Part III Logic Michaelmas Term 2017
Chapter 1: A Tutorial on Constructive Logic
Lectures 1-8 of 24

Thomas Forster

October 25, 2017

Contents

The intention is that this handout shall be the definitive course material for the first third of the course. I do not delude myself that it is bug-free, and I welcome *errata* and suggestions for clarification.

Much of this material has had earlier outings in front of less sophisticated audiences than your good selves, and many of the exercises will come across as over-cautiously elementary. Feel free to omit any of them.

The appearance made by λ-calculus here might seem excessively brief. Fear not! It will be treated in more detail later.

Preface

Some readers may already know the standard horror story about $\sqrt{2}^{\sqrt{2}}$. For those of you that don't—yet—here it is.

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, as desired. However, we haven't met the challenge. We were asked to *find* a pair $\langle\alpha, \beta\rangle$ of irrationals such that α^β is rational, and we haven't found such a pair. We've proved that there *is* such a pair, and we have even narrowed the candidates down to a short list of two, but we haven't completed the job.[1]

This example is admittedly a bit contrived. A more idiomatic example is Roth's theorem, which states that for an irrational algebraic number α and a given $\epsilon > 0$, the inequality

$$|\alpha - \frac{p}{q}| < \frac{1}{q^{2+\epsilon}}$$

can have only finitely many solutions in coprime integers p and q. Therefore there is a bound on the sizes of the solutions given α and ϵ. However the usual proof does not enable us to obtain such a bound.

But we can't really use that as illustration because we haven't got time to tease the proof apart to see where we have used excluded middle. If you like number theory you could try that as an exercise!

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.

By emphasising constructions we nurture intuition, give it something to get its teeth into. It's not quite correct to say always that an existence proof that

[1] 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 $(\sqrt{2})^{(log_{\sqrt{2}}(3))} = 3$.

constructs the desired object tells you why it exists, but in most cases it will tell you more than a proof that doesn't perform the construction.

However the constructive Mathematics that has come down to us is a curious beast. It is born of two errors, but these errors have given rise to some deep and important mathematics.

The two errors are (i) the confusion between truth and provability; and (ii) the unspoken assumption that two proofs that give different information cannot be proofs of the same thing.

(i) leads to a rejection of the law of excluded middle: if $A \vee B$ is true then one of A and B must be true, but in circumstances where we can prove neither p not $\neg p$, how can we be justified in asserting $p \vee \neg p$? One of them must be true but *ex hypothesi* neither is provable, violating the identification of truth with provability.

(ii) leads people to think that since an effective proof of $(\exists x)\text{Wombat}(x)$ tells us where to find the wombat, whereas an ineffective proof does not, the two proofs must surely be proofs of different things: they cannot both be proofs of $(\exists x)\text{Wombat}(x)$. Since the ineffective proof is typically a deduction of a contradiction from the assumption that there is no wombat then perhaps it is instead a proof of $\neg\neg(\exists x)\text{Wombat}(x)$, and perhaps this proposition is not the same as $(\exists x)\text{Wombat}(x)$. At any rate, that is the conclusion the constructivists draw. Let us see where it leads.

The axioms and rules of classical logic that you saw in Part II arise from the key thought that there are precisely two truth-vlues, true and false, together with the desire to formulate those rules of inference that preserve truth: give true conclusions if the premisses are true. The enterprise that we are embarking on here is interested in capturing all those inferences that preserve the property of corresponding to a construction. It has to drop excluded middle, so it sounds as if it's denying the dogma that there are precisely two truth values, but that's actually not what is going on. It doesn't *deny* the dogma, it merely refrains from asserting it.

1 Natural Deduction

If we are to take an interest in the difference between effective proof and ineffective proof, we are going to have to start taking proofs seriously as mathematical objects. Proofs in the *Natural Deduction* style will be our first port of call.

(Say something about why we don't use Hilbert-style proofs as in Part II)

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

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

(The idea that everything there is to know about a connective can be captured by an elimination rule plus an introduction rule has the same rather operationalist flavour possessed by the various *meaning is use* doctrines one encounters in philosophy of language. In this particular form it goes back to Prawitz, and possibly to Gentzen. See section 3.4.)

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 1

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:*[2]

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

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

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

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.)

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; $\frac{\bot}{A}$

If we were setting up a proof system for classical logic (which we aren't) we would insert here the rule of

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

but we include it anyway for the sake of completeness.

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 $B \to \bot$, so the inference

$$\frac{A \qquad \neg A}{\bot} \tag{1}$$

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

1.1 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 see 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 $A \to (A \vee C)$.

$$\frac{A}{A \vee C}\ \vee\text{-int} \tag{2}$$

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

$$\frac{A}{A \vee C}\ \vee\text{-int} \tag{3}$$

$$\frac{}{A \to (A \vee C)}$$

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.

$$\dfrac{\dfrac{[A]^1}{A \vee C}\ \vee\text{-int}}{A \to (A \vee C)}\ \to\text{-int (1)} \tag{4}$$

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. Indeed, if you are up for it, you can even set up the rule so that you are allowed to cancel nonexistent occurrences! I find this tends to frighten the horses, so that possibility is hived off as a separate rule, the *Identity Rule*.

1.2 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."

We often use $\vee$-elimination in Sudoku puzzles. Consider the following example:

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

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

1.3 The Identity Rule

Finally we need the identity rule:

$$\frac{A\ B\ C\ldots}{A} \tag{5}$$

(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'? One could do something like

$$\cfrac{\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A \wedge B}\ \wedge\text{-int}}{A}\ \wedge\text{-elim}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{6}$$

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

$$\cfrac{\cfrac{\cfrac{[A]^2 \qquad [B]^1}{A}\ \text{identity rule}}{B \to A}\ \to\text{-int (1)}}{A \to (B \to A)}\ \to\text{-int (2)} \tag{7}$$

If we take seriously the observation above concerning the rule of $\to$-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 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 7 to 8. It's a matter of taste.

$$\cfrac{\cfrac{[A]^1}{B \to A}\ \to\text{-int}}{A \to (B \to A)}\ \to\text{-int (1)} \tag{8}$$

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.

1.4 Rules for the Quantifiers

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

Rules for $\exists$

$$\frac{A(t)}{(\exists x)(A(x)}\ \exists\text{-int} \qquad\qquad \frac{\begin{array}{c}[A(t)]^{(1)}\\ \vdots\\ C\end{array} \qquad (\exists x)(A(x))}{C}\ \exists\text{-elim}(1) \tag{9}$$

Notice the similarity between $\vee$-elimination and $\exists$-elimination.

Rules for $\forall$

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

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.[3]

Theorem 1 *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.

We also need a rule of substitutivity of equality:

$$\frac{\phi(x) \qquad x = y}{\phi(y)}$$

[3]Thanks to the late Aldo Antonelli.

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 (alluded to earlier) that elimination rules are **use** rules:

2.1 The rule of $\to$-elimination

The rule of $\to$-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.

2.2 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 supposing that you know which of A and B is true? Well, **if** you know you can deduce C from A, and you ALSO know that you can deduce C from B, **then** as soon as you are told $A \vee B$ you can deduce C. One could think of the rule of $\vee$-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 **??**.

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

$$\dfrac{\begin{matrix}[A_1]^1 & [A_2]^1 & \dots & [A_n]^1 & \\ \vdots & \vdots & & \vdots & \\ C & C & & C & A_1 \vee A_2 \vee \dots A_n\end{matrix}}{C}\ \vee\text{-elim (1)} \qquad (10)$$

where we can cancel more than one assumption. That is to say we have a list $A_1 \dots A_n$ 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 $A_1 \vee \dots A_n$ of the assumptions $A_1 \dots A_n$ 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 2
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 $\neg(Q \to P)$. (needs double negation)

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.

Your first step—when challenged to find a natural deduction proof of a formula—should be to identify the principal connective. 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 $A \wedge B$.

3.1 The Small Print

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

3.1.1 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.

$$\dfrac{\dfrac{\dfrac{[A]^2 \qquad [A \to (B \vee C)]^3}{B \vee C}\;\to\text{-elim} \qquad \dfrac{\dfrac{[B]^1}{A \to B}\;\to\text{-int } (2)}{(A \to B) \vee (A \to C)}\;\vee\text{-int} \qquad \dfrac{\dfrac{[C]^1}{A \to C}\;\to\text{-int } (2)}{(A \to B) \vee (A \to C)}\;\vee\text{-int}}{(A \to B) \vee (A \to C)}\;\vee\text{-elim } (1)}{A \to (B \vee C). \to .(A \to B) \vee (A \to C)}\;\to\text{-int } (3) \tag{11}$$

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 15.

3.1.2 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 12.

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). (Of course this is not in a constructive context, but the moral is the same wherever we are If you attack the principal connective you will of course use $\vee$-int and generate the attempt

$$\dfrac{A}{A \vee \neg A}\;\vee\text{-int} \tag{12}$$

or the attempt

$$\dfrac{\neg A}{A \vee \neg A}\;\vee\text{-int} \tag{13}$$

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, 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. P=NP?

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.

3.2 Some Exercises

Exercise 3 *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. $((A \wedge B) \vee (A \wedge C)) \to (A \wedge (B \vee C))$;

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

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*

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.

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 4

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

Exercise 5 *Annotate the following proofs, indicating which rules are used where and which premisses are being cancelled when.*

$$\cfrac{\cfrac{\cfrac{P \qquad P \to Q}{Q}}{(P \to Q) \to Q}}{P \to ((P \to Q) \to Q)} \tag{14}$$

$$\dfrac{\dfrac{\dfrac{P \wedge Q}{Q}}{P \vee Q}}{(P \wedge Q) \to (P \vee Q)} \tag{15}$$

$$\dfrac{\dfrac{\dfrac{P \qquad \neg P}{\bot}}{Q}}{P \to Q} \tag{16}$$

$$\dfrac{\dfrac{P \vee Q \qquad \dfrac{P \qquad P \to R}{R} \qquad \dfrac{Q \qquad Q \to R}{R}}{R}}{(P \vee Q) \to R} \tag{17}$$

$$\dfrac{\dfrac{\dfrac{A \qquad B}{A \wedge B}}{B \to (A \wedge B)}}{A \to (B \to (A \wedge B))} \tag{18}$$

$$\dfrac{\dfrac{\dfrac{(A \to B) \to B \qquad A \to B}{B}}{((A \to B) \to B) \to B}}{(A \to B) \to (((A \to B) \to B) \to B)} \tag{19}$$

3.3 A First Look at Three-valued Logic

A warning! The fact that constructive logic eschews excluded middle does not mean that it denies it. But constructive logic does slightly more than not-actually-deny excluded middle. It denies that there are more than two truth-values! It is possible (tho' a lot of work) to constructively deduce a contradiction from the three assumptions $\neg(A \longleftrightarrow B)$, $\neg(B \longleftrightarrow C)$ and $\neg(A \longleftrightarrow C)$ that say that A, B and C take three distinct truth-values.

Notwithstanding this, the thought-experiment of having more than two truth-values can be helpful, as we shall now see, on our trip to planet Zarg.

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

3.4 Harmony and Conservativeness

3.4.1 Conservativeness

Recall the discussion on page 9 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 9 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 constructive logic is that it respects this æsthetic.

The point of exercise 6 part 3 was to establish that there can be no proof of Peirce's law using just the rules for '$\rightarrow$'.

3.4.2 Harmony

A further side to this æsthetic is the thought that, for each connective, the introduction and elimination rule 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

$$\dfrac{\dfrac{\begin{matrix}\mathcal{D}_1\\ \vdots\\ A\end{matrix} \qquad \begin{matrix}\mathcal{D}_2\\ \vdots\\ B\end{matrix}}{A \wedge B}\ \wedge\text{-int}}{B}\ \wedge\text{-elim} \tag{20}$$

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

$$\begin{matrix}\mathcal{D}_2\\ \vdots\\ B\end{matrix} \tag{21}$$

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

$$\dfrac{\dfrac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \to B}\,\to\text{-int} \qquad A}{B}\,\to\text{-elim} \tag{22}$$

to

$$\begin{matrix}A\\ \vdots \\ B\end{matrix} \tag{23}$$

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

$$\dfrac{\begin{matrix}[A_1]^1\\ \vdots \\ C\end{matrix} \qquad \begin{matrix}[A_2]^1\\ \vdots \\ C\end{matrix} \qquad \dfrac{A_1}{A_1 \vee A_2}\,\vee\text{-int}}{C}\,\vee\text{-elim (1)} \tag{24}$$

to

$$\begin{matrix}A_1\\ \vdots \\ C\end{matrix} \tag{25}$$

DEFINITION 1

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.

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

The rules of classical contradiction/double negation doesn't submit itself naturally to this harmoniousness analysis. This is one respect in which constructive logic is more attractive than classical logic.

3.5 Maximal Formulæ

The first occurrence of '$A \to B$' in proof 22 page 19 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 24 p. 19. One might think so, but the situation is complex and not entirely satisfactory. One way into this is to try the following exercise:

Exercise 7
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.

It turns out that, if our system of introduction and elimination rules is sufficiently complicated, then maximal formulæ cannot always be eliminated. The following famous example is probably the simplest.

Let us adopt a rule of $\in$-introduction and one of $\in$-elimination:

$$\frac{\phi(a)}{a \in \{x : \phi(x)\}}\ \in\text{-int}; \qquad \frac{a \in \{x : \phi(x)\}}{\phi(a)}\ \in\text{-elim}$$

(we are allowed parameters in ϕ of course) and let R be short for '$\{x : x \in x \to \bot\}$'. Of course 'R' is intended to connote the Russell class.

$$\cfrac{\cfrac{\cfrac{[R \in R]^1}{R \in R \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{R \in R \to \bot}\ \to\text{-int (1)}$$

So we have an outright proof that $R \notin R$. Indeed, by extending the proof with a $\in$-introduction, we have an outright proof of $R \in R$ as well:

$$\cfrac{\cfrac{\cfrac{\cfrac{[R \in R]^1}{R \in R \to \bot}\ \in\text{-elim} \qquad [R \in R]^1}{\bot}\ \to\text{-elim}}{R \in R \to \bot}\ \to\text{-int (1)}}{R \in R}\ \in\text{-int}$$

[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.

So we can join these two proofs with a $\to$-elim and obtain a proof of $\bot$. (I've abbreviated '$R \in R \to \bot$' to '$R \notin R$' to save space.)

$$\dfrac{\dfrac{\dfrac{\dfrac{[R\in R]^1}{{\color{red}R\notin R}}\,\in\text{-elim}\qquad [R\in R]^1}{\bot}\,\to\text{-elim}}{{\color{green}R\notin R}}\,\to\text{-int }(1)\qquad \dfrac{\dfrac{\dfrac{\dfrac{[R\in R]^2}{R\notin R}\,\in\text{-elim}\qquad [R\in R]^2}{\bot}\,\to\text{-elim}}{R\notin R}\,\to\text{-int }(2)}{R\in R}\,\in\text{-int}}{\bot}\,\to\text{-elim}$$

A close reading will reveal to the reader that the last occurrence of '$R \notin R$' on the left at the bottom (in green if you have colour) is a maximal formula, being both the conclusion of a $\to$-introduction and the premise to a $\to$-elimination. There is an obvious manipulation to get rid of the maximal formula, namely to make two copies of the right-hand proof of $R \in R$, and put them above the two occurrences of '$R \in R$' in the left-hand proof, cancel the $\in$-introduction with the $\in$-elim which immediately follows it and discard everything below the '$\bot$' on the left. The reader should check that if (s)he performs the obvious manipulation then a new formula becomes maximal ... namely the occurrence of '$R \notin R$' two lines up (in red if you have colour). Sadly there is no convenient static representation of this phenomenon.

(Notice that this proof is constructive: there is no use of excluded middle. You cannot evade the set theoretic paradoxes by monkeying with the Logic.)

Exercise 8 *(the nonexistence of the higher-degree versions of Russell classes)*
Give a proof in constructive first-order logic that

$$\neg(\exists d)(\forall x)(R(x,d) \longleftrightarrow \neg(\exists y)(R(x,y) \wedge R(y,x)))$$

Your proof does not have to be in sequent or natural deduction form, but a bottle of port is offered as a reward for LaTeX source code of such a proof.

You might think that the pathology we have just seen in this proof of Russell's paradox (namely the maximal formula that Will Not Go Away) arises because the system is inconsistent. $\in$-introduction and elimination give us naïve set theory which (as any fule kno) is inconsistent. However we can modify the rules so that we have *separation* (as in Part II set theory):

$$\dfrac{\phi(a) \quad a \in A}{a \in \{x \in A : \phi(x)\}}\ \in\text{-int;}$$

and two rules for $\in$-elimination

$$\dfrac{a \in \{x \in A : \phi(x)\}}{\phi(a)}\ \in\text{-elim and}\quad \dfrac{a \in \{x \in A : \phi(x)\}}{a \in A}\ \in\text{-elim.}$$

(which is consistent) and then we can prove

$$(\forall A)(\{a \in A : a \notin a\} \notin A)$$

and the proof has the same pathology.

4 Sequent Calculus

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

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

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

$$\dfrac{\begin{array}{c} A \to (B \to C) \\ \vdots \end{array}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{26}$$

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

$$\dfrac{\dfrac{\vdots}{A \to C} \to\text{-int}}{(A \to B) \to (A \to C)} \to\text{-int} \tag{27}$$

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 26) 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 27, which said that $A \to C$ can be deduced from $A \to B$ and $A \to (B \to C)$] comes out as

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

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

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.

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

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

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

gives us something that looks in part like

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

and this means we are using a rule

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

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

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

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

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

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

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

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

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

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

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. (If you are a crazy germanophobe you might prefer the locution *principal formula*; i am no germanophobe and in any case 'principal' has too many uses already.)

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 $p \to \bot$.

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 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

multisets

consider it will be fairly clear how to move from one approach to the other and back.

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

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

$$A \vdash A$$

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

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

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

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

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

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.

Display this properly

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

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

- 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 9 *Find sequent proofs for the formulæ in exercise 3 (page 14). 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.

Need exercises here

4.1 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:

$$\frac{A, B \vdash C}{A \wedge B \vdash C}$$

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

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

(of course for us Δ is going to be empty) which is to say (with some relettering)

$$\frac{\Gamma \vdash A \quad \Gamma, B \vdash C}{\Gamma, A \to B \vdash C}$$

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

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

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

(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 $\Gamma, A \wedge B \vdash D$

The one exception is $\neg$-R. ($\neg$-L is OK because of *ex falso*.)

The top line of an occurrence of $\neg$-R is a sequent with no formula on the right. How can we prove sequents with nothing on the right? Such a sequent doesn't seem to mean anything. But remember that the false is the empty disjunction, so we can always think of a sequent with nothing on the right as having $\perp$ on the right. That saves the day.

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:

4.2 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,

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

then the allegation made by the lower sequent will be true too,

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

[*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 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.

Illustration: turn a proof with cut on the conjunction $A \wedge B$.

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta, A \wedge B}\wedge\text{-R} \qquad \dfrac{A, B, \Gamma \vdash \Delta}{A \wedge B, \Gamma \vdash \Delta}\wedge\text{-L}}{\Gamma \vdash \Delta}\textsc{cut}$$

into a proof with two cuts, on the conjuncts:

$$\dfrac{\dfrac{\Gamma \vdash \Delta, A \qquad A, B, \Gamma \vdash \Delta}{B, \Gamma \vdash \Delta}\text{CUT} \qquad \Gamma \vdash \Delta, B}{\Gamma \vdash \Delta}\textsc{cut}$$

of course we could have obtained the following proof

$$\dfrac{\dfrac{\Gamma \vdash \Delta, B \qquad B, A, \Gamma \vdash \Delta}{A, \Gamma \vdash \Delta} \qquad \Gamma \vdash \Delta, A}{\Gamma \vdash \Delta}$$

instead.

5 Two tips

5.1 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

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

[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.

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. numbering not working properly

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.

Exercise 10 *Find a proof of the sequent:*

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

5.2 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!

5.3 Exercises

You can now attempt to find sequent proofs for all the formulæ in exercise 3 page 14.

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 the Hilbert-style proofs you met in Part II.

6 Lambda Calculus and the Decoration of Formulæ

Talk about the cosmic pun.

6.1 The rule of $\to$-elimination

Consider the rule of $\to$-elimination

$$\frac{A \qquad A \to B}{B} \to\text{-elim} \tag{28}$$

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 functions that give you a member of B when fed a member of A. Thus we can decorate 28 to obtain

$$\frac{a : A \qquad f : A \to B}{f(a) : B} \to\text{-elim} \tag{29}$$

which says something like: "If a is in A and f takes As to Bs then $f(a)$ is a B.[9] 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 $\to$-int later. For the moment we will look at the rules for $\wedge$.

6.2 Rules for $\wedge$

6.2.1 The rule of $\wedge$-introduction

Consider the rule of $\wedge$-introduction:

$$\frac{A \qquad B}{A \wedge B} \wedge\text{-int} \tag{30}$$

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 banana then I don't have a thing that is both an apple and a banana; what I do have is a sort of plural object that I suppose is a pair of an apple and a banana. 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 30 is

[9]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.

$$\frac{a : A \qquad b : B}{\langle a, b \rangle : A \times B} \wedge\text{-int} \tag{31}$$

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* 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.

6.2.2 The rule of $\wedge$-elimination

If you can do them up, you can undo them: if I have a pair-of-an-A-and-a-B then I have an A and I 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.

(Do not 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 bananas above.)

6.3 Rules for $\vee$

To make sense of the rules for $\vee$ 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 $\wedge$, we have **disjoint union** to correspond with $\vee$. 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 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*. 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.

$\vee$-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:

6.4 Propagating Decorations

The first rule of decorating is to decorate each assumption with a variable, a thing with no syntactic structure: a single symbol.[10] 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!

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.

6.5 Rules for $\wedge$

6.5.1 The rule of $\wedge$-elimination

$$\frac{A \wedge B}{B}\ \wedge\text{-elim} \tag{32}$$

We decorate the premiss with a variable:

$$\frac{x : A \wedge B}{B}\ \wedge\text{-elim} \tag{33}$$

…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`'.

[10] 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.

$$\frac{x : A \wedge B}{\mathtt{snd}(x) : B}$$

By the same token we will need a gadget '`fst`' which gives the first component of an ordered pair so we can decorate[11]

$$\frac{A \wedge B}{A} \wedge\text{-elim} \tag{34}$$

to obtain

$$\frac{x : A \wedge B}{\mathtt{fst}(x) : A}$$

6.5.2 The rule of $\wedge$-introduction

Actually we can put these proofs together and whack an $\wedge$-introduction on the end:

$$\frac{\dfrac{x : A \wedge B}{\mathtt{snd}(x) : B} \quad \dfrac{x : A \wedge B}{\mathtt{fst}(x) : A}}{\langle \mathtt{snd}(x), \mathtt{fst}(x) \rangle : B \wedge A}$$

6.6 Rules for $\to$

The rule of $\to$-introduction

Here is a simple proof using $\to$-introduction.

$$\dfrac{\dfrac{[A \to B]^1 \qquad A}{B} \to\text{-elim}}{(A \to B) \to B} \to\text{-int (1)} \tag{35}$$

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 '$f(x)$':

$$\dfrac{\dfrac{f : [A \to B]^1 \quad x : A}{f(x) : B}}{??? : (A \to B) \to B}$$

[11] Agreed: it's shorter to write 'x_1' and 'x_2' than it is to write 'fst(x)' and 'snd(x)' but this would prevent us using 'x_1 and x_2' as variables and in any case I prefer to make explicit the fact that there is a function that extracts components from ordered pairs, rather than having it hidden it away in the notation.

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

$$\dfrac{\dfrac{f : [A \to B]^1 \qquad x : A}{f(x) : B}\ \to\text{-elim}}{\lambda f.f(x) : (A \to B) \to B}\ \to\text{-int (1)} \tag{36}$$

Thus, in general, an application of $\to$-introduction will gobble up the proof

$$\begin{array}{c} \underline{x : A} \\ \vdots \\ \hline t : B \end{array}$$

and emit the proof

$$\begin{array}{c} \underline{[x : A]} \\ \vdots \\ \hline t : B \\ \hline \lambda x.t : A \to B \end{array}$$

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 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 $t[y/x]$.

6.7 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! [12]

[12] For any gluttons for punishment out there here is a message from my former student Nick Benton. "$\vee$-elim goes to a generalization of `if-then-else` called "`case`":

Exercise 11 *Go back and look at the proofs that you wrote up in answer to exercise 1, and decorate those that do not use '$\vee$'.*

6.8 Remaining Rules

6.8.1 Identity Rule

Here is a very simple application of the identity rule.

$$\dfrac{\dfrac{\dfrac{A \quad B}{B}}{B \to A}}{A \to (B \to A)}$$

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.

6.8.2 The *ex falso*

The *ex falso sequitur quodlibet* speaks of the propositional constant $\perp$. To correspond to this constant *proposition* we are going to need a constant *set*. 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!! That takes care of $\perp \to A$, the *ex falso*. In fact it would seem that *any function whatever*[13] is a function from the empty set to A!)

$$\dfrac{\vdash E : A + B \qquad x : A \vdash M : C \qquad y : B \vdash N : C}{\vdash \texttt{case}\ E\ \texttt{of}\ \mathrm{inl}(x) \Rightarrow M | \mathrm{inr}(y) \Rightarrow N : C}$$

Note 'x' bound in M, 'y' in N. The operational behaviour is to evaluate E, see if it turns into $\mathrm{inl}(a)$ for some $a \in A$ and if so evaluate M with 'x' bound to a, otherwise the symmetric thing. `if-then-else` is morally the special case where A and B are both just 1, the one element type, though binding a variable to a value of type 1 is a bit of a waste of time, so we simplify the syntax. Haskell, ML etc have `case in`, and that's what it's called there too, but they generalize the forms of pattern matching somewhat."

[13]We might have to be *very* careful with the small print here. Since we want the cardinal number 0^0 to be 1 we would prefer there to be only one function from the empty set to itself.

6.8.3 Double Negation

What are we to make of $A \to \bot$? 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?

$$((A \to \bot) \to \bot) \to A$$

- If A is empty then $A \to \bot$ is the set of all functions from the empty set to itself and contains at the very least the empty function and is not empty. So $(A \to \bot) \to \bot$ is the set of functions from a nonempty set to the empty set and is therefore the empty set, so $((A \to \bot) \to \bot) \to A$ is the set of functions from the empty set to the empty set and therefore contains at least the empty function, so it is at any rate nonempty.

- However if A is nonempty then $A \to \bot$ is empty. So $(A \to \bot) \to \bot$ is the set of functions from the empty set to the empty set and is nonempty as before)—so $((A \to \bot) \to \bot) \to A$ is the set of functions from a nonempty set (probably the singleton of the empty function) to the nonempty set A, and is nonempty; at the very least it contains Ka for every $a \in A$.

update chch-lectures at some point

So $((A \to \bot) \to \bot) \to A$ is not reliably inhabited, in the sense that it's inhabited but we don't know what by! 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. $((A \to \bot) \to \bot) \to A$ has no λ-term because it has no proof! There are two things that look as tho' they are λ-terms which might correspond to proofs, but you don't know which. This is where the correspondence breaks down.

We will revisit this subject in section **??**.

6.9 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 12 *Find λ-terms for*

1. $(A \wedge B) \to A$;
2. $((A \to B) \wedge (C \to D)) \to ((A \wedge C) \to (B \wedge D))$;
3. $(A \to B) \to ((B \to C) \to (A \to C))$;
4. $((A \to B) \to A) \to ((A \to B) \to B)$;
5. $(A \to (B \to C)) \to (B \to (A \to C))$;
6. $(A \to (B \to C)) \to ((B \wedge A) \to C)$;

7. $((B \wedge A) \to C)) \to (A \to (B \to C))$;

Finding λ-terms in exercise 12 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)

$$\begin{prooftree}
\AxiomC{$[A \to A]^1$}
\RightLabel{identity rule}
\UnaryInfC{$A \to A$}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$(A \to A) \to (A \to A)$}
\end{prooftree} \tag{37}$$

$$\begin{prooftree}
\AxiomC{$[A]^1$}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$A \to A$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$(A \to A) \to (A \to A)$}
\end{prooftree} \tag{38}$$

$$\begin{prooftree}
\AxiomC{$[A]^1$}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$A \to A$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$(A \to A) \to (A \to A)$}
\end{prooftree} \tag{39}$$

$$\begin{prooftree}
\AxiomC{$[A]^1$}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$A \to A$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$(A \to A) \to (A \to A)$}
\end{prooftree} \tag{40}$$

$$\begin{prooftree}
\AxiomC{$[A]^1$}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\AxiomC{$[A \to A]^2$}
\RightLabel{$\to$-elim}
\BinaryInfC{A}
\RightLabel{$\to$-int (1)}
\UnaryInfC{$A \to A$}
\RightLabel{$\to$-int (2)}
\UnaryInfC{$(A \to A) \to (A \to A)$}
\end{prooftree} \tag{41}$$

Exercise 13 *Decorate all these proofs with λ-terms. If you feel lost, you might like to look at the footnote[14] for a HINT.*

On successful completion of exercise 13 you will be in that happy frame of mind known to people who have just discovered **Church numerals**.

7 Half of a Completeness theorem

Our concept of constructively correct formula arose from syntax, from a notion of constructively acceptable proof. This is in contrast to the classical notion, which arose semantically, as in "true in all rows of the truth-table". However there are semantic characterisations of constructive correctness and we shall work towards them.

It should be evident from the preceding discussion that

Theorem 2 *Every constructively correct propositional formula has a lambda-term corresponding to it.*

(Well, here we take 'lambda term' in an extended sense, where we are allowed pairing and unpairing.)

This will set us up for Scott's [5] cute proof that Peirce's Law is not constructively correct, which we now exhibit.

Remark 1 *(Scott)*

Peirce's Law is not constructively correct.

Proof:

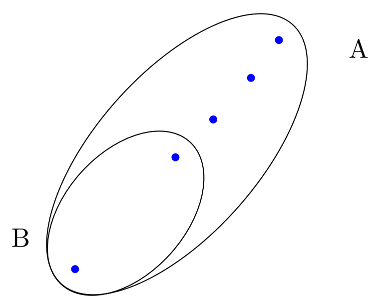
The attribution to Dana Scott here is this particular proof; the result itself is a lot older.

How do we prove that there is no lambda term for Peirce's law? Here we trade on the fact that a lambda term does not know what sets it is acting on: it is uniformly definable. Now uniform definability is clearly going to have something to do with invariance under permutations acting inside the sets we are considering . . . but what exactly do we mean by invariance? We need to get straight what it is in general for a permutation of A to act on some complex construct involving A and other things, and this we do by recursion on the structure of the complex construct. For $\pi \in \mathrm{Symm}(A)$, π acts on A as itself, and on any other atom it acts as the identity. How does π act on $X \to Y$? Clearly it must send $f \in X \to Y$ to $\{\langle \pi(x), \pi(y)\rangle : \langle x, y\rangle \in f\}$ where $\pi(x)$ is what the induced action of π does to x, told us by the recursion. 'Invariant' means *fixed by this action*. Observe that any λ-term is invariant in this sense.

[14] Notice that in each proof of these proofs all the occurrences of '$A \to A$' are cancelled simultaneously.. Look at the footnote on page 33.

Now for Peirce's Law: $((A \to B) \to A) \to A$.

Suppose *per impossibile* that there were a uniformly definable (and, accordingly, invariant) function P for Peirce's law. The idea is to cook up sets A and B such that the existence of such a global P led to Bad Things. Let B be a two-membered set, and let A be obtained from B by adding three new elements.



The pigeonhole principle now tells us that, for any function $f : A \to B$, there is a unique $b \in B$ such that $|f^{-1}``(\{b\}) \cap (A \setminus B)| \geq 2$. (A unique member of B that is hit by at least two members of $A \setminus B$). This defines a function from $A \to B$ to B, which is to say (since $B \subseteq A$) a function from $A \to B$ to A. Let us call this function F. F exists only because of the special circumstances we have here contrived, and it's not the sort of thing that P would normally expect to have to deal with, so we should expect P to experience difficulty with it ... which of course is what we want! At all events we must have $P(F) \in A$. In fact we can show that $P(F) \in B$. For suppose *per impossibile* that $P(F) = a$, for some $a \in A \setminus B$; with a view to obtaining a contradiction let π be a 3-cycle moving everything in $A \setminus B$ while fixing everything in B. We have

$P(F) = a,$	so
$\pi(P(F)) = \pi(a)$	which is to say
$\pi(P)(\pi(F)) = \pi(a)$	but P is fixed, whence
$P(\pi(F)) = \pi(a).$	

To obtain the desired contradiction we have to show that $\pi(F) = F$. We have $\pi(F) = \pi^{-1} \cdot F \cdot \pi$ by the recursion. So, for all $f : A \to B$, we obtain

$$\pi(F)(f) = (\pi^{-1} \cdot F \cdot \pi)(f) = (\pi^{-1} \cdot F)(\pi(f)) = \pi^{-1}(F(\pi(f))) =^{(1)} \pi^{-1}(F(f)) =^{(2)} F(f).$$

The first three equations hold by unravelling the recursion.

(1) holds as follows. $\pi(f) = \pi^{-1} \cdot f \cdot \pi$ and this is the same as $f \cdot \pi$ since π fixes both things in the range of f. Similarly $F(f \cdot \pi)$ must be the same as $F(f)$, since F looks only at the *range* of its argument not its *domain*, and everything in the range of F is fixed.

(2) holds because the output of F is in B, and π fixes both things in B.

That is to say, for all $f : A \to B$, $\pi(F)(f) = F(f)$; whence $\pi(F) = F$, giving

$$a \;=\; P(F) \;=\; P(\pi(F)) \;=\; \pi(P)(\pi(F)) \;=\; \pi(a) \;\neq\; a,$$

and the contradiction tells us that $P(F)$ was not in $A \setminus B$; it must have been in B as claimed.

So $P(F) \in B$. But this now means that we have a uniform way of finding a distinguished element in any two-membered set B. Simply add three new elements to B to obtain A, apply P to F to obtain a member of B; then throw away the new elements. In fact we have inferred the axiom of choice for sets of pairs! This is clearly absurd. The axiom of choice for pairs may be true, but it cannot be inferred from first principles. ■

8 Making Classical sense of Constructive Logic: Possible World Semantics

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.

The constructive conditional is not truth-functional: the truth-value of $A \to B$ depends on more than just the truth-values of A and of B. If we still want—in this new setting—to put to good use the idea of valuations giving truth-values to complex formulæ *despite* the new notions not being truth-functional (and, trust me, we do) then if we want to know what $v(A \to B)$ is we need more information than just $v(A)$ and $v(B)$. "Just what?" one might ask. One way of making sense of having valuations in this context of non-truth-functional connectives is to allow oneself to consult $v'(A)$ and $v'(B)$ where v' is some other valuation related somehow to v. 'Related somehow' introduces another degree of freedom to the semantics, another gadget. In principle this relation-between-valuations (always called the **accessibility relation**) could be absolutely anything under the sun, and in the full generality in which this kind of semantics can get studied it can, indeed, be anything under the sun. However in the particular setting of interest to us, with constructive logic, the relation is very specific.

We take our valuations to be functions from propositional letters to truth-values, as before, but this time they are allowed to be partial. If we think of these partial functions as sets of ordered pairs (a nasty unmathematical concretisation!) then it is easy to say what the accessibility relation is: it is simply $\subseteq$, set inclusion. However—confusingly—we write it '$\leq$'.

Our valuation-oriented semantics for constructive propositional logic now runs more-or-less like the classical case but with a few enhancements. We define a satisfaction relation (often written '$\models$') between valuations and complex formulæ by recursion

> When p a propositional letter we say $v \models p$ if $v(p) = 1$; $v(p) \models \neg p$ if $v(p) = 0$.
>
> If A and B are complex formulæ
>
> $v \models A \wedge B$ iff $v \models A$ and $v \models B$;
>
> $v \models A \vee B$ iff $v \models A$ or $v \models B$.

Now we see the violence inherent in the system:

> $v \models A \to B$ iff $(\forall v' \geq v)$(if $v \models A$ then $v \models B$).

We observe that $v \not\models \bot$, for all valuations v. Constructively we take $\neg A$ to be $A \to \bot$ so, altho' we determine whether or not $v \models \neg p$ by stipulation when p is a propositional letter, the question of whether or not $v \models \neg A$ when A is "molecular" is determined by an appeal to the recursion.

A "possible world model" for a bundle of formulæ over a propositional alphabet P is now a subset of the set $P \to \{\text{true, false}\}$ of all the partial valuations of P, with the accessibility relation of set inclusion $\subseteq$ written as $\leq$ as signalled above. It will have a $\leq$-minimum element, which will typically be the empty valuation. We say that a complex formaula A is true in such a model if it is satisfied by that minimum valuation.

In a more general setting we say:

Definition 2 *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 atomic formulæ, written '$W \models \phi$', subject to the stipulation that $W \models \bot$ never holds*[15]*;*
- *There is a* **designated** *(or 'actual' or 'root') world W_0^M.*

We may stipulate **persistence** of $\models$, namely that if ϕ is atomic, $W \models \phi$ and $W \leq W'$, then $W' \models \phi$. In the constructive case, from which we are generalising, persistence is clearly enforced by the partial ordering of the partial valuations.

Persistence is such an important idea that we'd better define it.

[15]Strictly speaking we do not stipulate this feature (we can't). It is our settled and unmoveable intention; we secure it by declaring the recursion in such a way that we can prove by induction that no world ever believes $\bot$.

Definition 3
We will say ϕ is **persistent** *if whenever $W \models \phi$ then $(\forall W' \geq W)(W' \models \phi)$.*

Later we will extend the concept of persistence to complex formulæ.

Next $\models$ is extended to a relation between worlds and arbitrary formulæ by recursion:

Definition 4

1. *$W \models A \wedge B$ iff $W \models A$ and $W \models B$;*
2. *$W \models A \vee B$ iff $W \models A$ or $W \models B$;*
3. *$W \models A \to B$ iff every $W' \geq W$ that $\models A$ also $\models B$;*
4. *$W \models \neg A$ iff there is no $W' \geq W$ such that $W' \models A$;*
5. *$W \models (\exists x)A(x)$ iff there is an x in W such that $W \models A(x)$;*
6. *$W \models (\forall x)A(x)$ iff for all $W' \geq W$ and all x in W', $W' \models A(x)$.*

Then we say

$$\mathfrak{M} \models A \quad \textit{if} \quad W_0^M \models A.$$

4 is a special case of 3: $\neg A$ 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 present purposes (tho' not in more general settings of the kind we are not currently interested in) that it is **transitive** and **reflexive**. Just for the record we note that '$A \leq B$' will sometimes be written as '$B \geq A$'.

[There is a subtlety here which we can probably safely ignore. In the general setting we can have distinct worlds which correspond to the same partial valuations. The way i introduced possible world semantics for constructive logic this can't happen. The way to finesse this is by introducing extra variables that do not appear in the formulæ for which we are doing semantics.]

End of digression to the general case!

The reader will observe that in the one case (the recursion for '$\to$') where we exploit the accessibility relation we exploit it in one direction only: for the truth-value of $A \to B$ at W we consult worlds related to W, not worlds to which W is related. In principle there would be nothing to prevent us declaring a recursive step for a connective £ s.t. $W \models A£B$ iff every world to which W is related that believes A also believes B but nobody ever uses accessibility that way round.

If our accessibility relation is a quasiorder (as it is in the constructive case) then truth-values of formulæ in "later" worlds are not controlled by truth-values in earlier worlds (they never get consulted)

can use this to motivate a root world. Chat to be supplied here

Theorem 3 *All formulæ are persistent*[16].

Proof:

We have taken care of the atomic case by stipulation. 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 $W' \models \neg\phi$. But W'' was arbitrary.

$\bigvee$ 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 $W'' \models (\forall x)(\phi(x))$.

$\rightarrow$ Finally suppose $W \models (A \rightarrow B)$, and $W' \geq W$. We want $W' \models (A \rightarrow B)$. 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 $A \rightarrow B$.

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. There are other more general settings where this assumption is not made, but (since our mission is constructive logic, where the accessibilty relation is transitive) we will not consider them here.

Now we can use persistence to show that this possible world semantics always makes $A \rightarrow \neg\neg A$ true. Suppose $W \models A$. 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 $W \models \neg\neg A$.

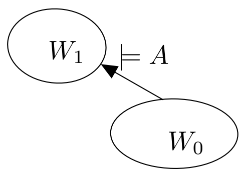
[16]This holds when [as here] we are using possible worlds to give semantics for *constructive* logic, and it follows from persistence for atomics. If, as more generally, we do *not* assume persistence for atomics, then of course persistence for complex formulæ does not follow.

8.1 Some Worked Examples

Challenge 8.1.1: Find a countermodel for $A \vee \neg A$

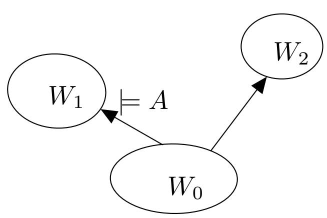
The first thing to notice is that this formula is a classical (truth-table) tautology. This means that any countermodel for it must contain more than one world.

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 8.1.2: 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 8.1.3: 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 8.1.2.

Challenge 8.1.4: Find a countermodel for $((A \to B) \to A) \to A$

You may recall from exercise 6 on page 17 that on Planet Zarg this formula is believed to be false[17]. 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 2)

$$(\forall W' \geq W)[W' \models X \;\; \to \;\; W' \models Y] \tag{42}$$

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 2 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 $(\exists W' \geq W_0)[W' \models A$ & $\neg(W' \models B)]$.)

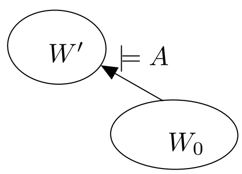
So our countermodel contains two worlds W_0 and W', with $W_0 \leq W'$. $W' \models A$ but $\neg[W_0 \models A]$, and $\neg[W' \models B]$.

[17] I have just corrected this from "You may recall from exercise 6 on page 17 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 8.1.1.



Challenge 8.1.5: Find a model that satisfies $(A \to B) \to B$ but does not satisfy $(B \to A) \to A$

We must have

$$W_0 \models (A \to B) \to B \tag{1}$$

and

$$\neg[W_0 \models (B \to A) \to A] \tag{2}$$

By (2) we must have $W_1 \geq W_0$ such that

$$W_1 \models B \to A \tag{3}$$

but

$$\neg[W_1 \models A] \tag{4}$$

We can now show

$$\neg[W_1 \models A \to B] \tag{5}$$

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

$$W_2 \models A \tag{6}$$

and

$$\neg[W_2 \models B] \tag{7}$$

From (7) and persistence we infer

$$\neg[W_1 \models B] \tag{8}$$

and

$$\neg[W_0 \models B] \tag{9}$$

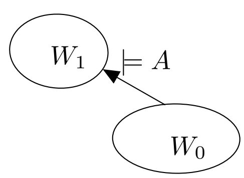
Also, (4) tells us

$$\neg[W_0 \models A]. \tag{10}$$

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:



8.2 Exercises

Exercise 14 *Return to Planet Zarg!*[18]

The truth-tables for Zarg-style connectives are on p 17.

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.

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 8.1.5) We will develop this thought in section 9.

Exercise 15

[18] Beware: Zarg is a planet not a possible world!

1. *Find a model that satisfies* $p \to q$ *but not* $\neg p \vee q$.

2. *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* $\neg p \vee \neg\neg p$.

3. *Find a model that satisfies* $(p \to q) \to q$ *but does not satisfy* $p \vee q$.

4. *Find a model that satisfies* $A \to (B \vee C)$ *but doesn't satisfy* $(A \to B) \vee (A \to C)$[19].

5. *Find a model that satisfies* $(A \to C) \wedge (B \to D)$ *but doesn't satisfy* $(A \to D) \vee (B \to C))$[20].

6. *Find a model that satisfies* $\neg(A \wedge B)$ *but does not satisfy* $\neg A \vee \neg B$

7. *Find a model that satisfies* $(A \to B) \to B)$ *and* $(B \to A) \to A$ *but does not satisfy* $A \vee B$.

8. *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* $A \vee B$. *Find a world that believes* $((A \to B) \to B) \wedge ((B \to A) \to A)$ *but does not believe* $A \vee B$.

9. *Is the following expression constructively correct? If it is then supply a natural deduction proof and a sequent proof. If it isn't then supply a countermodel and extract the corresponding Heyting Algebra.*

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

Exercise 16 *Find countermodels for:*

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

2. $(\forall x)(A(x) \to (\exists y)(B(x,y))) \to (\forall x)(\exists y)(F(x) \to B(x,y))$;

3. $(\exists x)(\forall y)(F(y) \to F(x))$;

4. $\neg(\forall x)\neg(\forall y)(F(y) \to F(x))$;

5. $(\exists x)(\forall y)\neg(F(y) \wedge \neg F(x))$;

Exercise 17 *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. $(\forall x)(F(x))$;

[19] You saw a fallacious attempt to prove this inference on page 13.

[20] 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".

2. $(\exists x)(\neg F(x))$*;*

3. $\neg\exists x \neg F(x)$*;*

4. $\neg\neg(\exists x)(\neg F(x))$.

Exercise 18 *Hard!!*

(i) Find a countermodel for $\neg\neg(\neg(\forall x)(F(x)) \rightarrow (\exists x)(\neg F(x)))$. *You may need the hint that in possible world semantics there is no overriding assumption that there are only finitely many worlds.)*

(ii) Find a countermodel for $\neg\neg((\forall x)(\neg\neg F(x)) \rightarrow \neg\neg(\forall x)(F(x)))$

The significance of this last example is that altho' it is the double negation of a classically valid formula it is nevertheless not itself constructively correct.

9 Heyting Algebras

In any possible world model, each possible world represents a decision about which primitive propositions are true. That is to say, a possible world is nothing more than a partial valuation on the primitive propositions under consideration. If we think of a valuation as a set of ordered pairs $\langle p, t \rangle$, where t is 0 or 1 and p is a primitive proposition (a letter), then the accessibility relation on the set of possible worlds is the subset relation among the corresponding partial valuations.

This puts a bound on the number of possible world models we can have, or need to consider. If we have n primitive propositions then we have 2^n valuations and 3^n partial valuations, and therefore 2^{3^n} sets of partial valuations.

Given a possible world model, we can think of the truth value $[[\phi]]$ of ϕ in that model as the set of worlds in which it is true. What can we say about these truth-values? Recall the phenomenon of **persistence.** A truth value must be an upper-closed subset of the quoset of worlds quasiordered by the accessibility relation. So our truth-values look like upper sets in quasiorders. What do we know about the poset of upper sets in a quasiorder? What structure can we infer?

Thus pitchforks us into the following definition.

Definition 5 *A Heyting Algebra is a distributive lattice with a* $\bot$ *which is an annihilator*[21] *for* $\wedge$*, and an extra operation* $\Rightarrow$ *s.t.* $a \Rightarrow b$ *is the* $\leq$*-maximum element* c *s.t.* $(a \wedge c) \leq b$.

There are other definitions in the literature. The way to navigate your way through them is to remember that (for us, at least) Heyting algebras arise as the posets of upper-sets-in-a-quasiorder.[22]

[21] So $(\forall x)(x \wedge \bot = \bot)$.

[22] "There are nine and sixty ways of constructing tribal lays, And every single one of them is right!"—Kipling

At all events we have to conceptualise Heyting Algebras in such a way that they have—either by stipulation as primitives or secondarily by definition—operations corresponding to the propositional connectives. This is so that we can "build truth-tables" with them.

Then we can say what it is for a Heyting algebra to "accept" a formula.

We'd better check that the operations of the Heyting algebra correspond nicely to the connectives in the Logic, as they do in the classical case.

(We need to distinguish between the '$\to$' of the logic and the '$\Rightarrow$' of the Heyting algebra; we already distinguish '$\wedge$' from '$\cap$' and '$\vee$' from '$\cup$'.)

It is a simple matter to check that

$[[A \vee B]] = [[A]] \cup [[B]]$,

$[[A \wedge B]] = [[A]] \cap [[B]]$ and

$[[\bot]] = \emptyset$.

Significantly harder is:
$[[A \to B]] = [[A]] \Rightarrow [[B]]$.

$$\begin{array}{llll}
(1)\ [[A \to B]] & = & \{W : W \models A \to B\} & \\
(2) & = & \{W : (\forall W' \geq W)(W' \not\models A \ \vee \ W' \models B)\} & \text{Expand '}\to\text{'} \\
(3) & = & \{W : (W \models A)\} \Rightarrow \{W : (W \models B)\} & \\
(4) & = & [[A]] \Rightarrow [[B]] &
\end{array}$$

The inference from (2) to (3) needs to be spelled out. We use extensionality. W is a member of the class in (2) iff $(\forall W' \geq W)(W' \not\models A \ \vee \ W' \models B)$. But this means that the principal upper set generated by $\{W\}$ consists exclusively of worlds which, if they believe A, also believe B. So the class in (2) is the union of all such upper sets, which is precisely to say that it is $\{W : W \models A\} \Rightarrow \{W : W \models B\}$.

Heyting algebras stand to constructive logic somewhat the way Boolean algebras stand to classical logic. A propositional formula is classically a tautology iff it is accepted by every Boolean algebra. (see my Part II notes on [3]). It will follow from theorem 4 that a propositional formula is constructively correct iff it is accepted by every Heyting Algebra. An important difference is that all boolean algebras accept the same formulæ, so being accepted by all boolean algebras is the same as being accepted by even one.

9.1 A Completeness Theorem

The reader may have the—reasonable—suspicion that this constructive logic lark is just a lot of ill-motivated nit-picking and not a serious part of mathematics. It has to be said that this suspicion is entirely reasonable because one way or another there has been quite a lot of ill-motivated nit-picking in Logic over the last 100 years or so. Name no names, least said soonest mended. However, constructive logic is not a lot of ill-motivated nit-picking but a substantial piece of mathematics, and we will underpin this bald assertion with a completeness theorem that is an analogue of the completeness theorem for classical logic that you saw in Part II.

In this section we tie together constructive proof, many-valued truth-tables (aka Heyting Algebras) and possible world semantics.

Lemma 1 *If there is a sequent proof of $\Gamma \vdash \phi$ satisfying the only-one-formula-on-the-right condition then every possible world model that satisfies everything in Γ also satisfies ϕ.*

Proof:

Suppose there is a sequent proof $\mathcal{D}$ of $\Gamma \vdash \phi$ satisfying the only-one-formula-on-the-right condition. We obtained this sequent by applying one of the rules to one (or possibly two) "simpler" sequents for which sequent proofs (satisfying the only-one-formula-on-the-right condition) can be found, and which, therefore (by induction hypothesis) hold in every possible world structure. So the proof is by induction on a simplicity relation (under which the upper sequents in every rule are simpler than the lower sequents). So all we have to do is show that every sequent rule preserves the property of being-satisfied-by-every-possible-world-model.

The Base Case

Let $\Gamma \vdash A$ be a sequent that does not match the output of any sequent rule (so there is no inductive descent available). Nevertheless it is a valid sequent. How can this be? It must be an initial sequent. Indeed, there being only one formula—A—on the right, it must be that $A \in \Gamma$. But clearly every initial sequent is true in every possible world structure.

There are several inductive cases to consider, depending on what the last step in $\mathcal{D}$ was.

$\vee$-R, $\wedge$-R and $\wedge$-L, $\vee$-R and $\vee$-L

These are routine. We illustrate with $\vee$-R.

Suppose our sequent was obtained by $\vee$-R. So it must be $\Gamma \vdash A \vee B$, and it was obtained from $\Gamma \vdash A$ or from $\Gamma \vdash B$

By induction hypothesis, **if** there is a sequent proof of $\Gamma \vdash A$ (resp. $\Gamma \vdash B$) satisfying [...] **then** every possible world model that satisfies everything in Γ satisfies A (resp. B)

So assume there is a sequent proof of $\Gamma \vdash A \vee B$ obtained by $\vee-$R. Then there is a sequent proof of $\Gamma \vdash A$ (or a sequent proof of $\Gamma \vdash B$). So (by induction hypothesis) either every possible world model satisfying everything in Γ satisfies A, or every possible world model satisfying everything in Γ satisfies B.

Either way we infer that every possible world model satisfying everything in Γ satisfies $A \vee B$.

→-R

This seems to be the only case where we need the single-formula-on-the-right condition.

If our sequent was obtained by $\to -\text{R}$ then it must be $\Gamma \vdash A \to B$ and was obtained from $\Gamma, A \vdash B$.

The induction hypothesis is that if there is a sequent proof of $\Gamma, A \vdash B$ satisfying [...] then every possible world model that satisfies $\Gamma \cup \{A\}$ also satisfies B.

We want to show that if there is a sequent proof of $\Gamma \vdash A \to B$ then every possible world model satisfying Γ also satisfies $A \to B$.

Under the assumption in play here (that our sequent was obtained by $\to -\text{R}$) we know that there is a sequent proof of $\Gamma, A \vdash B$ and by induction hypothesis this implies that every possible world that satisfies $\Gamma \cup \{A\}$ also satisfies B.

We want to infer from this that every possible world model satisfies $\Gamma \vdash A \to B$.

Possible worlds satisfying Γ come in two flavours: those that satisfy A and those that don't.

(i) Let W_0 be the root world of a possible world model that satisfies $\Gamma \cup \{A\}$. By persistence every W' in this model also satisfies $\Gamma \cup \{A\}$. So by induction hypothesis all these W' also satisfy B. So every $W' \geq W_0$ that satisfies A also satisfies B. So by the $\to$ clause in the recursive definition of $\models$ we can infer that $W_0 \models A \to B$. But W_0 was the root world of an arbitrary model that satisfies $\Gamma \cup \{A\}$.

(ii) The other case is where $W_0 \not\models A$. If no $W' \geq W_0$ believes A then vacuously every world $\geq W_0$ that believes A also believes B, so $W_0 \models A \to B$ as desired. But what if there is a $W' > W_0$ that *does* believe A? But then, by the analysis in case (i), $W' \models B$ so it's still the case that every world $\geq W_0$ that believes A also believes B, so $W_0 \models A \to B$ as desired.

→-L

Suppose our sequent was obtained by →-L. So it must be $\Gamma, A \to B \vdash C$, and came from $\Gamma, B, \vdash C$ and $\Gamma \vdash A$. The induction hypothesis will tell us that if there is a sequent proof of $\Gamma \cup \{B\}$ satisfying [...] then any possible world model that believes $\Gamma \cup \{B\}$ also believes C and that if there is a sequent proof of $\Gamma \cup \{B\}$ satisfying [...] then any possible world model that believes Γ also believes A.

So suppose there is a sequent proof of $\Gamma, A \to B \vdash C$. Then (beco's the last rule used was $\to -\text{L}$) there are sequent proofs of $\Gamma, B, \vdash C$ and $\Gamma \vdash A$ ... from which it follows by induction hypothesis that any possible world model that believes $\Gamma \cup \{B\}$ also believes C and that any possible world model that believes Γ also believes A.

We wish to infer that any possible world model that believes Γ and $A \to B$ must believe C. But any possible world that believes Γ also believes A, as we have just seen, and it believes $A \to B$ so it believes B. But we were told that any possible world that believes Γ and B also believes C.

¬-L and ¬-R

Can be treated as →-L and →-R, but we might write out the details later.

■

THEOREM 4 *Given a propositional formula ϕ the following are equivalent.*
(i) ϕ has a λ-term;
(ii) ϕ is satisfied by all Heyting algebras;
(iii) ϕ is true in all possible world models;
(iv) ϕ has a sequent proof satisfying the one-formula-on-the-right condition, or a constructive natural deduction proof.

Further, if ϕ belongs to the implicational fragment the above are equivalent to
(v) ϕ is derivable from K and S;

Proof:

(v) implies (iv). Recall the proof of the Deduction Theorem from Part II. Observe that it needs K and S but not axiom 3.

(i) implies (v) From a lambda-term one can obtain an **SKI** combinator. See, for example, [7]. From a combinator term we can obtain a Hilbert-style deduction from the two axioms K and S.

(iv) implies (i)

So (i), (iv) and (v) are equivalent. Well only for formulæ in the implicational fragment. Need (iv) implies (v).

(ii) → (iii) We prove the contrapositive. Any possible world countermodel for ϕ gives rise to a Heyting algebra that rejects ϕ.

(iii) → (ii)

Again, we prove the contrapositive. Suppose $\mathcal{H}$ is a Heyting algebra and our formulæ are taking truth-values in $\mathcal{H}$, and $\mathcal{H}$ refutes ϕ. We will obtain a possible world model that refutes ϕ.

By the representation theorem for distributive posets[23] every Heyting algebra is isomorphic to a Heyting algebra whose order relation is $\subseteq$, set-inclusion. (I lectured this in Part II in 2016/7—see [3].) So without loss of generality we may assume that the order relation of $\mathcal{H}$ is set-inclusion. So the elements of $\mathcal{H}$ are all subsets of some set V. The worlds of the possible world model we are building are going to be the members of V. For an atomic formula A and a world $W \in V$ we declare that $W \models A$ iff $W \in [[A]]$. Finally we define the accessibility relation $\leq$ by $W \leq W'$ iff $(\forall H \in \mathcal{H})(W \in H \to W' \in H)$. This

[23] This representation theorem for distributive posets is an old favourite of Professor Johnstone's: look at his old example sheets and old Tripos questions. It will do you no harm to prove it: you may use Zorn's lemma.

enforces persistence: if (i) $W \models A$ and (ii) $W \leq W'$ then (i) gives $W \in [[A]]$ which (with (ii)) implies $W' \in [[A]]$, which is to say $W' \models A$. Clearly if $[[\phi]]$ is not the top element of $\mathcal{H}$ then there will be worlds that do not believe ϕ, so in particular the root world does not believe ϕ.

So (ii) and (iii) are equivalent. So we have to join up the triple (i)-(iv)-(v) with the pair (ii)-(iii)

Lemma 1 gives us that (iv) implies (iii). (And (v) easily implies (ii) and (iii) by induction on proofs.)

So the triple implies the pair. It will suffice to show that (ii) $\wedge$ (iii) implies one of (i), (iv) and (v).

I think our best chance is to prove (iii) implies (iv):

Specifically we prove: Let $\mathfrak{M}$ be a possible world model; we prove by induction on natural deduction 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 $\mathcal{D}$.

Proof:

Since (as we established earlier) there is a proof of the sequent $\Gamma \vdash \phi$ obeying the one-formula-on-the-right–constraint iff there is a natural deduction proof of ϕ all of whose assumptions are in Γ, we shall regard these two assertions about $\Gamma \vdash \phi$ as interchangeable in the following proof.

$\rightarrow$-int

The induction hypothesis will be that every world that believes A (and the other premisses in $\mathcal{D}$) also believes B. Now let W be a world that believes all the other premisses in $\mathcal{D}$. Then certainly (by persistence) every $W' \geq W$ also believes all the other premisses in $\mathcal{D}$, so any such W' that believes A also believes B. But that is to say that any world that believes all the other premisses in $\mathcal{D}$ also believes $A \rightarrow B$.

Ex Falso Sequitur Quodlibet

Suppose the last line is B. So we have a deduction $\mathcal{D}'$ whose conclusion is $\perp$. By induction hypothesis every world that satisfies the assumptions of $\mathcal{D}'$ must satisfy the conclusion, namely $\perp$. But if they satisfy $\perp$ they surely satisfy B.

$\vee$-elim

Then one of the premisses is a disjunction $A_1 \vee \ldots A_n$, and there are proofs $\mathcal{D}_i$ of a conclusion B, say, from A_i. By induction hypothesis, for each i, any world W that believes the assumptions of $\mathcal{D}_i$ believes the conclusion B. But if W believes the disjunction $A_1 \vee \ldots A_n$ it must believe one of them and must therefore believe B.

There are some other cases ...

■

Exercise 19 *Look again at the possible world countermodels you found for the propositional formulæ in earlier exercises, and the worked examples, and extract Heyting Algebras from them.*

Exercise 20 *Prove the following*

1. *For any topological space $\mathcal{T}$, the poset of open sets of $\mathcal{T}$ under inclusion is a Heyting algebra.*
2. *Every Heyting algebra is (iso to) the poset of open sets of some topological space.*
3. *For any topological space $\mathcal{T}$, the poset of regular open sets of $\mathcal{T}$ under inclusion is a Boolean algebra.*
4. *Every Boolean algebra is (iso to) the poset of regular open sets of some topological space.*

10 Making Constructive Sense of Classical Logic: the Negative Interpretation

The way the constructive logician narrates this situation is something like the following. Here *grokking* is a propositional attitude [24] whose precise nature is known at any rate to the constructive logician but possibly not to anyone else. The constructive logician muses:

> "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 most certainly is an x which I grok to be W.

[24] A *propositional attitude* is any relation between an agent and a proposition: knowledge, belief, hope ... *grokking* comes from [6].

> 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 $A \to B$ 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 $\neg\neg A \to A$! Accordingly henceforth whenever I hear them assert $A \to B$ I shall mentally translate this into $\neg(A \wedge \neg B)$. That covers the $\neg\neg A \to A$ 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.

Definition 6 *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*	$(\phi^* \wedge \psi^*)$;
$(\phi \to \psi)^*$	*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 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 7 *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 2 *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.[25]

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[p]^2 \qquad [p \to q]^1}{q}\ \to\text{-elim}}{(p \to q) \to q}\ \to\text{-int }(1) \qquad [((p \to q) \to q) \to q]^3}{q}\ \to\text{-elim}}{p \to q}\ \to\text{-int }(2)}{(((p \to q) \to q) \to q) \to (p \to q)}\ \to\text{-int }(3) \tag{43}$$

…noting that $\neg p$ is just $p \to \bot$.

$\bigwedge$:

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 $\neg\neg(p \wedge q)$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[p \wedge q]^1}{p}\ \wedge\text{-elim} \qquad [\neg p]^2}{\bot}\ \to\text{-elim}}{\neg(p \wedge q)}\ \to\text{-int }(1) \qquad \neg\neg(p \wedge q)}{\bot}\ \to\text{-elim}}{\neg\neg p}\ \to\text{-int }(2) \tag{44}$$

and the derivation of $\neg\neg q$ from $\neg\neg(p \wedge q)$ is of course precisely analogous.

But both p and q are stable by induction hypothesis, so we can deduce both p and q and thence $p \wedge q$.

[25] This was part 6 of exercise 3 on page 14.

$\forall$:

First we show $\neg\neg\forall \to \forall\neg\neg$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[(\forall x)\phi(x)]^1}{\phi(a)}\ \forall\text{ elim} \qquad [\neg\phi(a)]^2}{\bot}\ \to\text{-elim}}{\neg(\forall x)\phi(x)}\ \to\text{-int (1)} \qquad [\neg\neg(\forall x)\phi(x)]^{(3)}}{\bot}\ \to\text{-elim}}{\neg\neg\phi(a)}\ \to\text{-int (2)}}{(\forall x)\neg\neg\phi(x)}\ \forall\text{-int}}{\neg\neg(\forall x)\phi(x) \ \to \ (\forall x)\neg\neg\phi(x)}\ \to\text{-int (3)} \tag{45}$$

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. This last remark is probably worth recording

Remark 2 *ϕ and ϕ^* are classically equivalent.*

Proof:
By induction on the subformula relation. ■

So the negative interpretation will send every formula in the language to a stable formula classically equivalent to it.

Lemma 3 *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

$$\dfrac{\dfrac{\dfrac{[\neg p^* \wedge \neg q^*]^1}{\neg p^*}\ \wedge\text{-elim} \qquad p^*}{\bot}\ \to\text{-elim}}{\neg(\neg p^* \wedge \neg q^*)}\ \to\text{-int (1)} \qquad\qquad \dfrac{\dfrac{\dfrac{[\neg p^* \wedge \neg q^*]^1}{\neg q^*}\ \wedge\text{-elim} \qquad q^*}{\bot}\ \to\text{-elim}}{\neg(\neg p^* \wedge \neg q^*)}\ \to\text{-int (1)} \tag{46}$$

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^*.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\begin{matrix}[p^*]^1\\ \vdots\\ r^*\end{matrix} \qquad [\neg r^*]^3}{\bot}\ \to\text{-elim}}{\neg p^*}\ \to\text{-int (1)} \qquad \dfrac{\dfrac{\begin{matrix}[q^*]^2\\ \vdots\\ r^*\end{matrix} \qquad [\neg r^*]^3}{\bot}\ \to\text{-elim}}{\neg q^*}\ \to\text{-int (2)}}{\neg p^* \wedge \neg q^*}\ \wedge\text{-int} \qquad \neg(\neg p^* \wedge \neg q^*)}{\bot}\ \to\text{-elim}}{\neg\neg r^*}\ \to\text{-int (3)} \tag{47}$$

. . . and we infer r^* because r^* is stable.

$\to$-introduction

Given a constructive derivation $\dfrac{p^*}{\begin{matrix}\vdots\\ q^*\end{matrix}}$ we can build the following

$$\dfrac{\dfrac{\begin{matrix}\dfrac{[p^* \wedge \neg q^*]^1}{p^*}\ \wedge\text{-elim}\\ \vdots\\ q^*\end{matrix} \qquad \dfrac{[p^* \wedge \neg q^*]^1}{\neg q^*}\ \wedge\text{-elim}}{\bot}\ \to\text{-elim}}{\neg(p^* \wedge \neg q^*)}\ \to\text{-int (1)} \tag{48}$$

which is of course a proof of $(p \to q)^*$.

$\to$-elimination

The following is a deduction of q^* from $(p \to q)^*$ and p^*:

$$\dfrac{\dfrac{\dfrac{p^* \qquad [\neg q^*]^{1)}}{p^* \wedge \neg q^*}\ \wedge\text{-int} \qquad \neg(p^* \wedge \neg q^*)}{\bot}\ \to\text{-elim}}{\neg\neg q^*}\ \to\text{-int (2)} \tag{49}$$

. . . q^* is stable so we can infer q^*.

$\exists$-introduction

Constructively $\exists$ implies $\neg\forall\neg$ so this is immediate.

$\exists$-elimination

We use this where we have a classical derivation

$$\dfrac{\phi(x)}{\dfrac{\vdots}{p}}$$

and have been given $\exists y\phi(y)$.

By induction hypothesis this means we have a constructive derivation

$$\dfrac{\phi^*(x)}{\dfrac{\vdots}{p^*}}.$$

Instead of $\exists y\phi(y)$ we have $\neg(\forall y)\neg\phi^*(y)$.

$$\dfrac{\dfrac{\dfrac{\dfrac{\begin{matrix}[\phi^*(a)]^2\\ \vdots\\ p^*\end{matrix} \qquad [\neg p^*]^1}{\bot}\ \to\text{-elim}}{\neg\phi^*(a)}\ \to\text{-int (2)}}{(\forall y)\neg\phi^*(y)}\ \forall\text{-int} \qquad \neg(\forall y)\neg\phi^*(y)}{\dfrac{\bot}{\neg\neg p^{*(1)}}\ \to\text{-int (1)}}\ \to\text{-elim} \tag{50}$$

and p^* follows from $\neg\neg p^*$ because p^* is stable.

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 2 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:

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[\neg\phi(y)]^1 \qquad [x = y]^2}{\neg\phi(x)}\,\text{subst} \qquad \neg\neg\phi(x)}{\bot}\,\to\text{-elim}}{\neg(x = y)}\,\to\text{-int (2)} \qquad \neg\neg(x = y)}{\bot}\,\to\text{-elim}}{\neg\neg\phi(y)}\,\to\text{-int (1)} \tag{51}$$

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 3 ■

10.1 What the Negative Interpretation Does

The Negative Interpretation is thus quite useful. It enables the constructive Logician, whenever (s)he hears the classical logician utter 'ϕ', that his/her interlocutor actually meant the subtly different ϕ^*, which—according to that very interlocutor—is logically equivalent to ϕ (so they can't complain about being misunderstood!). Further, ϕ is constructively correct [acceptable to the constructive logician] iff ϕ was classically valid [acceptable to the classical logician]. One could hardly ask for a more diplomatically satisfactory outcome!

10.1.1 Prophecy

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 21 *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.

This kind of analysis is one of the reasons why even hardened Quineans such as your humble correspondent take constructive mathematics seriously. The thinking behind it may be bonkers but the analysis that it leads us through gives Mathematics a very dynamic flavour which is immensely attractive to anyone who cares about Mathematics. Thus it is possible to believe (as your humble correspondent does in fact believe) that constructivists have an important insight to offer *even if they are mistaken about what that insight is.*

11 Negative Interpretation for Richer Syntaxes

Suppose we have a classical theory T in a language $\mathcal{L}$. T has a constructive version T^H obtained from T by weakening the logic in which T is embedded to constructive Logic. Is there then a map $* : \mathcal{L} \to \mathcal{L}$ satisfying (i) $T \vdash \phi \longleftrightarrow \phi^*$ and (ii) $T \vdash \phi$ iff $T^H \vdash \phi^*$? The existence of such a map doesn't follow from the foregoing because T has lots of theorems that are not classical tautologies, so we are making a claim about a larger set of formulæ.

The arithmetic of the natural numbers has a negative interpretation; so does ZF (see [4]). It is not known whether Quine's theory NF has a negative interpretation.

12 Doing some Mathematics Constructively

The classical concept of *nonempty set* multifurcates into lots of constructively distinct properties. Constructively x is **nonempty** if $\neg(\forall y)(y \notin x)$; x is **inhabited** if $(\exists y)(y \in x)$, and these two properties are distinct constructively: the implication $\neg\forall\phi \to \exists\neg\phi$ is not good in general.

A is **decidable** iff $(\forall x)(x \in A \vee x \notin A)$.
$A \subseteq B$ is a **detachable** subset of B iff $(\forall x \in B)(x \in A \vee x \notin A)$.
We inductively define **Kuratowski-finite** and **N-finite** sets:

> The empty set is Kuratowski-finite; if x is kuratowski-finite so is $x \cup \{y\}$.
>
> The empty set is N-finite; if x is N-finite and $y \notin x$ so is $x \cup \{y\}$.

We take our natural numbers to be the cardinals of N-finite sets.

Inferring Mathematical Induction from the definition of $\mathbb{N}$ as $\bigcap\{X : 0 \in X \wedge S``X \subseteq X\}$ is constructive.

Suppose $\{n \in \mathbb{N} : F(n)\}$ contains 0 and is closed under successor. Then

$$\mathbb{N} = \bigcap\{X : 0 \in X \wedge S``X \subseteq X\} \subseteq \{n \in \mathbb{N} : F(n)\}$$

whence $(\forall n \in \mathbb{N})(F(n))$.

The least number principle says that every inhabited set of naturals has a least member. In constructive logic the equivalence between mathematical induction and the least number principle is lost: the least number principle implies excluded middle.

[It's worth sparing a thought on why mathematical induction implies the least number principle and how you used excluded middle to prove it]

Remark 3 *The Least Number Principle Implies Excluded Middle*

Proof:

Let p be any proposition, and consider $A = \{n \in \mathbb{N} : n = 1 \ \vee \ (n = 0 \ \wedge p)\}$.

A is inhabited (since 1 is a member of it) so, by LNP, it has a least member. Every member of A is 0 or 1. If this least member is 0 then we must have p. If it is 1 we must have $\neg p$. ∎

Remark 4 *(Diaconescu [2])*
The Axiom of Choice implies Excluded Middle.

Proof:
Clearly if every family of nonempty sets is to have a choice function then if x is nonempty we can find something in it. This would imply that every nonempty set is inhabited. That would be cheating and we shall not resort to it. If we are to refrain from cheating we will have to adopt AC in the form that every set of *inhabited* sets has a choice function.

Let us assume AC in this form, and deduce excluded middle. Let p be an arbitrary expression; we will deduce $p \vee \neg p$. Consider the set $\{0, 1\}$, and the equivalence relation $\sim$ defined by $x \sim y$ iff $x = y \vee p$. Next consider the quotient $\{0, 1\}/\sim$. (The suspicious might wish to be told that this set is $\{x : (\exists y)((y = 0 \vee y = 1) \wedge (\forall z)(z \in x \longleftrightarrow z \sim y))\}$). This is an inhabited set of inhabited sets. Its members are the equivalence classes [0] and [1]—which admittedly may or may not be the same thing—but they are at any rate inhabited. Since the quotient is an inhabited set of inhabited sets, it has a selection function f. We know that $[0] \subseteq \{0, 1\}$ so certainly $(\forall x)(x \in [0] \to x = 0 \vee x = 1)$. Analogously we know that $[1] \subseteq \{0, 1\}$ so certainly $(\forall x)(x \in [1] \to x = 0 \vee x = 1)$. So certainly $f([0]) = 0 \vee f([0]) = 1$ and $f([1]) = 0 \vee f([1]) = 1$. This gives us four possible combinations. $f([0]) = 1$ and $f([1]) = 0$ both imply $1 \sim 0$ and therefore p. That takes care of three possibilities; the remaining possibility is $f([0]) = 0 \wedge f([1]) = 1$. Since f is a function this tells us that $[0] \neq [1]$ so in this case $\neg p$. So we conclude $p \vee \neg p$. ∎

12.1 "Fishy" Sets

The two proofs we have just seen involve ...

$A = \{n \in \mathbb{N} : n = 1 \ \vee \ (n = 0 \ \wedge p)\}$

and

$x \sim y$ iff $(x = y) \vee p$.

In the first example A is classically either $\{1\}$ or $\{1, 0\}$. In the second example classically $\sim$ is either the identity relation or the universal relation—neither of them things involving 'p'. Constructively we cannot prove that $A = \{1\} \vee A = \{1, 0\}$ nor can we prove that $\sim\ =$ identity $\vee \sim\ =$ the universal relation. Ian Stewart calls sets like A and $\sim$ *fishy*. A fishy set is something that classically is demonstrably one of two things, but constructively cannot be so demonstrated—and some mileage is extracted from it, as in these two proofs. I learnt this terminology from Douglas Bridges. It is not standard but it should be. How does it arise?

Classically we have two infinite distributive laws:

$$p \vee (\forall x)(A(x)) \text{ is equivalent to } (\forall x)(p \vee A(x))$$

and

$$p \wedge (\exists x)(A(x)) \text{ is equivalent to } (\exists x)(p \wedge A(x))$$

so we can "export" from the scope of a quantifier any subformula not containing any occurrence of the variable bound by that quantifier. This does not work constructively (constructively $p \vee (\forall x)(A(x))$ does not follow from $(\forall x)(p \vee A(x))$—though the converse is good) and where we have failures of exportation we find these "fishy" sets that—as we have seen—turn up in proofs that certain set-theoretic principles imply excluded middle.

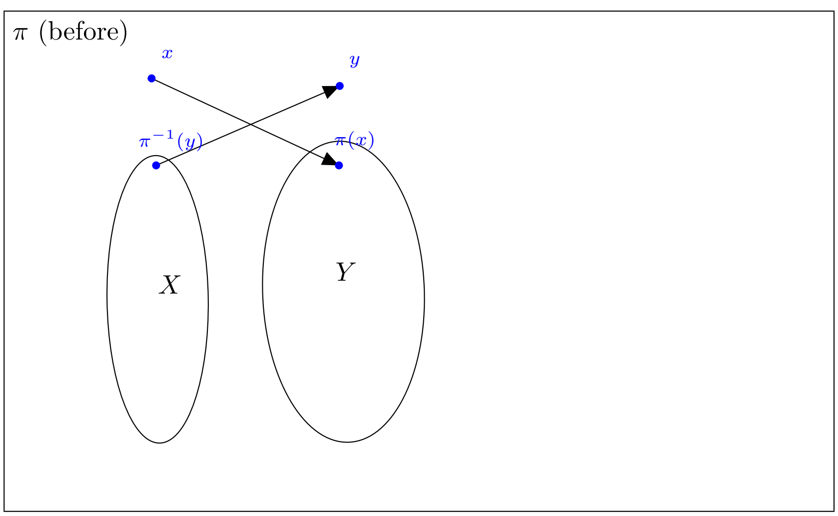
12.2 A Bit of Arithmetic

Heyting naturals the cardinals of N-finite sets. Heyting Naturals are the constructively correct concept of natural number. First we prove that it is...

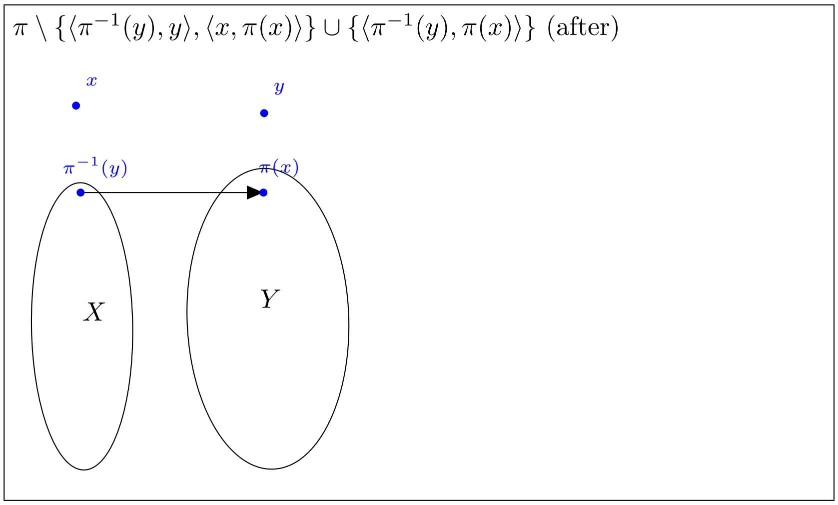
Lemma 4 *...decidable*[26] *whether or not two Nfinite sets are in bijection.*

Proof:

We prove by induction on X that $(\forall Y)(\text{Nfinite}(Y) \to (X \sim Y \vee \neg(X \sim Y))$. Clearly true for X empty. Suppose true for X, and consider $X \cup \{x\}$, with $x \notin X$. If Y is Nfinite then its either $\emptyset$ in which case the answer is 'no' or it's $Y' \cup \{y\}$ with $y \notin Y'$. By induction hyp on X we have $X \sim Y' \vee \neg(X \sim Y')$. If the first then $X \sim Y$. If $\neg(X \sim Y')$ then we can't have $X \cup \{x\} \sim Y' \cup \{y\}$. This is because any bijection $\pi : X \cup \{x\} \longleftrightarrow Y' \cup \{y\}$ will give rise to a bijection $X \sim Y'$, namely $\pi \setminus \{\langle \pi^{-1}(y), y\rangle, \langle x, \pi(x)\rangle\} \cup \{\langle \pi^{-1}(y), \pi(x)\rangle\}$.



[26] Different meaning of this word here!!



Then we prove trichotomy for Nfinite cardinals: given any two Nfinite sets one injects into the other.

LEMMA 5 $(\forall X)(\mathit{Nfin}(X) \to (\forall Y)(\mathit{Nfin}(Y) \to ((X \hookrightarrow Y) \vee (Y \hookrightarrow X))))$

Proof:

By induction on X. Base case (X empty) is easy.

Suppose true for X. Want it to be true for $X \cup \{x\}$. Let Y be Nfinite. By induction hypothesis either $Y \hookrightarrow X$ (in which case $Y \hookrightarrow X \cup \{x\}$) or $X \hookrightarrow Y$. Y is non empty so it is $Y' \cup \{y\}$. By induction hypothesis either $Y' \hookrightarrow X$ (in which case Y (which is $Y' \cup \{y\}$) $\hookrightarrow X \cup \{x\}$). On the other horn $X \hookrightarrow Y'$ which gives $X \cup \{x\} \hookrightarrow Y' \cup \{y\} = Y$. (NB for this to work we need both X and Y to be Nfinite not merely Kfinite). ■

The upshot is that, if we take our natural numbers to be cardinals of Nfinite sets, then $(\forall n, m \in \mathbb{N})(n = m \vee n \neq m)$. Interestingly the same doesn't go for reals: we cannot prove $(\forall x \in \mathbb{R})(x = 0 \vee x \neq 0)$.

12.3 Recursive Analysis

$\mathbb{R}$ has only two subsets that are detachable, itself and the empty set. This makes life difficult! See [1] pp 53*ff*.

Reals can arise as all sorts of things, from Dedekind cuts, or Cauchy sequences for example. But if we have the added dimension of computability to worry about then even if we have decided to think of computable reals as computable Cauchy sequences (in the rationals of course) we can wonder whether we think of those computable Cauchy sequences as functions-in-intension (programs) or as function graphs (functions in extension). Both make sense. If we do the first, then Rice's theorem will ensure that the equality relation between computable reals is undecidable.

Another thing we can do is say that a real is computable iff there is a Cauchy-sequence-in-intension whose limit it is. That way our computable reals aren't different things from reals, but delineate a subset $\mathbb{R}_c$ of $\mathbb{R}$; this is how Bridges does it.

Analysis is full of dependencies: If $f : \mathbb{R} \to \mathbb{R}$ is continuous then $(\forall x)(\forall \epsilon)(\exists \delta)(\ldots)$ But how does the δ depend on x and ϵ? Riemann's theorem: if f is integrable then $\forall \epsilon \exists \delta \ldots$ In the realistic cases we deal with in ordinary[27] mathematics we can obtain values for δ from the arguments x and ϵ in fairly explicit ways that one would like to be allowed to describe as 'computable'. People in Analysis don't make much of these dependencies but occasionally you will see the ϵs and δs equipped with subscripts, as in the following example (which is Q1 on Analysis 1 sheet 1):

> if $a_n \to a$ and $b_n \to b$ then $a_n + b_n \to a + b$.
>
> $a_n \to a$ so $(\forall \epsilon > 0)(\exists N_a(\epsilon))(\forall n > N_a(\epsilon))(|a_n - a| < \epsilon)$
>
> and
>
> $b_n \to b$ so $(\forall \epsilon > 0)(\exists N_b(\epsilon))(\forall n > N_b(\epsilon))(|b_n - b| < \epsilon)$
>
> Now set $N(\epsilon) := \max(\{N_a(\epsilon/2), N_b(\epsilon/2)\})$, and take it from there
>
> ...
>
> $(\forall n > N(\epsilon))(|(a_n + b_n) - (a + b)| < \epsilon)$

If you look carefully you can often see that these dependencies are in fact constructively provable.

Is it, in fact, OK to describe this process as 'computable'? There is an obvious *prima facie* problem in that the quantities x, ϵ and δ are *infinite precision objects*, so we cannot compute with them in the way we have been accustomed to so far. But that's not really a problem because we can always take these quantities to be rationals.

References

[1] Douglas. S. Bridges. "Computability, a Mathematical Sketchbook" Springer Graduate texts in Mathematics **146** 1994.

[2] Radu Diaconescu, "Axiom of Choice and Complementation". Proc. AMS **51** (1975) 176–178.

[3] `www.dpmms.cam.ac.uk/~tf/partiilectures2016.pdf`)

[4] Powell, William C. Extending Gödel's negative interpretation to *ZF*. *Journal of Symbolic Logic* **40** pp. 221−9.

[5] Scott D.S. Semantical Archæology, a parable. In: Harman and Davidson eds, Semantics of Natural Languages. Reidel 1972 pp 666–674.

[27] I know one shouldn't use the phrase 'ordinary mathematics' but sometimes temptation gets the better of one.

[6] Robert Heinlein. *Stranger in a Strange Land.* Putnam 1961.

[7] https://en.wikipedia.org/wiki/Combinatory_logic#Completeness_of_the_S-K_basis