x) \to (\forall y)(\texttt{feet-of}(x, y) \to \texttt{walks-on}(x, y)))$

   and of course

   $(\forall x)(\neg\texttt{wise}(x) \to (\forall y)(\texttt{hands-of}(x, y) \to \texttt{walks-on}(x, y)))$

   You might feel that the following are equally good formalisations:

   $(\forall x)(\texttt{wise}(x) \to (\exists y)(\exists z)(\texttt{feet-of}(x, y) \wedge \texttt{feet-of}(x, z) \wedge \neg(y = z) \wedge \texttt{walks-on}(x, y) \wedge$ $\texttt{walks-on}(x, z)))$ ... and the same for unwise men and hands.

   However that involves **two**-place relations and we haven't got to them yet!

10. No fossil can be crossed in love; an oyster may be crossed in love. Therefore oysters are not fossils.

    Lexicon:
    Unary predicate letters: `fossil`( ). `oyster`( ), `crossed-in-love`( ).

    Formalisation

$$\frac{\begin{array}{l}(\forall x)(\texttt{fossil}(x) \to \neg\texttt{can-be-crossed-in-love}(x));\\ (\forall x)(\texttt{oyster}(x) \to \texttt{can-be-crossed-in-love}(x));\end{array}}{(\forall x)(\texttt{oyster}(x) \to \neg\texttt{fossil}(x))}$$

11. All who are anxious to learn work hard; some of these students work hard. Therefore some of these students are anxious to learn.

    Lexicon:
    Unary predicate letters: `anxious-to-learn`( ), `works-hard`( ), `student`( ).

Formalisation

$$\frac{\begin{array}{l}(\forall x)(\texttt{anxious-to-learn}(x) \to \texttt{works-hard}(x));\\ (\exists x)(\texttt{student}(x) \wedge \texttt{works-hard}(x));\end{array}}{(\exists x)(\texttt{student}(x) \wedge \mathtt{anxious - to - learn}(x)).}$$

12. His songs never last an hour. A song that lasts an hour is tedious. Therefore his songs are never tedious.

    Lexicon:
    Unary predicate letters: `last-an-hour`( ), `song`( ), `his`( ), `tedious`( ).

    Formalisation

    $$\frac{\begin{array}{l}(\forall y)((\texttt{song(y)} \wedge (\texttt{his}(y)) \to \texttt{last-an-hour}(y));\\ (\forall x)((\texttt{song}(x) \wedge \texttt{last-an-hour}(x)) \to \texttt{tedious}(x));\end{array}}{(\forall z)((\texttt{song}(z) \wedge \texttt{his}(z)) \to \neg\texttt{tedious}(z)).}$$

13. Some lessons are difficult; what is difficult needs attention. Therefore some lessons need attention.

    Lexicon:
    Unary predicate letters: `lesson`( ), `difficult`( ), `needs-attention`( ).

    Formalisation

    $$\frac{\begin{array}{l}(\exists x)(\texttt{lesson}(x) \ \wedge \ \texttt{difficult}(x));\\ (\forall z)(\texttt{difficult}(z) \to \texttt{needs-attention}(z)).\end{array}}{(\exists x)(\texttt{lesson}(x) \wedge \mathtt{needs - attention}(x)).}$$

14. All humans are mammals; all mammals are warm blooded. Therefore all humans are warm-blooded.

    Lexicon:
    Unary predicate letters: `human`( ), `mammal`( ), `warm-blooded`( ).

    Formalisation

    $$\frac{\begin{array}{l}(\forall y)(\texttt{human}(y) \to \texttt{mammal}(y));\\ (\forall y)(\texttt{mammal}(y) \to \texttt{warmblooded}(y));\end{array}}{(\forall z)(\texttt{human}(z) \to \texttt{warmblooded}(z)).}$$

15. Warmth relieves pain; nothing that does not relieve pain is useful in toothache. Therefore warmth is useful in toothache.

    Lexicon:
    Unary predicate letters: `relieves-pain`( ), `useful-in-toothache`( );
    Constant symbol: `warmth`,

    Formalisation

    $$\frac{\begin{array}{l}\texttt{relieves-pain}(\texttt{warmth});\\ (\forall x)(\texttt{useful-in-toothache}(x) \to \texttt{relieves-pain}(x));\end{array}}{\mathtt{useful - in - toothache}(\texttt{warmth})}$$

spacing!?

You might want to break up `relieves-pain` by having a binary predicate letter `relieves`(,) and a constant symbol `pain`, giving

$$\frac{\begin{array}{l}\texttt{relieves}(\texttt{warmth}, \texttt{pain});\\ (\forall x)(\texttt{useful-in-toothache}(x) \to \texttt{relieves}(x, \texttt{pain});\end{array}}{\texttt{useful} - \texttt{in} - \texttt{toothache}(\texttt{warmth}).}$$

16. Louis is the King of France; all Kings of France are bald. Therefore Louis is bald.

    Lexicon:
    Unary predicate letters: `bald`( ), `King-of-France`( ),
    Constant symbol: `Louis`.

    Formalisation

    $$\frac{\begin{array}{l}\texttt{king-of-France}(\texttt{Louis});\\ (\forall x)(\texttt{king-of-France}(x) \to \texttt{bald}(x));\end{array}}{\texttt{bald}(\texttt{Louis}).}$$

    You might feel that `King-of-France` is not really a unary predicate but a binary predicate (`king-of`) with one argument place plugged by a constant (`France`).

Exercise 46

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.

1. Anyone who has forgiven at least one person is a saint.

   Lexicon:
   Unary predicate letters: `saint`( )
   Binary predicate letters: `has-forgiven`(,)

   Formalisation

   $$(\forall x)(\forall y)(\texttt{has-forgiven}(x, y) \to \texttt{saint}(x))$$

2. Nobody in the logic class is cleverer than everybody in the history class.

   Lexicon:
   Unary predicate letters: `is-in-the-logic-class`( ), `is-in-the-history-class`( )
   Binary predicate letter: `is-cleverer-than`(,)

   Formalisation

   $$(\forall x)(\texttt{is-in-the-logic-class}(x) \to (\exists y)(\texttt{is-in-the-history-class}(y) \wedge \neg(\texttt{is-cleverer-than}(x, y));$$

   Here you might prefer to have a two-place relation between people and subjects, so that you then have two constants, `history` and `logic`.

3. Everyone likes Mary—except Mary herself.

   Lexicon:
   Binary predicate letter: $L(,)$
   Constant symbol: m

   Formalisation

   $$\neg L(m, m) \wedge (\forall x)(x \neq m \rightarrow L(x, m))$$

4. Jane saw a bear, and Roger saw one too.

   Lexicon:
   Unary predicate letter: $B( )$
   Binary predicate letter: $S(,)$
   Constant symbols: j, r

   Formalisation

   $$(\exists x)(B(x) \wedge S(j, x)) \wedge (\exists x)(B(x) \wedge S(r, x));$$

5. Jane saw a bear and Roger saw it too.

   $$(\exists x)(B(x) \wedge S(j, x) \wedge S(r, x))$$

Supply an answer or delete

6. God will destroy the city unless there is a righteous man in it;

7. Some students are not taught by every teacher;

   Lexicon:
   Unary predicate letters: $\texttt{teacher}( )$, $\texttt{student}( )$.
   Binary predicate letter: $\texttt{taught-by}(,)$

   Formalisation

   $$(\exists x)\texttt{student}(x) \wedge \neg(\forall y)(\texttt{teacher}(y) \rightarrow \texttt{taught-by}(x, y))$$

   Of course you might want to replace '$\texttt{teacher}(x)$' by '$(\exists y)(\texttt{taught-by}(y, x))$'.

8. No student has the same teacher for every subject.

   Lexicon:
   Ternary predicate letter: $R(, ,)$
   Unary predicate letters: $\texttt{student}( )$, $\texttt{teacher}( )$, $\texttt{subject}( )$.

   Formalisation

   $$(\forall x)(\texttt{student}(x) \rightarrow \neg(\forall y)(\texttt{teacher}(y) \rightarrow (\forall z)(\texttt{subject}(z) \rightarrow R(x, y, z))))$$

9. Everybody loves my baby, but my baby loves nobody but me.

   Lexicon:
   Binary predicate letter[1]: $L(,)$;
   Constant symbols: b, m.

   Formalisation

[1] Observe that we do not have to specify that '=' is part of the lexicon. That's a given, since it is part of the logical vocabulary.

$(\forall x)(L(x, b)) \wedge (\forall x)(L(b, x) \rightarrow x = m)$;

Exercise 47

Match up the formulæ on the left with their English equivalents on the right.
(i) matches (a); (iii) matches (b); (ii) matches (c).

Exercise 48

(These involve nested quantifiers and dyadic predicates)

1. Everyone who loves is loved: $(\forall x)(\forall y)(L(y, x) \rightarrow (\exists z)(L(z, y))$.
2. Everyone loves a lover: $(\forall x)(\forall y)(L(y, x) \rightarrow (\forall z)(L(z, y))$.
3. The enemy of an enemy is a friend
4. The friend of an enemy is an enemy
5. Any friend of George's is a friend of mine
6. Jack and Jill have at least two friends in common
7. Two people who love the same person do not love each other.
8. None but the brave deserve the fair: $(\forall x)(\forall y)((F(x) \wedge D(y, x)) \rightarrow B(y))$.
9. If there is anyone in the residences with measles then anyone who has a friend in the residences will need a measles jab.

Exercise 49

Render the following pieces of English into Predicate calculus, using a lexicon of your choice.

1. There are two islands in New Zealand;
2. There are *three*[2] islands in New Zealand;
3. tf knows (at least) two pop stars: $(\exists xy)(x \neq y \wedge K(x) \wedge K(y))$.

   '$K(x)$' of course means that x is a pop star known to me.
4. If there is to be a jackpot winner it will be me.

   The lexicon is obviously going to have a one-place predicate `wins-the-jackpot`. The temptation is to write something like

$$(\exists x)\texttt{wins-the-jackpot}(x) \rightarrow x = \text{me}$$

[2] The third is Stewart Island

But nothing like that will work. Do you mean

$$(\exists x)(\texttt{wins-the-jackpot}(x) \to x = \text{me})$$

or

$$(\exists x)(\texttt{wins-the-jackpot}(x)) \to x = \text{me}$$

. . . ? Neither is correct. The first one is true as long as there is someone who does not win the jackpot (because that person is a witness to the 'x') or as long as there is someone equal to me; the second one is not a closed formula, because the occurrence of 'x' in '$x = \text{me}$' is not bound by the '$\exists x$'.

What we want is:

$$(\forall x)(\texttt{wins-the-jackpot}(x) \to x = \text{me})$$

Of course the following would also work

$$(\exists x)(\texttt{wins-the-jackpot}(x)) \to \texttt{wins-the-jackpot}(\text{me})$$

but one gets steered away from that by the thought that one needs equality.

5. You are loved only if you yourself love someone [other than yourself!];

$$(\forall x)(\forall y)(L(y, x) \to (\exists z)(z \neq x \land L(x, z))$$

$$(\forall x)((\exists y)(L(y, x)) \to (\exists z)(z \neq x \land L(x, z))$$

will do too.

But of course the sensible thought that the original English sentence is grasping elliptically is "you are loved *by someone else* only if you love someone else", and that is

$$(\forall x)((\exists y)(y \neq x \land L(y, x)) \to (\exists z)(z \neq x \land L(x, z)))$$

or

$$(\forall x, y)(y \neq x \land L(y, x) \to (\exists z)(z \neq x \land L(x, z)))$$

6. God will destroy the city unless there are (at least) two righteous men in it;

7. There is at most one king of France;

$$(\forall xy)(K(x) \land K(y) \to x = y)$$

8. I know no more than two pop stars;

$$(\forall xyz)((K(x) \wedge K(y) \wedge K(z)) \to (x = y \vee x = z \vee y = z))$$

9. There is precisely one king of France;

$$(\exists x)(K(x) \wedge (\forall y)(K(y) \to y = x))$$

   Notice that
$$(\exists x)(\forall y)(K(x) \wedge (K(y) \to y = x))$$
   would do equally well. **Make sure you are happy about this**.

10. I know three FRS's and one of them is bald;

11. Brothers and sisters have I none; this man's father is my father's son.

12. * Anyone who is between a rock and a hard place is also between a hard place and a rock.

Using the 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) y;
$F(x, y)$:	x and y are friends;
$R(x)$:	x lives in the residences;
$M(x)$:	x has measles;

Turn the following into English. (**normal** English: something you can imagine yourself saying—no xs and ys.)

$(\forall x)(F(\mathtt{Kim}, x) \to F(\mathtt{Alex}, x))$
Every friend of Kim is a friend of Alex.

$(\forall x)(\exists y)(F(x, y) \wedge M(y) \wedge Z(y))$
Everyone has a friend in the residences with measles

$(\forall x)(F(\mathtt{Kim}, x) \to R(x))$
All Kim's friends live in the residences

$(\forall x)((R(x) \wedge M(x)) \to F(\mathtt{Kim}, x))$
The only people in the residences with measles are friends of Kim

$(\forall x)(R(x) \to (\exists y)(F(x, y) \wedge M(y)))$
Everyone who lives in the residences has a friend with measles

$(\forall x)(S(x) \to (\exists yz)(T(y, x, z))$

Every student goes to at least one lecture course

$(\exists x)(S(x) \wedge (\forall z)(\neg A(x,z))$
There is a student that isn't going to any course

$(\exists x)(C(x) \wedge (\forall z)(\neg A(z,x))$
There is a course that nobody is taking

$(\exists x)(L(x) \wedge (\forall yz)(\neg T(x,y,z))$
One of the lecturers is giving a course that nobody is taking.[3]

$(\forall x_1 x_2)[(\forall z)(A(x_1,z) \longleftrightarrow A(x_2,z)) \rightarrow x_1 = x_2]$
No two students go to exactly the same courses

$(\forall x_1 x_2)[(\forall z)(A(z,x_1) \longleftrightarrow A(z,x_2)) \rightarrow x_1 = x_2]$
No two courses have exactly the same students going to them.

$(\forall xy)(x \neq y \rightarrow (\exists z)(F(z,x) \longleftrightarrow \neg F(z,y)))$
No two people have exactly the same friends.

Exercise 55 part 4

(Thanks to Matt Grice)

Exercise 58 part 1

$$
\begin{prooftree}
\AxiomC{$[F(a)]^1$}
\RightLabel{$\rightarrow$-int}
\UnaryInfC{$F(b) \rightarrow F(a)$}
\RightLabel{$\forall$-int}
\UnaryInfC{$(\forall y)(F(y) \rightarrow F(a))$}
\RightLabel{$\exists$-int}
\UnaryInfC{$(\exists x)(\forall y)(F(y) \rightarrow F(x))$}
\AxiomC{$[\neg(\exists x)(\forall y)(F(y) \rightarrow F(x))]^2$}
\RightLabel{$\rightarrow$-elim}
\BinaryInfC{$\perp$}
\RightLabel{ex falso}
\UnaryInfC{$F(b)$}
\RightLabel{$\rightarrow$-int (1)}
\UnaryInfC{$F(a) \rightarrow F(b)$}
\RightLabel{$\forall$-int}
\UnaryInfC{$(\forall y)(F(y) \rightarrow F(b))$}
\RightLabel{$\exists$-int}
\UnaryInfC{$(\exists x)(\forall y)(F(y) \rightarrow F(x))$}
\AxiomC{$[\neg(\exists x)(\forall y)(F(y) \rightarrow F(x))]^2$}
\RightLabel{$\rightarrow$-elin}
\BinaryInfC{$\perp$}
\RightLabel{$\rightarrow$-int (2)}
\UnaryInfC{$\neg\neg(\exists x)(\forall y)(F(y) \rightarrow F(x))$}
\RightLabel{double negation}
\UnaryInfC{$(\exists x)(\forall y)(F(y) \rightarrow F(x))$}
\end{prooftree}
\tag{12.5}
$$

$$
\begin{prooftree}
\AxiomC{$A(a) \vdash A(a)$}
\AxiomC{$B(f(a)) \vdash B(f(a))$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$A(a), A(a) \rightarrow B(f(a)) \vdash B(f(a))$}
\RightLabel{$\forall$ L}
\UnaryInfC{$A(a), (\forall x)(A(x) \rightarrow B(f(x))) \vdash B(f(a))$}
\RightLabel{$\exists$ R}
\UnaryInfC{$A(a), (\forall x)(A(x) \rightarrow B(f(x))) \vdash (\exists x)B(x)$}
\RightLabel{$\exists$ L}
\UnaryInfC{$(\exists x)A(x), (\forall x)(A(x) \rightarrow B(f(x))) \vdash (\exists x)B(x)$}
\end{prooftree}
\tag{12.6}
$$

[3] That's not right!

$$
\begin{prooftree}
\AxiomC{$b \in a \ \vdash b \in a$}
\AxiomC{$a \in b \ \vdash a \in b$}
\AxiomC{$b \in a \ \vdash b \in a$}
\RightLabel{$\neg$ L}
\UnaryInfC{$b \in a, b \notin a \ \vdash$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$b \in a, a \in b, a \in b \rightarrow b \notin a \ \vdash$}
\RightLabel{$\forall$ L}
\UnaryInfC{$b \in a, a \in b, (\forall z)(z \in b \rightarrow b \notin z) \ \vdash$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$a \in b, b \in a, b \in a \rightarrow (\forall z)(z \in b \rightarrow b \notin z), (\forall z)(z \in b \rightarrow b \notin z) \rightarrow b \in a \ \vdash$}
\RightLabel{$\wedge$ L$_1$}
\UnaryInfC{$a \in b, b \in a, b \in a \longleftrightarrow (\forall z)(z \in b \rightarrow b \notin z) \ \vdash$}
\RightLabel{$\neg$ R}
\UnaryInfC{$b \in a, b \in a \longleftrightarrow (\forall z)(z \in b \rightarrow b \notin z) \ \vdash a \notin b$}
\RightLabel{$\rightarrow$ R}
\UnaryInfC{$b \in a \longleftrightarrow (\forall z)(z \in b \rightarrow b \notin z) \ \vdash b \in a \rightarrow a \notin b$}
\RightLabel{$\forall$ L}
\UnaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)) \ \vdash b \in a \rightarrow a \notin b$}
\RightLabel{$\forall$ R}
\UnaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)) \ \vdash (\forall z)(z \in a \rightarrow a \notin z)$}
\AxiomC{$a \in a \ \vdash a \in a$}
\AxiomC{$a \in a \ \vdash a \in a$}
\AxiomC{$a \in a \ \vdash a \in a$}
\RightLabel{$\neg$ L}
\UnaryInfC{$a \in a, a \notin a \ \vdash$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$a \in a, a \in b \rightarrow a \notin a \ \vdash$}
\RightLabel{$\forall$ L}
\UnaryInfC{$(\forall z)(z \in a \rightarrow a \notin z), a \in a \ \vdash$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$a \in a \rightarrow (\forall z)(z \in a \rightarrow a \notin z), a \in a \ \vdash$}
\RightLabel{$\rightarrow$ L}
\BinaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z))(z \in y \rightarrow y \notin z)), a \in a \rightarrow (\forall z)(z \in a \rightarrow a \notin z), (\forall z)(z \in a \rightarrow a \notin z) \rightarrow a \in a \ \vdash$}
\RightLabel{$\wedge$ L$_1$}
\UnaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)), a \in a \longleftrightarrow (\forall z)(z \in a \rightarrow a \notin z) \ \vdash$}
\RightLabel{$\forall$ L}
\UnaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)), (\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)) \ \vdash$}
\RightLabel{Contract-L}
\UnaryInfC{$(\forall y)(y \in a \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)) \ \vdash$}
\RightLabel{$\exists$ L}
\UnaryInfC{$(\exists x)(\forall y)(y \in x \longleftrightarrow (\forall z)(z \in y \rightarrow y \notin z)) \ \vdash$}
\end{prooftree}
\tag{12.4}
$$

$$\dfrac{(\exists x)A(x) \qquad \dfrac{\dfrac{[A(a)]^{(1)} \qquad \dfrac{(\forall x)(A(x) \to B(f(x)))}{A(a) \to B(f(a))}\ \forall \text{ elim}}{B(f(a))}\ \to\text{-elim}}{(\exists x)B(x)}\ \exists\text{-int}}{(\exists x)B(x)}\ \exists\text{-elim}(1) \tag{12.7}$$

Exercise 43

Exercises from Chapter 5

Exercises from Chapter 6

Exercise 63 Part 2

Identify a row in which the formula does not take truth-value 1

Try $[[p]] = 2$ and $[[q]] = 3$. Then $[[p \vee q]] = 2$; $[[p \to q]] = 3$ whence $[[(p \to q) \to q]] = 1$, giving $[[((p \to q) \to q) \to (p \vee q)]] = 3$.

Exercises from Chapter 7

Exercises from Chapter 8

Exercise 73

	q	$\neg q$
p	A	B
$\neg p$	C	D

We want to say something like: p says $A \vee B$ and q says $A \vee C$. But that doesn't capture the exclusiveness, so we want p to say $(A \vee B) \wedge \neg(C \vee D)$. Similarly we want q to say $(A \vee C) \wedge \neg(B \vee D)$. Then $p \wedge q$ says A; $p \wedge \neg q$ says B; $\neg p \wedge q$ says C; and $\neg p \wedge \neg q$ says D.

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{(\mathbf{p} \to \mathbf{q}) \to \mathbf{q} \qquad [p \to q]^1}{q}\,\to\text{-elim} \qquad \mathbf{q} \to \mathbf{r}}{r}\,\to\text{-elim} \qquad [r \to p]^2}{p}\,\to\text{-elim}}{(p \to q) \to p}\,\to\text{-int (1)} \qquad ((p \to q) \to p) \to p}{p}\,\to\text{-elim} \qquad \mathbf{p} \to \mathbf{r}}{r}\,\to\text{-elim}}{(r \to p) \to r}\,\to\text{-int (2)} \qquad ((r \to p) \to r) \to r}{r}\,\to\text{-elim} \tag{12.8}
$$

Chapter 13

Indexes

13.1 Index of definitions

General Index

Bibliography

[1] Berkeley, G. Three dialogues of Hylas and Philonous.

[2] Andreas Blass, Nachum Dershowitz and Yuri Gurevich. When are two algorithms the same? Bulletin of Symbolic Logic **15** (june 2009) pp 145–168.

[3] George Boolos and Giovanni Sambin, Provability: the emergence of a mathematical modality, Studia Logica, **50** 1992 p 1-23

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

[5] F. H. Bradley *Appearance and Reality*.

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

[7] Carroll, Lewis. "What the Tortoise Said to Achilles." Mind, N.S. IV, (1895) 278–280.

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

[9] Chang, C. L. and Lee, R.C.T., Symbolic Logic and Mechanical Theorem Proving, Academic Press 1973.

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

[11] Roy Cook review of [41] Bull Sym Log **18** pp 128–31.

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

[13] Marcel Crabbé. L'Egálité et l'éxtensionalité. Logique et Analyse **185-188** dec 2004 pp 365–374.

[14] Michael Dummett: Elements of Intuitionism Oxford Logic Guides.

[15] Lance Endersbee "A voyage of discovery" published by the author 2005. ISBN 0-646-45301-7.

[16] `https://en.wikipedia.org/wiki/The_Fly_%281986_film%29`

[17] T. E. Forster. The Modal æther. `www.dpmms.cam.ac.uk/~tf/modalrealism.ps`

[18] Martin Gardner “The annotated Alice”

[19] Glivenko, V. “Sur la Logique de M. Brouwer”. Bulletin de L’Académie Royal de Belgique, série de la Classe des Sciences **14** (1928) pp 225–8.

[20] Glivenko, V. “Sur quelques points de la Logique de M. Brouwer”. Bulletin de L’Académie Royal de Belgique, série de la Classe des Sciences **15** (1929) pp 183–8.

[21] Gödel, K. “An interpretation of the intuitionistic sentential Logic”. Reprinted in Hintikka (ed) The Philosophy of Mathematics Oxford readings in Philosophy (1969) pp 128–9.

[22] Carl Hewitt. “Common sense for concurrency and strong paraconsistency using unstratified inference and reflection”. arXiv:0812.4852

[23] Wilfrid Hodges Logic Penguin

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

[25] Kalmár, L. “Über die Axiomatisierbarkeit des Aussagenkalküls”, Acta Scientarum Mathematicarum **7** 1934-5 pp 222–43.

[26] Norman Lindsay The Magic Pudding.

[27] Per Martin-Löf. “Truth of a Proposition, Evidence of a Judgement, Validity of a Proof”, Synthese **73**, pp.407–420.

[28] Per Martin-Löf. “On the meaning of the Logical Constants and the Justifications of the Logical Laws”, Nordic Journal of Philosophical Logic 1(1) (1996), pp.11–60.
`http://www.hf.uio.no/ifikk/filosofi/njpl/vol1no1/meaning/meaning.html`

[29] J.C.C. McKinsey and Alfred Tarski. “Some theorems about the sentential calculi of Lewis and Heyting”, Journal of Symbolic Logic **13** 1948 pp 1–15.

[30] Mendelson, E. (1979) Introduction to Mathematical Logic. (2nd edn) Van Nostrand

[31] Roger Penrose: The Emperor’s New Mind

[32] Graham Priest “Some Priorities of Berkeley”, Logic and Reality: Essays on the Legacy of Arthur Prior, ed. B.J.Copeland, Oxford University Press, 1996.

[33] Quine, Mathematical Logic

[34] Giovanni Sommaruga. “History and Philosophy of Constructive Type Theory” SYNTHESE LIBRARY Volume 290, (Feb 1, 2000)

[35] Arthur Prior “The runabout inference ticket”. Analysis **21**(1960) pp 38–9. Reprinted in Oxford Readings in Philosophy ed Strawson 1967.

[36] Quine, Ontological relativity and Other Essays

[37] Quine, Word and Object

[38] Quine, Predicate functor Logic in Selected Logic papers.

[39] Scott D.S. Semantical Archæology, a parable. In: Harman and Davidson eds, Semantics of Natural Languages. Reidel 1972 pp 666–674.

[40] Peter Smith: An introduction to Formal Logic CUP.

[41] G Sommaruga (eds) Foundational theories of constructive and classical mathematics.

[42] Florian Steinberger "On the Equivalence Conjecture for Proof-Theoretic Harmony", Notre Dame Journal of Formal Logic, **54** (1), 2013. 2012:

[43] Paul Taylor Practical foundations of mathematics

[44] Alfred Tarski and Steven Givant. A formulation of Set Theory without variables. AMS colloquium publication **41** 1987.

[45] Tversky and Kahnemann "Extensional versus Intuitive reasoning" Psychological Review **90** 293-315.

[46] Tymoczko and Henle. Sweet Reason: A Field Guide to Modern Logic. Birkhäuser 2000 ISBN