INTRODUCTION TO LOGIC

BY JOHN WORRALL

A: Truth-Functional Logic

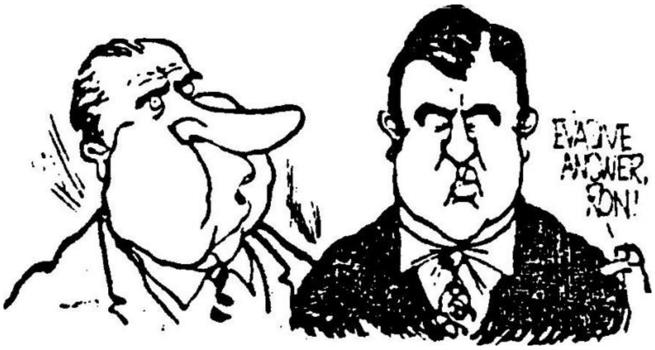
A1: Introduction:

Logic is about reasoning or arguing

"OK, Mr. Press Secretary, give me some answers!"

"If I knew about the Watergate Caper, what am I doing in the White House?"

"And if I didn't know anything about the affair…"



"...What am I doing in the White House?"

The cartoonist in this – thankfully dated – cartoon is implicitly landing Richard Nixon with an *argument* (or train of reasoning) – one that condemns him from his own mouth. More explicitly (and therefore draining it of any semblance of humour), the argument is that, since there are only two cases – that Nixon knew about the Watergate affair and that he didn't – and since in *either* case there would be grounds (different in the two cases of course) for inferringthat Nixon was unworthy of his presidential office, it *follows* that Nixon was indeed unworthy of his office. An argument consists of citing certain premises and showing (or claiming) that a certain *conclusion* follows from them. The premises here are that Nixon was unworthy of his office in *either* the case that he knew about the break-in *or* the case that he didn't. The conclusion is that he is indeed unworthy of his office.

We *argue* (in the intellectual rather than the 'falling out' sense) or reason or infer or make deductions all the time. This is true both in intellectual disciplines and, if often rather more loosely, in everyday life. For example, a scientist tests a particular theory by reasoning that if that theory is true then some other claim, one that can be checked observationally or experimentally, must also be true – that is, that some observationally checkable claim follows from the theory. For instance, Newton tested his theory of universal gravitation by inferring what followed from that theory about the motions of the planets – in particular that they describe (roughly) elliptical orbits around the sun. Einstein's general theory of relativity was tested by showing that you could infer from it that the stars would appear to be different distances apart during the day time than they were during the night (because of the effect of the sun on the trajectory of the rays of light from the stars). This prediction could only be tested in the special circumstances when the stars are visible during the daytime – during a total solar eclipse. When Eddington carried out the test, Einstein's prediction turned out to be correct. This testing process is an essential part of science in general. And of social science too: the Treasury tests its (theoretical) model of the economy by working out what it implies (what follows from it) about (observable) changes in the real UK economy.

Logic also plays a crucial role in mathematics. Mathematics is centrally concerned with proofs, which are in fact inferences or deductions or arguments. In formal mathematics, certain axioms are laid down (for example Euclid's axioms of geometry – basic assumptions that are accepted as givens, such as the 'parallel postulate' usually stated as: 'Given a line AB and a point C outside the line, there is one and only one line that goes through C and is parallel to AB') and proofs consist of showing that certain other assertions (theorems – for example, the theorem that the internal angles of any triangle sum to 180^{o}) follow from, or can be inferred from, those axioms.

Philosophy too is centrally concerned with arguments or deductions. For example, some philosophers argue that the presence of evil in the world is inconsistent with the existence of an all-powerful, all-knowing, benevolentGod as proposed in standard Judeo-Christian theology. They claim that if you assume that there is such a God, then it follows, or you could infer that, there would be no evil in the world. Hence, since there clearly is a lot of evil in the world, you can infer in turn that there is no such God.

Coming closer to more practical concerns, defence lawyers argue for the innocence of their clients, politicians argue for their policies, and, more mundanely, we reason, or make inferences, all the time – though we don't always think of it in that way. Suppose you wake up after an especially heavy night on the town and find yourself unable to remember what day it is. You might eventually reason: 'Well, yesterday was Saturday, so today must be Sunday'. This is a very basic inference, but it does fit the standard pattern; you eventually dredge up from your alcohol- (or other recreational drug-) soaked memory a **premise** (that yesterday was Saturday) and you make a very straightforward **inference**to the **conclusion** that today is Sunday. (Of course you are also implicitly assuming other premises – like that you were not so drunk that you slept for more than 24 hours!)

Finally, we reason in this way – that is, take certain 'information' as given, and work out what follows from that information – whenever we do IQ tests or try to solve complicated 'brainteasers' or 'logic puzzles'. Suppose – to take a real old chestnut – you are told that a certain man is standing in front of a portrait of another person and he says:

> 'Brothers and Sisters have I none, but that man's father is my father's son.'

You are asked whose portrait it is. What you must do is work out what assertion about his direct relationship to the person in the picture can be inferred fromthe information given (what the guy actually says). So whose picture is it?

To take an example from my favourite genre, suppose you are told that Alf has washed up on the Island of Knights and Knaves – a strange island inhabited exclusively by two separate, but intermingled tribes, Knights and Knaves. Knights always tell the truth, and Knaves always lie. In exploring the island, Alf comes to a fork in the road – one but only one of the forks leads to the Island's capital, which is where Alf wants to go. Luckily an inhabitant is standing at the fork and helpfully (if rather improbably) informs Alf:

> 'Either the correct fork is the left one, or I am a Knave (or both).'

The puzzle is: which road should Alf take? Try to work out the solution for yourself before reading further.

The solution is essentially the following argument or inference: Alf's informant is either a Knight or a Knave. If he were a Knave then he would be telling the truth when he said he was a Knave and hence he would be telling the truth when he said that either he was a Knave or the correct fork is the left one (because 'or' statements are automatically true if one part is true). But this is impossible, because Knaves always lie. So Alf's informant must be a knight. If so, what he says must be true, because Knights always speak the truth. But since the second part of his either/or statement is false, the first part must be true to make the whole either/or sentence true; hence the correct fork is indeed the left one. This is a correct inference – or as we shall say a **valid inference**.

The Island of Knights and Knaves was invented by the intrepid logician Raymond Smullyan. You can find out more about it and try out more puzzles here.

All of these pieces of reasoning – from scientific tests, mathematical proofs and philosophical arguments to logic puzzles and mundane bits of everyday reasoning – share the same basic structure (though they may differ greatly in complexity). Certain "information" is taken as starting point – we are 'given' Einstein's theory or Euclid’s axioms or that Alf is on this particular island and the inhabitant utters a certain sentence, or that yesterday was Saturday – and we are asked to work out *what follows from* or *what we can infer from* the given information. We shall refer to the information or assumptions from which a particular piece of reasoning starts as **premises** and the further claim that is inferred from those premises as the **conclusion**. So the reasoning or the **inference** or the **argument** itself is the *process* that takes us from a set of premises to a conclusion. All inferences, then, ultimately have the form:

PREMISES

Therefore,

CONCLUSION

The fact that an inference is being made is invariably signaled by some such word or phrase as ‘therefore’, ‘and so’, ‘it follows that’, ‘from which we may infer’, ‘ergo’ and so on.

Deductive logic is the study of such inferences in general – it has therefore an enormously broad scope and may be the most basic of all disciplines. Different disciplines have different ways of garnering information in the first place (i.e. coming up with premises). The way that we arrive at a scientific theory is different from the way that we arrive at an axiom in mathematics or a thesis in philosophy. However, the way that we *reason from* that information, the logic that we employ to draw further implicationsfrom that information is the same no matter what the discipline. Sologic studies inference and its main task is to give an explicit characterisation of those inferences that are correct, or as we shall say, **valid** (and hence differentiate them from those inferences that are **invalid**). Logic tells you exactly when some conclusion really does follow from some premises and when it does not.

This course will, then, investigate three main issues:

(1) What does it mean for a piece of reasoning or a deduction or inference to be ***valid***?

(2) How can we ***recognise*** valid inferences and hence distinguish them from **invalid** inferences?

(3) Are there any ***methods*** for making valid inferences?

A2: Validity and Soundness

Before getting down to work, let's pause to clarify right from the outset an issue that often confuses people. Put rather enigmatically we might say that while logic is centrally concerned with *truth-transmission*, it is not at all concerned with truth. Consider the following two inferences:

Inference 1:

Premise: Elvis Presley was a great rock singer

Conclusion: (So!) Marilyn Monroe was a great comedy actress

Inference 2:

Premises: All members of the Klu Klux Klan are intelligent.

All intelligent people are law-abiding

Conclusion: So, All members of Klu Klux Klan are law-abiding

Inference 1 has – or so I would (vociferously) assert – a true premise and a true conclusion. But clearly it's a ridiculous inference – the 'So' just isn't so: it might be true that Elvis was a great rock singer (of course before he joined the US army), and (independently!) true (very true) that Marilyn was a great (and underrated) comedy actress, but it surely *doesn't follow* that she was from the assertion about Elvis: there's just 'no connection' between the premise and the conclusion.

On the other hand, both the premises in Inference 2 are false: there are plenty of intelligent criminals (making the second premise false) and, although I don't know any personally, certainly members of the Klan quoted in the Media often do not seem too bright (so the first premise seems to be false). The conclusion is also false – members of the Klan have, historically, committed any number of criminal acts. Nonetheless, this second *inference* is **valid** – as I hope your intuitions will agree. The premises may not be true, and the conclusion might not be true, but nonetheless the conclusion clearly **follows from** the premises. (This situation is to be compared with inference 1 where the conclusion is true alright but it doesn't follow from the premise.) How does it follow despite being false? Well we'd be inclined to say that **IF** it were true that all the Klan members were intelligent and **IF** it were true that all intelligent people were law-abiding then **it wouldalso have to be true,** it would FOLLOW, that all the Klan members werelaw-abiding.

Logic is about what else has to be true, **supposing** that certain starting points are true – that is why it is about *truth-transmission* rather than about truth. Of course a scientist is not interested in drawing conclusions from any old theory – she must have reason to think that it at least *may be* true. But logic is indifferent – it will tell you what follows from *any* theory, no matter how ridiculous (that is it will tell you what else would have to be true *if* that theory were true). That seems, when you think about it, intuitively right: you can work out what follows from the (incorrect) theory that some electrons have positive charge just as you can from the (correct) theory that they all have negative charge. It's just that the conclusions you validly draw from the latter will all be true (i.e. borne out in experiments), while some at least of the conclusions you draw from the false theory that some electrons have positive charge will themselves be false – that is, run counter to what is actually observed.

Similarly, in the brainteaser case, you aren't interested in whether there really is an Island of Knights and Knaves and whether Alf really ever did raise his question. These are just ***assumptions: supposing*** ("for the sake of argument") that they were true, what else could you infer (what else would have to be true) about which road it is that leads to the capital?

Or consider again the Nixon cartoon we first looked at. There is one question logic can answer and one it *cannot.* The question it can answer is '*Suppose* it were true that if Nixon knew about the cover up then he is unworthy of his office and also true that if Nixon did not know about the cover up then he is again unworthy of office. Would it also then *have to be* true that he is indeed unworthy of office?' (The answer is, of course, 'yes'.) The question which logic *cannot* answer is whether or not these suppositions are true: deciding whether or not it is true, for example, that if Nixon knew of the cover-up then he is unworthy of office involves a complex of empirical and ethical issues.

Logic, then, is about which inferences are **VALID** (which ones have justified ‘therefore’s’ or ‘and so’s’) and this is independent of whether or not the premises of the inferences are true. Inferences which are not only valid, but which also have true premises are called **SOUND**. As ordinary reasoners or as scientists, soundness is, of course, a major concern – we would like the premises that we start from to be true (or at least arguably true). But logic, to repeat, is indifferent to soundness and involves only the issue of validity. The ‘premises’ are always just initial assumptions – logic will tell you what follows from those assumptions just as well if they are false as if they are true (or indeed if they are – as in the brainteaser case – merely assumptions about which the question of truth simply doesn’t arise).

The important connection between validity and soundness is that **if** the inference is indeed valid and **if** moreover it is sound (that is, if its premises are true) **then** its conclusion ***mustbe*** true as well. Exactly this same point can be read the ‘other way round’and is equally (perhaps even more) important when expressed in this negative way: if an inference is valid and its conclusion is false, then it ***cannot besound*** – that is, not all of the premises can be true, at least one must be false. (Peopleoften learn in this way: they begin by believing a certain set of assertions; and then realise (or are shown) that a certain conclusion (validly) follows from that set of assertions; and they acknowledge that that conclusion is false – hence logic dictates that not all the premises, that is, not all of the set of assertions they began by believing, can be true, at least one must be false and so must be rejected.)

A3: Truth-Functional Logic—An Introduction

So let's start investigating some inferences. Try not to be put off by the fact that all our early examples will be extremely simple – we have to learn to walk before we can run.

Someone might reason as follows:

> Either Uri Geller bends spoons because he possesses genuine psychokinetic powers or he bends them by standard magicians' trickery. He doesn't possess genuine psychokinetic powers (no one does). *Therefore,* Geller bends spoons by standard magicians' trickery.

This simple inference is **VALID**. It is, moreover, in my view, sound – its premises are true (and hence because the inference is valid, so is its conclusion). But, as we just saw, it doesn't matter at all from the point of view of validity if the premises are true or not. The validity stems – as always – from the fact that **IF** the premises were true, then **SO ALSO** would have to be the conclusion. The first premise asserts that one of two possibilities has to hold true. The second premise asserts that it isn't the first possibility that holds. It obviously follows that second possibility has to hold. Independently of the actual facts about Geller, it's just **NOT POSSIBLE** for the only possibilities to be A and B (genuine powers or magic tricks), for A not to be true and for B not to be true as well. To deny the conclusion of this inference while accepting both the premises would just be to **CONTRADICT** oneself. Or more pointedly: suppose you denied theconclusion, while you accepted the second premise (that he doesn't have real psychokinetic powers), then you would be contradicting yourself if you continued to hold the first premise: that the only two possibilities were real powers and magicians' tricks.

Or let's take a slightly more elaborate inference of similar form. Suppose that someone is trying to remember which London station the train to Edinburgh leaves from. She remembers going north to the station and this, together with her knowledge of London stations gives her as a first *premise:* 'Either the Edinburgh train leaves from Euston or it leavesfrom King's Cross'. She then remembers taking the Manchester train from Euston, and feels sure that the Edinburgh train leaves from a different station than the Manchester train. This in effect yields two further premises: 'If the Edinburgh train left from Euston then it would leave from the same station as the Manchester train' and 'The Edinburgh train doesn't leave from the same station as the Manchester train'. Taking all these premises together it follows of course that the Edinburgh train leaves from King's Cross. Although we can hardly imagine anyone spelling the argument or inference out in such gory detail, we can imagine that someone would infer where to go to catch the Edinburgh train *essentially* in this way (supposing she is not internet-connected and so has no need to rely on memory). Spelling out the inference fully we have:

1. *Either* the Edinburgh train leaves from Euston *or* it leaves from King's Cross.
2. *If* it leaves from Euston, *then* it leaves from the same station as the Manchester train.
3. It does *not* leave from the same station as the Manchester train.

Therefore: The Edinburgh train leaves from King's Cross.

Again the inference here is **valid** – I hope that this will be intuitively clear to you. If not, consider that it is just a slight elaboration on the Geller inference: the first premise states that one of two possibilities holds, while the second and third premises together rule out the first possibility (i.e. they rule out Euston). This leaves only the second possibility and the conclusion is simply the claim that it is this second possibility which holds. We will soon use the ideas elicited by these two examples, to produce a general characterisation of validity of inference. However, this general characterisation will be easier to grasp if we look first at a couple of inferences that are intuitively clearly **INVALID**.

Before the discovery of Australia, European ornithologists believed that all swans are white. Their evidence was of course a whole lot of observations of white swans. They were clearly making an *inference* of something like the following form:

a_1 is a swan and is white a_2 is a swan and is white

…

a_n is a swan and is white

Therefore, all swans are white.

This is an *invalid* inference. Even *had* it turned out that all Australian swans, like all other swans in the world are white (in other words, even if the conclusion here turned out as a matter of fact to be true), the ornithologists' grounds for holding it to be true were clearly not adequate. This is easily seen by reflecting that it is **POSSIBLE** for individuals $a_1, \dots a_n$ all to be white swans (that is all premises to be true), while some *other* swan is black (and so the conclusion that 'All swans are white' is false). It turned out that this possibility is actualised: in Australia there are black swans. But the inference would *still* not be deductively valid even if all swans *were* in fact white (though it might nonetheless be persuasive in some other sense – it is often referred to as an **inductive**, rather than a deductive, argument). You don't contradict yourself if you accept that all the swans you observed so far are white, but assert that some other swan is not white (and hence that it is false that all swans are white). Contrast this with the Geller case in which, as we noted, you would contradict yourself if you rejected the conclusion and continued to assert both premises.

Or, suppose someone is reading an Agatha Christie-style novel and, not being an expert in these matters (there's always a last page surprise), has come to the next to last page with the firm belief that the Butler 'did it'. His evidence is that the Butler had both the motive to kill the vicar (who was really a blackmailer who knew of the Butler's affair with the 'Lady' of the house, who was really...) and the means (the murder was committed with the Butler's machete, which he kept for 'deadheading' his roses and so to which he had access). Our non-expert reader again has made an inference. Something like the following one:

1. Anyone who murdered the vicar had the means and the motive.
2. The Butler had both the means and the motive.

Therefore, the Butler did it.

Again this inference is invalid; that is, again the conclusion is not *guaranteed* to be true simply because the stated premises are known to betrue: it is possible for both of the premises to be true while the conclusion is false. This is of course because it is possible for *more than one* person to have had the motive and the means. Indeed, we can suppose that our non-expert reader gets the customary shock on the last page when it turns out that little Miss Goody Two-shoes – in fact the "vicar's" former lover and accomplice – did it. But it wouldn't matter if this turned out to be a very boring whodunnit and the conclusion this reader had drawn was correct – the Butler really did do it. The inference would still have been invalid, as an inference, because it was still *possible* (even if, so it turned out, not-actual) for someone else to have had both the motive and the means and for that someone else, rather than the Butler, to have been the guilty party.

We seem to be heading toward the following general characterisation of what it is for an inference to be valid: An inference is valid if it is **NOT POSSIBLE** for the conclusion to be false and the premises to be true.

It's not possible for the only two explanations for Geller's spoonbending antics to be trickery and genuine psychic powers, for him not to have genuine psychic powers and, at the same time, for him *not* to be doing it by trickery. On the other hand, it *is* possible, whether or not it's true, for all observed swans to be white and yet not all swans to be white (because some so-far unobserved swans are some other colour). But surely we can't rest what I've argued is a crucially important and fundamental notion (of validity of inference) on the opaque notion of possibility – after all, pigs *might possibly* fly.

But let's for the moment, give ourselves the notion of 'possibility' (we will soon replace it with a much less mysterious notion) and summarise the important points that have been made so far:

Validity:

An inference is valid if it's *not possible* for the conclusion to be false and (all) the premises true at the same time.

Another way of thinking about this impossibility is that in a valid inference, you would contradict yourself if you held that the conclusion was false and all the premises true. In the case of an invalid inference, on the other hand, you might be wrong if you asserted that the conclusion was false while accepting the premises as true but you would not contradict yourself.

To make this clear, think about an analogous inference to the "swans" one: so far as I know, all ravens (at any rate all normal ravens, there are some albino ones) are as a matter of fact black. The inference from any number of observed black ravens to the assertion that **all** ravens (observed or so far unobserved) are black would nonetheless be invalid just the same as the swans one. If someone accepted that all observed ravens have been black, but denied that all ravens are, they would be (factually) wrong, but they would clearly not be contradicting themselves. Just as someone about to celebrate their 18th birthday could accept that they were under 18 yesterday and under 18 on the day before that, and on the day before that, etc., while accepting that they will not be under 18 tomorrow!

Hence a good way to think about what makes an inference **INVALID** is that it is invalid if it is **POSSIBLE** for the conclusion to be false even though the premises are all true.

To drive home this important lesson, consider finally the following train of reasoning: Someone who knows little about Opera is trying to recall which composer wrote *Tosca.* She remembers that the composer was Italian, so that it's a fair bet that it was either Puccini or Verdi. Something tells her that it wasn't Puccini, so she *infers* or concludes that it was Verdi. She has made the following inference:

1. Either *Tosca is* by Puccini or by Verdi.
2. *Tosca is* not by Puccini.

Therefore, it is by Verdi.

Here the conclusion is *in fact* false. Nonetheless there is a clear sense in which the reasoning is correct. *Had* both the premises been true then the conclusion *would have had* to be true as well. The assumption that the premises are trueand the conclusion false is, again, self-contradictory. It's just that *as a matter offact* the conclusion is false – thus showing that at least one of the premises mustbe false too (*Tosca is* by Puccini).

We must, then, as we agreed earlier, always sharply separate the two questions:

1. Are the premises or assumptions from which some piece of reasoning starts true?
2. Is the reasoning *valid?* That is, does the conclusion follow from the premises? (whether or not the conclusion is true)

In the *Tosca* case the conclusion *does* follow from the premises, but it is false (because one of the premises is false).

A3(A): LOGICAL FORM AND TRUTH-FUNCTIONAL VALIDITY

The *Tosca* inference and the Uri Geller inference are valid inferences for exactly the same reason. In fact, in logical terms they are the same inference. Although one talks about Opera composers and the other about spoonbending (so that their *contents* are radically different), both inferences have the ***same form***. The first premise of both inferences states that one of two possibilitiesholds. The second premise states that one particular possibility does *not* hold. The conclusion is that the other possibility holds. If we disregard the content of the two inferences by replacing single assertions by letters (different letters for different assertions), we can express both by the scheme:

Premises: *Either* p *or* q

Not - p

Conclusion: Therefore, q

There is nothing magical about the symbols: the p's and q's are simply place-holders for particular assertions. The above scheme is the **logical form** of the inference about *Tosca* and about Geller. Let's call it the *inference-scheme* of both these inferences.

Given any such inference-scheme we can of course turn it back into a particular inference by replacing p and q by ordinary sentences. Substituting'*Tosca* isby Puccini' for p, and '*Tosca* isby Verdi' for q, we arrive back fromthe scheme to the *Tosca* inference. If we substitute 'The *Genesis* account of the creation of the universe is wrong' for p, and 'The Darwinian theory of evolution is wrong' for q, we obtain the quite different inference:

1. *Either* the Genesis account of the creation of the universe *or* the Darwinian theory of evolution is wrong.
2. The Darwinian theory is not wrong.

Therefore, the Genesis account is wrong

Since there are infinitely many sentences in English, there are in fact infinitely many possible substitutions for p and q in our simple scheme. However, any such substitution must produce an inference that falls under just one of the following four headings:

(1) Both premises true, and conclusion true

For example, we might substitute the sentence 'The sum of two and two is four' for p and the sentence 'pigs can fly' for q, thus producing the inference:

1. *Either* the sum of two and two is four *or* pigs can fly.
2. Pigs can’t fly.

Therefore, the sum of two and two is four.

(2) At least one premise is false, and conclusion false

For example, substitute 'Mozart wrote *Fidelio*' for p and 'Beethoven wrote *DonGiovanni*' for q. This produces an inference whose first premise is false (sinceboth sides of the either/or are false) and whose conclusion (‘Beethoven wrote *Don* Giovanni’) is false as well.

(3) At least one premise false, and conclusion true

For example, substitute 'Newton was a great scientist' for p and 'Einstein was a great scientist' for q. Here the second premise (not-q) is false; but the conclusion (‘Einstein was a great scientist’) is true.

(4) Both premises true and conclusion false

??

It is no accident that I cannot cite any examples under heading (4). For this particular inference-scheme, **there are no such examples.** Do your best to find substitutions for p and q that might make both premises true and conclusion false - even your best will not be good enough!

This is in fact the key to replacing the vague talk of possibility with a clear notion and hence to producing our first precise notion of validity. We arrived intuitively at the idea that an inference is valid if the truth of the premises *would be* enough to guarantee the truth of the conclusion (even if the premises are as a matter of fact false) or equivalently if the conclusion could not possibly be false if the premises were true. We can now eliminate this rather tricky subjunctive notion ('would be's' are sometimes called ‘subjunctives’) and say that:

An inference is **VALID**, if and only if, ***NO INFERENCE OF THE SAME FORM has true premises and a false conclusion.***

The form of the inference, remember, is its symbolic representation found by replacing single assertions in the inference by letters – p, q, r, etc. – using a different letter for each different assertion. If this characterisation of validity is correct then an *invalid* inference must, of course, fail to meet it. That is, for an invalid inference, case (4), (true premises and false conclusion) *should* be possible. And it is.

Consider the inference that can be taken to underlie the reasoning of our earlier duped Agatha Christie reader:

1. If the Butler 'did it', then he had both the motive and the means.
2. He had the motive and the means.

Therefore, the Butler 'did it'.

Going through replacing single assertions by letters, as before, we obtain the *form* of this inference:

1. *If* p, *then* q
2. q

Therefore, p

(Here 'p' stands for 'The Butler did it' and ‘q’ for the sentence ‘The Butler had the motive and the means'.) We can assume that in the original inference we are unsure about the truth or falsity of p (we did take it that we knew q to be true). But whether or not p is true, the premises are not sufficient to establish its truth becauseof the *possibility* that the premises are true while the conclusion (p) is false. We can again now eliminate this rather vague talk of 'possibility': *the inference is invalid if (and only if) we can find at least onesubstitution for p and q in the inference-scheme which makes both premises true and the conclusion false.*

This is in fact easily done. Take, for example, p as 'Joe diMaggio was president of the US' and q as 'Joe diMaggio was born in the US'. This substitution into the inference scheme produces the inference:

1. If Joe diMaggio was president of the US, then he was born in the US.
2. Joe diMaggio was born in the US.

Therefore, Joe diMaggio was President of the US

Here the premises are true (the second just *is* true and the first is true in view of the fact that *anyone* who stands for President must have been born in the US), but the conclusion is of course false, though he did have the not-inconsiderable consolation of not only being a great baseball player but also of being married for a time to the wonderful Marilyn Monroe. This, then, is why the inference about the Butler is invalid. There is an inference of the same form as that inference which has true premises and false conclusion (the Joe diMaggio one).

(*Exercise:* Try to think yourself of other substitutions for p and q which do the same job – thatis make the premises true, but the conclusion false.)

Let's record our results so far in the form of a couple of important definitions:

***Definition:* Counterexample:**

Let I be any inference. An inference of the same logical form as Ithat has true premises and a false conclusion is called a ***COUNTEREXAMPLE* to I.**

***Definition:* Validity:**

An inference is **VALID** if and only if there are **no counterexamples to it**, and is **INVALID** if and only if**there is acounterexample** to it.

This eliminates the vagueness involved in the notion of 'possibility' but only at the cost of introducing the so far rather unspecified notion of the "form of an inference". In the next section this notion is made explicit (at any rate for a restricted range of inferences).

A3(B): TRUTH-FUNCTIONALLY COMPOUND SENTENCES

We need first to reflect on a couple of simple facts about language. *First*, some sentences might be called "**atomic declarative sentences**": declarative because they make an assertion which is either true or false, and atomic because they have no parts that are themselves sentences. So 'Logic is easy' is an atomic declarative sentence, and so is 'Donald Trump is crazy'. On the other hand, 'Shut the door!' and 'Is the door shut?' are not declarative (they aren't true-or-false assertions) and so automatically not atomic declarative. Meanwhile, 'If Trump wins, I will want to leave the planet' and 'Trump is crazy and Clinton is untrustworthy' are declarative alright but not atomic – since both contain parts ('Trump wins' and 'I will want to leave the planet' in the first case and 'Trump is crazy' and 'Clinton is untrustworthy' in the second that are themselves sentences).

Second, given a stock of such atomic declarative sentences, there are many ways in which we can use them to build new more complicated sentences. For example, we can form a single sentence by taking any two of them and sticking an 'and' between them, and another one by sticking an 'or' between them (usually with an 'either' in front). Indeed, the sentence 'Trump is crazy and Clinton is untrustworthy' is formed exactly by sticking an 'and' between the two separate atomic sentences 'Trump is crazy' and 'Clinton is untrustworthy'. Out of the sentences 'Tony Blair lied' and 'I am a bad judge of character' we can form the sentence 'Either Tony Blair lied or I am a bad judge of character'.

Also, given any single sentence such as 'I am a bad judge of character' we can form another by placing 'It's not the case that' in front of it to form: 'It's not the case that I am a bad judge of character'. This would more usually be expressed as: 'I am not a bad judge of character'. (As will become clear as we go along, very often an idiomatic English sentence does not display its logical form directly but employs various abbreviatory devices: so instead of saying 'Blair lied about weapons of mass destruction in Iraq and Blair misled the British people' we would say 'Blair lied about weapons of mass destruction in Iraq and misled the British people'.)

Another way of making '**compound**' sentences out of single ('atomic') sentences is by the '**if ... then**' construction. For example, out of the two 'atomic' sentences 'Logic is interesting' and 'I'm a Dutchman', we can form the single *compound* sentence: 'If logic is interesting, then I'm a Dutchman.' Or out of the sentences 'Einstein's theory is true' and 'Light rays are bent by gravitating bodies' we can form the single compound sentence 'If Einstein's theory is true, then light rays are bent by gravitating bodies'.

The 'and' construction is called **CONJUNCTION**. The compound sentence 'Trump is crazy and Clinton is untrustworthy' is the *conjunction* of its two component sentences (which are called 'conjuncts').

The 'or' construction is called **DISJUNCTION**. The compound sentence 'Either logicians are mad or the moon is made of green cheese' is the *disjunction* of the two atomic sentences 'Logicians are mad' and 'The moon is made of green cheese' (each individual sentence is a 'disjunct').

The 'it's not the case that' construction is called **NEGATION**. 'There is no life after death'' (which is an abbreviated form of 'It's not the case that there is life after death') is the negation of 'There is life after death'.

The 'if ... then' construction forms the **CONDITIONAL**. In 'If the Conservatives win the next election then I shall emigrate', the ***antecedent*** is the sentence 'The Conservatives win the next election' and the ***consequent***is 'I shall emigrate'.

One assumption that will be made throughout this course is that every atomic declarative sentence is indeed *either* true *or* false (not, of course, both). Talking in a way that will prove useful later on, we can say that every atomic declarative sentence has one of the two ***TRUTH VALUES***- '**true**' or '**false**'. There are atomic sentences ('God exists', 'Man and the apes share a common ancestor'etc.) whose truth-values have been a matter of heated debate. But the fact that we may not be able to *agree* on the truth value of a sentence does not mean that it doesn't have one. 'God exists' is, presumably, either true or false – even though there is no universally agreed way of deciding which. Other sentences – a favourite example is 'Colourless green ideas sleep furiously' – although grammatically correct, and clearly of a declarative form (not, for example, an injunction like 'Shut the door!') arguably have no truth value. Some philosophers have claimed that moral assertions like 'Lying is wrong' or 'You ought not to commit adultery' also do not have truth values – they are neither true nor false, since there are no moral facts and statements like this really amount to implicit injunctions 'Don't lie!', 'Don't commit adultery!' Or maybe statements expressing the feelings of the speaker: 'I don't approve of people who lie/commit adultery.' While still other philosophers have suggested that vague statements like 'This set of pebbles forms a heap' (100 pebbles form a heap, one or two don't, but how about 8?) may have a third 'truth value' (something like 'indeterminate'). But we will ignore these complications throughout this course and assume that all declarative sentences are either true or false.

So atomic sentences are either true or false and we can make various compound sentences using the constructions outlined above. The important point about all the compound sentences just considered is that ***they depend fortheir overall truth value on the truth values of their atomic components – what truth value the compound has is determined in a definite way by the truth values of the atoms.***

Case (1), CONJUNCTION, is particularly straightforward.

CONJUNCTION:

The sentence 'Humphrey Bogart starred in *Casablanca* and Fred MacMurray starred in *Double Indemnity*'is true because both of its components (both 'conjuncts') are true. The sentence 'Ingrid Bergman starred in *Casablanca* and Veronica Lake in *Double Indemnity*' is *false*, because the second conjunct is false, even though thefirst conjunct is true. The sentence 'Karl Marx was a great composer and Beethoven a great philosopher' is false because *both* conjuncts are false.

Nothing depends in the slightest on what the individual sentences are about (film stars, composers or whatever). We know the truth-value of the compound sentence once we know the truth-values of the components. Any conjunction is true if and only if **both conjuncts are true**, and is false otherwise (that is, the conjunction is false if *either* conjunct is false, or both are). If p and q are the individual atomic sentences, then the truth value of the sentence 'p and q' is 'true' if and only if the truth values of p and q are both 'true'. We can re-express this simple rule using a graphic device known as a ***truth***

table (this graphic device is due to the famous philosopher Ludwig Wittgenstein). In order to save space, we will from now on use the symbol '**&**' to mean '**and**'.

Truth Table for Conjunction:

p	**q**	**p&q**
T	T	**T**
T	F	**F**
F	T	**F**
F	F	**F**

There are four lines in this truth table corresponding to each of the different possible *combinations* of truth values of the conjuncts. The final column gives the overall truth value of the compound for the corresponding truth values of the components.

NEGATION:

The case of negation is just as straightforward. The sentence 'It's not the case that the moon is made of green cheese' is *true* because the atomic sentence 'The moon is made of green cheese' is false. The sentence 'It's not the case that Pavarotti was a great tenor' is false because the atomic sentence 'Pavarotti was a great tenor' is true. (Let's in order to avoid heated debate understand this sentence in a timeless sense so that you are a great tenor if you ever have been – so that an opera buff could readily consent to this sentence even while believing that the estimable Luciano was over the hill for several years before he died). Again, nothing depends on the particular sentence involved: the negation of *any* true statement is false, and the negation of *any* false statement is true. For any sentence p, **not-p** (we shall use the symbol **¬p**) is true, if and only if, p is false. Again we can re-express this in the form of a truth table:

Truth Table for Negation:

p	**¬p**
T	**F**
F	**T**

DISJUNCTION:

We can also form a compound sentence using the either/or construction. Out of the sentences ‘I shall go to visit my grandma in hospital today’ and ‘I shall go to visit my grandma in hospital tomorrow’ we can form the disjunction: ‘Either I shall go to visit my grandma in hospital today or I shall go and visit my grandma in hospital tomorrow’ (more idiomatically of course ‘I shall go and visit my grandma in hospital either today or tomorrow’). Here, however, we come across an ambiguity in ordinary language.

Sometimes we use ‘either/or’ in an **inclusive** sense, meaning ‘either/or, *or both’.* If an advertisement for a university lectureship specified that a candidate must have *either a* PhD *or* scholarly publications, an applicant who had *both* a PhD *and* scholarly publications would feel *most* aggrieved if she weretold that she did not meet the specification! Here the disjunction is used in the *inclusive* sense: the disjunction is true if *either or both* of the disjuncts are. (Asanother example, someone might say ‘To have done that he is either wicked or stupid’ – a statement which we would not normally take to exclude the possibility that the person concerned is *both* wicked *and* stupid.)

Sometimes (perhaps more often) we use either/or in the **exclusive** sense – meaning ‘one or the other *but not both*’. Suppose you were the unfortunate victim of a (slightly old-fashioned) mugger who threatened ‘Your money or your life’ (more explicitly ‘Either you give me your money or I will take your life’). You would feel *very* aggrieved if, having given him your money, he proceeded to shoot you anyway – insisting that he intended the either/or in the inclusive sense! (Though, assuming he was a good shot, at least you wouldn’t feel aggrieved for too long.) So we would normally take the ‘or’ in ‘Your money or your life’ as clearly to be understood in the *exclusive* sense.

Sometimes it is unclear whether 'or' is meant in the inclusive or the exclusive sense. Is the sentence ‘Either Lennon or McCartney wrote*A day in the life*’true or false? (They both did.) Would the earlier case of ‘Either I shall go to visit my grandma in hospital today or I shall go and visit my grandma in hospital tomorrow’ be true or false if you were extra nice and went to visit her on both days?

Logic cannot tolerate ambiguity and clearly it matters for precise logical purposes which sense we take 'or' in. If both p and q are true, *then in the inclusive sense* 'p or q' is *true* but in the *exclusive* sense 'p orq' is *false.* Logicians happen (for reasons that don't matter here) to have elected to take the *inclusive* sense as primary. (As we shall see – and this *does* concern us – we won't lose anything by making this conventional decision.) The shorthand symbol for '**or**' in this inclusive sense is '**v**'. So we have the following:

Truth Table for Disjunction:

p	**q**	**p v q**
T	T	**T**
T	F	**T**
F	T	**T**
F	F	**F**

(The reason why we don't lose anything by making the conventional decision to go for the inclusive sense of either/or as primary is that when we definitely mean an either/or sentence in the exclusive sense, we can express it formally using our symbolic apparatus by a simple further compounding using inclusive-or. Suppose I say (regretfully) 'Either Manchester United or Manchester City will win the Premiership this season'. This clearly means either/or in the exclusive sense (ties are not allowed). So spelling it out more fully I assert: 'Either Manchester United will win the Premiership this season or Manchester City will win the Premiership this season, though not of course both'. Taking 'p' to be 'Manchester United will win' and q to be 'Manchester City will win', then 'p [exclusive] or q' is equivalent to '(p v q) & ¬(p&q)', where p v q as always involves v in the inclusive sense.)

CONDITIONALS:

Another way in ordinary language of making a compound sentence out of simple atomic ones is by the 'if/then' construction. Suppose, for example, a suspect is being interrogated by the police and is asked where he was on the evening of the 23rd of last month. Claiming not to have a completely clear memory, the suspect replies, not categorically but **conditionally**: 'If the 23rd was a Wednesday, then I was at the greyhound races.' (Perhaps because it is his general habit, or so he claims, to go to the dog races on Wednesday evenings.) This is called a **conditional sentence** and, just to have some handy terminology, the sentence after the ‘if’ (here the sentence ‘The 23rd was a Wednesday’) is called the **antecedent** of the conditional and the sentence after the ‘then’ (here the sentence ‘I was at the greyhound races’) is called the **consequent**of the conditional. The truth of this conditional sentence isdependent on the truth values of its components (i.e. the truth values of its antecedent and consequent), just as conjunctions and disjunctions are. However, the form of the dependency in the case of the conditional is subtler.

Let's carefully consider the truth or falsity of our particular conditional assertion under all possible different suppositions about the truth or falsity of its components. *First,* suppose that the 23rd was indeed a Wednesday (antecedent true) and that the suspect did indeed go to the greyhound track that night (consequent also true). In that case we would surely regard the suspect as having spoken truly when he said 'If the 23rd was a Wednesday, then I was at the greyhound races', that is, we would regard his 'if/then' statement as being true.

Now suppose that the 23rd*was* a Wednesday (antecedent true), but the suspect was *not* at the dog track (consequent false). In that case (true antecedent, false consequent) the suspect surely spoke falsely – his 'if/then' statement was definitely *false.*

These are the two obvious cases (the two cases in which the antecedent is true) and they dictate two out of the four lines in the truth table for the conditional. But what if the antecedent was *false*? What if the 23rd was, say, a Friday rather than a Wednesday? Here intuitions are not altogether clear. But it *is* surely clear that if the 23rd was a Friday, then the suspect *did not lie* when he said that ‘If the 23rd was a Wednesday then I went to the greyhound races’ – and this is so *either* in the case that he went to the greyhound races on Friday the 23rd*or* he did not. That is, his conditional assertion is at least not outright false if the antecedent is false whatever the truth value of the consequent.

If therefore we are to stick by our decision that, for our purposes in this course, all (grammatically legitimate) sentences are to be regarded as either true or false, then it would seem that we are forced to the conclusion that the conditional ‘If the 23rd was a Wednesday, then I was at the greyhound races’ is*true* both in the case that the antecedent is false and the consequent true, (the 23rd was not a Wednesday, but he was at the dogs) *and* in the case that the antecedent is false and the consequent is false (the 23rd was not a Wednesday and he was, say, in fact shooting 'Dangerous Dan' McGrew somewhere far away from the greyhound stadium).

The reason why this case is not clear-cut is that it is not clear that we are intuitively happy to say that the conditional is indeed true in these two cases. It perhaps seems more natural, certainly in this instance, to regard the conditional as 'Not applicable' when the antecedent is false – the suspect's assertion only 'comes into play' if the antecedent is true (if the 23rd was indeed a Wednesday) and is then clearly true if the consequent is true (he *had* gone to the dogs) and false if the antecedent is false (he was somewhere else).

On the other hand, suppose – going back to the drunken stupor case we thought about earlier – I say: 'If today is Sunday, then yesterday was Saturday'. Suppose, moreover, that I am wrong about today being Sunday – in fact I had so many drinks on Saturday that I slept through the whole of Sunday and today is in fact Monday. In that case, the antecedent of my conditional assertion is false (it's not true that today is Sunday), so also as a matter of fact is the consequent (yesterday was Sunday not Saturday) – nonetheless most of us would still want to say that my conditional assertion was true. So there are at least some cases in which 'if/then' sentences with false antecedents are intuitively regarded as true.

Moreover, consider a sentence like: 'If Tony Blair really believed that there were WMDs in Iraq, then I am a Dutchman (or 'then I am Marilyn Monroe' or 'then pigs can fly')'. We actually use constructions like this (there are other ones used in other cultures and age groups) as an emphatic way of asserting a negation. What you would actually mean to imply if you asserted such a sentence is that (of course, in your opinion – let's not get into making any assertion about the actual facts here in case lawyers might become involved), Tony Blair definitely did *not* really believe that there were WMDs in Iraq. And you imply that by using an *obviously* false sentence (like 'I'm a Dutchman' or 'I'm Marilyn Monroe' or 'The Pope is Jewish') as consequent of a conditional that you assert as true. So in conditional sentences like these everyone knows the consequent to be false; *any* conditional, we all agreed, is unambiguously false if the antecedent is truebut the consequent false; hence, in this Blair case, your conditional would be false not true if the antecedent were true – that's why by asserting the overall conditional (asserting it to be true) you in effect assert its antecedent to be false.So again this is a case in which a conditional is intuitively true (rather than not applicable) when it has a false antecedent and a false consequent.

So, what are we to do? Clearly we cannot hope to capture ordinary usage directly since ordinary usage is not unambiguous and logic does not tolerate ambiguity. Again, as in the case of disjunction, we make a decision that captures some of the intuitions and hope that the others can be met by more sophisticated means (it's in fact a lot less clear-cut than in the case of disjunction if they can: the status of conditionals remains an issue of hot debate in analytic philosophy, but these debates will not concern us in this course). The decision taken in logic is to stick with the idea that all sentences are either true or false (that is, to avoid the 'not applicable' possibility). This means – since we agreed that our criminal suspect's sentence was certainly not false if the antecedent was false (if the 23rd was not a Wednesday) – that we must take the conditional as true whenever its antecedent is false. So, symbolising any sentence of the form 'if p then q' as '**p → q**', we have the following:

Truth Table for the Conditional:

p	**q**	**p → q**
T	T	**T**
T	F	**F**
F	T	**T**
F	F	**T**

BI-CONDITIONALS:

The final way of compounding atomic sentences that we shall consider is found more often in scientific and mathematical contexts than in ordinary discourse, but – maybe because of this – is another straightforward case like conjunction. This involves connecting sentences using the phrase **'if and only if'**. For example, someone might say 'Corbyn will survive as Labour leader if and only if Labour wins the next election' or an economist might predict ‘The economy will recover if and only if the interest rate is increased by a whole point’. (Synonymous phrases to ‘if and only if’ are 'exactly when' or 'just in case'.) The 'if and only if' connective (often abbreviated to '**iff**' and symbolized as ‘↔') is again clearly truth-functional: that is, the truth value of the compound ‘p ↔ q’ is dependent on the truth values of p and q. In fact, p ↔ q is true whenever p and q have the *same* truth value and false whenever they have *different* truth values.

So, for example, the statement about Corbyn will turn out to be true if one of two cases turns out to hold (a) Corbyn survives and Labour wins the next election (p and q both true) or (b) Corbyn does not survive and Labour loses (p and q both false). If, on the other hand, (c) Labour loses and yet Corbyn survives as leader (p true, q false) or (d) Labour wins but Corbyn is ousted (p false, q true) – that is in either of the two separate cases in which p and q have different truth values, then the assertion (prediction) that ‘Corbyn will survive as Labour leader if and only if Labour wins the next election’ clearly has turned out to be false.

Sowe have the following:

Truth Table for the Bi-conditional:

p	**q**	**p ↔q**
T	T	**T**
T	F	**F**
F	T	**F**
F	F	**T**

FURTHER COMPOUNDING:

The procedures of building more complicated sentences out of ‘atomic’ components are, of course, not restricted to single-step affairs: they can be iterated (indeed they can, in principle, be iterated any number of times). Given any compound sentence you have already created you can put any two of them together in any of the ways just indicated to form a more complex, though still single, sentence.

So, for example, we can form a conditional whose consequent is itself a conjunction: e.g. ’If the Tories win the next election, then income tax rates will decrease *and* social inequality will increase’. Or we can form a conditional both the antecedent of which and the consequent are themselves compounds: e.g. ‘If *either* Manchester United *or* Manchester City wins the premiership, then I shall be very unhappy *and* so will every other Liverpool fan.’ Things can get as complicated as you like: for example, having created the *conjunctions* ‘The US remains the only Western global superpower and the situation in the Middle East will get worse’, and ‘A Federal Europe will be created as a new Western global superpower and the situation in the Middle East will improve’, we can go on to form the *disjunction* of the two conjunctions: *'Either* the US remains the only Westernglobal superpower and the situation in the Middle East will get worse *or* a Federal Europe will be created as a new Western global superpower and the situation in the Middle East will slowly improve’.

Or you might be told at some airport: ‘If you are either a British citizen or a citizen of an EU country then you need not fill in a landing card and you should go through channel A.’ This compound sentence is a conditional whose antecedent is a disjunction and whose consequent is a conjunction. The combinations, the ways of compounding, are (literally) endless. The way to appreciate this is through looking at examples, of which you will be given plenty.

Unsurprisingly, any such compound, no matter how complex, is itself truth-functional: that is, its truth value depends in some clearly specifiable way on the truth-values of its components. It’s just a matter of applying the rules for the individual connectives (and, either/or, if/then, etc. in the appropriate order). In order to see this, we first need to think about how to formalise more complex compounds using symbols.

Formalising complex truth functional compounds – the need for brackets:

Let’s think again about a couple of examples we just had: starting with ‘If the Tories win the next election, then income tax rates will decrease *and* social inequality will increase’. Taking p as ‘The Tories will win the next election’, q as ‘Income tax rates will decrease’ and r as ‘Social inequality will increase’, this looks like it formalises as:

$p \rightarrow q \,\&\, r$

But this is ambiguous as it stands. It could also mean the – quite different and admittedly rather strange – assertion:

‘If the Tories win the next election then income tax rates will decrease *and* [in any event] social equality will increase.’

This indicates the need for brackets (brackets are very important in logic!) to disambiguate. What we really meant was:

$p \rightarrow (q \,\&\, r)$ [we meant to assert a conditional with a conjunctive consequent]

The alternative reading would be expressed as

$(p \rightarrow q) \,\&\, r$ [this unintended reading makes it a conjunction, the first of whose conjuncts is a conditional]

Or take ‘If *either* Manchester United *or* Manchester City wins the premiership then I will be very unhappy *and* so will every other Liverpool fan.’ Taking p as ‘Manchester United wins the premiership’, q as ‘Manchester City wins the premiership’, r as ‘I will be very unhappy’ and s as ‘Every other Liverpool fan will be very unhappy’, it looks like this formalises as:

$p \vee q \rightarrow r \,\&\, s$

But again, without brackets, this is ambiguous – multiply so in this case. It could mean what we wanted it to mean but it could equally (if oddly be read as):

Alternative 1: *Either* Manchester United wins the premiership *or* if Manchester City wins the premiership then I will be unhappy and so will every other Liverpool fan.

Alternative 2: If either Manchester United or Manchester City wins the premiership, then I will be very unhappy *and* [in any case] every other Liverpool fan will be very unhappy.

In order to disambiguate again we need brackets. What we really meant was:

$(p \vee q) \rightarrow (r \,\&\, s)$

That is, we wanted to assert a conditional with a disjunctive antecedent and a conjunctive consequent – that is exactly what the bracketing indicates.

Exercise: How we would correctly formalise alternatives 1 and 2? That is, how we would use brackets to make those 2 (aberrant) assertions?

Finally, let's think about our political example: *'Either* The US remains the only global superpower and the situation in the Middle East will get worse *or* a Federal Europe will be created as a new global superpower and the situation in the Middle East will slowly improve'. This is clearly a disjunction each of whose disjuncts is a conjunction. So, letting p be the sentence about the US, q the sentence that 'The situation in the Middle East will get worse', r the sentence about a Federal Europe and s the sentence that 'The situation in the Middle East will slowly improve', the correct formalisation is:

$$(p \,\&\, q) \vee (r \,\&\, s)$$

Exercise: where p, q, r and s retain this meaning, what do each of:

$$p \,\&\, (q \vee (r \,\&\, s))$$

$$p \,\&\, ((q \vee r) \,\&\, s)$$

mean in ordinary language?

You will soon get used to this by practising formalising ordinary language sentences. There are one or two wrinkles that turn up. For example, strictly speaking, we should if we wish to consider the negation just of the atomic sentence sentence p, write ¬(p) to indicate that the negation is only of p. But, in order to save typescript, we avoid brackets in that case and just write ¬p. So the consequent of the intended formalisation of our airport sentence, (¬r & s), is to be read as 'not r and [but!] s' (i.e. you need not fill in a landing card and you should go through Channel A). The formula ¬(r & s), on the other hand, says that it's not true that you *both* need to fill in a landing card *and* that you should go through Channel A.

Exercise:

1. Take a more straightforward case: Say p is the sentence 'Blair was a liar' and q is the sentence 'Bush was a liar': what do each of ¬p & q, p & ¬q and ¬(p & q) mean in ordinary language?
2. ¬(p & q) is actually equivalent to (i.e. says the same thing as) a certain disjunction, can you work out which one?

There are other cases where brackets become redundant – but it is best to learn of these through exercises, rather than through trying to understand a general explanation. Certainly the rule is ‘if in doubt leave the brackets in’.

By using the basic truth tables step by step we can build up a truth table for these more complicated sentences too. This will again tell us the overall truth value of the sentence for every possible combination of truth values of the atomic components. Let’s take our airport case again, which formalised, remember, as (p v q) → (¬r &s).

There are in this case four atomic sentences: p, q, r and s and hence a total of16possible combinations of truth values and hence 16 lines in the truth table. (It is easy to prove mathematically that if there are n atoms, there are 2^n possible combinations of truth values and 2^4 of course equals 16.) The relevant truth table is then:

p	**q**	**r**	**s**	**(p v q) → (¬r & s)**
T	T	T	T	t **F** f f
T	T	T	F	t **F** f f
T	T	F	T	**T**
T	T	F	F	**F**
T	F	T	T	**F**
T	F	T	F	**F**
T	F	F	T	**T**
T	F	F	F	**F**
F	T	T	T	**F**
F	T	T	F	**F**
F	T	F	T	**T**
F	T	F	F	**F**
F	F	T	T	**T**
F	F	T	F	**T**
F	F	F	T	**T**
F	F	F	F	**T**

Here, I have indicated the overall truth value of the compound in each row – employing, as usual, capital Ts and Fs and placing these overall truth values under the 'main connective', in this case the conditional (the sentence as we already agreed, and as we made the bracketing indicate, is a conditional with a disjunctive antecedent and a conjunctive consequent). I have also indicated in the case of the first two rows *and using lower case letters to indicate theintermediate steps* how the overall truth value for that row is to be worked out.

So, in the first row, all the atoms take the truth value true, hence (p v q) is true by the basic truth table for a disjunction (hence the 't' under that bit of the sentence), but ¬r is false (by the basic truth table for negation, given that r is true) and hence, although s is true, by the basic truth table for conjunction, (¬r & s) is false (hence the double bit of working out under (¬r&s) ending up with an f. So, finally, we have now worked out that for this particular assignment of truth values (all atoms true) we have, for the overall conditional, a true antecedent and a false consequent: and hence, by the basic truth table for the conditional, we have the overall truth value F (indicated by the capital F under the main connective, the '→').

Similarly, for the second row, the antecedent is again true, the consequent is false (both ¬r and s are false in that second row), and so again we have 'true → false' which again gives overall F by the basic truth table for the conditional.

Exercise: Go carefully through each of the remaining 14 rows and check that the overall truth value assigned to that row is correct. Although you should force yourself to go through all the working (just this once, on the assumption you find this stuff easy), you may notice that various short cuts are possible – for example you know the last 4 lines must all take the overall value T once you have seen that the antecedent is false in all those lines and independently of what happens with the consequent ¬r & s: this is because, by its basic truth table, a conditional, takes the overall value T if the antecedent is F, regardless of whether if the consequent takes T orif it takes F.

You should check that this method gives the right answers intuitively for each line. So remember our sentence read: 'If you are either a British citizen or a citizen of another EU country, then you need not fill in a landing card and you should go through channel A'. Of course we normally suppose that such notices are only put up in airports if the sentence they contain is true, but we are supposing for the sake of this argument that the sentence may be true or false (perhaps some joker has been putting up notices some of which are correct, some incorrect, just for his perverted enjoyment of the subsequent confusion). So let's consider line 1: if line 1 holds (that is if the possible assignment of truth values to atoms that it contains is the one which actually holds in the real world), then the facts of the matter are:

1. You are a British citizen.
2. You are a citizen of an EU country (so you have dual nationality).
3. You *do* need to fill in a landing card (remember r stood for the unnegated sentence 'you do need to fill in a landing card').
4. You should go through channel A.

The rule then is indeed false, you have been misinformed, and having supposed that you do not need to fill in a landing card, you will presumably be stopped at customs and required to do so.

Similarly, if line 2 reflects the real world (the real truth values) then you are again a dual British and EU country national, you again need to fill in a landing card, but now you also should not go through channel A (s is false). So again our truth table agrees with our intuitions that in this case the sentence on the notice was false – we shall be in trouble twice over with customs by following what it says: we both need to fill in the landing card and we went down the wrong channel.

Notice finally the ***systematic method for ensuring that you have considered all possible combinations of truth values*** as exemplified in this case but to beused in all cases:

1. Construct a table with 2^n lines, where n is the number of atomic sentences involved in the compound sentence under consideration (in this case $2^4 = 16$). This is because, as remarked, it is mathematically provable that with *n* atoms, there are always 2^n possible combinations of truth values.
2. The easiest systematic way to remember (there are of course lots of other systems) is to start with the first atom, p, and write the first half of the 2^n rows (in this case the first 8 rows) T, and the second half F. Then for the second atom, q, write the first half of those rows in which p got T as T (in other words, in our case the first 4 rows) as T and the second half (next 4) as F, then the next 4 (where p has the value F) as T for q, and the final 4 rows as F for q; then, halving again, for r go 2Ts, 2Fs, 2Ts, 2Fs; and finally for the final atom, s, alternate Ts and Fs. (This is one of those things that is easier to see than explain – just look at the truth table and see how it's done and generalise the obvious procedure: so if you had a compound with 5 atomic sentences p, q, r, s and t and therefore 32 rows, you'd go 16Ts, 16Fs, for p, 8Ts 8Fs 8Ts 8Fs for q and so on through to TFTF… for the last atom, t.)

Exercise: Again to underline the importance of bracketing, consider two alternative ways of understanding the unbracketed string of symbols

$p \vee q \rightarrow \neg r \& s$

(where p, q, r and s all mean the same as in the real airport example):

Alternative 1: $p \vee (q \rightarrow (\neg r \& s))$

Alternative 2: $p \vee (q \rightarrow \neg (r \& s))$

Construct truth tables for each of these alternatives and show *both* that they differ from one another (that is, there are at least some lines in which they take different overall truth values) *and* that they each differ from the truth table that we constructed for the real airport sentence that we have been considering.

A3(c): Tautologies, Contradictions and Contingent Sentences

We will return to our principal concern – validity of inference – soon. But first we will take what will seem like a digression but which will turn out to useful as concerns validity.

There is a tri-partite distinction regarding ***single sentences*** (so we have left inferences for the moment). If I tell you 'The *Genesis* account of the creation of the universe is true' then I have made a claim about the world – some of you may think it preposterous, but nonetheless it at least *purports* to carry some information about the world. Similarly, if I tell you: 'September 11th 2001 was a Monday', I tell you something that may be true or false and, if true (which it is), gives us some real information about the world (which it does).

If, on the other hand, I tell you just that 'Either the *Genesis* account of the creation of the universe is true or that account is not true', or 'Either September 11 2001 was a Monday or it was not', then what I tell you is true alright – indeed it is guaranteed to be true, but it is *trivially* or *vacuously* true. These sentences ***could notpossibly be false*** no matter what the state of the world may be. On the otherhand, whether or not the initial statement that the *Genesis* account is true or that '9/11' was a Monday is true does depend on facts about the world.

This distinction is reflected in a feature of the truth tables of some sentences. If we formalise the statement: 'Either the *Genesis* account of the creation of the universe is true or that account is not true', then we obtain of course:

$p \vee \neg p$.

This has the following very simple truth table:

p	**p v ¬p**
T	**T**
F	**T**

In fact, the only overall truth value in any row of this table is T. This reflects the fact that this sentence is true independently of the way the world is, it is true ‘in all possible worlds’. Such sentences are called (truth-functional) **tautologies** (tautologies, as we shall later, are a special case of a more generalspecies of **logical truths**).

So,

<u>Definition:</u> Tautology

A truth functional compound is a***tautology***if and only if ittakes the truth value ‘true’ in all lines of its truth table.

Slightly less obviously tautological tautologies than the Genesis or ‘9/11’ ones include:

If Farage is a liar, then either Farage or Duncan Smith is a liar.

Either if Farage is a liar then so is Duncan Smith *or* Farage is a liar but [and] Duncan Smith isn’t.

(*Exercise*: Formalise these two sentences (you will need as always to be careful about brackets, especially in formalising 2) and then show that their truth tables do indeed have the truth value true in all rows.)

On the other hand, statements like 'The *Genesis* acount of creation is true', or ‘Either the Tories will win the next election or I'm a bad judge of British politics’, are, if true at all, ***contingently*** true – their truth depends on, or is *contingent* on, the way the world is (or was or will be). This is reflected in thefact that their truth tables have **at least one line** 'true' and **at least one line** 'false'. The first statement formalises, of course, just as p. It has the trivial truth table:

p	**p**
T	**T**
F	**F**

The other can be formalised as p v ¬q (where p stands for ‘The Tories will win the next election’ and q for the assertion ‘I am a good judge of British politics’) and this has the truth table:

p	**q**	**p v ¬q**
T	T	**T**
T	F	**T**
F	T	**F**
F	F	**T**

Both of these last two tables, then, have at least one T and at least one F. Some people find it useful to think of the assignments of truth values to the atomic sentences as defining 'possible worlds' (e.g. the fourth line specifies the 'possible world' in which the Tories do not win the next election and I am a bad judge of British politics). Contingent sentences are ones whose truth value depends on the actual world – which line in the truth table, which 'possible world' corresponds to the real one. Tautologies on the other hand are sometimes said to be 'true in all possible worlds'. Hence

***Definition:* Contingent Sentence**

A truth functional compound is a **contingent sentence** if andonly if it takes the truth value true in at least one line of its truth table and the truth value false in at least one (other) line.

At the other end of the spectrum from tautologies are statements like 'The Tories will win the next election and they will not', or 'The Tories will win thenext election if and only if they do not', which are false but which don't just *happen* to be false, they are *necessarily* false. (It is logically impossible that they could be true.) These Sentences are called ***contradictions.*** When formalized their truth table takes 'false' inall lines. Take for example the second sentence. This has the truth table:

p	**p ↔ ¬p**
T	t ↔ f **F**
F	f ↔ t **F**

This sentence is logically, or trivially, false – "false in all possible worlds". So:

Definition**:** ***Contradiction***

A truth functional compound is a **contradiction** if and only ifit takes the truth value false in all lines of its truth table.

Elaborating a little on the difference between a contingent truth and a tautology or logical truth that has been introduced in this section, we might say: The sentence, for example, 'There are approximately 10^{11} stars in our own galaxy and approximately 10^{9} other galaxies in the universe' formalises as 'p & q' and happens to be (incidentally, mindbogglingly) *contingently true.* It is true, or if you like 'true in the world', because its formalisation is true when its atomic components take the truth value which they take in the 'real world' (namely both take the truth value 'true'). However, it is ‘only’ contingently true, not logically true, because there are other lines in its truth table which take the truth value false – in this sense, the sentence *might possibly* have been false, even though as a matter of fact it is true. On the other hand, the sentence 'Either there are 10^{11} stars in our galaxy or there aren't’ is *necessarily* or *logically true.* It formalises as 'p v ¬p' and it too is true 'in the real world' – that is, when its atomic components (in this case just 'p') take the truth value they actually take (in this case again p: true). But, in contrast, this sentence *could not possibly* be false – not only is its formalisation true 'in the real world', it is true in all other cases ('possible worlds') as well (in this instance 'all other cases' reduce to one case, *viz* p: false).

We now return to the central concern of logic: the question of when an *inference is valid.* You will remember that in giving an intuitive characterisation of the notion of validity, we invoked ideas of the *possible* truth/falsity of premises and conclusion. We are now in a position to give these a rigorous formulation and, as we will see, the notion of a tautology is involved in the most straightforward characterisation of validity of inference.

A3(D): TRUTH-FUNCTIONAL VALIDITY (AGAIN)

Look back over our earlier discussion of the notion of validity. We arrived at the idea that an inference is valid iff (remember this stands for: if and only if) there is no counterexample to it, where a counterexample is defined as an inference of the same logical form that has true premises and a false conclusion. We can now clarify these definitions further.

First of all, the ***form*** of an inference is specified, for the range of inferences we are currently considering, by its truth-functional formalisation. Two different inferences have the same form if they produce the same symbolic schema when formalised. So the two earlier inferences that we considered earlier:

A:

1. Either the Edinburgh train leaves from Euston or it leaves from Kings Cross**.**
2. It does not leave from Euston.

So, it leaves from Kings Cross.

And:

B:

1. Either *Tosca* is by Puccini or it is by Verdi.
2. *Tosca* is not by Verdi.

So, *Tosca* is by Puccini

Both have the same formalisation, namely:

1. p v q
2. ¬p

So, q

Hence inferences A and B, while of course different inferences (one about railstations and the other about composers) nonetheless **have the same form**.

If we construct a joint truth table for the three symbolic sentences concerned, we obtain the following:

p	**q**	**p v q**	**¬p**	**q**
T	T	T	F	**T**
T	F	T	F	**F**
F	T	T	T	**T**
F	F	F	T	**F**

where I have, in order to make the point, repeated the line for q since this is the conclusion of the inference.

Inspect this joint truth table closely: you will see that there are lines in it in which:

1. The premises are all true and so is the conclusion (line 3)
2. At least one of the premises is false, and the conclusion is true (line 1)
3. At least one of the premises is false, and the conclusions is false (lines 2 and 4)

However, there is **no line in the truth table in which all the premises aretrue and the conclusion is false**. That is, there is no counterexample and thisin turn means that both the train inference and the *Tosca* inference are ***valid***.

Now, consider, by way of contrast, two further inferences – the first a slightly simplified version of our earlier 'Agatha Christie' example and the second the one about Jo DiMaggio:

C:

1. If the Butler 'did it', then he had the motive.
2. The Butler had the motive.

Therefore, the Butler 'did it'

D:

1. If Joe DiMaggio was US President, then he was born in the US.
2. Joe DiMaggio was born in the US.

Therefore, Joe diMaggio was US President.

These are inferences of the same form, as is shown by the fact thatthey produce the same symbolic representation:

1. $p \to q$
2. q

So, p

But unlike **A** and **B**, these are both ***invalid*** inferences. In fact, **D** is a case in which the premises are true (in the 'real world') and the conclusion false (Joe DiMaggio was a US citizen but not president, so both premises are true (in the real world) and the conclusion false). So **D** supplies a *counterexample both to* **C**and (trivially) to itself. More formally, if we againdraw up a joint truth table for the two premises and the conclusion of the symbolic representation of either **C** or **D** we obtain:

p	**q**	**$p \to q$**	**q**	**p**
T	T	T	T	**T**
T	F	F	F	**T**
F	T	T	T	**F**
F	F	T	F	**F**

where I have again, redundantly, repeated q and p to the right since they appear as second premise and conclusion of the inference.

We see that in this case, unlike the case of the symbolic representation of either inference **A** or **B**, **there is a line in this truth table** (namely the third) **inwhich both premises are true and yet the conclusion is false.** Hence there isa ***counterexample*** to either of the inferences **C** and **D**. It may be true in the real world (in this case the “real world” of the fictional novel!) that the Butler did it, but that he did does not follow from the fact that if he did it then he had a motive, *and* he had a motive. There is a ‘possible world’ in which both premises are true and the conclusion false, or, putting it in a more down to earth way, there is an assignment of truth values to the atomic components of the inference – given by the third row, i.e. p:F, q:T – in which both premises are true and the conclusion false.

Put precisely, we have arrived at the following definition of (truth functional) validity:

***<u>Definition:</u>* Validity**

An inference form written in the language of truthfunctional logic is a VALID form iff there is no assignment of truth values to its atomic components which makes the premises true and the conclusion false.

As an old, and very boring, school teacher of mine used to say endlessly, this definition should be 'read, learned and inwardly digested'.

Correspondingly***, the original inference itself (written, like inferences A, B, C or D, in English or some other 'natural language') is truth functionally valid iff its symbolic representation in truth functional logic is a valid inference form according to the above formal definition***.

A3(E): DECISION PROCEDURES FOR TRUTH-FUNCTIONAL VALIDITY

The above considerations not only tell us what it means for an inference to be (truth-functionally) valid, they also indicate one method of *deciding* whether a given inference is valid or invalid. In fact, in the case of inferences in the language of truth-functional logic, we are in the happy position of being able to specify several different *algorithms* or *mechanical decision procedures* for ascertaining validity or invalidity. (They are called algorithms because they are guaranteed always to deliver the correct answer when applied to any inference.)

(i) The Truth Table Method

The first method is the essentially the one that we just used:

1. Formalise the inference.
2. Create a joint truth table for all the premises and the conclusion.
3. Check to see if there is a single line in which all the premises are true but the conclusion false.

   If there is such a line, then the inference is ***invalid.***

   If there is no such line, then inference is ***valid.***

Clearly this is indeed an algorithm – that is, it is bound to give an answer in all cases. The premises and conclusion of any inference, no matter how complicated, involve only finitely many atomic sentences and hence the joint truth table will have only finitely many rows (in fact, as we already noted, it will have 2^n rows where*n* is the number of atomic components). We just then need to look through all the rows and will either find a row in which all the premises are true and the conclusion false (in which case the inference is invalid) or wewill get to the end of the 2^n rows without finding one with this property (in which case the inference is valid).

The method can be given a rather more elegant form by considering something called the **associated conditional** of the inference:

1. The ***associated conditional*** of an inference is the ***singlesentence,*** conditional in form, which has the *conjunction of the premises* as *antecedent* and the *conclusion* as *consequent.*

So the associated conditional for both inference**A**and inference**B**is:

$((p \vee q) \,\&\, \neg p) \rightarrow q$

while the associated conditional for either inference **C** or **D** is:

$((p \rightarrow q) \,\&\, q) \rightarrow p$

(*Notice* bracketing is again important.)

2. Next, construct the *truth-table* for this associated conditional.

(Exercise: construct truth-tables for the above two 'associated conditionals'.)

3. Note whether the associated conditional is *a **TAUTOLOGY** (i.e.* all lines in table yield T) or *NOT* (at least one line is F).

If the associated conditional IS a tautology, then the *inference is* **VALID.**

If the associated conditional is NOT a tautology, then the inference is **INVALID**.

It is easy to see that this just is another way of checking whether or not there is a line in the joint truth table which makes all premises true and conclusion false by thinking a bit about step 4: any conditional $P \rightarrow Q$ (we here, and from now on, use capital Ps and Qs *etc* whenever these may themselves be *compound* sentences, reserving lower case ps and qs *etc.* for atomic sentences) is false just in case P is true and Q is false, and since a conjunction is true just in case *all* its conjuncts are, the associated conditional for an inference will have at least one F in its truth table (i.e. will *not* be a tautology) if and only if there is at least one assignment of truth values to the atomic components (i.e. at least one line in the truth table) which makes all the premises of the inference true and its conclusion false, that is, if and only if the inference is **invalid**.

So if you have done the latest exercise correctly you will have found that the *associated conditional* for either inference **A** or **B**, *viz.* $((p \vee q) \,\&\, \neg p) \rightarrow q)$ is a tautology (reflecting the fact that either inference is valid – no assignment of truth values to atomic components that makes all the premises true and the conclusion false).Meanwhile, the associated conditional for either inference **C** or **D**, *viz.* ((p→q) & q) → p) is **not** a tautology (note carefully that the lines in that conditional's truth table which show that it is not a tautology, i.e. in which it takes the value F, make both premises of the original inference true and its conclusion false). So, C and D are invalid.

(ii) The 'No Counterexample' Method

Since the truth table for any compound proposition with *n* atomic components has 2^n rows in it, the truth table method soon becomes fairly unwieldy with inferences of any complexity. Fortunately, it is possible to take a short cut by using the time-honoured method of 'indirect proof' or*'reductio ad absurdum'*: to prove ¬P is the case, assume P and derive a contradiction; this must mean that ¬P is true, since if P entails a contradiction it cannot be true. This soundscomplicated, but in fact turns out to be much quicker and more direct than the truth table method (especially asnoted, when many atomic sentences are involved). Let's first see the method at work and then describe it in general.

Go back again to inferences **A** and **B**, these, as we know, had the form

1. p v q
2. ¬p

So, q

This is a valid inference form. The 'no counterexample method' of showing this proceeds as follows:

1. *Assume* that the inference is ***INVALID***; i.e. that there *is* anassignment of truth values to the atomic components for which all the premises are *true* and the conclusion is *false.*
2. Since q is the conclusion, we're assuming that q is false.
3. Since ¬p is a premise, and we are assuming all premises are true, ¬p must be true and therefore p false.

4. But assumptions (2) and (3) together mean that the first premise ($p \vee q$) must be *false* (by the truth table for disjunction) and this ***CONTRADICTS*** our assumption that all premises were true.
5. Hence, the assumption that the inference is invalid has proved untenable (it turns out to impose contradictory, and therefore unfulfillable requirements), and so we conclude that the inference is ***VALID***.

So we have indeed used the argument of 'reduction to absurdity': we assumed there is a counterexample (inference invalid), derived a contradiction from that assumption and inferred that the inference cannot be invalid, i.e. that it is in fact valid.

Now consider inferences **C** and **D**. These formalised as:

1. $p \rightarrow q$
2. q

So, p

This, as we know, is an invalid inference. The no counterexample method establishes this as follows:

1. Assume the inference is ***INVALID***; i.e. that there is a case in which all premises are true and conclusion false.
2. p is the conclusion so p is false.
3. q is a premise so q is true.
4. This means that p is false and q is true in the first premise ($p \rightarrow q$), but that's OK, since by the truth table for the conditional F→T *is true.*
5. So, in this case, our assumption that the inference is invalid has *not* led to a contradiction, therefore, the inference is invalid, and the method has in fact led us to an actual ***COUNTEREXAMPLE***: viz. an assignment of truth values to the atoms (in fact p:F, q:T) which indeed makes all premises true and the conclusion false.

So, in general, the ***no counterexample method*** is as follows:

1. Take the inference at issue and assume that it is invalid.

2. Work out what this assumption requires by way of assignments of truth values to the atomic components.
3. You will either be led to a contradiction – that is, you won't be able consistently to assign truth values to atomic components given the assumption of invalidity – or you won't be led to a contradiction.

If you are led to a contradiction, then the inference cannot be invalid and therefore is **valid.**

If you are not led to a contradiction, then the inference is **invalid** (and you will in the process have constructed a counterexample).

Obviously with only two atoms, the no counterexample method is not greatly more efficient than the truth table method. But it *does* come into its own with more complicated inferences. Consider, for example, the moderately more complicated inference which has 5 atoms and therefore a 32-line truth table):

1. (p&q) ↔ (r→s)
2. s v ¬t

So, ¬s → ((r→ ¬p) & ¬t)

1. Assume that the inference is invalid.
2. This means in particular that the conclusion is false, but that means that ¬s is true (and so s is false), and either (r →¬p) or ¬t false.
3. But if t were true (as it would have to be to make ¬t false), then s would have to be true to make premise 2 true, and we already known (from step 2) that s is false. So it must be the case that t is false (to make premise 2 true).
4. So, from steps 2 and 3, we know (r → ¬p) is false, to make the conclusion false. That means r is true and ¬p is false, i.e. r:T, p:T.
5. So, we have p:T, r:T, s:F, t:F. Is this consistent with the truth of the first premise? Well, we have r:T and s:F so the right hand side of the biconditional is F. p is true, but we have no information yet on q, so we can consistently make q false, and then the LHS is false too, and hence the biconditional is true.
6. Hence the inference is invalid as p:T, q:F, r:T, s:F and t:F is a counterexample.

Or consider the following inference:

1. $\neg((p \to q) \vee (r \to s))$
2. $t \to s$

So, $\neg q$ & $\neg t$

1. Again assumethat the inference is invalid.
2. This means in particular that premise 1 is true, and hence that $(p \to q) \vee (r \to s)$ is false.
3. But that requires both disjuncts to be false, which in turn requires the following truth values: p:T, p:F, r:T, s:F.
4. But if s:F, then it must be the case that t:F in order for the second premise to be true.
5. But in order for the conclusion to be false at least one of ¬q and ¬t must be false, i.e. at least one of q and t must be true.
6. But we just worked out that both q and t are F.
7. Hence we have a contradiction. The assumption that the inference is invalid is untenable and hence the inference is **valid**.

(iii) The Method of Semantic Trees

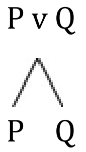
The third algorithm for truth functional validity is really just a systematic and more elegant version of the 'no counterexample' method.

An inference is truth functionally invalid, remember, if and only if, there is an assignment of truth values to the atomic sentences which makes the premises true and the conclusion false. Obviously this at the same time would be a truth value assignment which makes the premises of the inference and the *negation* of the conclusion of the inference *alltrue*.

The semantic tree method systematically explores whether it's possible for a given set of sentences to be true together (that is, whether there is at least one assignment of truth values to atomic components that makes all the sentences in that set true). To apply the semantic tree method to the question of whether a given inference is valid, therefore, we apply it to the set consisting of the premises of the inference together with *thenegation of its conclusion*.

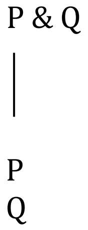
The basic idea of the tree method is that whenever there is more than one way for a sentence to be true, this is signified by a *branching* of possibilities – the tree branches at that point. For example, for a sentence of the form 'P v Q' to be true either P can be true or Q (or of course both, given that 'v' *is 'inclusive* or'). This is indicated by the basic schema for disjunction (remember we are using capital letters to indicate that the sentences concerned may be compound and hence have further truth-functional structure):

(i)



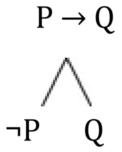
On the other hand, for 'P & Q' to be true, *both P and* Q must be true. There is no branching of possibilities, and hence the basic schema for a conjunction is:

(ii)



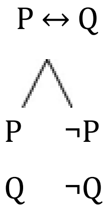
What about sentences involving the other connectives that we have introduced? Well, the sentence 'P → Q' is, it turns out, equivalent to '¬P v Q'. (*Exercise*: As we shall note in more detail later, logical equivalence means, in the case of truth functional logic, having exactly the same truth table. Show that 'P → Q' and '¬P v Q' do indeed have the same truth table.) Given this equivalence and the basic splitting idea, the schema for the conditional must be:

(iii)



As for biconditionals, a sentence of the form P ↔ Q basically says that the two sentences P and Q have the same truth value, i.e. there are two possibilities: both true and both false; but in order for a sentence to false, its negation must of course be true. So, we have the schema:

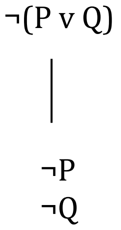
(iv)



Next we need a set of rules for negated sentences. Again these can easily be constructed by thinking about what the negated sentences mean and applying the basic idea of how many different ways such a sentence might be true.

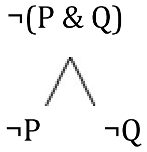
So, for example, in order for a sentence of the form ¬(P v Q) to be true, P v Q itself must of course be false and there is only one way for that to come about – namely both P and Q must be false (think about the truth table for P v Q, if this isn't already obvious). Hence we have the rule:

(i)'



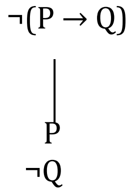
Similarly for ¬(P & Q) to be true, P & Q must be false but here there are *two* ways in which that can happen – *either* P *or* Q to be false (or of course both, but we don't need to take that into account, the fork in the tree method in effect represents inclusive or). So we have the rule:

(ii)'



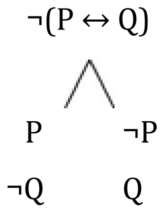
As for ¬(P → Q), the only way for a conditional to false and hence for ¬(P → Q) to be true, is for the antecedent to be true and the consequent to be false. So we have the rule:

(iii)’



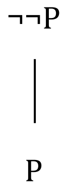
As for a negated biconditional, there are two ways in which a biconditional can be false (and hence its negation true) – that is, the two ways in which P and Q can have different truth values; so we have:

(iv)’



Then finally we have the obvious rule for double negation: the only way for a sentence of the form ¬¬P to be true, is for ¬P to be false, i.e. for P to be true.

(v)’



The idea behind the tree method is to keep on applying the above rules until we are left with simple sentences – either atomic sentences or the negations of atomic sentences. The method will, as we shall see, systematically lead us to a counterexample to any inference, ***if such a counterexample exists***.

Example 1:

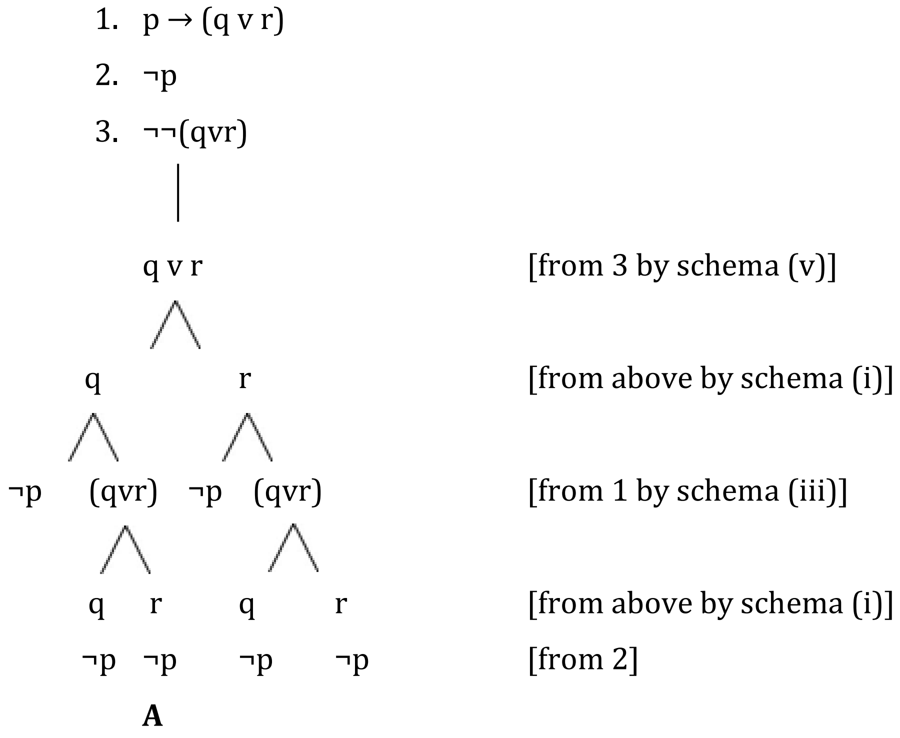
1. p → (q v r)
2. ¬p

Therefore, ¬(q v r)

As I said, a counterexample to this inference would be a set of truth value assignments to the atoms which makes the premises true and the conclusion false, that is, one that makes the premises and the *negation* of the conclusion true. The tree method will tell you whether it's possible for any set of sentences to be true together. So in this case we apply it to the following list:

1. p → (q v r)
2. ¬p
3. ¬¬(qvr)

We can then apply our tree rules to this list as follows:



Notice:

1. No further decomposition is possible since we are down to combinations of atomic sentences and negations of atomic sentences.
2. That where we have introduced new information from a so far unconsidered premise, we put that information on *all* the branches (strictly speaking we

should also put ¬p on the branches that already contain ¬p but that would clearly be redundant).

If we now look back up any of the branches in this tree from the bottom to the top, take for example the one marked **A**, we can read off an assignment of truth values that provides a counterexample to our inference, by applying the obvious rule:

> Wherever an atom appears unnegated, assign it the truth value **true**, and;
>
> Wherever an atom appears negated, assign it the truth value **false**.

So, looking along branch **A** we have ¬p, r and q. Hence the truth value assignment at issue is p:F, q:T, r:T. (*Exercise*: check that this does indeed supply a counterexample to our inference).

Now let's look at a second inference:

Example 2:

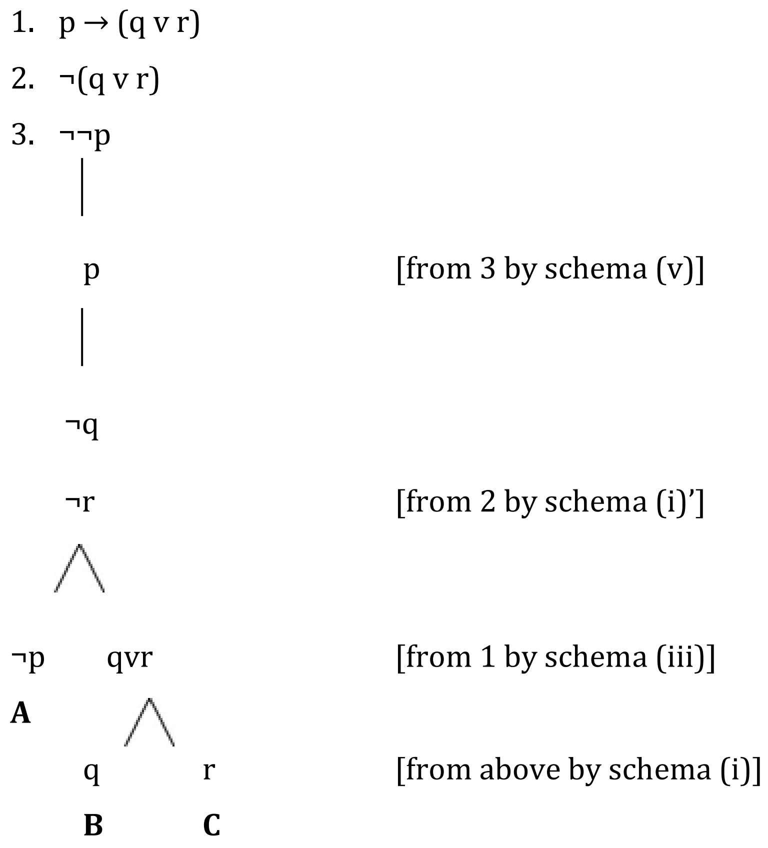
1. p → (q v r)
2. ¬(q v r)

So, ¬p

To apply our semantic tree method to it to try to find a counterexample, we list the set of sentences that would all have to be true to provide such a counterexample, i.e.

1. p → (q v r)
2. ¬(q v r)
3. ¬¬p

We can then proceed as before:



Once again we have decomposed all the sentences we are interested in, and no further decomposition is possible.

Now suppose that we try to do with this inference what we did with inference 1; that is, try to read a counterexample off any of the branches. We would get into trouble every time. Start with the branch marked A: we have ¬p on it so we should write p:F, but reading up towards the top, we also have p which would require the inconsistent assignment p:T. On branch B we have q so should assign T to q but then nearer to the top we have ¬q so should assign F to q; finally on branch C we have r and so r: T , but also ¬r so r:F!

Clearly if an atomic sentence and its negation *both* appear on any branch then we cannot read a consistent counterexample to the inference off that branch.Such a branch is called ***CLOSED*** (and we indicate the point at which it closes by drawing a double line under it). A branch that does *not* close, even though all the information has been processed, is, not surprisingly, called an ***open*** branch. We can only read counterexamples to an inference off an open branch.

Hence we have the ***fundamental result*** *that if* **all the branches of a tree foran inference close, then no branch can supply a counterexample, so there is no counterexample and so the inference is VALID.**

(I am here presupposing a result about the semantic tree method, namely that it is bound to find a counterexample if there is one. This seems intuitively obvious (and is indeed true) but it does need a proof, which I shan't pause to provide at this stage.)

A tree in which all branches close is, again unsurprisingly, called a ***closed tree***, so our fundamental result can be re-expressed as:

The inference from some set of premises to a conclusion C is valid iff the semantic tree for the set consisting of the premises and ¬C is closed.

There are some extra wrinkles about the tree method, particularly with respect to economy measures – you don't want to let your tree get unwieldy (too 'bushy'!) and the aim should be always to keep the branching to a minimum. So the general message is to use the information where the rules lead to no splitting (as in (ii) or (i)') first. But this is best learnt through practising with particular trees rather than attempting to lay down particular guidelines.

A3(F): CONSISTENCY

Although the central problem of logic has been taken to be that of characterising validity of inference, an almost equally important matter with which logic can deal is that of the ***consistency*** of a set of statements. The question often arises in mathematics, the sciences and the social sciences of whether it is consistent to make the assumption A' given that assumptions A_1... A_n have already been made. Moreover, in ordinary debate it is not unusual to charge someone with holding views that might separately be tenable but which taken together are inconsistent. Logic can supply a precise characterisation of the notions of consistency and inconsistency. (It turns out, as we shall see, that these notions have very close connections with the notions of validity and invalidity of inference – indeed in a sense they are identical notions.)

The basic idea again involves the notion of 'possible truth': a set of sentences is *consistent* if it is *possible* for them all to be true together (though they may all *as a matter of fact* be false). Restricting ourselves to truth functional logic thistranslates into the following:

***Definition:* Consistency**

A set of sentences in the language of truth functional logic is **TRUTH FUNCTIONALLY CONSISTENT** iff there is **at least one assignment of truth values to the atomic components of the sentences which makes them all true.**

Hence in order to decide whether a set of informal sentences is truth functionally consistent we translate them into the language of truth functional logic and look for an assignment of truth values to the atomic sentences that makes all the sentences in the set true.

Example 1:

The following set of sentences is truth functionally consistent:

{If the Butler is guilty then so is the Chauffeur; the Chauffeur is guilty and the Parlour Maid is not guilty; either the Butler is not guilty or the Parlour Maid is guilty.}

The set symbolises as {p →q, q & ¬r, ¬p v r}, where p, stands for the assertion that the Butler is guilty, q for the assertion that the Chauffeur is guilty and r for the assertion that the Parlour Maid is guilty. In fact, the following assignment of truth values makes these three sentences all true: p:F; q:T; r:F. (*Exercise*: check this using truth tables)

Example 2:

The following set of sentences is truth functionally inconsistent:

{If the Butler is guilty then so is the Chauffeur; either the Chauffeur is not guilty or the Parlour Maid is guilty; the Parlour Maid is not guilty, but the Butler is guilty.}

These formalise as {p →q, ¬q v r, ¬r & p}, where p, q and r are as in example 1. They form an inconsistent set because ***no assignment of truthvalues to p, q and r can make these true together***. [*Exercise*: try a fewassignments and note that they fail.]

As I indicated a few paragraphs ago, the question of consistency often arises in connection with proposed extensions of some set of assumptions: "Given that I am already assuming A_1 ... A_n would it be consistent to add the further assumption A'?". This question just translates into one of whether the whole set of assumptions {A_1 ... A_n, A'} is consistent. Let A be the original set of assumptions {A_1 ... A_n} and A' the single sentence, then generally we shall say that A' *is consistent* with A iff A U {A'} (i.e. the set of assumptions formed by adding A' to the assumptions already in A (here U stands for (set-theoretic) union) is *consistent.* (The question in its original form '... would it be consistent to *add* A'?*'* rather *presupposes* that the original set A is itself consistent, and this will be true in all cases of interest. Nonetheless this presupposition is not carried over into our formal characterisation: if A is itself inconsistent then so of course is AU{A'}, for *any* A'*;* and so A' is not consistent with A, if A is inconsistent.

(*Exercise:* show that this follows from the basic definition of consistency in terms of assignments of truth values to atomic sentences.)

Example 3:

If I already assume that 'The price of good G in economy E rises ifeither the demand for G rises or there is inflation in E' and that 'If there is tight monetary control in E then there is no inflation in E', and I know that in fact 'The price of G in E has risen', would it be CONSISTENT also to hold that 'The demand for G has not risen and there has been tight monetary control'?

Here the original set of assumptions A formalises as:

$A = \{p \leftrightarrow (q v r), s \rightarrow \neg r, p\}$

While the further assumption is $A' = \neg q \,\&\, s$

(*Exercise*: make sure you check exactly which sentences, p, q , r and s symbolise here.)

The question of whether it is consistent to add a to the set A reduces to the question of whether the set:

$A \cup \{A'\} = \{p \leftrightarrow (q v r), s \rightarrow \neg r, p, \neg q \,\&\, s\}$

is consistent.

(*Exercise*: the answer is that it is *not.* Again try a few truth value assignments – you won’t succeed in making all the sentences true.)

A3(G): DEMONSTRATING CONSISTENCY AND INCONSISTENCY

So now we know what it means for a set of sentences to be consistent and what it means for such a set to be inconsistent. How can we actually go about **deciding** which a particular set of sentences is? One method – essentially theequivalent of the truth table method of deciding validity or invalidity of inference – would be trial and error: try out all the various possible truth value assignments (write out the whole truth table) and see if any works. If at least one works (if there is at least one line in the joint truth table in which all sentences take the truth value true) then the set is consistent, if none works (no such line in the truth table) then the set is inconsistent.

A neat and systematic method, however, is to use ***semantic trees***. In this case we want to know if a given set of sentences might possibly all be true together. So we ***first***, list the various sentences in the set whose consistency is in question {N.B. we do **not** negate anything, as we did in deciding in/validity of inferences, where we negated the conclusion} and ***secondly***, construct a tree.

The rule in deciding consistency is then:

(a) IF ANY BRANCH REMAINS **OPEN** ***THE SET IS CONSISTENT AND AN ASSIGNMENT DEMONSTRATING ITS CONSISTENCY CAN BE 'READ OFF' THE OPEN BRANCH; while***

(b) IF THE TREE **CLOSES** ***(NO OPEN BRANCHES) THEN THE SET IS*** **IN*****CONSISTENT.***

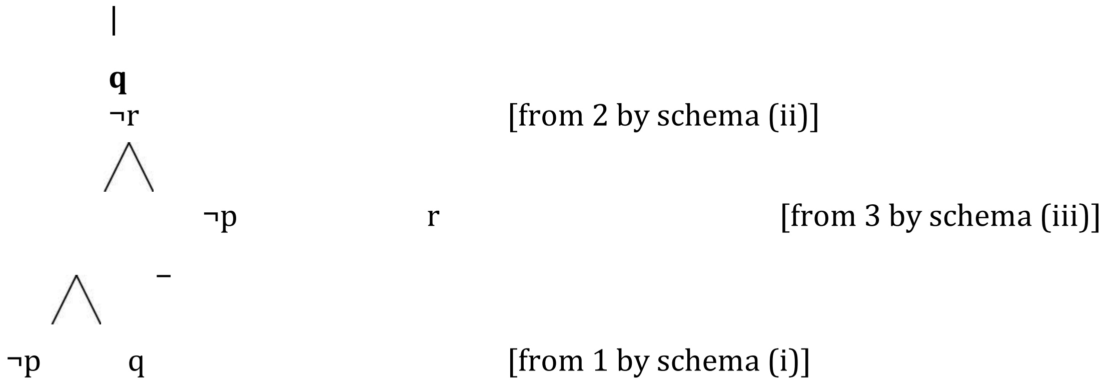
Examples 1 and 2 as above:

Example1:

1. p → q

2. q & ¬r

3. ¬p v r



q

¬r [from 2 by schema (ii)]

¬p r [from 3 by schema (iii)]

¬p q [from 1 by schema (i)]

All the information has now been exhausted and branches remain open. Hence the set of sentences is consistent and by following either of the two open branches we can read off an assignment that shows this (in this particular case we get the same assignment – viz. p:F, q:T, r:F – from both of the open branches, indicating that this is the only assignment which demonstrates consistency). (*Exercise:* Check this by substituting these truth values in the sentences 1, 2 and 3, and seeing that they all turn out true.)

Example 2:

1. p → q

2. ¬q v r

3.¬r & p

|

¬r

p [from 3 by schema (ii)]

∧

¬p q [from 1 by schema (iii)]

= ∧

¬q r [from 2 by schema (i)]

= =

Here all the branches ***close*** and hence the set of sentences is **inconsistent**.

One important wrinkle: *You* will sometimes find in using the tree methodeither to establish in/consistency or in/validity, that the tree you construct has open branches on which one of the atomic sentences fails to appear altogether, that is, it appears *neither* negated *nor un*negated. Here is a simple example:

Is {p → (q v r), ¬q} a consistent set of sentences?

1. p → (q v r)
2. ¬q

∧

¬p (q v r) [from 1 by schema (iii)]

¬q ∧ [from 2]

A q r [from above by schema (i)]

¬q ¬q [from 2]

= **B**

Here we have, even when all information has been taken into account, two open branches (which I have labelled **A** and **B** respectively). So the set of sentences is certainly consistent. Which truth value assignments show this?

On branch A we have ¬p and ¬q, so p and q must both be false. But how about r? It does not appear at all. **The answer is that if it doesn't appear then it doesn't matter – either truth value for r will do, so long as p and q have the truth values specified.** That is, branch **A** in fact supplies*two*assignmentswhich demonstrate consistency: p:F q:F r:T *and* p:F q:F r:F. So long as p and q are both false the sentences in the set are going to be true *whatever* truth value r has. (*Exercise*: check directly by substituting into the set of sentences that both truth value assignments make all (both) the sentences true.) Note that it is of course important that this be a really open branch, that is, that all the information has been put on in the (failed) attempt to close the branch. The above result does not, of course, hold for a branch that remains open in an *incomplete* tree.

Similarly for branch B we have r:T and q:F. But p does not appear. Again, this means that p can be *either T or F* (given that r:T and q:F). So again we have two assignments which demonstrate consistency: p:T q:F r:T and p:F q:F r:T.

This result is quite general and applies equally well when using semantic trees to decide validity or invalidity of inference. Strictly speaking, we need a proof of the result, but again we shall not pause to supply it here.

A3(h): The Connection Between (in)Consistency and (in)Validity

By reflecting on the two different uses we have made of semantic trees, it is easy to see the connection between the notion of consistency and that of validity of inference. From this new point of view (*i.e.* from the point of view of deciding whether a given set of sentences is consistent) what we were doing earlier was deciding whether the set of sentences formed **by the premises together with the negation of the conclusion is consistent**. (Remember thatin applying semantic trees in deciding validity of inference we constructed a tree for the premises and the *negation* of the conclusion.) If that set is indeed consistent we concluded that the inference is invalid, while if the set is inconsistent we concluded that the inference is valid.

Hence we have the following **important connection** between the two notions: If P is a set of sentences and C is a single sentence (all in the language of truth functional logic) then the set of sentences P U {C} (that is, simply the set of sentences formed by adding C to the original set P) is **inconsistent** *if, and onlyif* the inference from P as a set of premises to ¬C as conclusion is **valid**.

That is, an inference is valid iff it would be inconsistent to assume that the conclusion is false while assuming that the premises are true.

(Important Exercise: We arrived at this connection by thinking about the two uses of the semantic trees method; but show that it must indeed hold in virtue of the basic definitions of validity and consistency in terms of truth value assignments to the atoms.)

A3(i): Independence

In scientific and other intellectual disciplines, the question often arises of whether or not some particular assumption is ***independent*** of another set of assumptions. This is also a question that comes up in more ordinary, argumentative, situations such as in politics or the law. Someone might, for example, be criticised for holding a certain view and defend herself by saying 'That is not my view and it is quite independent of the position I *did* assert'. One of the most celebrated questions in the history of mathematics was that of whether Euclid's 5th (or Parallel) Axiom is or is not independent of the rest of his axioms. The 5th axiom states that for any line and any point outside the line there is one and only one line parallel to the given line through the given point. The question was whether this is an *independent assumption* (independent that is of the other axioms that Euclid laid down) or whether, on the contrary, the truth or falsity of the parallel axiom had already in effect been decided by the other axioms – that is, the assumptions about points and lines that Euclid had already made. (The general view was that the parallel axiom was so obviously correct that it must follow from the other axioms and therefore *not* be independent.) The question troubled mathematicians for over 2,000 years.

In ordinary cases at least, the question is usually that of whether or not the truth of the assumption in question is in fact already implied by other assumptions, but we would also surely say that if those other assumptions entailed the *falsity* of the statement, then that statement was *not* independent of that set of statements. Hence we get the following precise characterisation:

A single sentence a is **independent** of a set of sentences A iff ***neither***a***nor*** $\neg a$ can be validly inferred from the set A.

In terms of truth-functional logic, then:

***Definition:* Independence**

A single sentence$a$in the language of truth functional logic is(truth-functionally) **independent** of a set A of such sentences if ***neither*** the inference from A as premises to

a as conclusion ***nor*** the inference from A as premises to $\neg a$ as conclusion is truth functionally valid.

A3(I): DEMONSTRATING INDEPENDENCE

We saw earlier that an inference is invalid iff there is a counterexample to it. So, for **neither** the inference from *A* to *a***nor** the inference from *A* to ¬*a* to be valid there have to be TWO counter-examples (that is, two truth value assignments to the atomic propositions):

(a) one in which all of the sentences in A are true and a is true, and

(b) one in which all of the sentences in A are true and a is false.

Since case (a) is one in which all the sentences in *A* are true and so is a, while case (b) is one in which all sentences in A are true but ¬*a* is true, i.e. a is false, we can see that:

The single sentence* a *is independent of a set of sentences* A *iff* A *U* {a} *and* A *U {¬*a*}* *are both consistent sets of sentences.

Since we know how to decide the consistency of a set of sentences (by producing the relevant semantic tree) we see that the independence of *a* from *A* can be decided as follows:

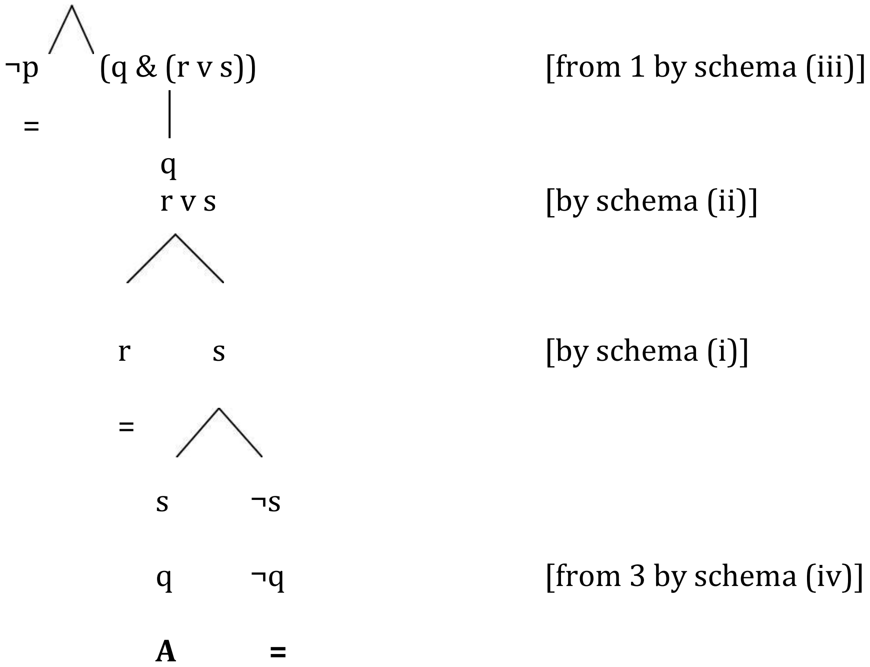
1. Construct the semantic tree for the set of sentences *A* U {*a*}; if it ***closes*** then *a* is **NOT** independent of *A*.
2. If that tree remains open, then construct another tree for *A* U {¬*a*}. If that second tree closes then again *a* is **NOT** independent of *A*, but if it too remains open then *a***IS*****independent****of A.*

Example 1:

Is p independent of {(p →(q & (r v s))), ¬r, (s ↔ q)}

To decide whether p is consistent with this set take the sentences 1-4 and proceed as below:

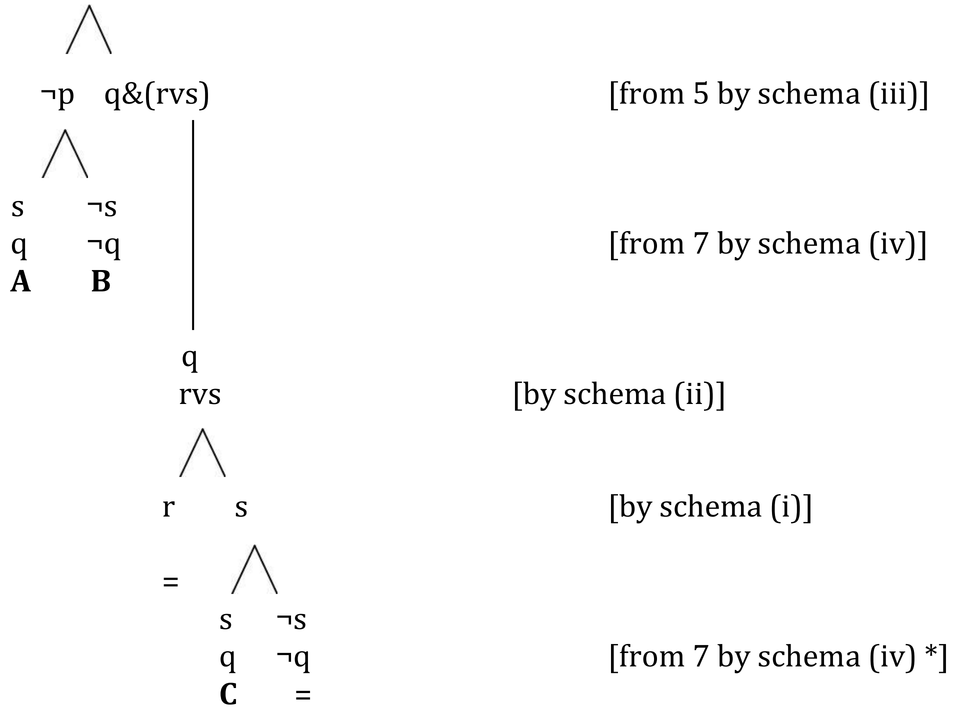
1. (p →(q & (r v s)))
2. ¬r
3. (s ↔ q)
4. p



One branch remains open (the branch marked **A**) and so the set is consistent as shown by the truth value assignment that we can read off **A** (p:T, q:T, r:F, s:T).

So next to decide whether ¬p is consistent with our set of sentences take the different set of sentences 5 to 8 below and proceed as follows:

5. (p →(q & (r v s)))
6. ¬r
7. (s ↔ q)
8. ¬p



(* Remember: all information on all open branches. This information has not been on the right hand branch created at the beginning and hence must be put on here – we must give the tree the best chance to close.)

We have three open branches (marked **A, B** and **C**) which between them supply two different assignments showing the consistency of sentences 5-8 (p:F, q:T, r:F, s:T and p:F,q:F, r:F, s:F).

These two trees together hence establish that p **is** independent of {(p →(q & (r v s))), ¬r, (s ↔ q)}

Example 2:

Is p independent of {p ↔q, ¬q}?

First decide whether p is consistent with {p ↔q, ¬q} *i.e.* is the set {p ↔q, ¬q, p} consistent?

1. p ↔q
2. ¬q
3. p

```
     /\
p         ¬p
q         ¬q                              [from 1 by schema (iv)]
=         =
```

Hence this tree closes, and so p is **not** independent of {p ↔q, ¬q} – in fact we could validly infer from {p ↔q, ¬q} as premises that p is false.

A3(k): Using Semantic Trees to decide the status of Sentences

We earlier noted that single sentences in truth-functional logic fall into one of three categories: tautology (all lines in its truth table are T), contradiction (all lines in its truth table are F) and contingent sentence (at least one T in its truth table and at least one F). As this characterisation indicates, the straightforward way to decide which of the three categories a particular sentence falls is by using truth tables. But for sentences of any complexity this can be a long process (remember that if there are *n* atoms in the sentence, then there are 2^n lines in its truth table). As in the case of validity of inference, the method of semantic trees can be applied and makes the decision simpler than writing out the whole truth table.

Here is how the method works (the exercises will give you practice in applying it):

1. ***Contradictions:*** The semantic tree method searches systematically for an assignment of truth values to atomic components that makes all the sentences in a set of sentences true. It fails to find such an assignment, and therefore this is no such assignment, just in case the tree closes (i.e**. all** branches close).

If the set consists of just a *single*sentence, then if there is no assignment of truth values to its atomic components that makes it true then that sentence is a *contradiction* (and vice versa). Hence:

A single sentence S is a contradiction iff and its semantic tree closes (again ***all*** *branches close).*

2. ***Tautologies:*** Fact: a sentence is a tautology iff its negation is a contradiction. Hence, applying the reasoning in (1) yields:

A single sentence S is a tautology iff the semantic tree for ¬S closes.

3. ***Contingent sentences***: Fact: S is contingent iff it is neither a contradiction (at least one assignment of truth values makes it true) nor a tautology (at least one assignment of truth values makes it false). So the reasoning in (1) and (2) yields:

A single sentenceS is contingent iff neither the tree for S nor the tree for ¬S closes (i.e both trees have at least one open branch).

A3(L): TRUTH-FUNCTIONAL EQUIVALENCE AND INTERDEFINABILITY

Truth functionality:

We now know how to do everything in truth functional logic that we need to: decide on validity/invalidity of an inference expressed in that logic; decide whether a particular compound sentence of the logic is a tautology, a contradiction or neither; decide whether a set of truth functional sentences is consistent or inconsistent; and finally decide whether a particular sentence *a* expressed in truth functional logic is or is not independent of a given set of such sentences *A*.

In the next couple of sections, we will investigate the system that we have produced – 'from the outside', so to speak. That is, we will do a little meta-logic: rather than using the logic to decide validity or whatever, we will produce some results about that logic.

First, why exactly is the branch of logic we are presently studying called *truth-functional* logic? The answer is because any compound sentence, no matter howcomplicated, has the following property: its truth-value (its truth value in the real world or in any 'possible world') is a function of the truth values of its atomic components.

The term “**function**” is used here in the mathematician's sense. A function takes objects one by one from some set of objects and "associates each with" or "maps each onto" another in some definite way. For example, in elementary arithmetic there is the doubling function f(x) = 2x which takes any number and maps it onto its double (2 onto 4, 3 onto 6, etc). A function is, if you like, a rule of *association* – one that always yields an outcome when applied to a particular input.

In the case of logic, any truth functional compound of any degree of complexity is characterised by a rule associating one of the two truth values (the overall truth value of the compound) with any given combination of truth values to the atoms. So, e.g., the truth functional compound ‘(p & q) v r’ defines the following *truth function f:*

f(T, T, T) = T

f(T, T, F) = T

f(T, F, T) = T

f(T, F, F) = F

f(F, T, T) = T

f(F, T, F) = F

f(F, F, T) = T

f(F, F, F) = F

The objects which the function applies to (in the jargon: 'takes as arguments') are ordered triples of truth values (‘ordered’ meaning that (T, F, F), for example, is a different triple from (F, F, T)). The function associates each ordered triple with one (and of course only one) truth value: it *maps* every possible triple of truth values *onto* a single truth value.

(*Exercise:* Make sure that you understand that the above truth function is the one associated with ‘(p & q) v r’; and also that you understand that it does no more than give another way of expressing the information contained in the relevant truth-table.)

More generally we can say that any truth-functional compound involving any number *n* of atoms characterises some truth-function *f* which maps in some definite way each of the different *n*-tuples of truth values onto a single truth value.

Truth-functional Equivalence:

The sentence ‘Pavarotti was a great tenor (p) and Bartoli is a great mezzo-soprano (q)’ clearly carries the same information, or "says the same thing", as ‘Bartoli is a great mezzo-soprano(q) and Pavarotti was a great tenor (p)’. It is also true (though a little less obvious) that either sentence carries the same information as ‘It's not the case either that Pavarotti was not a great tenor or that Bartoli is not a great mezzo-soprano’.

Although each of these three sentences is linguistically different from the others, so that they are not the *same* sentence, they are nonetheless **EQUIVALENT** sentences: they carry the sameinformation. (This is sometimes paraphrased as ‘the two sentences express the same proposition'.) If we formalise each of them – using the same atoms throughout, we have:

(p & q)

(q & p)

and

¬(¬p v ¬q)

Writing out a truth table for each, we have:

p	**q**	**p & q**	**q & p**	**¬(¬p v ¬q)**
T	T	**T**	**T**	**T** f f f
T	F	**F**	**F**	**F** f t t
F	T	**F**	**F**	**F** t t f
F	F	**F**	**F**	**F** t t t

The final truth value is the same in all three cases. The three sentences *have the same truth table.* (Notice that the first equivalence that between p & q and q & p, though simple andobvious, is *not* trivial: p → q and q → p, for example, have *different* truth tables.) This motivates the following *definition:*

Definition:* *Truth-Functional Equivalence:

Two truth functional compounds P and Q are **truth functionally equivalent** (which is written: P≡ Q) if and only if they have the same truth table, that is, for *any* given assignment of truth values to the atomic propositions, P and Q have the same overall truth value.

This means we can write

p & q ≡ q & p ≡ ¬(¬p v ¬q).

(*Important Exercise:* Not surprisingly, two compounds P and Q are equivalent if they are interderivable – that is, if either can be validly inferred from the other as premise.

Using the basic definitions of valid inference (given earlier) and of equivalence (given just now), show carefully that this is true.

Two compounds P and Q are equivalent iff the single sentence ‘P ↔ Q' is a tautology. Again use the definitions to show carefully that this is true. Show also that it follows from part (a). Would it be enough to say that P ≡ Q iff the single sentence ‘P ↔ Q' *is true* (rather than tautologically true)?)

We could also have expressed truth functional equivalence in terms of truth *functions* – two compounds being equivalent iff they determine the same truth function. So, for example, ‘p & q’ and ‘¬(¬p v ¬q)’ determine the same truth function *f* (*viz*. f(T,T) = T, f(T,F) = f(F,T) = f(F,F) = F).

Truth-Functional Interdefinability:

If two sentences are equivalent, they both 'say the same thing' or 'carry the same content'. For any sentence of the form P & Q (remember we use capital letters to indicate that the sentences may themselves be compound – so P might be (p → q) and Q might be (¬r & (s ↔ t))) – there is an *equivalent* sentence in which the connective '&' is eliminated in favour of the connectives '¬' and 'v', viz. ¬(¬P v ¬Q). (So, e.g., ¬(¬(p → q) v ¬(¬r & (s → t)) is equivalent to (p → q) & (¬r & (s → t)).)

(*Exercise:* make sure that you can work this through.)

Similarly, for any sentence of the form P → Q, there is an equivalent sentence in which the → is eliminated in favour of ¬ and v, namely ¬P v Q.

(*Exercise*: Explain carefully why this is true and notice that this equivalence justifies the semantic tree rule (iii) given earlier for P → Q.)

Finally, for any sentence of the form P ↔ Q there is also an equivalent sentence in which the '↔' is eliminated in favour of '¬' and 'v'. First, P↔ Q is equivalent to (P → Q) & (Q→ P). So, eliminating → as indicated earlier, we have P ↔ Q ≡ (¬P v Q) & (¬Q v P). And then eliminating & in favour of ¬ and v, we have P ↔ Q ≡ ¬(¬(¬P v Q) v ¬(¬Q v P)) – bracketing again being crucial.

This all means that by applying these equivalences sequentially we can eventually eliminate all occurrences of all the other connectives (or at least all the connectives that we know about) in favour just the two connectives: ¬ and v. Or in other words, for any sentence whatsoever in the language of truth functional logic that we know about so far, there is an equivalent sentence using just the connectives ¬ and v.

Example 1:

1. (p & q) → (r → s) ≡¬(¬p v ¬q)→ (¬r v s)
2. ¬(¬p v ¬q) → (¬r v s) ≡ ¬¬(¬p v ¬q) v (¬r v s)
3. ¬¬(¬p v ¬q) v (¬r v s) ≡ (¬p v ¬q) v (¬r v s)

So, (p & q) → (r → s) ≡ (¬p v ¬q) v (¬r v s)

Example 2:

1. p ↔ (q & r) ≡ p ↔ ¬(¬q v ¬r)
2. p ↔ ¬(¬q v ¬r) ≡ ¬(¬(¬p v ¬(¬q v ¬r)) v ¬(¬¬(¬q v ¬r) v p)

Clearly, elimination of connectives in this way leads to greatly increased complexity. But if we were interested in having *as few primitive notions* – that is, as little basic vocabulary – as possible (as we might be, for example, if we wanted to program a computer to do truth-functional logic for us), then the above result shows that, instead of taking all our connectives {¬, v, &, →, ↔} we could get by with just {¬, v} as basic and then define the other connectives in terms of these two.

The same story as I have just told for the set of connectives {¬, v} could also be told for the set {¬, &} and indeed for the set {¬, →}. That is, we could equally well find equivalents for any sentence using connectives from among {¬, v, &, →, ↔} which only used {¬, &} or which only used {¬, →}.

Exercise:

(a) Produce equivalents for P v Q, P →Q, P ↔Q, using only the connectives ¬ and &.

(b) Produce equivalents for P v Q, P & Q, P ↔ Q using only the connectives ¬ and →.

Single connectives: 'alternative denial ' and 'joint denial':

A technical question that arises naturally at this point is: can this process be taken one stage further? Is there a single connective in terms of which all our other connectives can be defined?

The answer to this question is 'yes'. Although there is no such connective available in ordinary English (at least not a direct connective), we can in fact produce two separate single connectives, characterised by their truth table, either of which will do the job. These are **'joint denial'** (symbolised '↓') and **'alternative denial'** (symbolised '|'). Theseconnectives are defined (as truth functional connectives must be) by their truth tables:

P	**Q**	**P ↓ Q**
T	T	**F**
T	F	**F**
F	T	**F**
F	F	**T**

(So 'joint denial' is the correct name: P ↓Q says in effect that P and Q are both false, and is therefore itself true exactly when they *are* both false – that is, only in the last line of the truth table. For this reason, joint denial is sometimes also known as 'nor', as in 'Neither P nor Q'.)

P	**Q**	**P | Q**
T	T	**F**
T	F	**T**
F	T	**T**
F	F	**T**

(So again "alternative denial" is the right name: P|Q says in effect that at least one of P and Q is false and hence is itself false only when they are both true.)

Since we know from our recent considerations that the connectives we started with, *viz.* {¬,v, &, →, ↔} can be defined in terms of {¬, v}, all we need to do to show that all thoseconnectives can be defined in terms of ↓ is to show that there are equivalents for any sentence of the form ¬P or any sentence of the form P v Q that contain only the connective ↓. In fact:

1. ¬P ≡ P↓P   and,
2. P v Q ≡ (P↓Q) ↓ (P↓Q)

Important Exercises:

(a) Show using truth tables that the equivalences 1 and 2 hold.

(b) Since we also showed that all the connectives can be defined in terms of just {¬, &} we could also have proved this result about ↓ by showing that there are equivalents to both ¬P and P&Q that involve only the connective ↓. We already know the one for ¬P, find the equivalence for P & Q.

As for 'alternative denial', | , (also known in the literature as 'the Scheffer stroke' after the logician who discovered it), the following equivalences show the same result for it:

1. ¬P ≡ P|P   and,
2. P v Q ≡ (P|P) | (Q|Q)

Important Exercises:

(a) Show using truth tables that the equivalences 1 and 2 hold.

(b) Find an equivalence using only '|' for (P & Q).

Adequacy of a set of connectives for truth functional logic:

***<u>Definition:</u>* Adequacy:**

A set of connectives S is said to be ***ADEQUATE*** *for truth functional logic iff any truth function can be characterised by a truth-functional compound where the only allowable connectives are those in S.*

It is important to realise that no set of connectives has been proved adequate for truth functional logic above. All that was established there is a series of *conditional* results:

that **IF** {¬,v, &, →, ↔} is an adequate set of connectives **THEN** so, are {¬, v} and the set {|} consisting of the single connective, etc.

That is, all we know so far is that *if* for any truth function there is a compound which characterises that function and which uses only connectives from {¬, v, &, →, ↔} *then* for any truth function there is a compound which characterises it which uses only connectives from {¬, v} and indeed only from {|}.

But isn't the antecedent in this last conditional sentence just obviously true? After all {¬, v, &, →, ↔} are the *only* connectives we know about, so isn't that set adequate by definition? Well, these certainly *are* the only connectives we talked about at the beginning of this course. But truth functional logic is about ANY WAY of compounding atomic sentences that is truth functional – that is, in which the truth value of the compound depends systematically on the truth values of the atoms.

In the exercises you will come across ordinary English connectives which are (or which can be construed as) *truthfunctional* but are not in the set (not, and, or, if ... then, if and only if). Three examples are 'but', 'unless' and 'only if'. It happens thatin those three cases there *are* straightforward equivalents for 'P but Q', 'P unless Q' and 'P only if Q' which use connectives from the usuallist. But how do we know that there aren't other connectives which *are* truth functional but which we haven't taken into account and which have no such equivalents?

Moreover, even if there are no such further connectives in ordinary English, this surely would just be a peculiarity of that language. For example, it just happens that all the ordinary connectives (aside from negation) are *binary* (connecting two propositions - each of which may itself contain connectives). But one can easily envisage a language in which onecan *say straight off* that, for example, at least two of three propositions p, q and r are false. Thatis in which there is a three-place connective - say 'plink', where 'Plink John, Jane and Joyce are invited to the party' means in ordinary boring English that atmost only one of the three is invited. Again it turns outthat we could easily produce an equivalent using the connectives we know about. The simplest is:

Plink (p, q, r) ≡ (¬p & ¬q) v (¬q & ¬r) v (¬p & ¬r)

But what guarantee do we have that *any* such imaginary connective (only imaginary in English, of course, perhaps real in other natural languages) has equivalents using only connectives from our list?

The answer is that so far we have no such guarantee. But we can in fact produce one via an important theorem about truth functional logic called the ***disjunctive normal form theorem****.* (We are now, remember, considering results *about* logic, rather than for example using it to decide in/validity of inferencesin ordinary language, so we are – briefly – in the domain of *meta*-logic.)

A3(M): THE DISJUNCTIVE NORMAL FORM THEOREM

The easiest way to understand this result is by starting with a simple particular case of a truth functional compound, let's (arbitrarily) say ¬(p → (q & r)). Next, construct the truth table for this:

p	q	r	¬(p → (q & r))
T	T	T	**F**
T	T	F	**T**
T	F	T	**T**
T	F	F	**T**
F	T	T	**F**
F	T	F	**F**
F	F	T	**F**
F	F	F	**F**

We can use this table, or indeed any such table, to construct an equivalent sentence involving ***just the connectives ¬, v and &*** in a completely mechanical way as follows:

1. Concentrate exclusively on those rows in which the truth functional compound takes the value T. (Note that this means that we are – temporarily – excluding truth functional contradictions which by definition have F in all lines of their truth table.)
2. For each such line form the ***conjunction*** ±p & ±q & ±r, where the '±' sign indicates that that atomic sentence is to be taken negated or un-negated depending on whether that atom takes F or T in that particular row: so, for example, the conjunction corresponding to line2 in the above truth table (where indeed the compound takes the overall value T) is p & q & ¬r, while the conjunction corresponding to line 4 is p & ¬q & ¬r.
3. Finally form the ***disjunction*** of all the conjunctions formed at step 2.

Following this rule for the above truthfunctional compound we get the following three-fold disjunction (lines 2, 3 and 4 are the lines that take the value T):

$(p \,\&\, q \,\&\, \neg r) \vee (p \& \neg q \,\& r) \vee (p \& \neg q \,\&\, \neg r)$.

This is called the ***disjunctive normal form*** of our original sentence ¬(p → (q & r)). If you think about it, this disjunction is *bound* by construction to be equivalent to the original. This is because it is bound to have the same truth table: each of its constituent conjuncts (e.g. (p & q & ¬r)) is true ***precisely*** in the line from which it was constructed (and *only* in that line); hence this new sentence takes the value T in just the same three lines as the original sentence (since T v F v F, for instance, is T), and it takes F in all other lines (since the disjunction is F v F v F (=F) in all those other lines).

The **disjunctive normal form THEOREM** basically consists in noting the fact that the above construction can be applied quite mechanically to any truth function (with the one exception of contradictions which we can deal with separately as we shall see). The theorem can be proved as follows:

Consider *any* such truth function with an arbitrary number, *n*, of argument places. ***And notice that weare talking about* any *truth function whatsoever whether or not it corresponds to some compound produced by using the connectives available to us in ordinary English.***By definition such a truth function corresponds to a truth table with 2^n lines. So consider any such truth table:

P_1	P_2	…	p_n	$F(\tau(p_1) \dots \tau(p_n))$
T	T		T	τ_1
…	…		…	
…	…		…	
…	…		…	
F	F		F	$\tau_2{}^n$

where $\tau(p_i)$ is the truth value of the i^{th} atom p_i and τ_i = T or F and is the overall truth value for the truth function *f* in the i^{th} row. Now take any row *i* in which τ_i = T and define for *j* = 1, …, n

$p_j{}' = p_j$ if $\tau(p_j)$ is T

$= \neg p_j$ if $\tau(p_j)$ is F

Then form the conjunction p_1'& p_2'& … &p_n'

Repeat this for all rows that have overall truth value T. Finally, form the *disjunction* of all such conjunctions.

The resulting disjunction of conjunctions is called the **DISJUNCTIVE NORMAL FORM** corresponding to the truth function *f*. It is bound to determine the given truthfunction *f*, because:

(1) For any row in which *f* takes over truth value T, so will its dnf. This is because the conjunction constructed from that row will take truth value T; all other conjunctions take F, but a disjunction with one true disjunct is true; and

(2) For any row in which *f* takes overall truth value F, so will its dnf. This is because for these rows *all* the corresponding conjunctions will be F and a disjunction all of whose disjuncts are F is itself F.

So let’s state the theorem formally:

Every truth functional sentence, so long as it is not a contradiction, has an equivalent in disjunctive normal form.

As for contradictions, consider any contradiction involving atomic components p_1, … p_n it will of course be equivalent to (i.e. have the same truth table as) any other contradiction involving the same atoms. But (p_1 & ¬p_1) v (p_2 & ¬p_2) v …. v (p_n & ¬p_n) is a contradiction and it involves only the connectives ¬, v, &. So finally for *every* truth functional sentence including contradictions there is an equivalent that just uses connectives from the set {¬, v, &}.

It follows that we finally do have an assurance that the set of connectives that we introduced, namely {¬, v, &, →, ↔}, is indeed an adequate set of connectives for truthfunctional logic. This is because the disjunctive normal form theorem, together with the above remark about contradictions, shows that {¬, v, &} is an adequate set. (Henceour earlier interdefinability results kick in to show that all of {¬, v}, {¬, &} {|} and {↓} are – unconditionally – adequate sets of connectives.

(*Exercise:* explain *carefully* why.)

B: INFORMAL REASONING

At the beginning of the course, I said that we would be studying the way that we reason both in systematic disciplines like science, social science or mathematics and in ordinary argument and reasoning. Yet the examples we have studied in connection with truth functional logic might seem very far away from ordinary reasoning. Of course this is in part because in logic we are formalising processes that we normally just use intuitively. But there is another reason – as we will see in this section, we often do not spell out arguments in full gory detail. This section is aimed to convince you that truth-functional logic does have important things to say about real arguments (and not just artificial examples about whether Peter, Quentin and Rita *etc.* do or don't go to the party!).

The argument from evil, 1:

In philosophy, ordinary language arguments do sometimes come close to 'straight' logical inferences. So someone might argue – admittedly a bit sketchily – 'If an all-loving God existed then there would be no evil in the world. But there is evil in the world. So no all-loving God exists.' This pretty immediately formalises as:

1. $p \rightarrow \neg q$
2. q

Therefore, $\neg p$

This is of course valid (*Exercise*: Check). The question then becomes whether or not it is sound. It seems difficult to challenge the second premise; but religious believers have of course very much questioned the first. One line is that this premise is false because much of the evil in the world results from human actions, which (allegedly) involve human free will: an all-loving God could have given humans free will even while foreseeing that evil might result because the overall benefits of our (allegedly) having free will outweigh the resulting evil. This does not of course tackle 'natural evils' (such as tsunamis or earthquakes). But here the line from those who challenge premise 1 is often that such evils allow humans to exhibit some of their finest characteristics – bravery, generosity, fortitude, etc. So, contrary to initial appearances, an all loving God might have allowed them. (Of course these responses have been challenged in turn.)

The argument from evil, 2:

A more systematic version of the same argument might go as follows. ‘Suppose there were a God as characterised within the Judeo-Christian tradition: omnipotent (all-powerful), omniscient (all-knowing) and omnibenevolent (all-loving or all-merciful). If God were all-knowing then he would foresee any evil before it occurred. If God were all-powerful then he could prevent that evil occurring, if he wished to. If he were all-loving then he would so wish. But if he foresaw the evil, wished to prevent it, and could prevent it, then there would be no evil; but there is; so, there is no such God.’

This is of course an elaboration of argument 1. It is naturally formalised as a *reductio*: that is, we suppose that there is a God of the alleged type, derive a contradiction and infer that our supposition is untenable: there is no such God. So it breaks down into a sequence of two arguments.

Let p be ‘God is all powerful’ q: ‘God is all knowing’ r: ‘God is all loving’, s: ‘God would foresee any evil’, t: ‘God would prevent any evil’, u: ‘God wishes to prevent evil’, v: ‘There is evil’:

1. $p \,\&\, q \,\&\, r$
2. $q \rightarrow s$
3. $p \rightarrow (u \rightarrow t)$
4. $r \rightarrow u$
5. $(s \,\&\, u \,\&\, t) \rightarrow \neg v$
6. v

Therefore $(p \,\&\, q \,\&\, r) \rightarrow (v \,\&\, \neg v)$

1. $(p \,\&\, q \,\&\, r) \rightarrow (v \,\&\, \neg v)$

Therefore, $\neg(p \,\&\, q \,\&\, r)$

Exercise: Check that both inferences are valid.

Note that we could simply infer ¬(p& q & r) as line 7 in the first inference. (*Exercise:* Check). So splitting it into two arguments is unnecessary: but doing so reflects the *reduction ad absurdum* character of the original and so rhetorically gives it perhaps a bit more force. (It is important to note that many 'real life' arguments break down naturally into a sequence of inferences rather than just a straight one step inference from the initial premise(s) to some conclusion. Indeed, this is true in more formal argumentation too: a proof in mathematics is (almost invariably) a *sequence* of inferences, each one of which is valid – ending in the proof's last line with the conclusion – i.e. the theorem to be proved.)

So this second argument is valid. Since it is an elaboration on argument 1, it is no surprise that those who would argue that it is not sound, have responded in similar ways to those recorded concerning argument 1. (Though whether human free will is even compatible with the assumption of an all-powerful and all-knowing god is controversial.)

Hume's argument about miracles:

Another famous argument in philosophy is Hume's argument to the effect that no rational person could ever believe that a "genuine miracle" had occurred. (Here a "genuine miracle" is an exception to the laws of nature, as opposed just to something that seemed in advance extremely unlikely.)

The main premise of the argument is that the probability that the evidence for the alleged "miracle" is faulty in some way (no matter how much apparent evidence there may be from however many sources) is always going to be higher than the probability of there actually being an exception to the laws of nature.

Hume writes, for example:

> *"When anyone tells me, that he saw a dead man restored to life, I immediately consider with myself, whether it be more probable, that this person should either deceive or be deceived, or that the fact, which he relates, should have really happened. I weigh the one "miracle" against the other; and according to the superiority, which I discover, I pronounce my decision, and always reject the greater "miracle" If the falsehood of*

> *the testimony would be more miraculous, than the event which he relates; then, and not till then, can he pretend to command my belief or opinion."*

He goes on to suggest that the probability of the testimony being false is indeed always higher than the probability that the allegedly miraculous "event" really occurred.

So, the argument might be formalised as follows:

IF there is evidence for some alleged "miracle", THEN EITHER that evidence is true AND the "miracle" did occur OR the evidence is false AND the "miracle" did not occur.

IF the probability that the evidence is false is higher than the probability that the "miracle" occurred, THEN a rational person believes that the evidence is false AND that the "miracle" did not occur.

Moreover, the probability that the evidence is false *is* (always) higher than the probability that the "miracle" occurred. Therefore, the rational person believes the evidence is false and that the "miracle" did not occur.

1. $p \rightarrow ((q \,\&\, r) \vee (\neg q \,\&\, \neg r))$
2. $s \rightarrow (t \,\&\, v)$
3. s

So, $t \,\&v$

Where:

p: There is evidence for some alleged miracle.

q: That evidence is true.

r: That miracle really did occur.

s: The probability that the evidence is false is higher than the probability that the miracle occurred.

t: The rational person believes the evidence is false.

v: The rational person believes that the miracle did not occur.

This is a valid argument – though you may notice that the validity results simply from 2 and 3. Premise 1 in fact is a near tautology just reminding you of the possibilities – it nonetheless, although strictly redundant, seems to add to the cogency of the argument. Again then attention turns to the question of whether or not the argument is sound: not surprisingly, since it is philosophy, opinions differ.

Informal arguments and hidden premises:

It is perhaps not surprising that arguments like the above ones from philosophy are 'fairly logical' – although presented in ordinary language, it seems pretty obvious how to formalise them and once formalised they turn out to be valid as they stand. But they do not represent the norm so far as informal arguments go. Consider the following examples:

1. 'How can anyone claim that the Tories' economic policies are working? Economic growth is near zero and inflation is on the increase.'
2. Old advert on London Underground for a computer software company: 'If your machine doesn't run BOS software, it's a fridge'. (The ad actually said '...it's *probably* a fridge', but that would complicate things unnecessarily for present purposes.)

They are both clearly intended to be arguments, but they look nothing like the sort of arguments you have practised on in studying truth functional logic. For one thing, in both of these arguments, the conclusion is not made explicit. However, you are clearly meant to infer something in each case: that the Tories'economic policies are *not* working in example 1 and that your machine *will* run BOS software in the case of example 2. So the first task in spelling out ordinary arguments is sometimes to make the conclusion explicit (if it is left implicit in the original).

If we make the initially implicit conclusions explicit, then the arguments become:

1´. Economic growth is near zero and inflation is on the increase

---

So, the Tories' economic policy is not working.

And:

2´. If your machine doesn't run BOS software, then it's a fridge.

So, your machine *will* run BOS software.

We can readily formalise these arguments in truth functional logic:

1´´. p & q

So, ¬ r

(where p is ‘Economic growth is near zero’, q: ‘Inflation is on the increase’ and r: ‘The Tories’ economic policy is working’.)

And:

2´´. ¬p →q

So, p

(where p is ‘your machine runs BOS software’ and q is ‘your machine is a fridge’.)

These arguments, as they stand, are of course invalid (*Exercise*: show this). But clearly the person making the remark in example 1 and the advertisers in example 2 intended the inferences to be valid.

So what is going on here? The fact is that, in ordinary arguments, we rarely spell out all the premises – instead we assume that the people responding to our arguments will share with us some background knowledge, the elements of which we expect to be taken for granted and so are not in need of spelling out, even though they are necessary for the validity of the argument. So some premises are left ‘hidden’. Clearly the person formulating argument 1 was expecting us to agree to the further, hidden premise, that ‘If an economy suffers near zero growth and an increasing inflation rate, then the policy governing it is not working’. In other words, the ‘hidden premise’ is ‘(p&q) → ¬r’.

And if we add this initially hidden premise to the argument then it becomes

1´´´.p & q [explicit premise]

(p&q) → ¬r [initially hidden premise, now articulated]

So, ¬r

And this is, of course, valid. (*Exercise*: check) Here the explicit premise was observably true at the time this argument was being presented, but an obvious reaction from a defender of Tory policy would be to reject the hidden premise and argue that, far from being part of background knowledge that we all accept, that premise is a highly debatable claim: sometimes low growth and inflation in the short term are necessary for longer term economic well-being.

As for the BOS software: in that case the 'machine' you are meant to be thinking about is a computer. So there is an obvious bit of background information or hidden premise: that your machine is not a fridge! (It is the obviousness of this hidden premise that is intended to give the advert its impact.) So the argument becomes:

2'''. $\neg p \rightarrow q$ [explicit premise]

$\neg q$ [initially hidden premise—your computer is not a fridge!]

So, p (i.e. your machine will run BOS Software).

And of course this is valid. (Usual *exercise*.)

There is a technical term for arguments in which one or more premise is left unstated – they are called ***enthymemes.***

A further example:

At the time of the 'dodgy dossier', Tony Blair said 'The BBC reports [that the dossier had deliberately exaggerated the threat posed by Sadaam Hussein and that Blair had insisted on the exaggerations being made, knowing them to be exaggerations] are absurd. If they were true, then I would have to resign.'

Analysis:

Conclusion: The BBC reports are absurd (r)

Explicit premise: If the BBC reports were true (p), then I (Tony Blair) would have to resign(q).

This formalises as:

1. $p \rightarrow q$

Therefore, r.

And this is obviously invalid as it stands. And yet clearly the saint-like Tony was expecting us to take the argument to be convincing, i.e. to be valid or at least not obviously invalid. So what is going on?

One possibility is that Blair had vaguely in mind a valid inference involving premises that he himself at least believed to be so obvious as not to need explicit articulation. His vague thought might have been that if he had done what the BBC reports said then he would have done something morally reprehensible and surely no one could think that he, of all people, was capable of that?!

So what was he expecting us to accept as hidden premises? Well, one obvious, and uncontentious, hidden premise is that 'If the reports were true then they were not absurd' (i.e. $p \rightarrow \neg r$). However, adding this uncontentious hidden premise is not enough to make the inference valid.

(*Exercise*: add that premise to the original and show that the resulting inference remains invalid.)

What Blair was expecting his hearers to take for granted, if this construal is correct, was something like that it was absurd to think that he would do anything so morally reprehensible that it would make resignation a moral requirement. Stripped of the rhetoric, this extra claim amounts to 'If I would have to resign (q) then I would have done something morally reprehensible (s), but I could never do anything morally reprehensible ($\neg s$)'. So he is expecting us to accept '$(q \rightarrow s) \& \neg s$' as a piece of 'background knowledge'. Moreover, he is expecting us to accept that if the assumption that a report were true entailed that he had done something morally reprehensible then it was not only false but absurd. That is $(p \rightarrow s) \rightarrow r$. If we were gullible enough to accept these two additional 'hidden premises' then the inference would indeed become valid. The inference would be

1. $p \rightarrow q$ [explicit premise]
2. $p \rightarrow \neg r$ [uncontentious hidden premise]
3. $(q \rightarrow s) \& \neg s$ [hidden premise]
4. $(p \rightarrow s) \rightarrow r$ [hidden premise]

So, r

This is indeed valid (*Exercise*: check), but some of us would regard the initially hidden premises 3 and 4 as far from being uncontentious parts of ‘background knowledge’.

A more important, and complicated example:

Let's now consider a somewhat more detailed and intellectually more important argument – one that might have been given in the 19th century in favour the claim that there must be an invisible, intangible medium that fills the whole of space, called the ‘luminiferous ether’, vibrations in which constitute light:

> *“The idea of an invisible, intangible "luminiferous aether" pervading the whole universe is a strange and unsettling one. Nonetheless it must be true. The mechanical world view states that the world (at least the ‘physical world’) consists fundamentally of matter in motion (and nothing else) and therefore implies that sources of light, such as the sun, can emit only one of two things: matter or motion (i.e. energy). If light consisted of matter, light sources would either emit a continuous stream or a succession of particles. If it were a continuous stream, two such streams could not cross one another without affecting one another. But light rays DO cross one another without any "interference" (think of two torch beams set at right angles to each other – each beam just carries on after the crossing as if the other one had not been there). If light sources emit material particles, then those particles would, when left to themselves, travel in strictly straight lines. But light does not travel in strict straight lines (even though it often appears to). So light sources must emit energy. But light is also known to have a finite velocity - that is, it takes a definite time to arrive from, say, the sun to the earth's surface. But where is the energy in between the sun and the earth? It must be stored in some medium that fills empty space. And that medium is the "luminiferous aether".”*

Analysis: The conclusion is that there is indeed a 'Luminiferous aether pervading the wholeuniverse'. Without a full account (which would be very lengthy in this case), we can see the main outline of the argument and its relationship to the truth functional logic we have studied. The truth of 'the mechanical world view' is taken as a 'hidden premise'. We are explicitly told that this implies that light is either matter or energy. *If* it is matter it would be *either (a)* a continuous stream *or(b)* a succession of particles. If

(a) then something would happen (viz. that the streams would affect one another when they cross) which does not in fact happen (another explicit premise). So it can't be (a). If it were (b) then light would travel in straight lines, but it *doesn't* (explicit premise again). So neither possibility holds, so the assumption that light is matter can't hold either. (This ought to remind you of the more elaborate version of the Edinburgh train example that we talked about very early on – it is a sort of extended 'disjunctive syllogism' *Exercise*: articulate this part of the argument explicitly in terms of propositional logic and show that it is valid.)

So we now have an intermediate conclusion, viz. that light consists of energy (we started with only two possibilities and have now eliminated the other). The rest of the argument is essentially as follows:

1. There is a finite time interval, Δt, when the energy is neither in the light source nor in the light receptor (This in turn is a consequence of the explicit premise that light has a finite velocity.)
2. Energy cannot be 'free' but must be stored in some matter at all times (implicit premise – but really a consequence of the initial premise i.e. the mechanical world view which entails that 'disembodied motion' is a nonsense).

Therefore, there must be some medium in between the sun and the earth that stores the light energy in the interval Δt.

Sometimes tracking down hidden premises can be tricky and lead to major intellectual breakthroughs:

Many major intellectual breakthroughs have been made in the following circumstances: An argument has been produced that seems to be valid and in which the explicit premises appear to be true, but whose conclusion is false. This must mean that some hidden premise has slipped in which is itself false. (*Important Exercise*: Explain carefully why this is true, using the definition of valid inference.)

If the argument is intuitively convincing, then it may be very hard to decide exactly what that hidden premise that is and why it is false. The breakthrough is made by discovering the hidden premise at issue and seeing why it is indeed false.

A good example is one of Zeno's famous paradoxes - the one about Achilles and the Tortoise. The two agree to have a race. To be fair, Achilles gives the Tortoise a start. Let Achilles start at A and the Tortoise at B:

A B C D E Finish

Zeno 'proved' that Achilles can never overtake the Tortoise.

His argument went as follows: By the time Achilles gets from A to B (the Tortoise's starting point), the Tortoise, no matter how slowly he moves, but given of course that he *is* moving, has gone some distance (let's say to C). So far, Achilles has not overtaken the Tortoise. Achilles eventually arrives at C, of course, but by that time the Tortoise has moved on (only a little bit, but he *has* moved on) say to D. Achilles eventually arrives at D, but by then the Tortoise has moved on to E, *etc.* Hence, Achilles *never* overtakes the Tortoise.

Clearly something is wrong with this argument since, if the race is long enough, the conclusion is empirically false – Achilles will overtake the Tortoise and win. But what exactly is it that's wrong? The answer turned out to depend on quite subtle features of a physical continuum. There is an implicit assumption that we can coherently talk of Achilles and the Tortoise being *AT* particular *POINTS* at particular times. But this assumption, however intuitively appealing, is not true of the sort of 'points' involved in the real number theory that underlines continuous processes in physics: to speak intuitively, no one is ever at such a point but always passing through it. Working out the correct theory of the continuum was of course an intellectual breakthrough of the highest order.

Plain bad arguments:

Of course some arguments are just plain bad arguments. From the point of view we have developed, we can see that there are basically two possible reasons for this:

(a) An *obviously false* explicit premise (even if the argument is valid, that is the conclusion *has* to *be* true if all the premises are, it will of course cut no ice if one or more of the premisesis or are obviously false – we called such arguments, remember, UNSOUND.

(b) The argument is invalid and any 'hidden premise' which might be invoked would itself be obviously false.

The following example illustrates both these possibilities:

In the 1970s a cult was built up around the Indian mystic Maharaj Ji – his (surprisingly many) followers believing that he was divine. One of these followers was an American professional tennis player of the time called Tim Galloway. Galloway wrote a book setting out his convictions about Maharaj Ji called *Inner Tennis*. The *New York Times* sent along a reporter to interview Galloway and published the following account:

> *“I asked Galloway how he had come to believe Maharaj Ji was God. [He replied:]*
>
> *"When I first heard him my only approach was to say to myself, 'He's either the real thing ora con artist'. Well the first time I saw him he just did too bad a job as a con-artist. A good con-artist wouldn't wear a gold wristwatch or give such stupid answers. When I was staying withhim in India I once asked him how much time I should spend on work* and *how much onmeditation* and *he just said get up an hour earlier and go to bed an hour later– hardly aprofound answer. I decided that if he was doing such a bad job of being a con-man he simply had to be genuine.”*
>
> *"Did it ever occur to you that he might be a* bad *con-man?"*
>
> *"Then how could he have six million followers?" the tennis pro replied.*

Here we have essentially two arguments. The first, which Mr. Galloway gave unprompted, has the following form:

1. p v q
2. r → ¬(s v t)
3. s &t

Therefore, p.

Here, p is 'Maharaj Ji is God', q is 'Maharaj Ji is a con-artist', r is (notice the important change here) 'Maharaj Ji is a *good* con-artist', s is 'He wears a gold watch', and t is 'He gives stupid answers'. (As usual things are not quite as simple as this and there is in the interview in addition a sort of sub-argument for t.)

This argument is *invalid* as it stands. (*Exercise*: Supply a counterexample.) In order to make it valid, we should have to add an ‘implicit or hidden premise’ to the effect that ‘If Maharaj Ji were a con-artist [at all] then he would be a *good* con-artist’, i.e. q → r. If we add this premise the inference indeed becomes valid (*Exercise:* check this by semantic trees). But there is of course no reason to believe that premise at all – there is no reason why Maharaj Ji could not indeed just have been a *bad* con-artist, as the reporter suggests.

Moreover, an explicit premise here is patently false – viz. premise 1. There is no reason at all why the two possibilities mentioned should be the only two possibilities. Maharaj Ji might for example be perfectly sincere but deluded (though no doubt one would suspect in that case that some of the men behind his organisation were con-artists). This is a frequent ploy in bad arguments: claim that the only two possibilities are p and q, quickly move on to an elaborate argument which (allegedly) refutes q, and infer p. The elaborate argument for q distracts attention from the otherwise obvious fact that p and q are *not* the only possibilities.

The second argument in this passage is generated as a result of the journalist's pointing out that the 'hidden premise' could obviously be challenged *["Did it ever occur to you that he might be a* bad *conman?"]*

Galloway's response is to produce an argument for the implicit premise. The argument can be formalised as follows:

Using 'u' for 'Maharaj Ji is a *bad* con artist' and 'w' for 'Maharaj Ji has six million followers' (and, remember, q means ‘MJ is a con artist’ and r that ‘MJ is a good con artist’):

1. $q \to (r \vee u)$
2. $u \to \neg w$
3. w

Therefore, q → r. (That is, if he were a con artist at all, he must be a good con artist.)

(As you will see from this, the distinction between 'hidden or implicit' and 'explicit' premises is not always totally clear-cut. Some 'hidden' premises are so little hidden as to verge on the explicit. For example, although Galloway never actually asserts w (i.e. that Maharaj Ji has six million followers) this is so clearly implied as to be “almost explicit”. Just as in formalising single sentences in truth-functional logic, so here in analysing formally ordinary informal arguments, we have the 'problem' that the logical system is totally precise, while ordinary discourse is often imprecise. It is therefore often a matter of judgement whether one's precise formal account 'captures' the informal one. Of course this is only a ‘problem’ in that it requires work from the logician seeking to apply his logical tools to ordinary argument, the total precision of logic is clearly a virtue.)

In this new argument, premise 1 seems reasonable enough, but the (more or less) explicit premise 2 is false. It is quite possible, given everything we know about people's psychological needs, that so many people could be taken in even by a bad con artist (indeed it seems overwhelmingly likely that Mr Galloway was one of them!). The *full* argument given by Galloway (obtained by replacing premise 2 in the initial argument by the argument for that premise that we have just analysed) is this:

1. $p \vee q$
2. $q \rightarrow (r \vee u)$
3. $u \rightarrow \neg w$
4. w
5. $r \rightarrow \neg(s \vee t)$
6. $s \,\&\, t$

Therefore, p

Although valid, this argument can hardly be taken to establish its conclusion (that Maharaj Ji is God) since it suffers from the embarrassment of including two premises – the first and the third – that, although we did need to articulate them in order for the reasoning to go through – are *quite patently* false.

C: FIRST ORDER PREDICATE LOGIC

C1 INTRODUCTION: THE NEED FOR A MORE POWERFUL SYSTEM OF LOGIC

Truth-functional logic is strong enough, as we have seen, to capture a wide range of intuitively valid inferences – and also, of course, to decide validity or invalidity where our intuitions are not so clear or, in the case of very complex inferences, non-existent. However, it is not difficult to see that truth-functional logic is not strong enough: there are lots of inferences that are obviously valid from the intuitive point of view – that is, inferences whose conclusions have to be true if their premises are – but which when formalised in truth functional logic produce invalid inference schemes. This is true in particular of all of Aristotle's famous syllogisms often considered the historical starting point of deductive logic.

A simple example is the 'daddy of them all':

1. All men are mortal.
2. Socrates is a man.

Therefore, Socrates is mortal.

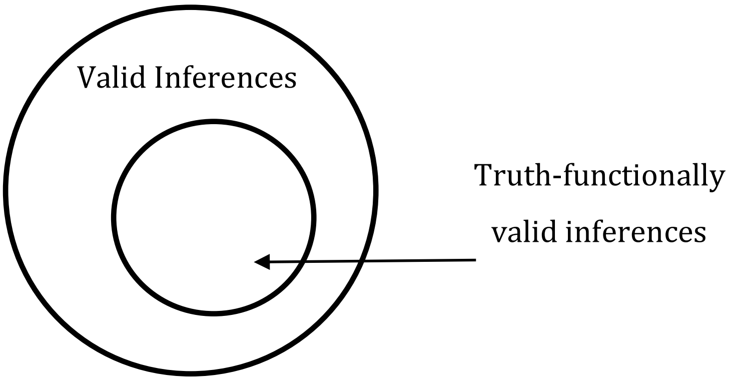
Try to think how this could be formalised in truth-functional logic. Neither any of the premises nor the conclusion breaks down into atomic sentences connected by truth functional connectives. There is something vaguely conditional about the first premise – we might think of it as saying something like 'If man then mortal'. But neither 'man' nor 'mortal ' is a sentence, so this is not a truth functional conditional. In fact, since none of the sentences involved is a truth functional compound of simpler sentences, the best we can do by way of formalisation in truth functional logic is simply:

1. p
2. q

Therefore, r

This inference scheme is obviously invalid since we can just assign p:T, q:T, r:F and this is a counterexample.

This simple example alone shows that the set of truth-functionally valid inferences is only a proper subset of the set of all valid inferences. In other words, the picture is:



Let's think a little more about the Socrates example since it can supply hints about how to move towards a more adequate, more extensive system of logic. The problem is that in formalising ‘All men are mortal’ as p, ‘Socrates is a man’ as q, and ‘Socrates is mortal’ as r, we lose the intuitive connection that exists between the various sentences as expressed in English. Once formalised in this way the sentences are regarded as totally independent from one another – any combination of truth values (in particular p:T, q:T, r:F) being possible. But clearly if it is true that 'All men are mortal' and true that 'Socrates is a man' then it can't be false that 'Socrates is mortal'. We have left out some important connections between the sentences in formalising them in truth functional logic in this way – but we can't produce any more elaborate formalisation within truth functional logic. Moral: we need a more refined language than that of truth functional logic in order to capture the intuitive connections between sentences such as these.

The method of syllogisms, developed some two thousand years ago by Aristotle, can of course easily deal with the Socrates example (which it was indeed constructed to deal with). But if we simply added Aristotle's syllogisms to truth functional logic we should soon find ourselves back in a similar situation to the one we are now in. There are lots of inferences that are intuitively valid but are neither truth functionally valid nor an instance of one of Aristotle's syllogisms. One simple example (suggested by a famous

19th Century mathematician and logician called Augustus de Morgan) is the following inference:

1. All horses are animals.

So, all heads of horses are heads of animals.

But there are many important inferences in science, social science and mathematics which fall in the same boat: clearly valid and yet there is no valid formalisation of them within truth-functional or Aristotelian syllogistic logic.

So we shall ignore Aristotelian syllogisms and jump instead straight to a much more powerful language and associated logic developed in the 19th and 20thcenturies that subsumes Aristotelian syllogisms as special cases and delivers much more besides. This is called **first-order predicate logic** (For the purposes of this course you can forget about the 'first-order' and just call it 'predicate logic'.)

The Socrates example shows that we must get within the structure of sentences that are truth-functionally simple. Remember that our original characterisation of a valid inference was as one that is valid because of its ***logical form***: a given inference is **valid iff there is no inference of the same logical form with true premises and a false conclusion.** In truth functional logic we take only the connectives as characterising the form of a sentence. The basic idea in predicate logic is to consider the 'all' and the 'are' in sentences like 'All men are mortal' as also constituting part of the logical form.

In fact, if we consider the equally ancient inference:

1. All Greeks are men.
2. All men are mortal.

So, all Greeks are mortal.

and take the form of this to be:

1. All A's are B's.
2. All B's are C's.

So, All A's are C's.

then we can give an explanation of the validity of this inference similar to that of the validity of truth functional inferences.

Here, initially A stands for ‘Greeks’, B for ‘men’ and C for ‘mortal’. As it stands the premises are true and so is the conclusion. But this in itself, as we now know, is no guarantee that the inference is valid. (The inference from ‘All electrons are negatively charged’ and ‘All protons are positively charged’ to the conclusion that ‘All neutrons are electrically neutral’ has true premises and a true conclusion, but obviously the conclusion, although true, does not follow from the premises. On the other hand, the inference from ‘All politicians always tell the truth’ and ‘All those who always tell the truth have blue eyes’ to ‘All politicians have blue eyes’ is valid, even though both its premises and its conclusion are of course false.) The question, as before, is: would the conclusion *have to be true* if the premises were true (whether or not they actually are). The answer in the ‘Greeks’ case is that it would, and that this is revealed by the fact that **no matter what we substitute for A, B and C in the above schema** (keeping the ‘all’ and the ‘are ’ fixed, just as we kept ‘and’, ‘or’, etc fixed in the case of truth functional logic),**we NEVER get true premises and a false conclusion**.

(*Exercise:* Try a few substitutions. Notice that we are now substituting common nouns (men, mortals, dogs… or whatever, rather than whole sentences as we did in truth functional logic.) You should be able to find substitutions which (i) make the premises true and the conclusion true, (ii) make the premises false and the conclusion true and (iii) make the premises false and the conclusion false. But you won’t find any that make the premises true and the conclusion false.)

Similar considerations apply to inferences like:

1. All the pieces that Mozart wrote are beautiful.
2. Some of the pieces that Mozart wrote are operas.

So, some operas are beautiful.

If we again regard any parts of the verb 'to be' as part of the form of a sentence, and also regard 'some' as part of the form then the form of this inference is:

1. All A's are B's.

2. Some A's are C's.

So, some C's are B's.

Here A means ‘is a piece that Mozart wrote’, B: ‘is an opera’ and C: ‘is beautiful’. But what they initially mean is irrelevant to the issue of validity: the inference about Mozart is valid because no matter what other meaning we substitute for the stuff about Mozart, operas, *etc.* – that is, no matter what we take A, B and C in this formalised inference scheme to mean – we never get true premises and a false conclusion.

(*Exercise:* Try a few substitutions, as before.)

C2: MONADIC PREDICATE LOGIC

I shall build up the system of predicate logic that we will study in two stages. First I shall take a restricted but very simple sub-system – so called **monadic logic** (don't worry about this term, it will become clear later). This sub-system has the advantage of allowing us to introduce all the main ideas in especially simple forms. I shall only then move on to the full and more complicated system.

C2(A): ASCRIPTIONS OF PROPERTIES TO INDIVIDUALS

The basic idea is going to be that an inference is valid if and only if there is no inference of the same logical form that has true premises and a false conclusion. How then should we express the logical form of sentences like 'Socrates is a man', 'Boris Johnson is a liar', 'All students are hardworking', 'Some logic lectures are boring'? Grammarians treat such sentences as *subject-predicate* assertions: each assertion attributes a certain 'property' or 'predicate' to a 'subject'. The first cases are especially straightforward: they consist in (truly or falsely) ascribing a property (e.g. that of being a liar) to a *particular individual* (namely Boris Johnson).

Using lower case letters from the beginning of the alphabet (a, b, c … or, if we need lots, *indexed* constants a_1, a_2, a_3 ...) as **names of individuals** and upper case letters (P, Q, R ... or P_1, P_2, P_3 ...) as **names of properties** or **predicates**, then it is natural to regard the ***form*** of this first sentence as just:

'a is P'

or

'a has the property P'.

And, simply to make the formalism as neat as possible, we will in fact formalise such sentences as:

Pa [but to be read, remember, as "a has the property P"].

Lots of other sentences ('Socrates is a man', 'Roman Abramovich has spoiled English football', 'Ken Livingstone is a pain in the neck', *etc.*) have exactly the same form. Though if we wanted to formalise two such sentences together, we should use different individual names and different predicate names to mark the differences. For example, the unlikely sentence 'Placido Domingo is the world's finest tenor and Roman Abramovich has spoiled English football' would be:

'a has the property P and b has the property Q'

or, more simply:

'Pa & Qb'.

(One thing you should note at this point is that predicate logic is indeed going to be an extension of truth functional logic – we will still use all the connectives we learned earlier and treat truth functional conjunctions, for example, as such; but it is just that what were, in truth functional logic, the un-analysed atoms that were conjoined are now given further structure – Pa, Qb, *etc.* rather than just p, q, *etc.*)

How about sentences like 'David Cameron was Prime Minister' or 'Humphrey Bogart starred in *Casablanca*'? Although in the past tense, these are just as much subject-predicate assertions as our earlier examples: there is nothing special here about the present tense of the verb 'to be'. We can in fact regard these sentences too as having the same form. The first can be taken as 'David Cameron *has [timelessly] the property of* having been Prime Minister. Similarly, 'Humphrey Bogart has the property of having starred in *Casablanca*'. (Some philosophers – who ought to have better things to do – have found difficulties with so-called 'posthumous predication' – that is, with sentences that ascribe properties to no longer living people. But we'll just take the naïve – and surely correct – view that all individuals, alive, dead, abstract (such as the number 4) and fictional (such as Hamlet or Santa Claus) can just be named and straightforwardly have properties attributed to them.)

So we shall in fact symbolise such sentences as those about David Cameron or Humphrey Bogart also as just Pa, Qb, or whatever, to be read as always as 'a has the property P', 'b has the property Q', *etc*.

C2(b): Universal Statements: The Universal Quantifier

Grammarians treat general or universal statements like 'All students are hard-working' or 'All logic lectures are interesting' (as before, the sentences we are considering do not need to be true!) as simply particular kinds of subject-predicate assertion. It's just that in these cases the 'subject', instead of being an individual (named by a 'proper noun'), is a whole class of individuals (named by a so-called 'common noun'). Since, however, there is no such entity as 'all students' (although there are of course lots of individual students) it is more accurate to regard the statement 'All students are hardworking' as a sort of ***indefinite conjunction**:* stating that any individual who happens to have one property (that of being a student) also has another (that of being a hardworker), or, more specifically, if more laboriously:

***For any object whatsoever*, if it's a student *then* it works hard**.

Similarly, 'All logic lectures are interesting' is best construed as saying:

For any object whatsoever, if it's a logic lecture then it's interesting.

The pronoun 'it' in the phrase, 'For any object whatsoever, if it's a P then it is also a Q', does not pick out a particular individual (unlike an individual name such as 'David Cameron'); instead it **VARIES OVER** individuals (not all of which may have individual names – for example we can say (truly) that 'all electrons are negatively charged', though electrons generally do not have names). This means we must introduce the idea of **VARIABLES –** and we denote these by lower case letters from the end of the alphabet (x,y,z, occasionally u, v, w, or again if we need lots of variables we use indices: x_1, x_2, x_3 ...).

(So, remember, a, b, c are ***individual constants*** that pick out or name *particularindividuals* – David Cameron, Boris Johnson, Roman Abramovic *or* whomever. x, y, zare *individual variables* varying *over* individuals, and do not name any particular individual).

So we can read our sentence about logic lectures as:

For any object x, if x has the property of being a logic lecture then x has the property of being interesting.

Using our predicate symbols and remembering the truth functional connective ‘→’, this becomes:

For any object x, $Px \rightarrow Qx$.

Finally, we introduce as shorthand for the phrase 'For any' the symbol '∀' (upside down A, as in ‘for All’ or ‘for Any’ - called the **UNIVERSAL QUANTIFIER**) and write:

$\forall x(Px \rightarrow Qx)$

The sentences 'All students are hardworking', 'All jabberwocks are dangerous', 'All *Sun* journalists are morally good' all share this same logical form; although again if wewanted to formalise several at once we should, of course, use different predicate-letters for the different properties. (So the sentence, ‘All jabberwocks are dangerous and all bandersnatches are frumious' could be formalised as ‘$\forall x(Px \rightarrow Qx) \& \forall x (Rx \rightarrow Sx)$’.)

C2(c): 'Some' Statements: The Existential Quantifier

How about statements like *'Some* lectures are interesting' or *'Some* politicians are liars'? These are again grammatically of subject-predicate form. The 'subjects' again do not name particular individuals. But nor do they refer to a whole class of individuals. They simply assert that, for example, *amongst* the class of all politicians ***there are some*** who have the property of being liars. So, using our notions ofvariables and predicates, we shall read this statement as:

There are some objects x, such that x has both the property of being a politician and the property of being a liar.

That is,

There are some objects x, such that Px & Qx.

(Exercise: 'There are some objects x such that *if Px then Qx'* would be the WRONG way to read 'some politicians are liars'. Can you explain why?)

How many count as 'some'? Would three lying politicians be enough to make 'Some politicians are liars' true? How about two? Or even one? The answer is that, just as we found for example with the connective 'or', ordinary usage is vague and ambiguous. We need to be precise in our logic, however, and logicians have found it best to take the minimal understanding of the phrase 'some' and take it as meaning merely 'at least one'. So that a 'some statement' is false if and only if there are **no objectsat all** of the kind described (no interesting logic lectures, no lying politicians, orwhatever). We shall see later that nothing is lost by making this decision. So we have the following 'formalisation' of 'Some Ps are Qs':

There exists at least one object x, such that Px & Qx.

Introducing the abbreviation ∃ (backwards E for 'there Exists') for the cumbersome phrase 'There exists at least one object…such that… ', we have finally:

∃x(Px & Qx).

The symbol ' ∃' is called the **EXISTENTIAL QUANTIFIER.**

C2(D): THE EXPRESSIVE POWER OF MONADIC PREDICATE LOGIC

The whole expressive power of truth functional logic carries over into predicate logic. If a sentence is a truth functional compound then this is maintained in predicate logic – but we ADD the ability to capture the form of what were in truth functional logic the simple, un-analysable, atomic sentences. So, for example, the sentence 'All politicians are liars and some voters have been fooled' has truth functional form ‘p **&** q' and so its predicate logic formalisation is:

$\forall x(Px \rightarrow Qx)$ **&**$\exists x(Rx \& Sx)$

(where Px means x is a politician, Qx means x is a liar, Rx means x is a voter and Sx means x has been fooled).

Similarly, 'If all politicians are liars then some voters have been fooled' has truth functional form p → q, where p is the sentence ‘All politicians are liars’ and q the sentence ‘some voters have been fooled’, hence its formalisation in predicate logic is:

$(\forall x(Px \rightarrow Qx)) \rightarrow (\exists x(Rx \& Sx))$

(It is important to understand that the p, q, r ... *etc.* of truth functional logic are full sentences, making an outright assertion and hence having a truth value. The Px, Qx, ...*etc.* that we use in predicate logic are *not* sentences but conditions or predicates: ‘is a politician’ is not a sentence which is either true or false, rather it ***produces*** a true-or-false sentence when some particular object is substituted for the variable.)

A sentence like 'All lying politicians are either wicked or stupid' is ***NOT*** a truth-functional compound. (It is not, in particular, equivalent to 'Either all lying politicians are wicked or all lying politicians are stupid' which *is* truth functionally compound. *Exercise:* Explain carefully why. You may find a diagram helps.) The sentence does however have more structure than merely 'All Ps are Qs'. In fact, it says: For all x, if x is BOTH a politician AND a liar, then x is EITHER wicked OR stupid.

So using our connectives, variables and the predicates P, Q, R and S (for ‘is a politician’, ‘is a liar’, ‘is wicked’ and ‘is stupid’, respectively) we get:

$\forall x((Px \& Qx) \rightarrow (Rx \vee Sx))$

All sorts of quite complicated sentences can be expressed in these ways.

(*Exercise*: have a go at:

1. Some lying politicians are wicked.
2. All confident lying politicians are successful.
3. All lying politicians are either successful or not confident.
4. If not all confident lying politicians are successful then some confident lying politicians are not successful.
5. If Boris Johnson is a non-wicked politician, then if he is a liar then not all lying politicians are wicked.)

So it is important to note then that once we have introduced the idea of predicates, we can use our truth functional connectives in obvious ways, not only to link fully fledged sentences but also predicates. That is, not only can we conjoin the full sentences ‘All politicians are liars’ and ‘All philosophers are honest’ as $\forall x((Px \rightarrow Qx)\ \&\ \forall x\ (Rx \rightarrow Sx))$, we can also conjoin (or disjoin or whatever) ***predicates*** to form more complex conditions – so the complex predicate ‘Px & Rx’ with the same understanding of Px and Rx is the predicate that holds of those and only those objects that are both politicians and philosophers. (There are a few, usually bad ones – both bad politicians and bad philosophers, that is!) The complex predicate Px v Rx is the condition that holds of all those objects that are either politicians or philosophers (or both – since, remember, we have taken ‘v’ as the *inclusive* sense of either/or).

C3: Validity of Inference: The Idea of an Interpretation

C3(a): The Intuitive Idea of Validity

We shall extend our system beyond monadic predicates shortly. But all the important logical notions – validity and invalidity of inference in particular – are relatively straightforward if we restrict ourselves to 'monadic predicates' (you will understand precisely what these are only later when we discuss non-monadic predicates).

As we already indicated, the basic idea of validity of inference is, as before, that an inference is valid if it is ***impossible for the premises to be true and the conclusionfalse***. In the case of truth-functional logic, we cashed out 'impossible' in terms of truthvalue assignments to the atomic propositions – impossible meant ‘no assignment of truth values to atomic components which makes the premises true and the conclusion false’. What is the corresponding notion in the case of predicate logic?

Well, let's think about the two ancient inferences I already mentioned.

A. All men are mortal	**B.** All Greeks are men
Socrates is a man	All men are mortal
So, Socrates is mortal	So, All Greeks are mortal

As before, to decide validity we first formalise the inference, only now we do it in the language of predicate logic. Using 'Px' for 'x is a man', 'Qx' for 'x is mortal', ' Rx' for 'x is a Greek' and the individual constant 'a 'as the name of the individual Socrates, we have:

A’:

1. $\forall x((Px \to Qx)$
2. Pa

So, Qa

And:

B’:

1. $\forall$x((Rx→Px)
2. $\forall$x((Px→Qx)

So, $\forall$x((Rx→Qx)

A’and **B’** are, then, the schematic **forms** (in predicate logic) of the originalintuitive inferences **A** and **B.** The basic idea is that, again as before, the ordinary language inferences **A** and **B** are both valid because no inference with either the form **A**' or the form **B**' has true premises and a false conclusion. That is, no matter what we substitute for 'men', 'mortals' and ‘Socrates' in **A**, or for 'Greeks', 'men' and 'mortals' in **B**, we **never get true premises and a false conclusion**. This means, more formally, that whatever predicates we substitute for the predicate letters P, Q, R in either **A'** or **B**' and whatever object we substitute for the individual constant *a* in inference **A'** we never get both premises true and the conclusion false.

So, for example, just considering **A'**, we might substitute 'x is an aardvark' for 'Px'; 'x is a quadruped' for 'Qx'; and Alf (a particular aardvark in London zoo) for 'a'. We would thus get:

A":

1. All aardvarks are quadrupeds
2. Alf is an aardvark

So, Alf is a quadruped

(True premises, true conclusion – I'm assuming that 'quadruped' means 'having 4 legs in the natural, complete state', so that we needn't worry, for example, about whether Alf might be an unfortunate aardvark-amputee).

Or we might substitute 'x is an LSE student' for 'Px', 'x is hardworking' for 'Qx' and Bert (an especially slothful sloth in London zoo) for a. We would thus get

A’’’

1. All LSE students are hardworking
2. Bert is an LSE student

Therefore, Bert is hardworking.

Here the first premise is (sadly) false and so is the second and the conclusion is also false**.**

If you try some further substitutions, you will also find cases in which the premises are false (remember this means: ***not all*** premises are true) and the conclusion true (try – Px: x is a football hooligan; Qx: x is a devout Christian; a: Justin Welby). But you can play this game all day long– and what you will NEVER find ***is a substitution that gives TRUE premises anda FALSE conclusion.***

This ('no substitution which give true premises and false conclusion') is going to form our characterisation of a valid inference in predicate logic. Clearly, if it is the correct characterisation, then intuitively invalid inferences ought to *fail* to satisfy it. This means that, in the case of intuitively invalid inferences, there should be substitutions for the predicates that ***do*** make the premises true but the conclusion false. Well, let's think about the following example:

C:

1. Some football players are very skillful**.**
2. Luis Suarez is a football player.

So, Luis Suarez is very skillful.

Here the premises are true and so is the conclusion. But the inference is invalid – not because the conclusion isn't true but because the conclusion is not *guaranteed* to be true just because the premises are. It ***MIGHT*** have been true that some football players are very skillful and Suarez is a footballer player but that he just happens to be one of those who are not very skillful.

So, checking more formally that inference **C** fails to satisfy our criterion for validity, we first formalise the inference. Using Px and Qx for the predicates 'x is a footballer' and 'x is very skillful', and a for the individual Luis Suarez, we get:

C':

1. $\exists$x(Px & Qx)

2. Pa

So, Qa

The original inference **C** is invalid on our proposed criterion if there are substitutions for Px, Qx and *a* in the inference *scheme***C'** which yield true premises and a false conclusion. But such substitutions are easy to supply. For example, let Px be 'x is a natural number' (the natural numbers are the counting numbers 1, 2, 3 ...), Qx be 'x is even' and *a* be the number 5. Then, under this substitution, the scheme **C'** yields:

1. Some natural numbers are even.
2. 5 is a natural number.

So, 5 is even.

where we do indeed have true premises and false conclusion.

Similarly, consider the following inference:

D.

1. Some judges are wealthy.
2. Some wealthy people are out of touch with 'ordinary life'.

So, some judges are out of touch with ordinary life.

Here, as I think every unbiased observer (this excludes judges) would agree, we have true premises and a true conclusion. The inference however is invalid – the truth of the conclusion is not guaranteed by the truth of the premises. It MIGHT have been true that there are wealthy judges and out-of-touch wealthy people, while it just happened that none of the out-of-touch wealthy people were also judges. The conclusion may actually be true, but it *could* have been false even when the premises were true. Let's again check that our characterisation of validity captures this idea by pronouncing inference **D** invalid.

We can formalise the inference using Px, Qx and Rx for the respective predicates 'is a judge', 'is wealthy' and 'is out-of-touch with 'ordinary life'', and obtain:

D':

1. ∃x(Px & Qx)
2. ∃x(Qx & Rx)

So, ∃x(Px & Rx)

The original inference **D** is invalid iff there is a substitution for Px, Qx and Rx in **D'** which makes both premises true but the conclusion false. Obviously to make the conclusion false we shall need Px and Rx to be incompatible properties: say, 'is male' and 'is female' in humans, or 'is less than 10' and 'is greater than 20' in the natural numbers. We then need Qx to be a property compatible with either – say, 'is Russian' or 'is even'. The two sets of properties would give the following inferences when substituted into **D'**

1. Some males are Russian
2. Some Russians are female

So, some males are females

1. Some numbers less than 10 are even
2. Some even numbers are greater than 20

So, Some numbers less than 10 are greater than 20

The first has true premises and a false conclusion – at any rate if we idealise away certain difficult arguable borderline cases (e.g. of hermaphrodites). One advantage of using properties of natural numbers is that everything is clear cut: there are no borderline cases. So in the second substitution it is completely clear that we have true premises and a false conclusion. Hence this second inference forms a clear-cut **COUNTEREXAMPLE** to our original inference D. (The first would count asclear-enough-cut for me, so don't worry if you are happier dealing with 'ordinary' properties than with numerical ones.)

So, our account of validity in predicate logic does indeed characterise intuitively invalid inferences as invalid. In order to put the account into precise form, we need first to give a precise account of the idea of substituting one set of properties and individuals for another and for this we in fact use the central notion of an **INTERPRETATION**.

C3(B): THE IDEA OF AN INTERPRETATION OF A SET OF PREDICATE LOGIC SENTENCES

Basically, an interpretation is what turns a set of symbolic sentences (like those involved in the inference-schemas **A', B', C'** and **D'** above) back into ordinary language true-or-false assertions – about people, numbers or whatever.

For technical reasons (to be discussed later) it is not sufficient to interpret 'for all x' as meaning *'for anything at all'.* Instead we need to specify some set (the set of all (past, present and future) humans, the set of natural numbers, or the set of all physical bodies in the universe, or whatever) as the so-called ***domain*** of the interpretation**.**Hence 'for all x' will mean 'for all objects in the domain' (whateverthe domain set might be in the particular interpretation concerned) and 'some x' will mean 'for at least one object in the domain'.

How about the constants? There are two types of constants in our system: ***individual constants*** naming individuals, and ***predicate constants*** standing forproperties. Obviously we must pick out some particular element of the domain to associate with each individual constant (perhaps Socrates or Marilyn Monroe if the domain is humans, the number 4 or the number 77 if the domain is natural numbers). And we must associate with each predicate some particular property that makes sense when applied to the domain (so it might be 'is male' if the domain is humans, or 'is even' if the domain is natural numbers). In sum, an **INTERPRETATION** of a set of sentences of predicate logic consists of:

(a) A specification of the ***domain*** over which the variables range.

(b) A meaning within that domain to the constants (individual and predicate) occurring in the symbolic sentences.

Example 1:

{∀x (Px → Qx), Pa, ∃x (Px & Qx)}

Interpretation A:

Domain: Natural numbers

Px: x is even

Qx: x ≥ 2 (i.e. x is greater than 2)

a: 4

Under this interpretation the set of sentences reads:

{Every natural number which is even is ≥2; 4 is even; there are natural numbers that are both even and ≥2.}

Of course, we are only allowed to specify one meaning for a constant in any given interpretation. P can't mean one thing in the first sentence in Example 1 and something else in the second.

There is nothing in the notion of interpretation that requires the interpreted sentence to be *true* – the interpretation just turns the symbolic formulae back into ordinary true *or* false assertions. E.g., the following would also be an interpretation of the set of sentences in Example 1:

Interpretation B:

Domain: Natural numbers

Px: x is even

Qx: x is odd

a: 5

Under this interpretation the first two sentences in the set are false. (*Exercise:* Write out the interpreted sentences.)

As I just emphasised, an interpretation of a set of predicate logic formulas need not produce sentences all of which are true. IF*,* however*,* an interpretation turns all the formulas in some set into true sentences, then it is called a **MODEL** of that set of formulas. So interpretation A***is***a model of the set ofsentences in Example 1, while interpretation B is ***NOT*** a model of that set. Interpretations simply turn the symbolic formulas back into ordinary true or false assertions. Models are a proper subclass of interpretations – they are those interpretations that happen to make the interpreted set of sentences all true.

Example 2:

{∀x((Px v Qx) →¬Rx), ∃x(¬Px & Rx), ¬Pa & ¬Ra, Qb & ¬Rb}

Interpretation A:

Domain: UK citizens of voting age

Px: x is a criminal

Qx: x is a member of the Royal Family

Rx: x has a vote

a: Prince Charles

b: The Queen

Under this interpretation the sentences read:

{Any UK citizen of voting age who is either a criminal or a member of the Royal Family has no vote; Some UK citizens of voting age who are non-criminals have a vote; Prince Charles is not a criminal and does not have a vote; The Queen is a member of the Royal Family and does not have a vote}

This interpretation happens to be a model (or so I believe – legal experts may want to correct me).

Interpretation B:

Domain: Natural Numbers

Px: x is even

Qx: x is prime

Rx: x is odd

a: 2

b: 4

Under this interpretation the sentences read:

{Any natural number which is either even or prime is not odd; There are some natural numbers which are both not even and odd; 2 is not even and not odd; 4 is prime and is not odd}

This interpretation is not a model of the set of sentences – the second sentence is true in this interpretation, but to be a model of the set S the interpretation must make *all* the formulas in S true.

Given these notions of an interpretation and of a model of a set of sentences, we can make our earlier characterisation of valid inference in (monadic) predicate logic precise:

***Definition:* Validity:**

Let S be a set of formulas of first order predicate logic andsa singlesuch sentence. The inference from S to s is ***VALID*** iff **there is NO interpretation** of the set of sentences S U{s} which makes **all the sentences in S true but s false**. (Or, more briefly, such an inference is valid iff **every model of S is also a model of s**, i.e. iff there is no model of S U {¬s}.)

(*Exercise*: This is a central definition, take time to think it through carefully and check that you understand why the various alternative formulations of the notion are equivalent.)

An inference in ordinary language will be deemed valid if there is a formalisation of it in first order predicate logic that is valid according to the above definition. Our ancient Socrates example is, for example, *valid* because there is no interpretation of the sentences '(∀x(Px → Qx))' and 'Pa' which makes them true which does not also make 'Qa' true. That is, there is no model of the set {∀x(Px→Qx), Pa, ¬Qa}; i.e. **no COUNTEREXAMPLE** to the inference.

C3(c): Demonstrating Validity and Invalidity

Although it is true that there are no counterexamples to our Socrates inference, it is not at all clear how we could actually ***demonstrate*** this. We would, it seems, have to check in turn **EVERY POSSIBLE interpretation** of the formal sentences in the inference scheme **A'** above to see that no interpretation is a model of the premises but at the same time makes the conclusion false. However, since there are infinitely many possible interpretations, this is clearly an impossible task.

(You might wonder why this problem does not arise in truth functional logic. There too the reason for the validity of the inference from, say, the set of premises {p → q, p} to the conclusion q is that no matter which ordinary sentences are substituted for p and for q, we never have true premises and a false conclusion. But again it would be impossible to check *all possible* substitutions (p could be 'today is Wednesday', 'The Moon is made of green cheese', 'all electrons have negative charge' and so on indefinitely and the same for q.) In the truth functional case, however, we can readily bring the problem down to manageable proportions: all that we need to ask of given sentences which are taken as interpretations of p and q is whether they are true or false, hence we can partition the infinite set of all possible interpretations of, in this case, p and q, into finitely many (in this case four) sub-sets: (1) those in which p and q, whatever particular sentences they happen to be – whether about the day of the week, the moon, or electrons or whatever), are both true; (2) those in which p is true and q false; (3) those in which p is false but q true; and (4) those in which both are false. This is precisely what we do in employing the truth table method, or one of its derivatives, to decide truth functional validity. However, there is no obvious way of reducing the problem for predicate logic to finite proportions in a similar fashion. In fact, as we will see, not only is no such way apparent, no such way exists! So we will have to approach the issue of demonstrating validity in predicate logic in a different way.)

So the basic definition, although it tells us precisely what it *means* for a predicate logic inference to be valid, gives us no hint of how we might ***demonstrate*** validity. On the other hand, we ***can*** specify conditions under which we would have demonstrated **in*validity***: all that it takes to show that an inference is invalid is to produce a single interpretation in which the premises are true and the conclusion false, or as we shall say, a single counterexample.

Example 1:

Take again inference**D**(the one about the judges)*above*. Thisformalised, remember, as:

1. $\exists x\ (Px \ \&\ Qx)$
2. $\exists x\ (Qx \ \&\ Rx)$

So, $\exists x\ (Px \ \&\ Rx)$

The following interpretation is a ***counterexample***– since it makes the premises true and the conclusion false. It thus establishes the invalidity of the inference – that is the original inference about the judges.

Domain: {Natural numbers}

Px: x is even

Qx: x > 10

Rx: x is odd

Under this interpretation the inference becomes:

1. Some natural numbers are even and > 10.
2. Some natural numbers that are > 10 are odd.

So, some natural numbers are both even and odd.

This is indeed obviously a counterexample – the premises are true and the conclusion is false. Hence the original inference **D** (which was remember about judges being out of touch and stuff) is invalid – because it has an invalid form and this shown by the fact that there is a content-wise completely different inference which nonetheless has the same form and which has true premises and a false conclusion.

Example 2:

1. All elementary particles are either positively charged, negatively charged or electrically neutral.

2. All neutrinos are elementary particles and are not positively charged.

So, all neutrinos are electrically neutral.

Here both the premises and the conclusion happen to be true, but the inference is (rather obviously) invalid. To show that it is invalid we first formalise it:

1. $\forall x(Px \rightarrow (Qx \vee Rx \vee Sx))$
2. $\forall x(Tx \rightarrow (Px \,\&\, \neg Qx))$

So, $\forall x(Tx \rightarrow Sx)$

And then produce an interpretation in which the premises remain true but the conclusion is false. For example:

Domain: {physical objects}

Px: x is an elementary particle

Qx: x is positively charged

Rx: x is negatively charged

Sx: x is electrically neutral

Tx: x is an electron

In other words, we just substitute 'electrons' (in our (re-)interpretation) for 'neutrinos' (in the original inference). The premises are still true (electrons are elementary particles and are not positively charged). But the conclusion is *false* – since, as a matter of fact, electrons are negatively charged, not electrically neutral.

The outcome of this section, then, is that IF we can actually produce a counterexample to an inference, then it must be invalid. The next question is whether there is some ***SYSTEMATIC*** way of producing counterexamples to inferences that are in fact invalid? That is, is there something analogous in predicate logic to the truth-table, or semantic tree *decision procedures* we developed for truth functional logic? These methods for truth-functional logic were, remember, completely algorithmic – you could apply either method completely automatically to any inference in the language of truth functional logic and the method would give you an answer – valid or invalid – in a finite number of steps. The answer to the question – for predicate logic generally - is **NO**: there is no general algorithmic method for producing counterexamples to invalid inferences. Instead you have to exercise your ingenuity to some extent.

Looking at our two examples, however, it is clear that we are not thrown back simply on undirected trial and error. In ***Example 1***, we can reason as we did earlier: The premises *require* there to be some Ps that are Qs and some Qs that are Rs but leave open the possibility that none of the Ps that are Qs are also Rs. It is just, then, a question of actualising this possibility by finding incompatible predicates Px and Rx, each of which is however separately compatible with Qx. This is what the interpretation we found does (go back and check).

In ***Example 2*** it is still clearer why the inference is invalid. The premises taken together leave open two possibilities for the neutrinos – the conclusion asserts that *one* of these holds, but it is clearly possible that it might have been the other (namely negative charge). Again it is a question of finding an interpretation that actualises this possibility.

If, however, you have no immediate intuitions about how to produce a counterexample, or indeed about whether or not a given inference is invalid, then you just have to try a few interpretations and hope that eventually light will dawn. (*With practice*, it will.)

C4: CONSISTENCY AND INDEPENDENCE

As I already indicated, the definition of validity of inference tells us *what it means* for an inference to be valid, but it doesn't tell us how to *demonstrate* the validity of any inference and it doesn't even tell us when we could justifiably assert that an inference is valid. (Clearly simply trying to find a counterexample and failing is not sufficient. There are infinitely many possible counterexamples and we may simply not yet have looked hard or long enough.) The natural next step would be to remedy this defect. But this requires the introduction of some new ideas, and there are some other important logical notions which – like that of *in*validity of inference – can be dealt with directly using the ideas of interpretations and models that we have already introduced. So let's pause to introduce them and turn to the new ideas needed to demonstrate validity later.

The further important notions at issue are those of the **CONSISTENCY** of a set of sentences and of the **INDEPENDENCE** of a single sentence from a given set of sentences.

We came across these notions of consistency and independence in truth-functional logic (you should go back and refresh your memory). The characterisations of these notions for sentences in the language of predicate logic are straightforward generalisations of the truth functional notions.

***Definitions:* Consistency and Independence**

(1) A set of sentences S in the language of (first order) predicate logic is **consistent** iff there is a model of S – i.e. a single interpretation under which all the sentences in S are true.

(2) Letsbe a single sentence and S be a set of sentences (all in the language ofpredicate logic). s is **independent** of S iff neither s nor ¬s is validly inferrable from S as premises (this requires there to be two models: one of Sand ¬ s (showing the invalidity of the inference from S to s) and one ofS and s (i.e. ¬¬s) (showing the invalidity of the inference from S to ¬s)).

(Notice that, just as in truth functional logic, the notions of consistency and independence are closely related: s is independent of S iff both the set S U {s} and the set S U {¬s} are consistent.)

(Important Exercises:

(1) Explain carefully why it is true that, in predicate logic, s is independent of S iff both S U {s} and S U{¬s} are consistent.

(2) '**ANY** sentence in predicate logic is validly inferrable from an inconsistent set of such sentences.' True or false?

(3) 'A set of predicate logic sentences S U {s} where s is a single sentence, is inconsistent if ¬s is validly inferrable from S'. True or false?))

Example 1:

Consider the set of sentences {All philosophers are either rationalists orempiricists; Some rationalists are obscure; No philosophers are obscure}. Not all of these sentences are true in the real world (notably the last!). The set is, however, **consistent** – all the sentences *could* be true together (even though as a matter of factthey aren't). (All it would take, as some of you might see intuitively, is for some rationalists not to be philosophers and for those non-philosopher rationalists to be the obscure ones.) Consistency is demonstrated by *first* formalising the sentences: {∀x(Px → (Qx v Rx)), ∃x(Qx & Sx), ¬∃x(Px & Sx)).

Then by producing an interpretation of these symbolic sentences in which they all turnout true. As in the case of demonstrating invalidity of inference, there is – in general – no algorithm for finding such an interpretation. You just have to exercise a little ingenuity.

Here, if you think about it, we clearly need Px and Sx to be incompatible predicates within whatever domain we select, while Qx and Sx must be compatible, and all Ps must be either Qs or Rs. The following interpretation will in fact work:

Domain: {animals}

Px: x is a dog

Qx: x is male

Rx: x is female

Sx: x is human

Under this interpretation, our symbolic sentences yield the following set:

(All dogs are either male or female; There are some male humans; There are no dogs that are human)

All these sentences (including the last – no matter what some dog owners may think) are of course true. Hence the *original* set of sentences – that was about philosophers – is consistent.

Example 2:

Consider the set of sentences {All students, except those studying logic, are lazy; Anyone who is lazy will do badly in the exams; All students who frequent the Three Tuns will do badly in the exams; All students either study logic or frequent the Three Tuns}.

Again not all sentences in this set are true! Are they, nonetheless, jointly consistent? First we formalise and obtain: {∀x((Px & ¬Qx) → Rx) , ∀x(Rx → Sx) , ∀x((Px & Tx) →Sx), ∀x(Px → (Qx v Tx))}

(Here Px stands for 'x is a student', Qx for 'x is studying logic', Rx: 'x is lazy', Sx: 'x will do badly in the exams', Tx: 'x frequents the Three Tuns' – note the construction for 'except' in the first sentence.) The following interpretation is a model of this set of symbolic sentences and hence demonstrates the consistency of the original set:

Domain: {positive and negative whole numbers}

Px: x > 0 Qx: x is odd

Rx: x is divisible (without remainder) by 2

Sx: x is the sum of two odd numbers

Tx: x is even

Under this interpretation the set reads:

(All whole numbers greater than zero, except the odd ones, are divisible by 2; Any whole number divisible by 2 is the sum of two odd numbers; Any even whole number greater- than zero is the sum of two odd whole numbers; Any whole number greater than zero is either odd or even)

These are all true (the 2nd and 3rd sentences being two forms of an elementary theorem of arithmetic).

Example 3:

Iss= "Some judges are out of touch with 'ordinary life'" ***independent*** ofthe set S = {Some judges are wealthy; Some wealthy people are out of touch with 'ordinary life'}? This, if true, would mean we can infer *neither* s *nor* ¬s from S. First, formalise:

S = {∃x(Px & Qx), ∃x(Qx & Rx)}

s = ∃x(Px & Rx)

{For the obvious choices of predicates Px, Qx and Rx.}

We already know that s is not validly inferrable from S, via the interpretation given earlier when we were establishing that certain inferences were invalid:

Domain: {humans}

Px: x is male

Qx: x is Russian

Rx: x is female

If the inference from to ¬s were valid, then there would be no counterexample, i.e. no interpretation in which the sentences in S are true and ¬s is false, but if ¬s is false, then s is true so this would require there to be no interpretation in which the sentences in S U {s} are all true. So if we can show that there *is* such an interpretation, then ¬s is not validly inferrable from S. And in fact the following interpretation fits the bill:

Domain: {natural numbers}

Px: x is even

Qx: x is divisible by 4

Rx: x is divisible by 8

Under this interpretation, S reads {Some even numbers are divisible by 4; Some numbers divisible by 4 are divisible by 8} and s is 'Some numbers are divisible by 8'. All these sentences are of course true.

Hence this interpretation together with the Russian/male/female one above demonstrate that **s *is* independent of S.**

Example 4:

Is the set {All logic lectures are interesting; Some logic lectures are notinteresting} consistent?

It formalises, for the obvious meanings for Px and Qx, as: {$\forall x(Px \rightarrow Qx)$, $\exists x(Px \& \neg Qx)$}. The answer from intuition is obviously 'no'. But so far we don't know how to show this. If we tried to produce an interpretation that was a model, we would of course fail. But failure (so far) to produce a model doesn't entail that *there isn't one.*

Example 5:

Is the sentence s = 'All Greeks are mortal' independent of the set S: {All Greeksare men; All men are mortals}? The answer is again obviously not – since, as we already remarked when discussing validity of inference, s here is validly inferrable from S. But again as already remarked, we can't show this to be true on the basis of consideration of interpretations.

The lesson is that in order to demonstrate **INCONSISTENCY** and **DEPENDENCE** we need some new idea. In fact, it turns out to be the same new idea as is needed to demonstrate validity of inference – we shall introduce it immediately after considering a way of making our investigations of consistency and independence more systematic.

C5: FINITE INTERPRETATIONS/MODELS

If you are asked to show that a set of sentences in predicate logic is consistent then you have to produce a model of that set. As in the case of finding counterexamples to inferences, there is no general algorithmic method. However, some consistent sets of sentences – and all the ones that you will be asked about – admit of a ***finite model***. With finite models, it is possible to work in an almost completely systematic way. A **finite *interpretation*** is simply one in which the domain set (*i.e.* the set of all the objects over which the variables range) instead of being infinite (like the set of all natural numbers) or indefinite (like the set of all humans) is finite (perhaps the set: {1, 2, 3}). And a finite model of a set of sentences is, of course, just a finite interpretation under which all the sentences turn out true.

The second step toward the finite model method technique concerns the predicates. Philosophers like to say that predicates can be specified either ***intensionally orextensionally.*** The **INTENSION** of a predicate is its meaning – so 'x is red' *means*x (whatever it is) is red and 'x is an even number' means x is an even number. The **EXTENSION** of a predicate, on the other hand, is the set of all individuals that possess the property concerned – so the extension of 'is red' will be a whole big set of things including various Ferraris, Rita Hayworth's hair, various Liverpool football shirts and so on. The extension of the predicate 'x is an even natural number' is the set {2, 4, 6, 8, ... }.

Now we said that when supplying an interpretation, we first specify a basic set as the domain of the interpretation. Within such an interpretation, each predicate will have as its extension some ***subset*** of the domain set. So if the domain is the set of all past and present cars then 'x is red' determines a subset of that set – *viz.* the subset which contains all and only all the red cars. If the domain is the set of all natural numbers, then the extension of 'x is even' is again the set of all even numbers, which of course is a subset of the set of *all* natural numbers. More importantly for present purposes, if the domain set is the set of all natural numbers up to and including 10, i.e. {1,2,3,4,5,6,7,8,9,10} then the predicate 'x is even' determines the subset {2, 4, 6, 8, 10} of that set.

The basic ideas of the finite model technique are:

1. Specify the domain as a *finite* set.
2. Forget about intensions altogether and regard *any* subset of the domain-set as a legitimate interpretation of any predicate.

So, if we are considering two predicates, say Px and Qx, an individual constant *a*, and a finite domain, say {1,2,3), then our finite interpretation might make the extension of the predicate Px the subset {1,2}, the extension of the predicate Qx the subset {2,3}, and might make the individual constant *a* stand for 3. This means that any individual x in the domain has the property P just in case it is a member of the subset {1,2}, that is, just in case x=1 or x=2, and has the property Q just in case it is a member of the subset {2,3}, that is, just in case x=2 or x=3. Given that the individual constant *a* has been interpreted as naming 3 then the *sentence* Pa would be *false* in this interpretation, since 3 is not an element of the extension of Px (i.e. ¬(3 ε {1,2}); while the *sentence* Qa would be true, since 3 ε {2,3}.) Here in line with general set theory we use the Greek lower-case letter epsilon, **ε**, to stand for '**is a member of**'.

How are quantified sentences to be interpreted in finite domains?

Universally quantified sentences:

Consider, for example, the sentence ∀x(Px → Qx).This means that everything that has theproperty P also has the property Q. In terms of extensions this means, of course, that everything in the extension of P is also in the extension of Q. (So 'All men are mortal', if true, means that the set of all men is a subset of the set of all mortals.) This is easy to check in the case of a finite domain.

Given the way that we have interpreted P and Q in the finite interpretation just specified, (namely with P having the extension {1,2} and Q the extension {2,3}), ∀x(Px →Qx) is in fact FALSE, since 1 is in the extension of P but not in the extension of Q (i.e. it is not true that {1,2} is a subset of {2,3}).

Existentially quantified sentences

The sentence ∃x(Px & Qx), for example, means that at least one thing has ***both*** the property P ***and*** the property Q.That is, it says that the extensions of P and Q have at least one element in common. Again this is easy to check in the case of finite interpretations.

In the interpretation just given $\exists x(Px \;\&\; Qx)$ is in fact *true* since the element 2 is of course an element *both* of the extension of Px *viz.*$\{1,2\}$ *and* of the extension of Qx, *viz.* $\{2,3\}$.

Other quantified sentences are just as straightforward. For example, $\forall xPx$ means that the extension of P coincides with the whole domain set. The finite interpretation we have been considering as a simple example has domain $\{1,2,3\}$ while Px was given the extension $\{1,2\}$ – this of course means that the sentence $\forall xPx$ is *false* in this interpretation, since 3 is in the domain set but not in $\{1,2\}$.

$\forall x(Px \vee Qx)$ means that everything in the domain is either the extension of P or in the extension of Q (or of course in both). This sentence is in fact *true* in our interpretation: because when we put together the elements in the sets $\{1,2\}$ and $\{2,3\}$ (when we 'form the **union**' of these two sets, as set-theoreticians say) we get the whole domain set $\{1,2,3\}$.

How about a sentence involving negation, such as $\forall x(Px \rightarrow \neg Qx)$? This means that everything in the extension of P *fails* to be in the extension of Q (in set theoretic terminology, the two sets 'have an empty **intersection**'). The sentence is therefore *false* under the interpretation we are considering, since 2 is in the extension of P but also in that of Q, i.e. 2 does *not* fail to be in the extension of Q. If this seems a bit opaque, think of it this way: the sentence $\forall x(Px \rightarrow \neg Qx)$ amounts, in the domain which has only the 3 members 1,2,3 to the finite conjunction $(P1\rightarrow \neg Q1) \;\&\; (P2\rightarrow \neg Q2) \;\&\; (P3\rightarrow \neg Q3))$; in order for this to be true all the conjuncts have to be true individually but in fact $(P2\rightarrow \neg Q2)$ is *false* since it has a true antecedent ($2 \,\varepsilon\, \{1,2\}$) but a false consequent – it is $Q2$ that holds not $\neg Q2$ since $2 \,\varepsilon\, \{2,3\}$.

Thinking in terms of a universally quantified sentence reducing, in finite interpretations, to a finite conjunction is often helpful.

Using finite interpretations to establish the consistency and independence:

This is best explained via examples.

Example 1:

Is the set S = {∀x(Px→Qx), ∃x(Qx & ¬Px), ∃x(Qx & Px), Pa} consistent?

Let, arbitrarily, the domain be the first three natural numbers {1, 2, 3} and use **P** and **Q** (bold face) as names of the sets (extensions) associated with the predicates Px and Qx.

Let, arbitrarily, the interpretation of 'a' be '1'. Then in order to make the final sentence in the set true 1 must be an element of **P**.

To make the third sentence true there must be at least one thing that is in **P** that is also in **Q.** We may as well let that be '1' also (for the time being – we can always come back and change it if things don't work out). So we put 1 in **Q.**

For the second sentence to be true there has to be at least one thing in **Q** which is not also in **P** - let that be '2'.

Thus, so far, we have ***Interpretation***:

Domain: {1, 2, 3)}

P: {1}

Q: {1, 2}

a: 1

In fact, the first sentence is also true in this interpretation since everything that is in **P** (just '1') is also in **Q.** So this interpretation is indeed a model of this set of sentences S and this demonstrates S is consistent.

(How many elements you need in the domain will depend on the particular sentences involved. There's nothing magic about having three elements – indeed in this case two would clearly have been sufficient. A good approach would be to start with 2 members in the domain set and add others only if that becomes necessary.)

Example 2:

Is the set S = {∀xPx, ∀x(Px→(QxvRx)), ∃x(Px&¬Qx), ∀x(Qx ↔ ¬Rx)} consistent?

Clearly the first sentence requires that everything in the domain be in **P** – so 1, 2 & 3 are all in **P**.The second sentence, given the first, requires everything to be either in **Q** or in **R** – let's say (this is just a decision) that 2 and 3 go in **Q** and 1 in **R**. This also makes the 3rd sentence –∃x(Px&¬Qx) – true since the sentence says that there is some element of the domain that is in **P** that is not also in **Q** and, as things stand, this is true of the element 1. Finally, the fourth sentence requires that any element from the domain is in **Q** just in case it is not in **R**– that is, in effect, that **Q** and **R** exhaust the domain between them and with 1 in **R** and 2 and 3 in **Q** that is exactly the case. (Again if this is opaque, it may be useful to think in terms of finite conjunctions: given that there are only three elements in the domain what it takes for the sentence ∀x(Qx ↔ ¬Rx) to be true is for (Q1 ↔ ¬R1) & (Q2 ↔ ¬R2) & (Q3 ↔ ¬R3) to be true and if you work through the conjuncts (on the basis of the interpretation) you will find that they all true (e.g. 2 is in Q but not in R so it is Q2 (True) ↔ ¬R2 (True) and True ↔ True is of course true.*Exercise*: work through the other conjuncts.)

So in sum the following is an interpretation which provides a model of S and hence shows that S is consistent:

Domain: {1,2,3}

P: {1,2,3}

Q: {2,3}

R: {1}

Example 3:

Is the set {∀x (Px → Qx), ∀x (Px →¬Qx)} consistent?

(It might be the formalisation of {All ravens are black, No ravens are black (≡ all ravens are not black)}.)

This is certainly a funny pair of sentences but it is **NOT an inconsistent one**. The two sentences are consistent so long as there are no P's (no ravens in the example). One permissible interpretation for any predicate is as the **EMPTY SET** – the set with no members, written ϕ (Greek lower case phi). So, e.g., the following is a finite model:

Domain: {1, 2}

P: ϕ

Q: {1, 2}

(Q doesn't in fact matter – it could be ϕ too, if you liked.)

Again thinking in terms of finite conjunctions should remove any lingering feeling of unclarity. ∀x(Px → Qx) is equivalent in our domain to (P1 → Q1) & (P2 → Q2) and both of these conditionals are true by the truth table for conditionals (in both cases the antecedent is false – nothing is a P and so 1 isn't, i.e. P1 is false and 2 isn't, so P2 is false; both the consequents are true (Q1 and Q2 are both true) but 'false then true is true'. As for ∀x (Px → ¬Qx), this amounts to (P1 → ¬Q1) & (P2 → ¬Q2) and again this is true – now both conjuncts have false antecedents and false consequents (since both 1 and 2 are Q) but false then false is also true.

Notice, by the way, that the set of sentences {Pa, ∀x(Px →Qx), ∀x(Px →¬Qx)} cannot be given a model in this way (*Exercise*: Explain carefully why.)

Finite interpretations can also be used to show that a given single sentence s in the language of predicate logic is *independent* of a set S of such sentences.

Example 4:

Is s = '∀x((Px & Qx) → (Sx v ¬Rx))' independent of S = {∃x(Px & Qx & Rx), ∃x(Px & Qx & ¬Rx), ∃x(Sx & Px), Pa & Ra, Qb & Sb}?

Independence holds if we can construct two interpretations: one of which is a model of S U {s} and the other of which is a model of S U {¬s}.

Interpretation (a):

Domain: {1,2,3}

P: {1,2,3}

Q: {2,3}

R: {1,2}

S: {2}

a: 1

b: 2

This is a model of S U {s}:

∃x(Px & Qx & Rx), is true because P2 & Q2 & R2 is true; '∃x(Px & Qx & ¬Rx)' is true because P3 & Q3 & ¬R3 is true; S2& P2 is true which means that∃x(Sx & Px) is true; and Pa & Ra and Qb & Sb are both true given the interpretation of a and b; finally s is true since all those things which are in both **P** and **Q** (*viz.* 2 and 3) are either in **S**(in the case of 2) or not in **R** (in the case of 3)*.

*Rather than saying 3 is not in **R**, we could define **¬R**as the complement of **R**– that is, the set of all objects in the domain which are not in **R**, in this case **¬R**={3} and of course 3 is in **¬R** since 3 is in {3}.

Interpretation (b):

Domain: {1,2,3}

P: {1,2,3}

Q: {2,3}

R: {1,3}

S: {2}

a: 1

b: 2

This is a model of S U {¬s}:

We now need to make s false: this means there must be at least one thing which is both in **P**and **Q**but is not either in **S** or in **¬R** (i.e. the complement of **R**– the set of all things in the domain but not in **R**). I have made this true of 3: 3 is in **P** and **Q**but it is neither in **S** nor in **¬R** (since it is in **R**). The sentences inS, however, are all true again. ∃x(Px & Qx & Rx) holds because P3 & Q3 & R3 is true; ∃x(Px & Qx & ¬Rx) holds because P2 & Q2 &

¬R2 is true; ∃x(Sx & Px) holds because S2 & P2 is true; Pa & Qa holds because P1 & R1 and Qb & Sb holds because Q2 & S2.

Interpretations (a) and (b) together demonstrate that s is independent of S. s can be either true or false consistently with all of the sentences in S being true.

The finite model technique allows us to be much more systematic in the search for models of various sets of sentences. However, for the full predicate calculus (to be introduced later) the technique is incomplete. That is, it can be shown that **not every set of sentences which has a model has a *finite* model**. The method works only one way: IF we can find a finite model, THEN the set of sentences is consistent; but the set may be consistent without there being a finite model of it. This will be the case iff there are consistent sets of sentences which **only have infinite models** – and there are. But all of the cases that you will be asked to deal with in the exercises can be worked using finite interpretations/models.

C6: Demonstrating Validity (Monadic Predicate Calculus)

So we now know how to use interpretations – both infinite and finite – to establish **invalidity** of inference, **consistency** of a set of sentences and independence of a single sentence from a set of sentences. We can't use them, however, to show ***VALIDITY*** of inference, **inconsistency** and **dependence** (that is, lack of independence). (Make sure that you fully understand why not.) We therefore resort to a different idea to tackle these problems. The idea is that of a **FORMAL PROOF**. The validity of an inference, for example, will be established by deriving its conclusion from its premises using certain permitted ***RULES OF PROOF***.

Let's plunge straight in by giving a formal derivation, even though it won't make much sense initially, and then analyse it so that it does make sense. And let's stick to our time-worn examples.

Example 1:

1. All men are mortal
2. Socrates is a man

Therefore, Socrates is mortal

This, as we know, formalises as:

1. $\forall x(Px \rightarrow Qx)$
2. Pa

Therefore, Qa

The following is a formal proof of its conclusion from its premises (and therefore a demonstration of the validity of the inference).

1.	$\forall x(Px \rightarrow Qx)$	Premise
2.	Pa	Premise
3.	$Pa \rightarrow Qa$	US, 1

4.	Qa	TI, 2, 3

A proof is an ordered, numbered, sequence of formulas (I shall say precisely what a formula is later, for now a vague idea is enough). Each formula in the proof requires a *justification* – written on the right. There are only two permitted types of justification (1) thatthat line has been given as a **Premise** or (2) that that line has been ***derived from one ormore previous lines in the proof using one a small list of permitted rules of inference***(inwhich case the justification is represented by the name of the rule of proof and the number(s) of the previous lines in the proof to which that rule has been applied). So in the above very simple proof, steps 1 and 2 are justified in that both are premises (and we are of course allowed to write down a premise at any stage – the premises are "given"); step 3 applies a rule of proof (called Universal Specification and abbreviated "US") to line 1; step 4 applies another rule of proof (called the rule of Tautological Implication – "TI") to lines 2 and 3. Since line 4 is the required conclusion, this very simple proof is complete and has established the validity of the inference.

Example 2:

1. All Greeks are men $\forall x\,(Px \rightarrow Qx)$
2. All men are mortal $\forall x\,(Qx \rightarrow Rx)$

So, All Greeks are mortal $\forall x(Px \rightarrow Rx)$

Proof:

1.	$\forall x(Px \rightarrow Qx)$	Premise
2.	$\forall x\,(Qx \rightarrow Rx)$	Premise
3.	$Px \rightarrow Qx$	US, 1
4.	$Qx \rightarrow Rx$	US, 2
5.	$Px \rightarrow Rx$	TI, 3, 4
6.	$\forall x\,(Px \rightarrow Rx)$	UG, 5

As before (and as always in a formal proof), the justification for each step is given on the right; so lines 1 and 2 are justified by the fact that they are given to us as premises, steps 3 and 4 are both application of the rule 'US' (to lines 1 and 2 respectively) step 5 uses TI (tautological implication) on lines 3 and 4, and finally the rule applied at line 6 and based on line 5 is called the 'rule of universal generalisation' (UG).

This second example indicates the form of very many proofs in predicate logic: we use rules (like US) to drop quantifiers (lines 3 and 4), manipulate the unquantified formulas essentially truth-functionally via the single but (as we will see multi-faceted) rule of tautological implication; and then apply generalisation rules, as at step 6 – in this case 'universal generalisation' (UG) – to restore the quantifier(s).

Of course the required rules are not just any old rules. We clearly want them to have the property that ***whenever we properly apply them to derive a particularconclusion from some premises then the inference from those premises to that conclusion is valid*** (in the sense that we have specified – no interpretation in whichthe premises are true and the conclusion false). Any set of rules of proof which has this property is called a **SOUND set of rules**. We would also like the rules of proof to have the further property (the converse of the one just stated) that ***whenever aninference from some premises to a conclusion is in fact valid then there is a formal proof of the conclusion from the premises using the specified rules of proof***. Any set ofrules which has this property is called a **COMPLETE set of rules**. The rules of proof that I shall introduce are indeed ***both*** **sound** ***and*** **complete** – though for the purposes of this course, we shall take this for granted rather than proving it (the proof is quite complex).

C6(A): THE RULE OF TAUTOLOGICAL IMPLICATION (TI)

Let's start with the ***RULE OF TAUTOLOGICAL IMPLICATION (TI).*** One formula, **G**, is **tautologically implied** by some number of other formulas **F_1 ... F_n** iff the single formula **‘(F_1& ... & F_n) → G’** is a tautology (in the sense specified in truth-functional logic). So, you pretend that the formulas involved (again I'll say more precisely what a formula is soon) are truth-functional atoms and then apply the above test.

So, consider step 4 in ***Example 1*** above: there we use TI to justify the step from 'Pa' and 'Pa → Qa' to 'Qa', and this is a correct application of the rule since the sentence '(Pa & (Pa → Qa)) → Qa' is indeed a *tautology* (that is**,** if we regard Pa and Qa as just truth functional atoms, say p and q, the resulting sentence '(p & (p → q)) → q’ is a truth-functional tautology). This means that Qa *is indeed* tautologically implied by Pa and Pa → Qa. Hence, since the proof already has the steps Pa and Pa → Qa, we are allowed to derive Qa from lines 2 and 3 of that proof by TI.

Similarly, in ***Example 2***, the justification of the step from lines 3 and 4 to Px → Rx as line 5 is that the formula ((Px → Qx) & (Qx → Rx)) → (Px & Rx) is a tautology: (again pretend that the Px, Qx and Rx are just truth functional atoms p, q and r, then ((p →q) & (q → r)) → (p → r) is indeed a tautology.) (Just what funny formulas like 'Px → Rx’ mean we shall consider later.)

Although the rule of tautological implication is all we need for these truth-functional manipulations (it's a sort of single grand rule covering all cases), certain special cases of it are used more frequently than others and correspond to certain classical logical rules with established names. Some students prefer to remember the special cases as well as the general rule. Here are some of them (the capital letters **F**, **G**, etc. indicate ***any*** formulas of predicate logic):

Rule of *Modus Ponens*:

The formula **G** can be inferred from the formulas **F** and **F** → **G**: so, for example, Qa follows from Pa and Pa → Qa and (Pa & (Pa→Qa)) → Qa and (Px & (Px → Qx)) → Qx are tautologies; since (p & (p → q)) → q is atautology.

Rule of *Modus Tollens*

The formula **¬F** can be inferred from the formulas **F** → **G** and **¬G.** So, e.g., ¬Pa can be inferred from Pa → Qa and ¬Qa (again, ((Pa → Qa) &¬Qa) → ¬Pa is a tautology).

Rule of Hypothetical Syllogism

The formula **F** → **H** can be inferred from the formulas **F** → **G** and **G** → **H**. So, e.g., Px → Rx is derivable from Px → Qx and Qx → Rx.

Rule of Simplification

*Both*the formula **F** *and*the formula **G** can be inferredfrom the formula **F** &**G**. So, e.g., Pa can be inferred from Pa & Qa, and so can Qa; (Px → Qx) could be inferred from (Px → Qx) & Rx, and so could Rx.

Rule of Disjunctive Syllogism

The formula **F** can be inferred from the formulas **F** v **G** and ¬**G**. So, for example, Pa follows from (Pa v Qa) and ¬Qa.

(***Important Exercise***: Show that each of these rules of proof is a special case of the rule of tautological implication.)

One further point to notice is that the rule of tautological implication applies equally well to formulas that are fully-fledged quantified sentences. Consider, for example, the inference:

1. If all LSE students are hardworking then some pigs can fly.
2. No pigs can fly.

So, Not all LSE students are hardworking.

The inference can already be shown to be valid in truth-functional logic – taking p as 'All LSE students are hardworking' and q as 'Some pigs can fly' we have:

1. p → q
2. ¬q

So, ¬p

However, predicate logic being an extension of, or an elaboration on, truth-functional logic, we could certainly also express the inference in predicate logic and, with our new idea of a proof, also establish its validity there. In predicate logic the inference would be formalised as:

1. $(\forall x(Px \rightarrow Qx)) \rightarrow (\exists x(Rx \,\&\, Sx))$
2. $\neg\exists x(Rx \,\&\, Sx)$

So, $\neg(\forall x(Px \rightarrow Qx))$

Its validity is very simply established by the following proof:

1.	$\forall x(Px \rightarrow Qx) \rightarrow \exists x(Rx \,\&\, Sx)$	Premise
2.	$\neg\exists x(Rx \,\&\, Sx)$	Premise
3.	$\neg\forall x(Px \rightarrow Qx)$	TI, 1,2

(The form of TI being here, of course, *Modus Tollens*: $((\forall x\ (Px \rightarrow Qx) \rightarrow \exists x\ (Rx \,\&\, Sx)) \,\&\, \neg \exists x\ (Rx \,\&\, Sx)) \rightarrow \neg\forall x(Px \rightarrow Qx)$ is a tautology – because if we replace the constituent sentences ($\forall x\ (Px \rightarrow Qx)$ and $\exists x\ (Rx \,\&\, Sx)$) in this formula by the atoms p, and q (while still retaining the truth functional structure) then we get $(p\rightarrow q) \,\&\, \neg q) \rightarrow \neg p$ which is a tautology (*Check!).*

So the basic idea for most proofs in predicate logic is going to be: start with premises, drop the quantifiers using the appropriate rules, manipulate the resulting quantifier-free formulas using the rule of tautological implication (or – equivalently – the appropriate one of its special forms, like *Modus Ponens*), then replace the quantifiers (if necessary – that is, assuming that theconclusion is a quantified sentence) using the appropriate generalisation rules. There are basically four extra rules, then, that we need now to introduce: one for dropping universal quantifiers, one for putting them back, one for dropping existential quantifiers and one for putting them back.

C6(b): The Rule of Universal Specification (US)

The rule for dropping universal quantifiers is called the rule of **UNIVERSAL SPECIFICATION (US).** The simple basic underlying idea is that***if some property holds of every individual then it must hold of any individual in particular***.

So if it is true for anything at all that if it's a man then it's mortal, then it follows that if Socrates is a man then he's mortal, *i.e.* assuming *a* is our individual constant for Socrates, we can infer $Pa \rightarrow Qa$ from $\forall x(Px \rightarrow Qx)$ by universal specification. We could equally well infer $Pb \rightarrow Qb$, $Pc \rightarrow Qc$ *etc*. for any individual named by an individual constant.

That is, one type of application of the rule of US is to just the step from $\forall x\mathbf{F}$ to $\mathbf{F}[a_j|x]$, where $\mathbf{F}[a_j|x]$ means the formula obtained from **F** by replacing alloccurrences of x by the individual constant a_j. (So if, e.g., **F** is $Px \rightarrow (Qx \vee Rx)$, $\mathbf{F}[a|x]$ is $Pa \rightarrow (Qa \vee Ra)$; $\mathbf{F}[c|x]$ is $Pc \rightarrow (Qc \vee Rc)$ *etc.*)

However, we also often need, as in **Example 2** above, to drop a universal quantifier – but without wishing to infer anything about any particular individual (in that inference we are just talking about Greeks, men, and mortals in general: no particular individual is mentioned). Instead we go from $\forall x(Px \rightarrow Qx)$, say, to $Px \rightarrow Qx$. This latter formula is best interpreted as saying ***of any arbitrary, unspecified, but single, individual*** if it's a P then it's a Q. It is indeed clearly valid to infer from the assertion that 'All triangles have internal angles that add up to 180°' that: 'If any arbitrary object is a triangle then its internal angles add up to 180º'. This would be the step from $\forall x(Px \rightarrow Qx)$ to $Px \rightarrow Qx$ (or indeed to $Py \rightarrow Qy$ or $Pz \rightarrow Qz$) by US (any individual variable could be used to name an arbitrary object).

So, thinking about it purely formally or syntactically (and this is the ***BEST WAY*** to think about these rules until you have fully got the hang of them), the rule of US says:

> *You can take any formula with a universal quantifier on some variable, say x,* ***at the front*** *(and 'governing' the rest of the formula – that is, the quantification on x extends over the whole formula) and infer from it the formula obtained by dropping the quantifier and, if*

you like, replacing the occurrences of the originally quantified variable x by any variable or by any individual constant.

So, for example, the formulas Pa → Qa, Pb → Qb, Px → Qx, Py → Qy can all be obtained from ∀x(Px → Qx) by US.

C6(c): The Rule of Universal Generalisation (UG)

How about putting universal quantifiers back? This is the role of the rule of **UNIVERSAL GENERALISATION**. Clearly it would be quite*wrong*to infer fromthe fact that 'Socrates is a man' that 'Everything is a man', or from the fact that ‘Boris Johnson is a liar’ that ‘everyone is a liar’. Nor, e.g., should we infer from ‘If Boris Johnson is a liar, then he is unelectable’, that ‘Everyone who is a liar is unelectable’ (it might just apply to Boris and some others but not to everyone.) So we do **NOT** allow the inference from Pa to ∀xPx or from (Pa → Qa) to ∀x(Px → Qx).

However, it is standard (especially in mathematics) to reason using an 'arbitrary' object – an arbitrary triangle or an arbitrary prime number say – *establish* some result about this arbitrary object and then infer that the resultholds generally (of *all* triangles or *all* prime numbers). Thus we ***do*** allow (subject to some restrictions considered later) the inference from Px to ∀xPx or from Px → Qx to ∀x(Px → Qx), *etc*. This rule is called universal generalisation (**UG**). We will give a precise formulation of it shortly, but basically the idea is:

If you have a formula of the form ***F(x)*** *involving some unquantified variable x (we shall later call these “free variables”) then you may infer the formula ∀x**F(x)**, in which that initially unquantified variable is universally quantified over.*

Because the ‘x’ is genuinely arbitrary (that is you suppose, in the geometry case, simply that the object you are considering is a triangle, and you do not suppose that it has any particular other properties – say that of being equilateral – it is intuitively ok to generalize.

C6(D): SOME SIMPLE DERIVATIONS:

Using these three rules (TI, US, UG) we can demonstrate the validity of a wide range of valid inferences. We already have seen a couple of these (1 and 2 above, you should go over them again now); and here are a couple more.

Example 3:

1. All philosophers are either empiricists or rationalists.
2. No rationalist knows science.

So, Any philosopher who knows science is an empiricist.

This formalises as:

1. $\forall x\ (Px \rightarrow (Qx \vee Rx))$
2. $\forall x(Rx \rightarrow \neg Sx)$

So, $\forall x((Px \ \& \ Sx) \rightarrow Qx)$

Here, Px: x is a Philosopher; Qx: x is an empiricist; Rx: x is a rationalist and Sx: x knows science. Notice that the conclusion might also have been formalised as ($\forall x(Px \rightarrow (Sx \rightarrow Qx))$. This should seem intuitively clear and the intuition is underwritten by the fact that this second formalisation of the conclusion is logically equivalent to the first. (Basically because $(p \ \& \ q) \rightarrow r$ is tautologically equivalent to: $p \rightarrow (q \rightarrow r)$. *Exercise*: check this.)

Here is proof of validity:

1.	$\forall x(Px \rightarrow (Qx \vee Rx))$	premise
2.	$\forall x(Rx \rightarrow \neg Sx)$	premise
3.	$Px \rightarrow (Qx \vee Rx)$	US, 1
4.	$Rx \rightarrow \neg Sx$	US, 2
5.	$(Px \ \& \ Sx) \rightarrow ((Qx \vee Rx) \ \& \ \neg Rx)$	TI, 3, 4*

6. $(P\ x \ \&\ Sx) \rightarrow Qx$ TI, 5*

7. $\forall x((Px \ \&\ Sx) \rightarrow Qx)$ UG, 6

*Think carefully about the – quite complicated – tautologies involved in these twosteps. And – *exercise* – check that the tautologies involved are indeed tautologies.

***(Further important exercise**: I could also have collapsed steps 5 and 6 into one step – using of course a still more complicated application of TI. What tautology would be involved in that single step? This is a general feature of predicate logic proofs – any series of successive applications of TI could be replaced by a single application of that rule: how far you breakdown applications of TI into intermediate steps is always a question of taste and of which tautologies you intuitively "see" as tautologies.)

Example 4:

1. All leopards are spotted and all tigers are striped.

So, Anything that is either a leopard or a tiger is either spotted or striped.

That is:

1. $\forall x\ (Px \rightarrow Qx)\ \&\ \forall x(Rx \rightarrow Sx)$

So, $\forall x\ ((Px \vee Rx) \rightarrow (Qx \vee Sx))$

Proof:

1. $\forall x\ (Px \rightarrow Qx)\ \&\ \forall x(Rx \rightarrow Sx)$ Premise
2. $\forall x\ (Px \rightarrow Qx)$ TI, 1*
3. $\forall x(Rx \rightarrow Sx)$ TI, 1
4. $Px \rightarrow Qx$ US, 2
5. $Rx \rightarrow Sx$ US, 3
6. $(Px \vee Rx) \rightarrow (Qx \vee Sx)$ TI, 4, 5
7. $\forall x((Px \vee Rx) \rightarrow (Qx \vee Sx))$ UG, 6

*It would be wrong to go straight from 1 to, say, (Px → Qx) & (Rx → Sx). US only allows you to go from a formula with a quantifier *IN FRONT* which *governs the whole rest of the sentence* to that formula with the quantifier dropped – so you need to 'detach' the two conjuncts in 1 first – at steps 2 and 3 (by TI) – and only then apply US to them separately.

C6(E): THE RULE OF EXISTENTIAL SPECIFICATION (ES)

So far we have dealt only with universally quantified formulas. How about the existentially quantified ones? If we are going to drop quantifiers in this case too (we are!) then we need to be careful. Clearly it would be wrong to infer from the statement 'Some natural numbers are prime' that some *particular* number is prime (that number *might* actually be prime but the inference would still clearly be invalid). So we can't infer, for example, from ∃xPx to Pa. ('Some numbers are prime. So 4 is a number that is prime' is a counterexample.) Similarly, it would be clearly mistaken to infer from 'Some triangles are isosceles' or from 'Some lectures are interesting' to 'An arbitrary triangle is isosceles' or 'An arbitrary lecture is interesting'. That is, the inference from ∃xPx to Px is also NOT sanctioned.

What we **ARE** entitled to infer from 'Some triangles are isosceles', for example, is, given the minimal interpretation of 'some' that we have adopted, simply that there is at least one isosceles triangle. Generally, from ∃xPx we are entitled to infer only that there is at least one (possibly unknown) individual with property P – we cannot say that any particular object has P, only that *some* particular object does (even though we may not be able to name it). The idea, then, behind the rule of **EXISTENTIAL SPECIFICATION (ES)** is to introduce a new sort of name – a so-called ***ambiguous name.*** As 'ambiguous names' we use letters from the beginning of the Greek alphabet:

α, β, γ, *etc*., or (α1, α2, ... if we need lots).

If, for instance, we know ∃xPx, then we know that at least one individual has property P and we can "ambiguously christen" that individual 'α'. All we know about α is that it's a particular individual that has property P.

So the rule of ***existential specification*** states:

The formula **F**[αIx] can be inferred from the formula ∃x**F**x (where **F**[αIx] is the formula you obtain from **F** by substituting α for x, so you can read it as 'F with α for x').

So, for example, we can infer Pα & Qα from ∃x(Px & Qx), and Rα v Sα from ∃x(Rx v Sx).

C6(F): The Rule of Existential Generalisation (EG)

The rule for putting existential quantifiers back – the **rule of EXISTENTIAL GENERALISATION (EG)** – is intuitively more straightforward. Although, as we just noted, we can't infer '5 is a prime number' from the fact that 'There are some prime numbers' (even though it is true), it certainly does work the other way round: that is, it does follow from the fact that '5 is a prime number' that 'There are prime numbers' (i.e. on our understanding that 'there is at least one prime number'); it also clearly follows from the fact that some ambiguously named entity α has property P that there is at leastone P; finally it follows from the fact that any arbitrary object has property P that thereare some Ps. That is, Pα entails ∃xPx, Pa entails ∃xPx and Px entails ∃xPx.

(More formally):

Rule of EG: *if **G** is a formula that results from a formula **F** by at most replacing either an ambiguous name or an individual constant by a variable x then ∃x**G** can be inferred from **F**.*

So, for example, ∃x(Px & Qx) can be derived from (Pα & Qα) or from (Px & Qx); ∃x((Px v Qx) & Sx) can be derived from (Pa v Qa) & Sa; and so on.

C6(g): Some More Derivations

The only way to get straight about these rules is to get used to employing them to make valid derivations. So here are a few more examples.

Example 5:

1. Everything Mozart wrote is beautiful.
2. Some of the things Mozart wrote are operas.

So, Some operas are beautiful.

The premises formalise as 1 and 2 below (with Px meaning 'x was written by Mozart', Qx meaning 'x is beautiful' and Rx meaning 'x is an opera') and then the proof takes us to $\exists x\ (Rx \ \&\ Qx)$ as the required conclusion:

1.	$\forall x(Px \to Qx)$	Premise 1
2.	$\exists x(Px \ \&\ Rx)$	Premise 2
3.	$P\alpha \ \&\ R\alpha$	ES, 2
4.	$P\alpha \to Q\alpha$	US, 1*
5.	$P\alpha$	TI, 3
6.	$Q\alpha$	TI, 4,5
7.	$R\alpha$	TI, 3
8.	$R\alpha \ \&\ Q\alpha$	TI, 6,7**
9.	$\exists x(Rx \ \&\ Qx)$	EG, 8

*There is obviously no reason to exclude ambiguously named entities from the range of the US rule. If everything has a certain property, then any ambiguously named entity certainly has it. We will note this formally below.

**The tautological implication used in line 8 (sometimes called the rule of conjunction) is $p \to (q \to (p \ \&\ q))$. As before, I have broken down the TI's into "simple" steps, but we could equally well have gone straight from lines 3 and 4 to $R\alpha\&\ Q\ \alpha$ – the tautology involved then being $((p \ \&\ r) \ \&\ (p \to q)) \to (r \ \&\ q)$.

Example 6:

1. Every war is terrible, but some wars are justified and anything that is justified is morally acceptable.

So, some terrible things are morally acceptable.

Formalisation:

1. $\forall x(Px \to Qx) \,\&\, \exists x(Px \,\&\, Rx) \,\&\, \forall x(Rx \to Sx)$

So, $\exists x(Qx \,\&\, Sx)$

Proof of validity:

1. $\forall x(Px \to Qx) \,\&\, \exists x(Px \,\&\, Rx) \,\&\, \forall x(Rx \to Sx)$	Premise
2. $\forall x(Px \to Qx)$	TI, 1
3. $\exists x(Px \,\&\, Rx)$	TI, 1
4. $\forall x(Rx \to Sx)$	TI, 1
5. $P\alpha \,\&\, R\alpha$	ES, 3*
6. $P\alpha$	TI, 5
7. $P\alpha \to Q\alpha$	US, 2
8. $Q\alpha$	TI, 6,7
9. $R\alpha$	TI, 5
10. $R\alpha \to S\alpha$	US, 4
11. $S\alpha$	TI, 9,10
12. $Q\alpha \,\&\, S\alpha$	TI, 8,11
13. $\exists x(Qx \,\&\, Sx)$	EG, 12

*Notice again that it would be wrong to start dropping quantifiers directly from line 1 (i.e. from the single premise in this inference). The specification rules only tell you that you can drop a quantifier in a certain way IF it is a quantifier governing the ***whole*** of the rest of the formula – i.e. there is a bracket directly after that quantifier and then the corresponding closing bracket comes at the very end of the whole formula. This is ***not*** true of any quantifier in 1 (in particular it is not true of the first one '$\forall$x'); but it is true of the quantifiers in lines 2, 3 and 4.

C6(H): THE NEED FOR RESTRICTIONS ON THE RULES SO FAR FORMULATED

Most of the time, intuition will guide you to employ these rules of proof in such a way that they sanction only genuinely valid inferences. However, they are faulty as they stand and need a few restrictions so that they do the job properly (and so that intuition can be eliminated entirely and a genuine rock solid foundation can be found for our logic).

Here, for example, is a very clear cut example of an ***INVALID*** inference which indicates the need for an obvious restriction on the rule ES:

1. Some numbers are odd
2. Some numbers are even

So, some numbers are both odd and even

This inference has true premises and a false conclusion – and so is a counterexample to itself. It formalises as:

1. ∃x(Px & Qx)
2. ∃x (Px & Rx)

So, ∃x (Px & Qx & Rx)

Here Px means x is a number, Qx means x is odd and Rx means x is even. Since the inference is invalid, it should, of course, have no proof. But what, then, is wrong with the following?

"Proof":

1. ∃x(Px & Qx)	Premise 1
2. ∃x(Px & Rx)	Premise 2
3. Pα& Qα	ES, 1
4. Pα& Rα	ES, 2

5. $P\alpha \& Q\alpha \& R\alpha$ TI 3,4

6. $\exists x(Px \& Qx \& Rx)$ EG, 5

Steps 5 and 6 (as well of course as 1 and 2) are perfectly OK. The problem occurs at step 4 – but ONLY *given that step 3 has already been made.* Certainly premise 1 guarantees the existence of at least one thing that has both the property P and the property Q: we can call it α. Similarly premise 2 guarantees the existence of at least one thing which has both property P and property R – we can, however, by no means guarantee that the thing which has properties P and R *is the same thing* as that which has properties P and Q. We should *not* therefore use the ambiguous name again.

So we clearly need in general to impose a ***restriction*on ES** – namely that ***whenever ES is applied, a NEW ambiguous name must be used.*** (Here “new” means ‘an ambiguous name that has not already been used in the proof that we are providing.’)

So the correct step at line 4 would be$P\beta$ & $R\beta$(or $P\gamma$& $R\gamma$or whatever – that is, anyambiguous name aside from α), and hence the invalid inference would be blocked. (Once this restriction is in place, we could not then legitimately use EG to get from $P\alpha$& $Q\alpha$ and $P\beta$ & $R\beta$to $\exists x(Px \& Qx \& Rx)$, since the EG rule requires that the introduced variable replaces the *same* ambiguous name throughout.)

All the rules we have introduced need further tightening to be fully logically acceptable. However, it is best first to get a good idea of what is involved by using them – most of the time the need for restrictions will not arise. We have seen how to use them to establish validity of some inferences. What other purposes can they serve?

C7: LOGICAL TRUTH AND LOGICAL FALSEHOOD

C7(A): THE IDEA OF LOGICAL TRUTH

One important notion in truth-functional logic for which we have not yet produced an analogue in the case of predicate logic is that of a *tautology.* A truth-functional tautology, remember, was a compound sentence which turned out *true no matter whattruth values were assigned to its atomic components.* This meant that, like 'Either it isFriday or it is not' it was **triviallytrue** – true because uninformative. Consider a sentence like 'Everything is either physical or not-physical'. This is *not* a tautology (it is truth functionally simple and so would formalise just as p) but, unlike 'Everything is physical', which, if true, is informative (it rules out, for example, the existence of immaterial 'souls'), it is bound to be true because uninformative. It formalises in predicate logic as ∀x(Px v ¬Px). (The truth functionally disjunctive ∀xPx v ∀x¬Px is a quite different sentence from the one we are considering. *Exercise*: What does that disjunction say? Is it also trivial?)

The reason why the original sentence 'Everything is either physical or not-physical' is *bound tobe true is* easily seen if we consider its formalisation. No matter what we take thevariable to range over (implicitly material objects in the original sentence, but it could be numbers, people, animals or whatever) and no matter which property we take for the predicate Px (it could be 'x is even', 'x is male', 'x is feline' or whatever) we would always get a true sentence. ('All numbers are either even or not-even', 'All animals are either feline or not-feline, *etc*.)

We shall use the term ***logical truth*** for the generalisation to first order logic of the truth functional notion of tautology. And the above consideration gives us our characterisation of the notion:

***Definition:* Logical Truth**

*A single sentence s in the language of first order predicate logic is a **logical truth** iff it is **true in all interpretations** (i.e. it has only models).*

C7(B): DEMONSTRATING LOGICAL TRUTH

Just as in the case of our definition of validity, we can readily demonstrate that a sentence is **NOT** a logical truth. This simply requires us to produce ***a single interpretation in which the sentence is false***.

For example, the sentence 'All swans are either white or black', although true in the actual world, is *not logically true*. This is why it gives (limited but) genuine information about the world. The sentence can be formalised as:

$\forall x\ (Px \rightarrow (Qx \vee Rx))$

(where Px: x is a swan, Qx: x is white, and Rx: x is black).

The following interpretation is *not a* model, and hence shows that the sentence is *not* a logical truth:

Domain: {Animals}

Px: x is human

Qx: x is English

Rx: x is Austrian

Under this interpretation the sentence reads 'All humans are either English or Austrian', which is (thankfully) not true.

So: to demonstrate that a single sentence s in the language of predicate logic is NOT a logical truth, we simply need to produce **one single interpretation** under which **it is false.**

On the other hand, the sentence 'All swans are either white or not white' *is* logically true. But clearly we can't *show* this directly from our definition – this would require checking ***all possible*** interpretations and the set of all possible interpretations is ***infinite***.

So just as with invalidity and validity, we have an asymmetry: to show that an inference is **in**valid or that a single sentence is **not** a logical truth, we need only produce one interpretation – a counterexample to show the invalidity of an inference and an interpretation which is not a model of the sentence at issue to show that that sentence is not a logical truth. But trying to use interpretations to show either validity or logical truth would be an unachievable, because infinite, task.

This is more than an analogy as we will immediately see: the same strategy that solved the validity problem – namely the notion of a formal proof – **also solves the logical truth problem.**

First notice that, just as in truth functional logic, whenever we demonstrate the validity of an inference in predicate logic we thereby establish that a particular sentence is logically true. This sentence is the ***"associated conditional".***

If the inference from premises P_1..., P_n to conclusion C is valid, then there is no counter-example to it, i.e. no interpretation in which *all* of the premises are true and C is false. But this means, because of the truth-table definition of '→', that the single sentence:

$(P1\&... \& P_n) \to C$

must be a logical truth, on our definition. (*Exercise*: think this through carefully.) So, for (the old boring) example, since 'Socrates is mortal' follows from 'All men are mortal' and 'Socrates is a man', then single sentence: 'If all men are mortal and Socrates is a man then Socrates is mortal', which formalises as $((\forall x(Px \to Qx) \& Pa) \to Qa)$, must be a logical truth.

Moreover, *if* a conclusion C can be derived from a single premise P that is itself a truth-functionaltautology (and remember we do have a decision procedure for tautologousness) then C itself must be logically true. Why? Well, first if C can be validly inferred from P then the sentence 'P → C' must be logically true (via the considerations in the preceding paragraph). Next, if C were *not* itself logically true there would be at least one interpretation, say **I**, in which it is false; but since P is a tautology it must be true in *all* interpretations and so in particular true in **I**, but then 'P → C' must be false in **I** (true antecedent, false consequent) contrary to our assumption that 'P → C' is a logical truth.

This means that C must be a logical truth if it can be proved (using the rules of proof) from a tautological premise. But that would be effectively ***from no premises at all***– because the rule of tautological implication in any case allows us to write down any truth functional tautology at any stage (any tautology is tautologically implied by *any* sentence). This gives us the following result which allows us to demonstrate logical truth:

Result: A sentence of first order logic is a logical truth iff itis***provable from nopremises.***

But that sounds weird. How could a conclusion possibly be provable from no premises? The answer is 'pretty easily'. Such proofs invariably involve a very useful additional rule of proof that we have not yet introduced. (I won't pull this stunt again; this is the last rule we will need). This further rule is:

The rule of conditional proof:

If the formula **G** can be derived from a set of premises {P_i} together with the formula **F,** then the formula **F** → **G** can be derived just from the premises {P_i}.

We shall later need to make our system more sophisticated in order to accommodate fully this rule, but for the moment we will operate with it in an intuitive way. The rule may sound a little unintuitive initially, but in fact, once you think about it, it is just commonsense (as it needs to be given that it is a rule of logic). Suppose, for example, we are considering the Treasury Model of the UK Economy – basically a collection of economic theories and facts. The Bank of England is considering increasing the base rate of lending by 1% – this has not been decided, so it stands as a possible assumption. We would then be interested in what the Treasury Model predicts *would* happen *if* the Bank took that decision. So, we add the assumption that the Bank will increase the rate by 1% to the 'premises' already embodied in the model. Suppose we can deduce, for example, that 'Inflation will increase by 2%'. So, we have deduced that inflation will rise by 2% from the Treasury Model *plus* the extra assumption that the Bank raises the lending rate by 1%. But this is obviously equivalent to (is just another way of saying) that the Treasury Model itself (that is without any further assumption) entails the conditional 'If the Bank raises the interest rate by 1%, then inflation will rise by 2%'. The rule of conditional proof is just a formal statement of that obvious equivalence.

We can recognize this by seeing the rule of conditional proof (CP) in action. Say we want to show that 'All intellectuals are a drain on society' follows from 'All intellectuals are economically unproductive', and 'All economically unproductive people are a drain on society.' (Margaret Thatcher came close to believing that all three statements were true!)

Formalising, with Px, Qx, Rx having the obvious meanings, we obtain:

1. $\forall x(Px \to Qx)$
2. $\forall x(Qx \to Rx)$

So, $\forall x(Px \to Rx)$

And we can construct the following proof:

Proof:

1.	$\forall x(Px \to Qx)$	Premise
2.	$\forall x(Qx \to Rx)$	Premise
3.	$Px \to Qx$	US, 1
4.	$Qx \to Rx$	US, 2
5.	Px	Assumption (A)
6.	Qx	TI, 3,5 (Modus ponens)
7.	Rx	TI, 4,6 (Modus ponens)
8.	$Px \to Rx$	CP, 5-7
9.	$\forall x(Px \to Rx)$	UG, 8

What we have shown in this proof by line 7 is that the formula 'Rx' follows from the original premises *plus the extra formula* Px (introduced "by assumption" at line 5). The rule of conditional proof then tells us that we could have inferred the conditional '$Px \to Rx$' just from the original premises – and this is how we applied the rule at line 8. This should seem like an intuitively sound way of proceeding.

Of course, we already know that we can establish the validity of this inference *without* invoking CP (we could just go straight from lines 3 and 4 to line 8 by TI –in fact by the rule of 'hypothetical syllogism': the relevant tautology being $((p\to q) \& (q\to r)) \to (p\to r)$. (*Usual exercise.*) The advantage is that CP has allowed us to break down the truth-functional steps in the proof to *very* simple ones – namely two applications of *modus ponens*. In other cases, the rule of CP is essential.

It is of particular value in establishing logical truth, as the following examples show:

Example 1:

Show that the sentence: 'If Newton was a genius then there are some geniuses' is a logical truth. The sentence formalises as Pa → ∃xPx.

Proof:

1.	Pa	Assumption (hereafter just A)
2.	∃xPx	EG, 1
3.	Pa → ∃xPx	CP, 1-2

So Pa → ∃xPx has indeed been proved from no premises. Make sure you fully understand that there are no premises in the above proof. There was – at line 1 – an *assumption*, but that assumption was 'discharged' at line 3 when we applied CP. Notice, then, the crucial difference between a premise and an assumption. A premise is a statement that is given for the purposes of deducing some conclusion from it (together with whatever other premises are involved, that is, are also given). The premises remain 'given' throughout. An assumption is a claim that you – *temporarily* – "give yourself" for the purposes of producing a proof; and it disappears, so to speak, when discharged.

Example 2:

Show that the sentence: 'If Newton was a genius then not everyone is not a genius' is a logical truth. The sentence formalizes as: Pa → ¬∀x¬Px.

Proof:

1.	∀x¬Px	A
2.	¬Pa	US, 1
3.	∀x¬Px → ¬Pa	CP, 1-2
4.	Pa → ¬∀x¬Px	TI, 3

Notice, then, that we have here proved a sentence of the form 'F → G' *not* by assuming F, proving G and then using CP to establish F → G, but *instead by* assuming ¬G and proving $\neg F$. This gives $\neg G \rightarrow \neg F$ by CP, but $\neg G \rightarrow \neg F$ tautologically implies $F \rightarrow G$ (since $(\neg q \rightarrow \neg p) \rightarrow (p \rightarrow q)$ is a tautology). (*Exercise:* show that this is true.)

Example 3 (a bit trickier):

If some lectures are interesting, then not all lectures are uninteresting.

Formalisation: $(\exists x(Px \,\&\, Qx)) \rightarrow \neg\forall x(Px \rightarrow \neg Qx)$

This sentence is logically true as is again shown by the fact that we can prove it ‘absolutely’ – i.e. without invoking any premises.

Proof:

1.	$\exists x(Px \,\&\, Qx)$	A
2.	$\forall x(Px \rightarrow \neg Qx)$	A
3.	$P\alpha \,\&\, Q\alpha$	ES, 1
4.	$P\alpha \rightarrow \neg Q\alpha$	US, 2
5.	$Q\alpha\& \neg Q\alpha$	TI, 3,4*
6.	$\forall x(Px \rightarrow \neg Qx) \rightarrow Q\alpha \,\&\, \neg Q\alpha$	CP, 2-5
7.	$\neg(\forall x(Px \rightarrow \neg Qx))$	TI, 6
8.	$\exists x(Px \,\&\, Qx) \rightarrow \neg(\forall x(Px \rightarrow \neg Qx))$	CP, 1-7

*The tautology invoked here is $((p \,\&\, q) \,\&\, (p \rightarrow \neg q)) \rightarrow (q \,\&\, \neg q)$, but you may feel happier breaking that single step down into a couple of steps: first getting $Q\alpha$ from 3, then $\neg Q\alpha$ from that plus 4 by Modus Ponens, then putting the two together. Once again, there is no right or wrong about this – it is just a question of which instances of tautological implication you are confident in identifying as genuine, i.e. backed up by genuine tautologies.

This proof illustrates a couple of very important points. **First**, that we can iterate applications of CP – i.e. make more than one assumption (so long as we remember to "discharge" all the assumptions through CP before completing the proof – indeed, the proof is *not* complete until **all assumptions *have been* discharged**). **Secondly,** we can, by using CP, imitate the time-honoured informal proof technique of *reductio ad absurdum.* Our two assumptions (lines 1 and 2) amount to the assumption that the conditional sentence we are out to prove is in fact false. But the assumption that it is false entails a (truth-functional) contradiction (line 5) – hence it is false that it is false, that is, it must be true. The last part of this reasoning is formally captured in lines 6-8. ***Notice inparticular***the very important application of TI at line 7 – this heuristic is used time and timeagain. The particular rule is that:

¬**F** follows from the single sentence **F** → (**G** & ¬**G**) for any formulas **F** and **G**.

(The sentence ((p→ (q & ¬q)) → ¬p) is an important tautology. *Revision exercise:* demonstrate this *both* by constructing the truth table *and* by using the method of semantic trees.)

(*Further Exercise*: Although it is easiest to see how CP can be used to prove that conditionals are logically true, its use is not confined to conditionals (or rather to explicit conditionals – we know from truth functional logic that any truth-functionally compound sentence is equivalent to one that just uses '¬' and '→'). For example, CP can be used to prove the disjunction ∀xPx v ∃x¬Px is a logical truth (as it obviously is). Can you work out how?)

C7(c): The Idea of Logical Falsity

In truth functional logic, we had the notion of a ***Contradiction.*** ‘Today is Friday and it is not the case that today is Friday’ is one such. It is "uninformatively (because necessarily) false" – false "in all possible worlds" (that is, no matter what truth values are assigned to its atomic components).

Similarly, a sentence like ‘Some numbers are both even and not-even’ – though *not* a truth-functional contradiction (notice that it is NOT a truth-functional conjunction but is truth-functionally simple or atomic) *is* necessarily or ***logically false.*** This sentence formalises in first order predicate logic as ∃x(Px & ¬Px) (taking Px to mean ‘is even’ and 'numbers' as implicit – remember that if all of a set of sentences talk about the same things, then this is covered, in any interpretation, by the Domain). This sentence is indeed a ***logical falsehood*** because *every* interpretation makes it false. This is, then, our characterisation of a logical falsehood.

Definition: ***Logical Falsehood:***

A single sentence in the language of first order predicate logic is **logically false** iffsis false in all interpretations (equivalently of course iff ¬sis*true* in all interpretations, i.e. if ¬s is a logical truth).

C7(D): DEMONSTRATING LOGICAL FALSEHOOD

As in the case of logical truth, we can show that a sentence is **NOT** logically false by producing a single interpretation– in this case, a single interpretation in which is not false but instead true. So, for example, the sentence '∃xPx &∃x¬Px' is NOT logically false. This is because it is not false in all intepretations, since it is for instance true in the following interpretation (under which it just says 'Some natural numbers are even and some natural numbers are not even'):

Domain: {Natural Numbers}

Px: x is even

But again, since we cannot inspect all possible interpretations, we cannot establish that a sentence **is** a logical falsehood directly on the basis of this definition.

However, as should be clear from the above definition, and especially the bracketed addition:

Fact 1:

sis logically false iff ¬sis logically true

Hence we can show a sentence s is logically false by showing that its ***negation*** is logically true – and we already know from the last subsection how to prove that a sentence is logically true: we prove it 'absolutely' – that is, without invoking any premises. Formally speaking, we have:

Fact 2:

A sentencesof first order predicate logic is logically false iff its **negation** canbe proved from the empty set of premises.

Example 1:

∃x(Px & ¬Px) is logically false:

Proof:

1.	∃x(Px & ¬Px)	A
2.	(Pα& ¬Pα)	ES, 1
3.	∃x(Px & ¬Px) → (Pα& ¬Pα)	CP, 1-2
4.	¬(∃x(Px & ¬Px))	TI, 3

Study this proof carefully. We are in effect again using the method of *reductio adabsurdum*. We have proved that∃x(Px & ¬Px) is logically false by proving ¬∃x(Px& ¬Px) – but in order to prove *that* we have assumed *its* negation (which of course takes us back to ∃x(Px & ¬Px)) and then derived a contradiction from its negation.

Example 2:

Show that: '∀x(Px → ¬Px) &∃xPx' is logically false.

Proof:

1.	∀x(Px → ¬Px) &∃xPx	A
2.	∃xPx	TI, 1
3.	Pα	ES, 2
4.	∀x(Px → ¬Px)	TI, 1
5.	Pα → ¬Pα	US, 4
6.	¬Pα	TI, 5
7.	Pα& ¬Pα	TI, 3,6
8.	(∀x(Px → ¬Px) &∃xPx) → (Pα& ¬Pα)	CP, 1-7
9.	¬(∀x(Px → ¬Px) &∃xPx)	TI, 8

C8: Demonstrating the Inconsistency of Sets of Predicate Logic Sentences

You will remember that we were able to show that a set of sentences of predicate logic is consistent by producing a ***joint model,*** i.e. a single interpretation in which all the sentences in the set are true. But how can we demonstrate that such a set is inconsistent?

In truth-functional logic, remember, a finite set of sentences is inconsistent iff the conjunction of the sentences is contradictory. So we would expect that in predicate logic a finite set of sentences is inconsistent iff the conjunction of the sentences is logically false.

(Exercise: explain carefully why, on the basis of our characterizations in terms of interpretations and models, this is indeed true – remember that it is an 'if and only if' statement.)

Hence for finite sets of sentences ***we can demonstrate inconsistency by deriving the negation of the conjunction of the sentences.***

Fact:

A finite set of sentences of predicate logic S= {s_1, … s_n} is**inconsistent**iff ***the single sentence s_1& … & s_n is logically false*** – i.e. its negation is derivablefrom the empty set of premises.

Example 1:

The set {$\forall xPx$, $\exists x\neg Px$} is – unsurprisingly – inconsistent since $\forall xPx \;\&\; \exists x\neg Px$ is logically false.

Proof:

1. $\forall xPx \;\&\exists x\neg Px$ A
2. $\forall xPx$ TI,1
3. $\exists x\neg Px$ TI,2
4. $\neg P\alpha$ ES,3

5.	$P\alpha$	US,2
6.	$(P\alpha \& \neg P\alpha)$	TI, 4,5
7.	$(\forall xPx \& \exists x\neg Px) \rightarrow (P\alpha \& \neg P\alpha)$	CP, 1-6
8.	$\neg(\forall xPx \& \exists x\neg Px)$	TI, 7

Another – perhaps more intuitive – characterization can be given of what it means for a set of predicate logic sentences to be inconsistent. Suppose we could derive a truth functional contradiction from a set of sentences as premises – that is, a formula of the form **F**& ¬**F** for some formula **F**. Since this would mean that the contradiction was validly inferrable from the premises, then there could be no counterexample to that inference. But this can *only* mean that *no* interpretation makes all the premises true. (*Exercise*: Explain carefully *why.*) Thus we also have the following:

Result:

A set of sentences $\{s_1, \dots s_n\}$ in the language of predicate logic is **INCONSISTENT** if, taking $s_1, \dots s_n$as premises we can **derive a truth-functional contradiction**.

This means that, using the same example as before, the proof of inconsistency is slightly rejigged and can stop sooner.

Example 1 (again):

The set $\{\forall xPx, \exists x\neg Px\}$ is inconsistent:

1.	$\forall xPx$	Premise*
2.	$\exists x\neg Px$	Premise*
3.	$\neg P\alpha$	ES, 3
4.	$P\alpha$	US, 2
5.	$(P\alpha \& \neg P\alpha)$	TI, 4,5

*Notice that these are now premises, in this way of showing inconsistency, *not* assumptions.

Hence we have derived a truth functional contradiction from the set of sentences as premises – hence that set is inconsistent.

Let's then underline these results by taking a second example.

Example 2:

The setS= {Every politician is either mendacious or incompetent; Noone who is incompetent is capable of running the country; Only politicians are capable of running the country; There are some non-mendacious politicians who are capable of running the country} is an inconsistent set of sentences.

It formalises as:

S = $\{\forall x(Px \rightarrow (Qx \vee Rx); \forall x(Px \rightarrow \neg Sx); \forall x(Sx \rightarrow Px); \exists x\,(Px \,\&\, \neg Qx \,\&\, Sx)\}$

(where: Px: x is a politician, Qx: x is mendacious, Rx: x is incompetent, Sx: x is capable of running the country)

(Exercise: check that you are happy with the formalisation of the third sentence.)

Proof of inconsistency (using the second method):

1.	$\forall x(Px \rightarrow (Qx \vee Rx))$	Premise
2.	$\forall x(Rx \rightarrow \neg Sx)$	Premise
3.	$\forall x(Sx \rightarrow Px)$	Premise
4.	$\exists x(Px \,\&\, \neg Qx \,\&\, Sx)$	Premise
5.	$P\alpha\& \neg Q\alpha\& S\alpha$	ES, 4
6.	$P\alpha \rightarrow (Q\alpha \vee R\alpha)$	US, 1
7.	$R\alpha$	TI, 5,6
8.	$R\alpha \rightarrow \neg S\alpha$	US, 2
9.	$\neg R\alpha$	TI, 8,5
10.	$R\alpha\& \neg R\alpha$	TI, 7,9

Hence we have derived a truth functional contradiction (a formula of the form **F**& ¬**F** – here **F =** Rα) from the sentences in our original set S taken as premises, and so we have proved that S is inconsistent.

C9: Full First-Order Predicate Logic

C9(a): The insufficiency of monadic predicate logic: the need to introduce Relations.

So far we have restricted attention to simple, so-called *monadic* predicates – 'is a man', 'is a liar', 'was a film star', or whatever. These are monadic or unitary because they apply (or fail to apply) to **single individuals**: Socrates, Boris Johnson, Marilyn Monroe or whomever. (Of course we can then go on to *quantify* – but always over single individuals having 'monadic' properties). But how about statements like 'Cain was a son of Adam' or 'Mrs. Thatcher was politically to the right of Attila the Hun', or 'Liverpool is north of Watford'? These seem intuitively not to be assertions that a single individual has a certain property, but instead assertions that a certain **RELATION** holds between**TWO** individuals, or, as we shall usually express it, that two individuals***stand in a certain relation.*** The two individuals Cain and Adam stand in the relation that the first is the son of the second; the two individuals Mrs Thatcher and Attila the Hun stand in the relation that the first is politically to the right of the second, and so on.

Of course, we *could* just treat each of these sentences (and others like them) as straightforward subject-predicate assertions (and so formalise them in Monadic Predicate Logic) via a suitable choice of predicates. For example, we could introduce the monadic predicate Px, 'x is a son of Adam', and using *a* as the individual constant naming Cain, formalise the first sentence: Pa.

Similarly introducing the monadic predicate Qx, meaning 'x is politically to the right of Atilla the Hun' and using *b* as a name for Mrs Thatcher, the second sentence would formalise as Qb.

Finally, using Rx to mean 'x is north of Watford', Rc would do for 'Liverpool is north of Watford' – using *c* as the name of the city of Liverpool. This possibility explains why logicians for 2000 years saw no need to go beyond the simple subject-predicate form and so no need, in effect, to go beyond Monadic Predicate Logic.

However, it is clear that ***we lose something***– and lose something important – in formalising these ordinary English assertions this way. For example, consider our sentence about Liverpool and Watford, formalised as Rc; next consider the sentence:‘Manchester is north of Watford’. This would formalize, using d for Manchester as Rd – reflecting the fact that this second sentence 'says the same thing" about d (i.e. Manchester) as the first does about c (i.e. Liverpool). But now consider the sentence ‘Liverpool is north of Luton’.If we only had monadic predicates, we would have to introduce a new monadic predicate, say Sx, for 'x is north of Luton', and this last sentence about Liverpool would formalise as Sc. But then the fact that the two original ordinary language sentences about Liverpool said the "same *sort* of thing" about Liverpool would be completely lost in our formalisation – Rc and Sc are completely different assertions about the entity c, and our rules for interpreting predicates in the search forcounterexamples, consistency proofs, *etc*, would allow us to interpret the predicates Rx and Sx as we liked.

Still worse: if we consider the sentences ‘Liverpool is north of Watford’ and ‘Manchester is north of Luton’, we would have to formalise these as Rc and Sd respectively – losing entirely the similarity of form of the two ordinary language sentences.

Relatedly, various intuitively valid inferences could not be shown to be valid (at least not at the same time) if we had only monadic predicates. For example, from 'Liverpool is north of Watford', we could validly infer 'Something is north of Watford' (as you may know, some Londoners regard this conclusion as contentious). The inference simply being from Rc to ∃xRx – an inference that is demonstrated to be valid by one application of the rule EG. However, we could *not* validly infer from that same premise that ‘Liverpool is north of something’ (equivalently: ‘There is something to the north of which Liverpool lies’). But why should one of these inferences be demonstrable as valid and not the other?

We are, in fact, in an analogous situation here to the one we were in with 'Socrates is a man', and 'All men are mortal' vis à vis truth-functional logic. There the best we could do was to formalise the two sentences as p and q, respectively, hence losing the intuitive connection between them, and hence we were unable to capture the validity of certain intuitively valid inferences (see the very beginning of section **C** of these notes

– this was the reason that we introduce predicate logic in the first place rather than remaining content with truth-functional logic).

The remedy for our current problem is the same as it was then: to increase the expressive power of our language. In this case, the suggestion is to acknowledge the intuitively **relational** character of sentences like the ones that we have been considering, by allowing ***two-place*** or binary or dyadic predicates (**relations**) like **Rx,y**: x is to the north of y; Sx,y: x is politically to the right of y, etc. (We shall usually use the term ‘two-place relation’ but you will see the others used elsewhere.)

Hence using Rx,y as our relation for x is to the north of y, a for Liverpool, b for Manchester, c for Watford and d for Luton, our geographical assertions become:

(i) Liverpool is north of Watford Ra,c
(ii) Manchester is north of Watford Rb,c
(iii) Liverpool is north of Luton Ra,d
(iv) Manchester is north of Luton Rb,d

The similarity of form between the four original sentences is thus completely captured by our formalisations. Similarly introducing Sx,y to mean 'x is the son of y' and taking a to mean Cain and b to mean Adam, 'Cain is a son of Adam' is formalised as Sa,b, 'Cain is a son of Eve' could be Sa,c (where c of course is the individual constant naming Eve), and so on. (Whereas, remember, if we had only monadic predicates we would have to have separate predicates Px and Qx, say, for ' x is a son of Adam' and ' x is a son of Eve').

Introducing relations not only allows us to reflect more faithfully the logic of ordinary language, it also permits an immensely better representation of reasoning in the formal disciplines like the various sciences and, above all, mathematics. You should carefully remember these advantages when wrestling with the complications that the introduction of relations undeniably brings in its wake.

Notice that with relations in general *the* ***order*** *in which we take the individuals involved is ofvital importance.* 'Liverpool is north of Watford' is true while 'Watford is north ofLiverpool' is false: hence Ra,c (sticking to our original individual constants) is an importantly different assertion from Rc,a.

There is no reason in principle why we should stick at two-place relations. Admittedly, in ordinary language, three (and higher) place relations are thin on the ground – "lies in between" is an exception: 'Watford lies in between Liverpool and London' could be formalised using the three place predicate Rx,y,z which holds just in case x lies in between y and z, so that, using the same constant as before and e for London, we would have Rc,a,e. Similarly, 'is an immediate descendant of' is a relation that holds between 3 people: for example, Prince Charles, the Queen and Prince Philip.

In mathematics, three (and higher) place relations are much easier to find – for example we could characterise a three-place predicate, say Sx,y,z, which held between three numbers x, y and z just when x was the sum of y and z. This would mean, for example, that if the individual constant a stood for the number 1, b for the number 2, and c for the number 3, Sc,a,b and Sc,b,a were both true sentences, while Sa,b,c, for example, is false (it asserts that 1=2+3.) Although we shall be primarily concerned with two-place predicates, all the results we will arrive at apply generally to predicates with any (finite!) number of places.

C9(B): QUANTIFIED SENTENCES INVOLVING RELATIONS

Once we have relations in our formal language, we can build up quite complicated sentences by quantification. For example:

Example 1:

'Everybody has a father' (≡ For anyone at all, there is someone who is that first person's father) formalises as:

$$\forall x \exists y Ry,x$$

where Rx,y is the relationship which holds between x and y just when x is the father of y – hence Ry,x says y is the father of x.

What, then, would $\forall x \exists y Rx,y$ mean with R understood in the same way? It would say that 'Everyone has fathered someone' – a very different proposition from the one we wanted to formalise. Evidently the order of the variables is again vitally important.

If we wanted to include in our formalisation the fact that our original sentence (Everybody has a father) is evidently meant to be about *people* we could give it the fuller formalisation:

$$\forall x(Px \rightarrow \exists y(Py \ \& \ Ry,x))$$

where Px means x is a person and Rx,y as before stands for x is the father of y.

Example 2:

How about the sentence 'Some people have no daughters' (≡ There is at least one person such that no person is the first person's daughter)? This formalises as:

$$\exists x \forall y \neg Ry,x$$

where Rx,y ≡ x is a daughter of y, so Ry,x means y is a daughter of x)

Or, again putting in the 'persons':

$$\exists x(Px \ \& \forall y(Py \rightarrow \neg Ry,x))$$

(Exercises:

1. Make sure you understand why this formalisation is correct and ***not**,e.g.;* $\exists x(Px \,\&\forall y(Py \,\&\, \neg Ry,x))$
2. What does $\exists x(Px \,\&\forall y(Py \rightarrow \neg Rx,y))$ mean in ordinary English with the same understanding of Px and Rx,y?)

Example 3:

How about: 'For every natural number there is one greater but it is not true that there is a natural number greater than all natural numbers'?

Let Rx,y ≡ x is greater than y (usually written $x > y$), and Nx be 'x is a natural number', then our sentence formalises as:

$$\forall x(Nx \rightarrow \exists y(Ny \,\&\, Ry,x)) \,\&\, \neg\exists x(Nx \,\&\forall y(Ny \rightarrow Rx,y))$$

Notice a few things about this formal sentence:

(a) It is compound in the truth-functional sense: there are two fully-fledged sentences either side of the '&' (formalising 'but'). Truth-functional logic just carries over into our predicate logic.

(b) If we formalised this without worrying about saying that we are exclusively concerned with natural numbers (as we well might if we were formalising a whole stock of sentences which were *all* about natural numbers – so that we took it for granted that everything being talked about was a natural number) then it would formalise more simply as:

$$\forall x\exists yRy,x \,\&\, \neg\exists x\forall yRx,y.$$

Notice that so long indeed as all the sentences we are interested in are talking about objects of the same kind (here all natural numbers) we don't lose anything in suppressing direct reference to that kind – whenever we interpret such a set of formal sentences so that they again make intuitive sense, we specify a domain set.)

(c) The *order* of the quantifiers and variables is again crucially important, as the next example shows still more clearly.

Example 4:

'Everybody has a father but no one is the father of everyone'.

Taking the reference to people as given, this formalises as:

$\forall x\exists yRy,x \,\&\, \neg\exists x\forall yRx,y$ (where $Rx,y \equiv$x is the father of y.)

The second conjunct here could equally well be formalised as $\neg\exists y\forall xRy,x$ (variables are just 'placeholders' with no intrinsic significance) in which case we would have the sentence:

$\forall x\exists yRy,x \,\&\, \neg\exists y\forall xRy,x$.

This sentence, far from being inconsistent, is *true* (remember, it says that everyone has a father but no one has fathered everyone). So, this can only mean that $\forall x\exists yRy,x$ and$\exists y\forall xRy,x$ are very different sentences.

In fact, under the above interpretation the first says 'everyone has a father', the second 'someone has fathered everyone'. Similarly if we interpret Rx,y as meaning 'x is greater than y' with the natural numbers as domain then $\forall x\exists yRy,x$ says 'for every number there's one greater' (*true*) while $\exists y\forall xRy,x$. says 'there's a number greater than any number' (obviously *false*).

Example 5: Mixed Quantification

A sentence involving mixed quantification is of course one in which there is at least one universal, and at least one existential quantifier. It can be a bit tricky to read these correctly – but *with practice* you will get there. Let's consider the four sentences:

(i) $\forall x\exists yRx,y$
(ii) $\exists y\forall xRx,y$
(iii) $\forall x\exists yRy,x$
(iv) $\exists y\forall xRy,x$

What do each of these mean with Rx,y interpreted as 'x (strictly) greater than y' in the natural numbers? (iii) and (iv) we just dealt with towards the end of example (**4**) – *check themover again.*

(i) says: 'For every number x there's a number y such that x is greater than y' – that is, since x is greater than y iff y is less than x, (i) says 'for every natural number there is one less than it'. (This is *false. Exercise:* Explain *carefully* why.)

(ii) says: 'There's at least one number y such that every number x is (strictly) greater than it', that is: 'There's a y which is *less* than every number x' (again in view of the fact about numbers that x is greater than y if y if y is less than x). This is false too – though less obviously. I shall explain why later.

Example 6:

Once you've got the knack, you can really get fancy. Let's try 'Everybody loves a lover' and Lincoln's famous 'You can fool all of the people some of the time, and some of the people all of the time, but you can't fool all of the people all of the time'.

The first says 'Everyone loves anyone who loves someone' (because 'having someone that you love' is what it means to be a lover) and so, taking 'x is a person' as *implicit* and Lx,y, for mnemonic purposes, as 'x loves y', we have:

$$\forall x(\exists yLx,y \rightarrow \forall zLz,x)$$

That is, for anyone x at all, if there is someone y whom x loves (i.e. if x is a lover) then anyone z at all loves x.

(Exercise: How would 'Some people are not loved by anyone','No one loves everyone', 'Some people love no one who loves them', and 'No lover is loved by everyone' formalise?)

Lincoln's famous remark requires monadic predicates Px for 'x is a person' and Qx for 'x is a moment in time', and a two place relation Rx,y which holds just when x can be fooled at y. The remark then formalises as:

$$\forall x(Px \rightarrow \exists y(Qy \mathbin{\&} Rx,y)) \mathbin{\&} \exists x(Px \mathbin{\&} \forall y(Qy \rightarrow Rx,y)) \mathbin{\&} \neg\forall x\forall y((Px \mathbin{\&} Qy) \rightarrow Rx,y)$$

(Make sure you understand how exactly to read each of these conjuncts. Do NOT try to bring all quantifiers to the front of an expression, this *may* (and usually does) lead to disaster – just put them where they occur naturally in the formalisation. For instance $\forall x\exists y(Px\rightarrow (Qy \mathbin{\&} Rx,y))$, with the above meaning for predicates P, Q and R, means something very different from 'You can fool all of the people some of the time'. *Exercise:* what does it mean?)

So now we have for the second time in this course made a move towards a much more expressive formal language. We started with truth-functional logic and decided that, although quite powerful, it needed to be extended to monadic predicate logic, if it was to capture a wider range of intuitively valid inferences (such as the hoary one about Socrates) as valid. Monadic predicate logic involved introducing new ideas about interpretations and rules of proof. Now we have *relations* in our language as well as monadic predicates and so the next items on the agenda are to show that we can again capture a greater range of valid inferences once we have this greater expressive power; and showing that involves in turn revisiting our notions of (a) an interpretation (the key, remember, to invalidity, consistency and independence), and (b) a formal proof (the key to validity, logical truth, logical falsity, and inconsistency) in order to extend them to take account of our enriched language, including relations as well as monadic predicates. You will be comforted to know that all the basic ideas remain the same – though some of the details get a little trickier.

C9(c): Full First Order Predicate Logic: Interpretations – Invalidity and Consistency

The extension of our notion of an ***interpretation*** to include sentences involving relations is straightforward. The notion is exactly the same as before (see *above*), except that we now need to include meanings for the ***relational* constants** within whichever domain is specified. (In fact the term 'predicate' is used to cover *all* predicates, monadic, two-place, three-place or whatever). We then use our notion of an interpretation and the related ones of a model and a counterexample to characterize:

(1) What it **means** for an inference to be **in/valid**

(2) How to **demonstrate** that the inference is invalid (produce a counterexample – i.e. an interpretation in which the premises are true and the conclusion false)

(3) What it **means** for a set of sentences to be **in/consistent**

(4) How to **demonstrate** that a given set is consistent (produce a **model**)

(5) What it **means** for a single sentence to be **independent** of a set of sentences

(6) How to **demonstrate** that a given sentence is independent of a given set of sentences

(7) What it **means** for a sentence to be **logically true/logically false**

(8) How to **demonstrate** that a given sentence is **not** logically true (or **not** logically false)

(*Exercise:* take this opportunity to revise all of the above notions as they came up in monadic predicate logic.)

Example 1:

Show that the following inference is invalid:

1. Everyone who has a father has a mother

Therefore, Anyone who has a mother has a father

This inference has a true premise and a true conclusion (though somewhat surprisingly, you might think, some believing Christians would deny that the conclusion is true if 'everyone' is restricted to 'every human being'!) Nonetheless, intuitively, (as I hope you will agree) the truth of the conclusion doesn't *follow* from the truth of the premise (N.B. from that premise ***alone:*** we could readily produce an augmented inference involving some further premises about the biology of reproduction that would have the conclusion as a valid consequence – though we might have to get fancy about biological parent vs some other kind, worry about test tubes and egg donations *etc*.)

To show that the inference is indeed invalid, *first* we formalise (taking reference to 'persons' as implicit):

1. $\forall x(\exists yRy,x \rightarrow \exists zSz,x)$

So, $\forall x(\exists ySy,x \rightarrow \exists zRz,x)$

where Rx,y: x is a father of y; Sx,y: x is a mother of y.

Consider the following interpretation:

Domain: Humans

Rx,y: x is a son of y

Sx,y: x is a direct descendant of y

Under this interpretation the inference reads:

1. Everyone who has a son has a direct descendant

So, everyone who has a direct descendant has a son

The premise is true but the conclusion (in view of the very fortunate existence of daughters) is false. Hence the original inference is invalid.

Example 2:

Show that the following inference is invalid:

1. Everyone who likes Sartre likes Camus, but not everyone who likes Camus likes Sartre.
2. Some people who like Camus like Flaubert.

Therefore, some people who like Flaubert do not like Sartre.

First, formalise:

1. $\forall x(Lx,a\rightarrow Lx,b)$ & $\neg\forall x(Lx,b\rightarrow Lx,a)$
2. $\exists x(Lx,b$ & $Lx,c)$

So, $\exists x(Lx,c$ & $\neg Lx,a)$

Here:

Lx,y: 'x likes y',

a: Sartre

b: Camus

c: Flaubert.

Now consider the interpretation **I**:

Domain: {natural numbers}

Lx,y: x > y

a: 5

b: 4

c: 6

Under this interpretation the inference reads:

1. Any number strictly bigger than 5 is strictly bigger than 4 but not every number strictly bigger than 4 is strictly bigger than 5.
2. Some numbers are strictly bigger than 4 and strictly bigger than 6.

Therefore, some numbers are strictly bigger than 6 and not bigger than 4.

The premises are true (the second conjunct is true since 5 is strictly bigger than 4 but not strictly bigger than itself); the conclusion is false.

Hence there is a counterexample to the original inference about the French writers and that inference is therefore invalid – the conclusion may perhaps be true but its truth would *not* be *guaranteed* by assuming the truth of the premises.

Example 3:

Show that the set of sentences $\{\forall x\exists yRx,y, \exists x\forall yRx,y, \forall x\exists yRy,x)\}$ is consistent.

Consider the interpretation **I**:

Domain: {natural numbers}

Rx,y: $x \leq y$ (x less than or equal to y, or equivalently, y greater than or equal to x)

Under this interpretation we have:

{For every natural number there is one greater than or equal to it; there is a natural number less than or equal to every natural number; for every natural number there's one less than or equal to it}

This is *a****model*** of the set since all these sentences are true – even the last. (*Exercise:* Think this through carefully – even for the *least* natural number (0 or 1 depending on whether or not 0 is included as a natural number) there is one less than *or equal to* it – viz. itself. (Hence the interpretation with Domain: {natural numbers}and Rx,y: $x < y$ (x strictly less than y) would not be a model – because this last sentence would then be false.)

Because **I** is a model, our set of sentences **is** consistent.

Example 4:

The set of sentences S = (Everybody is his own father; If one person is the father of a second and the second the father of a third then the first person is the father of the third; Everybody has fathered someone} is clearly false – since every sentence in the set is false. But is the set (taken together) ***necessarily*****false** i.e. **inconsistent**?

To decide this, formalise S using the two place relation Rx,y: x is the father of y, obtaining:

$S = \{\forall xRx,x; \forall x\forall y\forall z((Rx,y \,\&\, Ry,z) \rightarrow Rx,z); \forall x\exists yRx,y\}$.

This set *is* consistent as is shown by the fact that the following interpretation **I'** is *amodel:*

Domain: {Natural Numbers}

Rx,y: x ≤ y

(*Exercise:* Write out the set of sentences S as interpreted by **I'** and check that they are all true.)

Example 5:

Let S be as in example 4.

Is the sentence s = ∃x∀vRx,y independent of S?

The same intepretation **I'** as in example 4: (i.e. Domain: {Natural numbers}; Rx,y: x ≤ y) is a model of S U {s} since ∃x∀y x≤y is true in the natural numbers – because ∀y (0 ≤ y) is true. Hence s is consistent with S.

The following interpretation, **I''** is a model of S U {¬s}

I'':

Domain: {positive and negative integers}

Rx,y: x ≤ y.

Under this second interpretation the set of sentences S reads: Any integer is less than or equal to itself (*true*). For any three integers, if the first is less than or equal to the second and the second is less than or equal to the third, then the first is less than or equal to the third (*true*). For any integer there is an integer greater than or equal to it (*true*).

While s reads:

There is an integer less than or equal to every integer i.e. there is a *least* integer – and this is *false*, since the positive *and negative* integers stretch out "infinitely far" backwards as well as forwards.

C9(D): FINITE INTERPRETATIONS

How are the interpretations that do the various jobs in the above **exercises 1-5** arrived at? So far as you are concerned at least they are simply pulled out of the blue and written down for your inspection. Is there some systematic way of arriving at interpretations which do the jobs we want them to do (show invalidity, consistency, *etc.*)? The answer, as indicated earlier, is that there is no fully systematic way ofproceeding here. Although you will, through practice, become adept at producing suitable interpretations, there is no algorithmic procedure, analogous to the truth table method or its derivatives, for producing models and hence deciding issues like invalidity or consistency in full First Order Predicate logic. (This is actually provable as a consequence of some really deep theorems *about* logic that we shall touch on again later.) However, as in the case of monadic predicate logic, we can produce a sort of quasi-systematic method by exploiting the idea, introduced earlier for the restricted case of monadic logic, of ***finite interpretations/models.***

How can we extend that idea to include relational predicates? Well, a two-place relation, such as that of 'father/direct descendant' or of 'being a greater natural number than' holds, or fails to hold, not of single individuals but of ***pairs ofindividuals considered in a particular order*** –or, as we shall say for brevity, of **an ordered pair** of individuals.

So, for example, the relation Rx,y: 'x is a son of y' *holds* of, amongst many others of course, the ordered pair (Prince Charles, Prince Philip), just as the predicate Px: 'x is an even number' holds of, again amongst many others, the individual number 2.

For any interpretation with domain D, monadic predicates, as we saw, determine a ***subset*** *of*D, so that in the domain {1, 2, 3, 4, 5} the predicate Px 'x is even'determines the subset {2, 4}. Similarly, two-place relations determine a subset *of* the set of all ordered pairs that can be formed from the members of the domain. For example, the relationship 'x is (strictly) less than y' (x< y) interpreted in the domain {1, 2, 3} determines the following set of ordered pairs {(1,2) (1,3) (2,3)} –*i.e.* all those ordered pairs that we can form out of elements of the domain in which the first member of the pair is indeed less than the second.

So, as before, the idea with finite interpretations is to forget *intension (i.e.* meaning) altogether and just consider the ***extensions**of* the predicates – so, in particular, *any* set of ordered pairs formed from the domain is a legitimate interpretation of any two-placerelation: we needn't concern ourselves with what 'natural' relation if any has that particular extension in that particular domain. This means that we can be much more systematic in searching for models – as the following examples will illustrate.

Example 1:

Show that S = (∀x(Px→∃yRx,y), ∀x(Px →∃yRy,x), ∃xPx) is consistent.

We are looking to see if we can construct a ***model*** of S, i.e. an interpretation in which all the sentences in S are true, and we are proposing to do this using just a finite number of elements in some domain.

Well let's give ourselves the domain {1,2,3} – as before there's nothing special about three elements, it just works reasonably well in most cases that we'll think about (basically because we'll only deal with fairly simple cases).

In order to make the last sentence in S true (and one of the tricks of the trade here is to start with the existentially, rather than universally, quantified, sentences), there has to be something in the extension of P (which as before we'll denote by **P**). So let's – arbitrarily – put the first element of our domain, 1, in **P** – leave it at that for the time being and see how things go with the other sentences in S. (Another rule of thumb in dealing with finite models is 'always do the minimum that it takes to make a sentence true'.)

So now let's turn to the first sentence in S. It says that everything that's a P has something that's R-related to it. So, given that we just decided that 1 has the property P, there must – at least – be an ordered pair (1, blank) in **R**– the extension of the relation R. So again let's – arbitrarily – make 2 the at least one thing that's R related to 1, *i.e.* let's put (1,2) in **R**. Again we do the minimum as a first step – we may have to come back and revise these assignments if we run into trouble with other sentences.

So, finally, we have the second sentence in S. This says that everything that's a P has something *to which* it's R-related. (Note carefully the difference with how to read the first sentence in S) Again given that we made 1 have the property P, this means that there must be another ordered pair in **R** of the form (blank, 1). So let's, since we haven't yet used 3, make 3 the necessary blank, *i.e.* let's put (3,1) in R. This doesn't mess anything up that we dealt with earlier – obviously not with the third sentence (1 is still a P so $\exists xPx$ is true) and also not with the first sentence. (Think through why not.) So we have a finite interpretation **I** which is a model of S and hence demonstrates that S is consistent:

I:

Domain: {1,2,3}

P: {1}

R: {(1,2), (3,1)}

Intuitively, there is something with property P, *viz.* 1. For everything that's P (i.e. just 1), there's something, *viz*. 2, that it's R-related to, and also something, *viz*. 3, to which it is R-related. Hence all three sentences in S are true in **I**.

Example 2:

Show that s = $\exists x\forall yRx,y$ is *independent* of the set S = {$\forall x\exists yRx,y$, $\forall x\forall y\forall x((Rx,y \,\&\, Ry,z) \rightarrow Rx,z)$}

For independence, remember, we require two models, one of S U{s} and one ofS U{¬s}:

(a) Model of S U{s}:

Let the domain again (for no good initial reason) be {1, 2, 3}**.**

For $\forall x\exists yRx,y$ to be true requires that every element of the domain be R-related to something (not necessarily the same thing); that is, in terms of ordered pairs, that every element of the domain occur as 1st coordinate in at least one ordered pair in the extension **R** of the relation R. So let's try (1,2), (2,3) and (3,1) and then see how we go. (This step is again partially arbitrary: so long as we have (1, blank), (2, blank), (3, blank) in **R** then the blanks can be filled in as we like –∀x∃yRx,y will still be rendered true).

For ∀x∀y∀z((Rx,y & Ry,z) → Rx,z) to be true requires that whenever there are two ordered pairs in **R** such that the 2nd coordinate of one is the same as the 1st coordinate of the other then the ordered pair whose 1st coordinate is the 1st coordinate of one and whose 2nd coordinate is the 2nd coordinate of the other must also be in R. (You may need to read this several times, but it *does* make sense!) Hence, since we already have put (1,2) and (2,3) into **R** we must also have (1,3) in as well, since – to repeat, ∀x∀y∀z((Rx,y & Ry,z) → Rx,z) requires that if R1,2 holds and so does R2,3 then so must R1,3. Similarly since (1,3) and (3,1) are now in R, so must be (1,1) – that is, 1 must be R-related to itself. Since (2,3) and (3,1) are in R so must be (2,1), which, given that (1,2) is already there, means that so must be (2,2). And finally since (3,1) and (1,3) are in there so, to satisfy this transitivity requirement must be (3,3).

So:

R: {(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)}

(This set in fact contains allpossible ordered pairs made up of the 3 elements 1,2,3 – so in this interpretation everything is R-related to everything.)

Interpreting **R** is this way makes both sentences in S true. What about s? For s to be true there must be a *single x* which is R-related to *all y, including itself.*

(Exercise: make sure you understand clearly that this is what, formally speaking, s says.)

This means, in terms of ordered pairs, that there must be pairs (blank, 1), (blank, 2), (blank, 3) where the blank is filled in by the *same* member of the domain in all cases. In fact, since we already have in **R**, e.g., (1,1), (1,2) and (1,3), s is automatically made true. That is, *there is* at least one element of the domain, 1, which is R-related to every element in the domain including itself (i.e. to each of 1, 2 and 3). (In fact, as already noted, this is true of all three elements.)

(b) Model of S U{¬s}:

In order to make s false (i.e. ¬s true) no element can be R-related to every element. This means that our original choice of ordered pairs to satisfy ∀x∃yRx,y (the first sentence in the set S) will need to be modified – for, given that choice and given the requirements of the second sentence in S, we were *forced* to satisfy s (as we just saw).

So let's try (1,1), (2,1) and (3,1) to satisfy ∀x∃yRx,y – this set of ordered pairs does satisfy this sentence since every element of the domain occurs as 1st coordinate of some ordered pair; we have just in this second interpretation made it the same thing – viz. 1 – that is R-related to each of 1, 2 and 3.

Now, the second sentence in S is already satisfied by this assignment: we have (2,1) and (1,1) but this requires only an ordered pair with the same first coordinate as the first and same second coordinate as the second, but this is the ordered pair (2,1) which we already have, and similarly with (3,1) and (1,1). But if we stick with just these three ordered pairs in R then s is false, i.e. ¬s is true: since there is *no single element* of the domain which is R-related to every element: 1 is only R-related to itself and not to 2 or 3, 2 is only R-related to 1 and not to either itself or 3, and 3 is similarly only R-related to 1 and not to either itself or 2.

So the following interpretation **I'** is a model of S U {¬s}:

Domain: {1, 2, 3}

R: {(1,1), (2,1), (3,1)}

(*Exercise*: The interpretation **I''** with the same domain, and with **R**: {(1,2), (2,1), (3,1), (1,1), (3,2)} is also a model of S U {¬s}. Show carefully that this is true.)

Example 3:

Show that the inference:

1. ∀x∃yRx,y

So, ∀x∃yRy,x

is an invalid inference.

We could easily do this via an "ordinary" infinite model [Domain: {natural numbers}, Rx,y: x<y would do it, since the premise would then read 'For every natural number there is one greater'(*true*), while the conclusion would read 'For every natural number there's one less.' (*false* – since there is no natural number less than 0)]. But we can equally well, and more systematically, construct a finite model.

Let Domain = (1, 2, 3). We need to make the premise true, so everything in D has to occur as first coordinate of some member of **R** – let's say (1,2), (2,3) and (3,3).

We also need to make the conclusion false in our interpretation to produce a counterexample to the inference. In order to make the conclusion *true*, everything in D would have to occur in the *second* place of some ordered pair in **R** – but if we leave just those three pairs in **R** that we put in already to make the premise true, this conclusion will be false – 1 does not occur as the second coordinate of any ordered pair in {(1,2), (2,3), (3,3)}. Hence the following interpretation **I** is a counterexample to the inference:

D: {1,2,3}

R: {(1,2), (2,3), (3,3)}

(Exercise:

1. ∃y∀xRx,y

So,∀x∃yRx,y

is a VALID inference. Hence no counterexamples exist. It is, however, a good exercise to attempt to construct a finite counterexample and convince oneself, by failing, that it is impossible. This will also give you a greater understanding of what the two sentences – premise and conclusion – mean.)

I said, when first introducing **finite models** (in connection with monadic predicate logic), that it is *not* true that every set of sentences that has a model at all has a finite model. The technique only works one way round: if we can find a finite model the set must be consistent, but it is not true that if a set of sentences is consistent then it has a *finite* model (obviously when you think about it this must mean that such a set has models but they are all infinite – that is the domains must be infinite sets). Now that we have two-place relations in our language we can in fact give an example of a consistent sets of sentences which has no finite model. One such is the set:

$$S = (\forall x\exists yRx,y,\ \forall x\forall y\forall z((Rx,y\ \&\ Ry,z) \rightarrow Rx,z),\ \forall x\neg Rx,x\}.$$

S is clearly consistent as is shown by the fact that the following (infinite) interpretation **I** is a model:

D: {Natural numbers}

Rx,y: x strictly less than y ($x < y$).

(*Exercise:* Check that you understand that this is indeed a model.)

However, S has no finite model. (*Exercise:* Although a proof of the fact that S has no finite model is beyond the scope of this course, you will in fact convince yourself that it is a fact (and discover the basic idea underlying the proof) if you try to construct a finite model – and take notice of the reason why you are continually frustrated.)

C9(E): FURTHER CLARIFICATION OF THE MEANING OF MULTIPLY QUANTIFIED SENTENCES

When we read sentences like ∀x∀y(Rx,y) back into something more intuitive, we say things like 'Any two things are R-related'. Does this include the case in which the 'two things' are one and the same – i.e. does it require that every object is R-related to itself?

Similarly if we say ∀x∀y∀z((Rx,y & Ry,z) → Rx,z) which we start to read as: 'For any three things x,y,z, if x is R-related to y, and y is R-related to z, then x must be R-related to z', does this include the case in which the three variables x,y,z take on the same value or in which two out of the three of them do?

The answer is that it *does*. The phrase ∀x∀y∀z is to be read as 'For any three – not necessarily distinct – things'. This means that, for example, a sentence like ∀x∃yRx,y when interpreted in the domain D = {1,2,3} would be made true by the interpretation **R** = {(1,1), (2,2), (3,3)}. Under this interpretation for any individual in the domainthere is "another" individual in the domain – *namely itself* – that is R-related to it. It also means that when we employ the finite model technique and we are, say, attempting to make the sentence ∀x∀y∀z((Rx,y & Ry,z) → Rx,z) true, then if we have the ordered pairs (1,2) and (2,1) already in our interpretation **R** of R, then we ***must*** also have the ordered pair (1,1) in **R** – even though in that case the variables x and z take on the same value: 1 is R-related to 2 and 2 is R-related to 1; hence this sentence – to be read remember as 'for any three – *not necessarily distinct* – individuals x, y, z …' – requires that 1 be R-related to itself.

As before when we have taken decisions of this kind, nothing is lost by taking it. If we really want to say, e.g. for any 3 ***definitely distinct*** objects x,y,z then we do so explicitly by introducing the identity relation (x = y) and explicitly requiring distinctness.

So 'for any three ***definitely distinct*** objects x,y,z, (Rxy & Ry,z) → Rx,z) is formalised in full as:

∀x∀y∀z((¬(x=y) & ¬(y=z) & ¬(x=z) & Rx,y & R,y,z) → Rx,z).

Similarly, if we want to say that for any object there is a *different* further object that is R-related to it, we must say not just $\forall x \exists y(Rx,y)$, since this is compatible with the y that is R-related to any x being the same as x, but must instead formalize it as:

$$\forall x \exists y((\neg(x=y) \;\&\; Rx,y)).$$

C9(F): PROVING VALIDITY – FULL FIRST ORDER PREDICATE LOGIC

As already noted, we cannot show that an inference in first order predicate logic is valid using the definition of validity – this is because it would in principle involve looking at infinitely many possible interpretations to check that none provided a counterexample. Instead, as we already saw in the case of monadic predicate logic, we demonstrate the validity of an inference by producing ***a proof*** of its conclusion from its premises. The introduction of ***relations*** into our language requires no new rule of proof. The list (US, ES, UG, EG, TI and CP) remains the same as before. However, as we shall discover, introducing relations does force us to give modified, tighter versions of some of those rules. Before introducing those qualifications, however, it is best to get used to the ideas by looking at a couple of examples of proofs in full predicate logic.

Example 1:

1. No sensible person likes anyone who supported Brexit.
2. Some people who supported Brexit live North of Watford.

Therefore, there are people who live North of Watford whom no sensible person likes.

This formalises as:

1. ∀x∀y((Sx & By) → ¬Lx,y)
2. ∃x(Bx & Wx)

So, ∃x(Wx &∀y(Sy → ¬Ly,x))

Where:

Sx: x is a sensible person

Bx: x supported Brexit

Wx: x lives North of Watford (no advantage using a relational term here)

Lx,y: x likes y

Proof:

1.	$\forall x \forall y((Sx \,\&\, By) \rightarrow \neg\, Lx,y)$	Premise
2.	$\exists x(Bx \,\&\, Wx)$	Premise
3.	$B\alpha \& \, W\alpha$	ES, 2
4.	$\forall y((Sx \,\&\, By) \rightarrow \neg Lx,y)$	US, 1
5.	$(Sx \,\&\, B\alpha) \rightarrow \neg Lx,\alpha$	US, 4
6.	$Sx \rightarrow \neg Lx,\alpha$	TI, 3,5
7.	$\forall x(Sx \rightarrow \neg Lx,\alpha)$	UG, 6
8.	$Sy \rightarrow \neg L\, y,\alpha$	US, 7
9.	$\forall y(Sy \rightarrow \neg Ly,\alpha)$	UG, 8
10.	$W\alpha \& \forall y(Sy \rightarrow \neg Ly,\alpha)$	TI, 3, 9
11.	$\exists x(Wx \,\&\, \forall y(Sy \rightarrow \neg Ly,x))$	EG, 10

Example 2:

1. Any two numbers are either equal or one is less than the other.
2. If one number is less than a second, the second is not less than the first.
3. 5 and 3 are not equal

So, either 5 is less than 3 and 3 not less than 5, or 3 is less than 5 and 5 not less than 3.

Leaving 'numbers' as implicit, this formalises as:

1. $\forall x \forall y(Ex,y \text{ v } Lx,y \text{ v } Ly,x)$
2. $\forall x \forall y(Lx,y \rightarrow \neg Ly,x)$
3. $\neg Ea,b$

So, $(La,b \,\&\, \neg Lb,a) \text{ v } (Lb,a \,\&\, \neg La,b)$

Where:

E(x,y): $x=y$

L(x,y) : $x < y$

a: 5

b: 3

Proof:

1.	∀x∀y(Ex,y v Lx,y v Ly,x)	Premise
2.	∀x∀y(Lx,y → ¬Ly,x)	Premise
3.	¬Ea,b	Premise
4.	∀y(Ea,y v La,y v Ly,a)	US, 1
5.	Ea,b v La,b v Lb,a	US, 4
6.	La,b v Lb,a	TI, 3,5
7.	∀y(La,y → ¬Ly,a)	US, 2
8.	La,b → ¬Lb,a	US, 7
9.	(La,b & ¬Lb,a) v (Lb,a & ¬La,b)	TI, 6,8*

**Exercise*: state and check the tautology involved here.

C9(G): THE NEED FOR FURTHER QUALIFICATIONS OF THE RULES

Normally – especially if you can 'see' what is going on in a proof – using the rules in the loose form in which we have expressed them so far will lead to satisfactory proofs. However, what we are after is a *foolproof* set of rules that allows ***only*** proofs of conclusions that really do follow validly from the set of premises concerned and which could, for example, be programmed into a computer in such a way that the computer always gave the right answer.

The rules as we presently have them are, however, far from foolproof. As they stand they can lead to trouble when relations are involved by sanctioning inferences that are clearly invalid. Here is one simple example.

1. Everyone has a father

So, Someone has fathered himself

This is obviously invalid – since as it stands the premise is true (more or less, there is a bit of vagueness if we go back far enough in evolutionary time if our implicit domain is members of *homo sapiens*, and even more of a problem, if this is our domain, for a Christian who believes that Jesus really was the son of God) and the conclusion definitely false. But here is a simple "proof" of its conclusion from its premise (Rxy means x is the father of y and reference to persons is taken to be implicit):

"Proof":

1.	$\forall y \exists x\ Rx,y$	Premise
2.	$\exists x\ Rx,x$	US, 1

In the form in which we have so far expressed the rule this application of US is entirely legitimate. Line 1 says something (viz. $\exists x Rx,y$) holds for ***all*** individuals, so it must hold (mustn't it?) for the 'arbitrary individual' x. Well, clearly not since under the intended interpretation Rx,y: x is the father of y, the premise is true and the conclusion false. (If we wanted to eliminate the slight vagueness in the premise under this intended interpretation we could as always turn to the crisp unambiguous natural numbers. If Rx,y is taken to mean x>y in the natural numbers, then the premise truly states that for any natural number y there's another natural number x strictly bigger than it, whereas the conclusion falsely asserts that there is a natural number x which is strictly bigger than itself.)

In order to fend off this and some other fallacies, we need to express our rules of proof much more precisely than we have so far done, and this in turn requires a more detailed specification of the ***language*** of first order logic.

C9(H): THE LANGUAGE OF FIRST ORDER LOGIC

The ***basic logical symbolism*** of the language consists of:

(1) The **truth functional connectives** ¬, &, v, →, ↔ (if we are interested in conceptual economy we could of course do with fewer connectives, e.g. with just ¬ and & or even just with ↓) (*Exercise*: check back with Section A to remind yourself of the necessary results about the adequacy of various sets of connectives.)

(2) The **two quantifiers, ∀ and ∃**, and **brackets:** ( ).

(3) Various **individual variables** x, y, z (or if we need lots $x_i$for any i ε {natural numbers}; various **individual constants** a, b c or again a_i, and finally various **ambiguous names**α,β,γ or again α_i

(4) **Predicate symbols**: one-place, or monadic, predicates for properties like 'is male' or 'is even', two-place **relational predicates** like 'is bigger than', 'is a child of' or 'is north of', possibly three-place relations for 'is the sum of', 'isthe remainder on division of ... by ...' and so on in principle for *n*-place relations for any *n*.

*We will add one further linguistic item later, but this is plenty to be going on with.

Formulas:

In order to turn predicates into formulas, we need to apply the predicates to the appropriate number of entities (where this includes individual variables). So, e.g., we could apply the predicate 'is even' (let's say our symbol for this is P) to the individual constant *a* (it might be the number 3) or to the individual variable x or to the ambiguous name α to get the formula 'a is even' (expressed as Pa) or the formula 'x is even' (expressed as Px) or 'the unknown but specific entity α has the property P' (expressed as Pα).

Similarly if R is a two-place predicate symbol for 'is bigger than' it needs to be predicated of two entities, perhaps two variables x and y to make the formula Rx,y – x is bigger than y; or two individual constants a and b to make Ra,b – particular individual a is bigger than particular individual b; or one variable and one constant, say Rx,a - meaning 'the variable individual x is bigger than the particular individual a'; or two ambiguous names to get Rα, β – saying that the two possibly different specific but unknown entities α and β stand in the relation R.

Terms:

The linguistic entities that stand (perhaps 'variably' or 'ambiguously') for**individuals** (as opposed to properties or relations) are called ***TERMS***. So, all ***individual variables*,** x, y, z …, all ***individual constants*** a, b, c …, and all ***ambiguous names***α, β, γ, *etc.*are **terms.**

Atomic Formulas:

As indicated, we create our initial, or **ATOMIC FORMULAS** by substituting the appropriate number of terms in a predicate. So:

<u>Definition:</u> Atomic Formula

If P is an *n*-place predicate and t_i, ..., t_n are terms, then $Pt_1 \dots t_n$ is an atomic formula and only such formulas are atomic.

The rest of the formulas (i.e. the rest of the meaningful expressions) can all be built up in a step-by-step ("recursive") way from these atomic formulas. For example we can apply our truth functional connectives to atomic formulas: to create the formula ¬Pa, for example, from the atomic formula Pa; or the formula Pa & Qb from the atomic formulas Pa and Qb; or Pa → Rx,a from the atomic formulas Pa, and Rx,a etc. Moreover we can iterate these procedures any (finite) number of times to create formulas like Pa → ¬(Rx,a v Qb), or (Pa & Qb) ↔ (Rx,a v ¬Sa,x), *etc.* Brackets are used in obvious ways to indicate the method of construction.

We can also apply our quantifiers to create new formulas – for any formula F and any individual variable x we can create the formulas:∀xF or ∃xF. So, for example, from the atomic formula Py we can create the formulas ∀yPy and ∃yPy; or from the formula Pz → Rx, z we can create the formula ∀z (Pz → Rx, z). Notice that this means that weird expressions like $\exists xPy$ or $\forall zRx,y$ count as formulas – it's just that they are not very interesting formulas. (In fact the first is equivalent to Py and the second to Rx,y.)

Again the process can be iterated to create formulas like $\forall x(Px \rightarrow \exists yRx,y)$ or $\forall x_i\exists x_j((Px_i\& Rx_i,x_j) \rightarrow\forall x_k(\neg Px_k\rightarrow\neg Rx_i,x_k))$.(Remember we can use indexed variables like $x_i$whenever we need lots of variables.) Again the brackets are used to indicate how exactly the overall formula has been built out of its ultimately atomic constituents. (Brackets are dropped whenever no confusion could result, so, e.g. it is usual to write $\forall xPx$ rather than $\forall x(Px)$; but it is necessary to write $\forall x(Px \vee Qx)$ if we mean to say that every individual is either P or Q, in order to distinguish this from the quite different formula $\forall xPx \vee Qx$ (really $\forall x(Px) \vee Qx$ – which means either everything is P or x is Q (though quite what this second disjunct means we will only understand fully a little later).

We have, then, the following so-called recursive definition of a formula in first-order predicate logic:

***Definition:* Formula**

(1) Any atomic formula is a formula.

(2) If **F** and **G** are both formulas so are ¬**F**, **F** & **G**, **F** → **G**, etc.

(3) If **F** is a formula and x_i an individual variable, then $\forall x_i$**F** and $\exists x_i$**F** are both formulas.

(4) F is a formula ***only if*** it is either an atomic formula or can be built up from atomic formulas by a finite number of applications of the operations in (2) and (3).

Sentences (or Closed) and Free Variable (or Open) Formulas:

Certain formulas make ***assertions***. For example Pa, or $\forall x(Px \rightarrow Qx)$ or $\forall x\exists yRx,y$ – these state respectively that some particular individual a has property P; that everything which is P is Q; and everything is R-related to something.

In a given interpretation these statements will all be true or false. We call such formulas **sentences** (or **closed formulas**).

But how about formulas like ‘Px’ or ‘$Qy \rightarrow \forall zPz$’ or ‘$\exists yRx,y$’? These are so-called ***FREE VARIABLE** (or ‘**OPEN**’) **FORMULAS**.* We shall need to take particular carewith free variable formulas when amending our rules of proof. So, first let's characterise them more generally and then say how they should be dealt with.

Free variable formulas can be recognised purely syntactically – purely, that is, in terms of the way that the symbols are put together. First, the **SCOPE *of a quantifier*** is that part of a formula which is governed by that quantifier – usually to be recognised by looking for the right bracket corresponding to the left bracket immediately after the quantifier. So, e.g., the *scope* of the quantifier ‘$\forall x$’ in ‘$\forall x(Px \vee Qx)$’ is the formula ‘$Px \vee Qx$’.The *scope* of the quantifier ‘$\exists z$’ in ‘$Px \rightarrow \exists zRx,z$’ is just '$Rx,z$' (obvious brackets having been dropped).The *scope* of the quantifier ‘$\exists y$’ in ‘$\exists y\forall x(Px \rightarrow Rx,y)$’ is the formula '$\forall x(Px \rightarrow Rx,y)$' (again a pair of brackets having been omitted here, for the full original formula would read: $\exists y(\forall x(Px \rightarrow Rx,y))$).

Now, distinguish between two types of occurrence of an individual variable: a ***BOUND*** occurrence and a ***FREE*** occurrence.

***Definitions:* Bound and Free Variables**

1. An occurrence of variable x_i is ***BOUND*** iff it is ***either*** the variable in a quantifier ($\forall x_i$ or $\exists x_i$) ***or*** it lies **within the scope of a quantifier on x_i.**
2. An occurrence of a variable which is **NOT bound** is ***FREE.***

So in $\exists x(Px \& Qx)$ all variables are bound. Similarly, in $\forall x(Px \rightarrow \exists yRx,y)$. In Px, the only variable is free. In $Px \rightarrow \exists yRx,y$ both occurrences of x are free, while both occurrences of y are bound. In $\exists x\forall yRx,y \rightarrow Py$ both occurrences of x are bound, while the first two occurrences of y are bound and the final occurrence of y is free – if we intended that all occurrences of y be bound we should have used brackets to write $\exists x\forall y(Rx,y \rightarrow Py)$.

So, finally then we have the further:

***Definitions:* Sentences and Free Variable Formulas**

(1) A formula in which **all variables, if any, are bound** is called a **CLOSED FORMULA** or a **SENTENCE**.

(2) A formula in which there is **at least one free occurrence of at least one variable** is called a **FREE VARIABLE (or OPEN) FORMULA**.

Examples:

So Pa, Pα, and$\forall$xPx are sentences, while Px and Py are free variableformulas. $\forall$x$\exists$yRx,y is a sentence, $\exists$yRx,y is a free variable formula, and so on.

Sentences, in a given interpretation, make an assertion and are therefore either *true* or *false.* So, Pa interpreted in the natural numbers with a as 5 and Px as 'x is even' is thefalse assertion '5 is even', while $\forall$x(Px → Qx) in the same interpretation with Px as ‘x is even’ and Qx as 'x is divisible by 2 (without remainder)' makes the true assertion that all even numbers are divisible by 2.

But how about free variable formulas like Px or $\forall$xRx,y? What do these ‘say’?

The Meaning of Free Variable Formulas:

In the same interpretation (D = {Natural Numbers}, Px; x is even) Px is not an assertion and is correspondingly neither true nor false – instead of making an assertion it represents a **condition**: one that is ***satisfied by some substitutions*** for its free variable x (*viz*. the substitutions 2, 4, 6, 8, *etc.*) and ***not satisfied by other substitutions*** for its free variable (the substitutions 1, 3, 5, *etc.*).

Similarly, while the sentence $\forall$x$\forall$y(x ≥ y) makes the false assertion that for any two numbers the first is bigger than or equal to the second, the free variable formula $\forall$x(x ≥ y) is neither true nor false, but instead is satisfied by some substitutions for the free variable y (i.e. it becomes true, or better yields a true sentence, when some individuals from the domain – in this case just the number '0' – are substituted for its free variable) and it is not satisfied by other substitutions (that is, it yields a false sentence when other substitutions are made for the free variable); in this case any substitution apart from 0, since $\forall$x(x ≥ 1), $\forall$x(x ≥ 2), $\forall$x(x ≥ 3) , *etc.*, are all false).

Finally, the free variable formula ‘x ≥ y’ in the same interpretation is again **neither true nor false**, but instead is ***satisfied by or holds of ordered PAIRS of individuals*** from the domain – pairs considered in a particular order. It is satisfied by (1,1) (2,1) (3,1) *etc.,* but not by (1,2) (2,3) *etc.* That is, when you substitute any of the first set of values (in order) for its two free variables it yields a true sentence (1≥1, 2≥1 *etc.*), while if you substitute any of the second set of values (again in order – that's why these pairs are ordered pairs) you get a false assertion 1≥2, 2≥3, *etc.*

Some free variable formulas are true:

IN GENERAL, then, free variable formulas are neither true nor false – they lay down conditions that are sometimes satisfied, sometimes not. But how about the formula Px→ Qx, interpreted in the natural numbers with Px meaning 'x is even' and Qx meaning 'x is divisible (without remainder) by 2', or the same formula interpreted in humans with Px meaning 'x is male' and Qx meaning 'x has the Y chromosome'.

In both these cases, the free variable formula would be true for all substitutions for its free variable. Concentrating on just the numerical case, P0 → Q0, P1 → Q1, P2 → Q2, … are all true. In such a case, the free variable formula can, as I suggested earlier, be interpreted as claiming that **any arbitrary individual** has a certain (complex) property. And in such a case (and only in such a case) we say that ***the free variable formula is itself true.*** **(Notice then that a formula F(x) with one freevariable x is true iff its universal quantification ∀xFx is true).**

Take another, slightly more complicated example:

The free variable formula ∃yRx,y is true in the interpretation D = {natural numbers}, Rx,y: y >x – it says, if you like, that an *arbitrary* natural number x is such that there is one bigger than it. (Again, the truth of the free variable formula is reflected in the fact that this formula's universal quantification –∀x∃yRx,y – is also true.)

These considerations about free variable formulas may seem a bit 'finicky' but in fact clarification of the status of free variable formulas is, as we shall see, ***a necessary pre-requisite for producing watertight versions of the rules of proof.***

C9(I): IMPROVED VERSIONS OF THE RULES OF PROOF

The Rule of Universal Specification (US):

This rule, remember, basically said that you can drop a universal quantifier from the front of a formula and substitute anything you like – an individual constant, individual variable or ambiguous name – for the variable that previously had been quantified. The intuitive justification of this rule is that if everything in a certain domain has a certain property then any individual element of the domain must have that property.

We already know, however, that we need to qualify this rule. This is because, as it stands, and as we saw earlier, the rule allows us to infer ∃xRx,x from ∀y∃xRx,y. But this inference is invalid as the interpretation I (Domain = {humans}, Rx,y: x is the father of y) or the interpretation I' (Domain = {natural numbers}, R x , y: x > y) showed.

Consider, then, ∀y∃xRx,y, and consider first the free variable formula that results from simply dropping the universal quantifier ∀y – *viz.* '∃xRx,y'. In, say, the arithmetical interpretation (Rx,y: x > y) this says that there is something which is bigger than y. If we substitute more or less anything for y in that free variable formula ∃xRx,y we get a formula that "says the same thing" about the substituted entity as the original does about y. So, e.g., substituting the individual constant *a* or the ambiguous name α for y, though it turns the free variable formula back into a sentence produces a sentence which says the same thing about a or about α as the free variable formula did about y: *viz.* that there is something bigger than it.

The exception is if we substitute x for y – because this produces the formula (actually a sentence since it contains no free variables) ∃xRx,x which says something different, *viz.* that there is something bigger than itself.

We need to ban such meaning-changing substitutions since we obviously require our rules of proof to be truth-transmitting (i.e. to produce truths if applied to truths) and there is no guarantee that US will be truth-transmitting if we allow such applications. In order to produce a ban which applies to all cases of this kind, we must introduce a rather complicated-sounding idea and some associated notation.

First remember that the **terms** are those linguistic items that stand (possibly 'variably' or 'ambiguously') for **individuals**, as opposed to the predicates which stand for **properties** of individuals. So the terms (at least as far as we are presently concerned) are:

(i) individual variables
(ii) individual constants, and
(iii) ambiguous names.

Now consider the following:

Definition:

A **term t is free for the variable x_i** in a formula **F** iff **no free occurrence of x_i in F** lies **within the scope of any quantifier** on a variable x_j, where x_j is a variable in t.

(This is the general notion needed, as we shall see, when we slightly extend our notion of 'term'. However, as we currently understand them the only way for a variable x_j to be 'in' a term t is for t to **in fact be the variable x_j.** For this case, the definition reduces to:

A term t is **NOT free for the variable x_i** in a formula **F** iff **t is the variable x_j** and a **free occurrence of x_i in F lies within the scope of some quantifier on x_j.**

Examples:

The term x is **not** free for y in ∃xRx,y since the free occurrence of y in this formula ***does*** lie within the scope of a quantifier on x.

The term z is not free for y in the formula Py → ∃z(Qy → Ry,z) since *two free occurrences of y lie within the scope of a quantifier on z*.

On the other hand, z is free for y in Py → ∃y(Qy → Ry,z). Moreover, the individual constant a is free for y in Py → ∃y(Qy → Ry,z) or indeed for any variable in any formula (*Exercise:* explain why.)

Test for whether the term is free for a variable:

The question, then, only arises when we are asking if one variable, say x, is free for another variable, say y, in some formula. This question can be answered in a purely mechanical way as follows:

Look to see whether the variable you are intending to substitute for another would be 'captured' by some quantifier already in that formula – if it would then that first variable is **not free** for the second in that formula.

All that we need to do to amend the rule US is to require that the term substituted for the free variable created by dropping the universal quantifier is **free for that variable** in the formula thus created. (*Read it slowly – it does make sense!*)

So, e.g, ***y is not free for x*** in $\exists yRx,y$ (substituting y for x would mean that that occurrence of x was captured by the existing quantifier), hence it is **not permitted** to infer $\exists yRy,y$ from $\forall x\exists yRx,y$.

Similarly, since z *is notfree for x* in $(Px \rightarrow \exists z(Sz \vee Rx,y,z))$ (because of the second free occurrence of x), we **cannot** use US to infer from $\forall x(Px \rightarrow \exists z(Sz \vee Rx,y,z))$ to $Pz \rightarrow \exists z(Sz \vee Rz,y,z)$; although we *could* in this second case perfectly well infer $Py \rightarrow \exists z(Sz \vee Ry,y,z)$ – since the variable y is free for x in $Py \rightarrow \exists z(Sz \vee Rx,y,z)$.

Remember that if **F** is any formula, then $F[t|x_i]$ is the formula obtained from **F** by substituting the term t for any occurrence of the variable x_i. So, e.g., if **F** is the formula $Px \rightarrow Qx$, then $F[a|x]$ is $Pa \rightarrow Qa$; if F is $\exists yRx,y$, F[zlx] is $\exists yRz,y$.

Given this notation and the restriction just indicated, we have the following form of the US rule, which is in fact the final form – no further restrictions being needed:

Rule of Universal Specification (Final Form):

For any formula **F** and any term t, **F**$[t|x_i]$ may be inferred from $\forall x_i$**F**, ***PROVIDED*** t is free for x_i in **F**.

Modifications of the rules caused by the introduction of the rule of Conditional Proof:

You will remember (refresh your memory if not!) that the rule of conditional proof permits the introduction of an "extra assumption" which, however, is subsequently "discharged". Consider the following inference:

1. All of those financing the Conservative Party (Px) are from big business (Qx).
2. No one from big business gives a damn about the Environment (Rx).

So, none of those financing the Conservative Party gives a damn about the Environment.

Formalising appropriately, we can prove that this inference is valid as follows:

1.	$\forall x\ (Px \rightarrow Qx)$	Premise
2.	$\forall x(Qx \rightarrow \neg Rx)$	Premise
3.	$Px \rightarrow Qx$	US, 1
4.	$Qx \rightarrow \neg Rx$	US, 2
5.	Px	A
6.	Qx	TI, 3,5 (Modus Ponens)
7.	$\neg Rx$	TI, 4,6 (Modus Ponens)
8.	$Px \rightarrow \neg Rx$	CP 5-7
9.	$\forall x(Px \rightarrow \neg Rx)$	UG, 8

This proof is perfectly OK. But what about the following “proof”?

Steps 1-5 as before, followed by:

6.	$\forall x Px$	UG, 5
7.	$Px \rightarrow \forall x Px$	CP, 5-6
8.	$\forall x(Px \rightarrow \forall x Px)$	UG, 7

There is nothing wrong with the step from 7 to 8 here. And nothing wrong at all with this “proof” so far as our rules of proof stand at the moment.

However, the conclusion of this variant proof, line 8, means:

"For anything at all, if it's a financer of the Conservative Party then everyone is a financer of the Conservative Party”.

This rather strange assertion nonetheless makes perfect sense when you think about it carefully – it is in fact equivalent to the (admittedly odd) assertion that if there's at least one financer of the Conservative Party then everyone finances the Conservative Party.This is obviously false and clearly does not follow from the premises of the argument.

The invalid step here occurs in fact at line 6 of the variant "proof". The general message is that we mess things up if we allow ourselves to generalise on variables that are involved in assumptions introduced for purposes of Conditional Proof (or rather we mess ourselves up if we so generalise *before* the point at which we have used Conditional Proof to 'discharge' the relevant assumption). In order, then, to prevent problems like this one, we introduce a notational convention and a corresponding restriction on the rule of Universal Generalisation.

Notational convention:

Any variable that is free in any assumption introduced into a proof must be ***FLAGGED***; this means that we record the variable on the RHS of the proofalong with the justification for that line. The variable is then flagged in any line that depends for its justification on a line in which it is already flagged, and the flagging ends only when all the assumptions in which it was introduced as a free variable have been discharged by applying the rule of Conditional Proof.

The ***Restriction on Universal Generalisation*** is then easy:

You can't universally generalise on flagged variables: that is, you can infer the formula $\forall x_i$**F** from **F**, if, but ***only if,*** x_i is not flagged in **F**. (We'll need to add a further restriction in a moment.)

Let's then amend our two most recent attempted proofs in accordance with our new notational convention.

Proof 1:

1.	$\forall x(Px \rightarrow Qx)$	Premise
2.	$\forall x(Px \rightarrow \neg Rx)$	Premise
3.	$Px \rightarrow Qx$	US, 1

4.	$Qx \rightarrow \neg Rx$	US, 2
5.	Px	A x
6.	Qx	TI, 3,5 (MP) x
7.	$\neg Rx$	TI, 3,5 (MP) x
8.	$Px \rightarrow \neg Rx$	CP 5-7
9.	$\forall x(Px \rightarrow \neg Rx)$	UG, 8

We have started the flagging on x at line 5 where an assumption is made in which x occurs free. Lines 6 and 7 depend on line 5 and hence x is flagged in both of these too. The flagging is dropped at line 8 when the assumption in which x occurred free is discharged.

Notice that this proof is not only intuitively valid, it is also perfectly kosher so far as our latest restriction is concerned. In particular, applying UG at line 9 breaks no rules since x is no longer flagged in line 8 (the flagging having stopped with the discharging of the assumption at line 7).

“Proof” 2:

1.	$\forall x(Px \rightarrow Qx)$	Premise
2.	$\forall x(Qx \rightarrow \neg Rx)$	Premise
3.	$Px \rightarrow Qx$	US, 1
4.	$Qx \rightarrow \neg Rx$	US, 2
5.	Px	A x
6.	$\forall xPx$	UG, 5 x
7.	$Px \rightarrow \forall xPx$	CP 5-6’
8.	$\forall x(Px \rightarrow \forall xPx)$	UG, 7’

This, remember, is the aberrant proof with the conclusion (‘If the Conservative Party has a single financer then everyone finances the Conservative Party’) which clearly doesn’t in fact follow from the premises. Once we introduce flagging, we see that the aberrant proof does indeed involve – at line 6’ – an application of UG to a flagged variable. Hence this proof is ruled out by our restriction on UG – no universal generalising is allowed on flagged variables.

A somewhat similar ***restriction on the Rule of Existential Generalisation (EG***) is also required. To see why here is how to "prove" that 'Everyone ishappy' from the premise 'Some people are happy'. Needless to say this "proof" is not valid – the conclusion does not follow from the premise. (Let's use Hx to mean 'x is happy')

"Proof" A:

1.	$\exists xHx$	Premise
2.	$H\alpha$	ES, 1
3.	$\neg Hx$	A x
4.	$H\alpha \& \neg Hx$	TI, 2,3 x
5.	$\exists x(Hx \& \neg Hx)$	EG, 4 x
6.	$H\beta \& \neg H\beta$	ES, 5 x
7.	$\neg Hx \rightarrow (H\beta \& \neg H\beta)$	CP 3-6
8.	Hx	TI, 7
9.	$\forall xHx$	UG, 8

(Here, the move from line 7 to line 8 is our old friend the formal equivalent of a *reductio ad absurdum*: any formula of the form $P \rightarrow (Q \& \neg Q)$ tautologically implies $\neg P$, since any formula of the form $(P \rightarrow (Q \& \neg Q)) \rightarrow \neg P$ is a tautology. **ALSO** remember that there was one restriction on ES that was so obvious that we introduced it right away: namely that we must always use a *new* ambiguous name whenever we use ES more than once. *Exercise*: remind yourself why. We have followed that restriction in this proof by using β at line 6 rather than the already used α.)

Here, although you wouldn't think up this proof unless you had a sick mind, every step is legitimate so far as our rules as presently formulated are concerned (you should check this carefully, paying attention to what the rules allow you to do *formally* and forgetting, for the moment, about what each line means intuitively). In particular, we have obeyed the notational convention on flagged variables, and have not transgressed the new condition on UG – this is because the only application of UG occurs at line 9, *after* the flagging has correctly stopped (the assumption in which x was introduced free has been discharged at line 7).

There is however something fishy about line 5 (which is in fact the only line in this “proof” that is faulty). Intuitively, making the step at line 5 forces us to identify the ‘arbitrary’ object x with the *particular*, if ambiguously named, entity α. An obvious restriction that would ban line 5 is the counterpart of our restriction on UG – namely to ban ***existential*** generalisation on ***any*** flagged variable, as well as universal generalization.

This, however, turns out to be overly restrictive: it would leave our system of rules of proof incapable of demonstrating the validity of certain inferences that are in fact valid. Here is one ***important example***:

The inference from ¬∃xPx to ∀x¬Px is clearly valid (indeed intuitively they ‘say the same thing’ since they are two equivalent ways of saying that nothing is a P; and so they should in fact be inter-derivable (as they indeed are).) Here’s how to show that the inference from ¬∃xPx to ∀x¬Px is valid:

Proof B:

1.	Px	A x
2.	∃xPx	EG, 1 x
3.	Px →∃xPx	CP 1-2
4.	¬∃xPx	A
5.	¬Px	TI, 3,4
6.	∀x¬Px	UG, 5
7.	¬∃xPx →∀x¬Px	CP 4-6

Notice that x is not flagged at step 4 since it is not *free* in the assumption made there. Everything else is done in accordance with our rules. The only questionable step is at line 2. If we were to ban existentially generalising on flagged variables, line 2 would be rendered illegitimate and – as it turns out – there is no other way to prove in our system that this valid inference is indeed valid.

Hence to deal with the problem highlighted by “Proof” A, we must introduce a less demanding restriction than a blanket ban on existentially generalizing on flagged variables. It turns out that it can be proved that with only this restriction the rule sanctions all and only all valid inferences (but we cannot give the proof of this in this course). For the present you should just learn the restriction and apply it correctly:

Restriction on Existential Generalisation:

Let **F'** be the formula obtained from the formula **F** by substituting the variable x_i for all occurrences of a name (that is, *either* an individual constant *or* an ambiguous name) **IF ANY.** Then the step from **F** to$\exists x_i$**F'** is legitimated by EG ***provided*** that, IF goingfrom **F** to **F'** actually involves dropping an ***ambiguous name***, THEN x_i is not flagged.

Pending a general proof that the rule as thus restricted is 'sound' (i.e. permits only valid inferences), this complicated restriction is bound to look *ad hoc*. However, notice that at least it does the job so far as our two most recent alleged derivations –***"Proof"A*** and ***Proof B*** – are concerned.

The step in (genuine) Proof **B**– line 2 – that was under suspicion is in fact exonerated, that is it does **not** run afoul of this restriction on EG: since no ambiguous name is dropped in the process. But step 5 in "proof" **A** is disallowed by the restriction since in that "proof" α was dropped in favour of $\exists x$... and x *was* flagged. Since the inference in Proof **B** is valid and the one in "Proof" A is invalid, the restriction definitely gets it right in these two cases. (The (meta-)proof that it gets it right in *all* cases is more complex and will not be given in this course.)

So just underlining what the restriction means a little more sharply: If some step in a proof is from $R\alpha,y$ to$\exists xRx,y$ then x ***must not*** be flagged if the step is to be legitimated by EG; but if the step is from, say, Rx,y to $\exists xRx,y$ then this is legitimated by EG (as restricted) ***evenif*** x is flagged, since no ambiguous name is dropped in making the step.

Further Restrictions on the Rules:

Consider the following derivation:

1.	$\exists x\forall yRx,y$	Premise
2.	$\forall yR\alpha,y$	ES, 1
3.	$\exists y\forall yRy,y$	EG, 2
4.	$\forall yRy,y$	ES, 3

This is invalid reasoning. Take the interpretation:

Domain: {natural numbers}

Rx,y: the product of x and y (x*y) is y.

The premise is true, since the number 1 is such (1*y = y). But the conclusion that ∀y (y*y = y), which says that every number is equal to its own square, is false.

Step 4 looks suspicious but is in fact OK – an existential quantifier has been dropped and an ambiguous name introduced for the free variable thus created (as required by ES) – it's just that in this case (because of the double quantification on y in line 3), no free variable *has* been created by dropping the first quantifier.

The fallacy in fact occurs at line 3: α occurs in the formula ∀yRα,y within the scope of a quantifier on y; and this means that substituting the variable y for α in (apparently) applying the rule EG yields a **bound occurrence** of that variable. We must restrict EG exactly by banning such substitutions:

Modified EG 1:

$\exists x_i$**F'** may be inferred from **F**, where **F'** is the same formula as **F** except that all the occurrences of some ***name*** (ambiguous or constant) have been replaced by the variable x_i, if but **only if** the name does not occur in **F** within the scope of a quantifier on the variable x_i.

Next, consider the following derivation:

1.	∀x∃yRx,y	Premise
2.	∃yRx,y	US, 1
3.	Rx,α	ES, 2
4.	∃xRx,x	EG, 3

Yet if we interpret Rx,y as x<y in the natural numbers, the premise here is a true statement about numbers and the conclusion 4 is a false statement (*Exercise*: make sure you understand why), so obviously something is wrong. To see precisely what, consider what is going on intuitively. The move from 1 to 2 is clearly correct: if for *every* number there is one greater than it, then this applies to any arbitrary number –

which is what 2 says. It is also true (step 3) that, given an arbitrary number x, we can pick a number greater than it; but the crucial point is that *whichα's make this true will depend on which numberx we have picked* (think about the two Hungarians in the 'joke' I usually give in the lecture). We ought to signalthis dependency of α on the prior choice of x by writing x as a subscript to α (i.e. as α_x.)

In fact, we adopt the following general **terminological rule**:

Any ambiguous nameintroduced by ES has as subscripts all the free variables occurring in the formula to which ES is applied.

Thus, e.g., if we apply ES to the formula $\exists xRx,y,z$ we must write $Rx,y,\alpha_{y,z}$. And given this convention we block the fallacious step 4 in the above proof by restricting EG as follows:

Modified EG 2:

It is *NOT* legitimate to apply ***existential generalisation using avariable that occurs as a subscript in the formula***.

Thus step 3 in our latestderivation should, in accordance with our terminological convention read '$x < \alpha_X$' and hence step 4 is blocked by this restriction on EG.

A similar restriction must also be applied to UG as the following fallacious "proof" shows:

1.	$\forall x\exists yRx,y$	Premise
2.	$\exists yRx,y$	US, 1
3.	Rx,α_x	ES, 2
4.	$\forall xRx,\alpha_x$	UG, 3
5.	$\exists y\forall xRx,y$	EG, 4

But in the same interpretation of Rx,y (*viz.* $x < y$) again the premise is true about numbers and the conclusion false (*Exercise*: make sure you understand why). The incorrect step here is 4 and this points to the required modification on UG: ***it is NOT permissible to universally generalise using a variable thatoccurs as a subscript in the formula.***

Final statement of the Rules of Proof:

So, here are the correct versions of the rules in all their glory:

1. **Rule of universal specification (US):**

F[tIx_i] may be inferred from $\forall x_i$**F**, where t is any term free for x_i in **F**.

2. **Rule of universal qeneralisation (UG):**

$\forall x_i$**F** may be inferred from **F** so long as x_i is ***NEITHER*** **flagged** in **F** ***NOR*** occurs as a **subscript** in **F**.

3. **Rule of existential specification (ES):**

F[α_j|xi] may be inferred from $\exists x_i$**F***,* where α_j is any ***NEW*** **'ambiguous name'** (i.e. one which does not occur already in some earlier line of the proof).*
(*Remember we already introduced and justified this restriction earlier – since it is so obvious.)

4. **Rule of existential generalisation (EG):**

Let **F**' be the same formula as **F** except POSSIBLY that some *name* (that is *either* some ambiguous name *or* some individual constant) in **F** is replaced throughout by some variable x_i in **F'**, then $\exists x_i$**F'** may be inferred from **F**, *PROVIDED* that the replaced name (if any) does not occur in **F***within the scope of a quantifier on* x_i; **AND** provided that ***if*** the replaced name in F is an ambiguous name, then x_i *neither* occurs as **a subscript** *nor* is **flagged**.

The other two rules remain straightforward, as originally introduced:

1. **Rule of tautological implication (TI):**

If the formula **F** → **G** is a truth functional tautology, then **G** may be inferred from **F.**

2. **Rule of Conditional Proof (CP):**

If the formula **G** can be inferred from some set of premises S*plus* the extra assumption formula **F** then the formula **F** → **G** can be inferred from S alone.

C10: Logical Truth, Logical Falsehood, and Inconsistency – Full Predicate Logic

C10(a): Logical Truth

The notion of a logical truth remains, of course, the same as it was in the restricted case of monadic predicate logic:

A single sentence s in the language of predicate logic is ***logically*****true iff** it is true in every interpretation **I**.

The method of demonstrating logical truth also remains the same: we show that s is a logical truth by proving it 'absolutely' – that is, without invoking any premises. As example 1 will show, even in the case of monadic predicate logic, we need, when producing proofs, to pay attention to the restrictions on the rules of proof that we just introduced; it is just that, as we will see as we go along, attention to the restrictions is much more often necessary when relations are involved.

Example 1:

$\forall xPx \leftrightarrow \neg \exists x \neg Px$

is a logical truth. (It's a sort of monadic predicate logic equivalent of the truth-functional law of double negation: 'Everything has property P' says the same thing as 'Nothing fails to have property P'.)

Proof:

1.	$\forall xPx$	A
2.	$\exists x \neg Px$	A
3.	$\neg P\alpha$	ES, 2
4.	$P\alpha$	US, 1
5.	$P\alpha \,\&\, \neg P\alpha$	TI, 3,4
6.	$\exists x \neg Px \rightarrow (P\alpha \,\&\, \neg P\alpha)$	CP, 2-5
7.	$\neg \exists x \neg Px$	TI, 6

8. $\forall xPx \rightarrow \neg\exists x\neg Px$ CP, 1-7

At this stage, we have established one "direction" of the biconditional. Now for the other part:

9. $\neg Px$ A x
10. $\exists x\neg Px$ EG x
11. $\neg Px \rightarrow \exists x\neg Px$ CP, 9-10
12. $\neg\exists x\neg Px$ A
13. Px TI, 11,12
14. $\forall xPx$ UG, 13
15. $\neg\exists x\neg Px \rightarrow \forall xPx$ CP, 12-14

That's the second direction dealt with, so we put them together via TI:

16. $\neg\exists x\neg Px \leftrightarrow \forall xPx$

Comments:

1. This is the first and most straightforward of a series of 'interdefinability of quantifiers' proofs, so you should pay it particular attention. The way that the proof works (the "heuristic") carries over to more complicated cases.
2. Proving an if and only if statement invariably involves using the tautological equivalence underpinning line 16: *viz.* $F \leftrightarrow G$ iff $(F \rightarrow G)$ & $(G \rightarrow F)$. That is, we break the proof of the biconditional down into proofs of two conditionals: the left to right conditional (lines 1 to 8) and the right to left conditional (lines 9 to 15). And then the two halves are tied together using the indicated tautology in the final line.
3. The left to right conditional is proved in a "natural" way. We want to prove a conditional – so we draw upon an almost invariable "heuristic rule": assume the antecedent of the conditional, derive the consequent and then use conditional proof. So, the antecedent is assumed (line 1);we now want to derive $\neg\exists x\neg Px$; but the (almost invariable) way to prove a negated sentence is to assume its negation (i.e. the unnegated form), derive a contradiction and again use conditional proof (this is mimicking *reductio ad absurdum*). So, at 2 we assume the negation of what are now out to prove; derive a contradiction by line 5, we

then discharge the assumption made at line 2 at line 6 and that gives us $\neg\exists x\neg Px$, via the important tautology we have often used (*viz.* $(F \rightarrow (G \& \neg G)) \rightarrow \neg F$). So now, at line 7, we discharge the assumption made at line 1 by CP and the first half of the proof is complete.

4. If we tried to operate in this "natural" way to prove the remaining right to left conditional, then we would soon be stymied. It would involve taking $\neg\exists x\neg Px$ as an assumption and setting out to derive $\forall xPx$. And if we tried to do that in the same way as the first half we would assume $\neg\forall xPx$ and try to derive a contradiction; but that would give us two negated sentences and negated sentences are bad news: US and ES only allow us to drop **quantifiers that begin a formula** and whose scope is the rest of the formula; they do not tell us what to do when a formula begins with a negation sign: in particular, it would be an error to just drop the quantifier in $\neg\exists x\neg Px$ and infer, say, $\neg\neg P\alpha$.
5. You should note very carefully what we do instead. This, and similar steps in proofs we will come across later, show the importance of the detailed restriction on EG. Rather than a blanket ban on existentially generalizing on flagged variables, the restriction only bans such generalization in cases where an ambiguous name is dropped as a consequence of applying EG. So we assume $\neg Px$at line 9 – although x is flagged we are still entitled to existentially generalize at line 10 and then discharge the assumption by CP to give us line 11 (where notice the flagging has ended). Line 11 is in effect a little logical truth: it says if x is P then (naturally) there must be at least one P! Now the assumption that we wanted to make all along (*viz.*$\neg\exists x\neg Px$ – the antecedent of the conditional we are out to prove) does "talk", i.e. allow us to derive something, but not via any quantifier rule (for the reason stated earlier), but instead via TI. That assumption together with the logical truth we derived at line 11, gives us Px (with no flagging) and then the rest of the proof is straightforward.

Example 2:

The sentence:

$\exists x\forall yRx,y \leftrightarrow \neg\forall x\exists y\neg Rx,y$

is a logical truth.

Proof:

1.	$\exists x\forall yRx,y$	A
2.	$\forall x\exists y\neg Rx,y$	A

Comment 1: So, just as in the 'easy' half of Example 1, we have now assumed, contrary to what we will end up proving, that the LHS of the biconditional at issue is true, while the RHS is false.

3.	$\forall yR\alpha,y$	ES, 1
4.	$\exists y\neg R\alpha,y$	US, 2
5.	$\neg R\alpha,\beta$	ES, 4
6.	$R\alpha,\beta$	US, 3
7.	$R\alpha,\beta \ \& \ \neg R\alpha,\beta$	TI, 5,6
8.	$\forall x\exists y\neg Rx,y \rightarrow (R\alpha,\beta \ \& \ \neg R\alpha,\beta)$	CP, 2-7
9.	$\neg\forall x\exists y\neg Rx,y$	TI, 8
10.	$\exists x\forall yRx,y \rightarrow \neg\forall x\exists y\neg Rx,y$	CP, 1-9

Comment 2: So, this ends the 'first half' of the proof: just as in Example 1, we are out to prove a biconditional, so it's natural to split the proof into two 'halves': first F→G, then G → F.

11.	$\forall yRx,y$	A x

Comment 3: We are now setting out to prove the 'second half' of the result (ie RHS→LHS) by assuming not that the RHS is true and then deriving the LHS, but instead by assuming that the LHS is false and showing that in that case the RHS is also false (and then 'turning it back round' which is what we will do at line 26, below). However just assuming $\neg\exists x\forall yRx,y$ straight off (as we will do at line 14) would leave us stymied (for essentially the same reason as explained in Comment 4 for Example 1) so we again take the analogous steps as in Example 1. We are going to assume $\neg\exists x\forall yRx,y$, so make the assumption that leaves off the negation and the existential quantifier, i.e. $\neg\forall yRx,y$ (noting that x is free and so must be flagged);we can still apply EG (line 12) because although x is flagged, no ambiguous name is dropped. Discharging by CP gives us again

a logical truth at line 12 and now the assumption we were going to make all along does again "talk" via the application of TI at 15. Here we go:

12. $\exists x\forall yRx,y$	EG, 11 x
13. $\forall yRx,y \rightarrow \exists x\forall yRx,y$	CP, 11-12
14. $\neg\exists x\forall yRx,y$	A
15. $\neg\forall yRx,y$	TI, 13,14

Comment 4: We know via the intuitive result that $\forall y = \neg\exists y\neg$ that this line is in fact equivalent to $\exists y\neg Rx,y$, which we will eventually get as line 23. It may be better to jump, the first time you try to grasp the proof, from here directly to line 23. However, this intuitive equivalence needs to be proved properly which is what lines 15-23 achieve. Study that sub-proof carefully; it again involves that little riff where you assume some sentence with a free variable and existentially generalize on it (lines 16-18)

16. $\neg Rx,y$	A x,y
17. $\exists y\neg Rx,y$	EG, 16 x,y
18. $\neg Rx,y \rightarrow \exists y\neg Rx,y$	CP, 16-17
19. $\neg\exists y\neg Rx,y$	A x
20. Rx,y	TI, 18,19 x
21. $\forall yRx,y$	UG, 20 x
22. $\neg\exists y\neg Rx,y \rightarrow \forall yRx,y$	CP, 19-21
23. $\exists y\neg Rx,y$	TI, 15,22
24. $\forall x\exists y\neg Rx,y$	UG, 23
25. $\neg\exists x\forall yRx,y \rightarrow \forall x\exists y\neg Rx,y$	CP, 14-24
26. $\neg\forall x\exists y\neg Rx,y \rightarrow \exists x\forall yRx,y$	TI, 25
27. $\exists x\forall yRx,y \leftrightarrow \neg\forall x\exists y\neg Rx,y$	TI, 10,26

C10(B): LOGICAL FALSITY

Again the definition of a logical falsehood is, of course, the same for the full predicate logic as it was for the restricted case of monadic predicate logic. Namely:

> A single sentence s in the language of predicate logic is **logically false*iff*** it is false in every interpretation **I**.

And the method of demonstration also remains the same:

> We show that s is logically false, by showing that ***its negation ¬s is logically true*** – i.e, by deriving ¬s from no premises.

Example:

The sentence:

$\forall x(Px \rightarrow \exists yRx,y) \& \exists x(Px \& \forall y \neg Rx,y)$

is logically false.

Proof:

1. $\forall x(Px \rightarrow \exists yRx,y) \& \exists x(Px \& \forall y \neg Rx,y)$ A
2. $\exists x(Px \& \forall y \neg Rx,y)$ TI, 1
3. $\forall x(Px \rightarrow \exists yRx,y)$ TI, 1
4. $P\alpha \& \forall y \neg R\alpha,y$ ES, 2
5. $P\alpha \rightarrow \exists yR\alpha,y$ US, 3
6. $P\alpha$ TI, 4
7. $\exists yR\alpha,y$ TI, 5,6 (Modus Ponens)
8. $R\alpha,\beta$ ES, 7
9. $\forall y \neg R\alpha,y$ TI, 4
10. $\neg R\alpha,\beta$ US, 9
11. $R\alpha,\beta \& \neg R\alpha,\beta$ TI, 8,10
12. $\forall x(Px \rightarrow \exists yRx,y) \& \exists x(Px \& \forall y \neg Rx,y) \rightarrow (R\alpha,\beta \& \neg R\alpha,\beta)$ CP, 1-9
13. $\neg(\forall x(Px \rightarrow \exists yRx,y) \& \exists x(Px \& \forall y \neg Rx,y))$ TI, 10

Notice, then, that we have shown that $\forall x(Px \rightarrow \exists yRx,y)$ &$\exists x(Px$ &$\forall y\neg Rx,y)$ is logically false by showing that its negation is logically true (the further twist being that we have proved the latter by assuming *its* negation – i.e. the original sentence and deriving a contradiction).

C10(c): The Inconsistency of Sets of Sentences – Full Predicate Logic

Finally, exactly the same considerations apply to showing that a particular (finite) set of sentences S is inconsistent in full predicate logic as they did in the simple monadic case – i.e. that S is an inconsistent set if the sentences it includes are never all true together. So, as before, S is inconsistent iff it has no models, i.e. there is not a single interpretation in which all the sentences in S are true. Moreover, the methods for demonstrating that S is inconsistent remain the same. There were, remember, two such methods that, though formally different, are clearly intuitively equivalent.

***Method 1*:**

(s_1, s_2, … s_n) is inconsistent iff s_1 & s_2 & … & s_n is a logical falsehood.

Hence, using this method we would show that ¬(s_1 & s_2 & … s_n) can be proved from no premises.

***Method* 2**:

(s_1, s_2, … s_n) is inconsistent iff a truth functional contradiction can bederived from (s_1, s_2, … s_n) as premises

Hence, using this method, we would take (s_1, s_2, … s_n) as premises and set out to deduce a truth functional contradication.

Example:

The set {∃x(Px → ∃y∃zRy,z), ¬(∀xPx → ∃y∃zRy,z)} is inconsistent.

Proof via Method 2:

1.	∃x(Px → ∃y∃zRy,z)	Premise
2.	¬(∀xPx → ∃y∃zRy,z)	Premise
3.	Pα → ∃y∃zRy,z	ES, 1
4.	∀xPx & ¬∃y∃zRy,z	TI, 2
5.	∀xPx	TI, 4

6.	$P\alpha$	US, 5
7.	$\exists y \exists z R y,z$	TI, 3,6
8.	$\neg \exists y \exists z R y,z$	TI, 4
9.	$\exists y \exists z R y,z \;\&\; \neg \exists y \exists z R y,z$	TI, 7,8

(*Exercise*: prove that this same set is inconsistent by the first method – this will simply involve collapsing the first two steps into one (and relabelling) and adding one further final step.)

C11: First Order Predicate Logic with Identity

C11(a): Is identity descriptive or logical?

The whole of logic is dependent on a sharp distinction being made between *logicalterms* and *descriptive terms.* The former consist of the truth-functional connectives andthe quantifiers 'all' and 'some' (and brackets). The whole notion of validity depends on regarding the logical terms as *fixed* in meaning, that is, as ***not*** reinterpretable. The *descriptive* terms on the other hand are exactly those that are reinterpretable – those that are, or may be, given different meanings from interpretation to interpretation.

If for example we were allowed to reinterpret what 'all' means – if we could consider interpretations in which it meant 'some', for example – then there would be no interesting valid inferences. Even:

1. All Greeks are men
2. All men are mortal

So, all Greeks are mortal

would not meet the test of validity if 'all' were not fixed from interpretation to interpretation. This is because if we *were* allowed to reinterpret all as 'some', the following would be a counterexample (i.e. an inference of the 'same form' (under this proposed laxer notion of form) with true premises and a false conclusion)

1. Some males are British
2. Some British people are females

So some males are females

But why *exactly* should 'all' be regarded as specifying part of the *form* of an inference, while 'is Greek', for example, is part of the *content* and therefore reinterpretable? This raises difficult and deep issues in the foundations of logic, that we shall not be able to go into in this course. However, we can investigate one *particular* issue that arises in this connection – the issue of which side of the line the *equality* or *identity* relation x = y falls. Is identity *logical* or *descriptive?*

In first-order logic as we have studied it so far, we have not given any special role to the relation of identity or equality – which has therefore implicitly been understood as a descriptive term, as an ordinary two-place relation: 'Every object is identical to itself', for example, just formalises as ∀xRx,x. This means, as we know, that when we start to reinterpret some set of sentences, for example in search of a counterexample to some inference, then we are perfectly at liberty to reinterpret what had started out as the identity relation in any way that we like: let's say we have formalised it as Rx,y then we can of course set up interpretations **I** in which Rx,y means, say, x is the father of y, or x loves y or whatever. And this would mean that 'Everything is identical to itself" – which sounds like it should be a logical truth – would be no such thing: since 'Everyone is the father of him/herself' would be a sentence of the same logical form as 'Everything is identical to itself', and hence that identity statement cannot be a logical truth.

What would happen if, on the contrary, we added identity to our list of *logical* terms so that '=' was required to *mean* identity or equality whatever the interpretation of the rest of the terms? It should be clear from our basic characterisation of validity that the effect of any addition to the list of logical terms will be to ***extend*** the class of valid inferences by ***restricting*** the class of possible counterexamples. Consider, e.g., the following intuitively valid inference in mathematics:

1. a = b
2. b = c

So, a = c

If we just treat identity as a two-place relation like any other, then the inference in first order logic is just:

1. Ra,b
2. Rb,c

So, Ra,c

As thus formalised, the inference is obviously invalid. Take, e.g., the interpretation **I:**

Domain: {natural numbers}

a: 1

b: 2

c: 3

Rx,y: y is the immediate successor of x (i.e. y = x + 1).

Interpretation **I** is a counterexample to the inference.

If, on the other hand, we treat identity as a ***logical*** term and hence require that its meaning remain the same in any interpretation, then the inference will obviously be valid. In that case, the only variable items, are the domain and the assignments of particular elements of the domain to individual constants, a, b and c; but whichever elements these are, clearly a = c clearly ***cannot be false***, when a = b and b = c are both true. (That is, if we have assigned the same element of the domain to both the constant a and the constant b, and we have also assigned the same element to both the constant b and the constant c, then we have *ipso facto* assigned the same element of the domain to the two constants a and c.)

Of course everyone would regard the above inference as intuitively valid. But this is not a knock-down argument for regarding identity as a logical term. After all, the inference:

1. Liverpool is north of Watford
2. Watford north of London

So Liverpool is north of London

is also intuitively valid. Yet it formalises – using, say, Nx,y for x is to the north of y – as:

1. Na,b
2. Nb,c

So, Na,c

which is invalid. (*Exercise:* Supply a counterexample).

The intuitive validity of *thislatte*r inference is clearly best explained, however, **NOT** by saying 'is north of' is a logical term (this seems obviously wrong), but by reflecting that weall carry round with us certain background information that we intuitively import as extra premises in ordinary arguments like this geographical one: so-called ***implicit* or *hidden* premises.**

In the geographical case, we all know that *whenever* one place is north of a second and the second north of a third, then the first is north of the third. (To put it in logician-speak, we all know that ‘is to the north of’ is a ***transitive* relation**.) When we "articulate" this hidden premise and add it as an *explicit* premise the inference becomes:

1. Na,b
2. Nb,c
3. $\forall x \forall y \forall z((Nx,y \,\&\, Ny,z) \rightarrow Nx,z)$ Initially implicit premise

---

So, Na,c

And this *is* valid (*Exercise:* supply the easy proof!).

Similarly, (in fact, logically speaking, identically) in the earlier inference we all know that whenever one thing equals a second and the second a third then the first equals the third. Adding $\forall x \forall y \forall z((Rx,y \,\&\, Ry, z) \rightarrow Rx,z)$ as an explicit premise turns the 'a = b, b = c, So, a = c' inference into a formally valid one, without needing to regard identity as a logical term (indeed formally it’s the same inference as the geographical one).

There are, however, at least ***two arguments*** for regarding identity, as distinct from ‘is to the north of’, as a logical notion. The first argument is that it does seem to be a genuinely general notion: the laws of identity are always the same whether one is talking physics, arithmetic, real analysis, economics or whatever.

The second argument is that it seems arbitrary that the notion 'There is at least one...' should be a logical notion (as it is of course in Predicate Logic where it is represented by ∃x…), while ‘There are at least two…’, ‘There are at least three...’ *etc.* are not. Such notions **all become purely logical if we treat identity as a logical term**. *E.g.* 'There are at least 2 things with property P' is:

(*) $\exists x \exists y(Px \,\&\, Py \,\&\, \neg(x = y))$.

Notice carefully that the last conjunct is necessary – just $\exists x\exists y(Px \,\&\, Py)$ is, as we stressed earlier, consistent with there being just one thing that has property P. Just involving two variables does NOT mean that there have to be 2 different things that are P. (In fact $\exists x\exists y(Px \,\&\, Py)$ is logically equivalent to just $\exists xPx$.)

(*Exercise: (perhaps a surprisingly tricky one)* show that this logical equivalence holds.)

Notice just as carefully that everything in this expression (*), aside from the predicate P, is logical, if identity '=' is regarded as such. "At least two" – despite being a slightly more complicated expression – would stand on a logical par with 'there is at least one'.

Moreover, it seems difficult to understand why, *e.g.,* there is at least one thing (such that P, or whatever) should count as a logical notion while '**There is exactly one thing'** (such that P or whatever) does not. And again 'There is exactly one thing such that...' also becomes a purely logical notion if identity is treated as logical. This is because it is expressed by $\exists x(Px \,\&\forall y(Py \rightarrow y = x))$ (or, equivalently, $\exists x\forall y(Py \leftrightarrow y=x)$)

(*Difficult) Exercise:* show that these two formulations are indeed equivalent.

Again, everything in this expression aside from the predicate P is logical if '=' is. If we do regard '=' as logical, then, similarly, 'There are ***exactly two*** things with property P' is also a logical notion (aside from the descriptive P), since it formalises as $\exists x\exists y(Px \,\&\, Py \,\&\, \neg(x = y) \,\&\forall z(Pz \rightarrow (z =x \vee z = y)))$. And so on, with 'exactly n' for *any* n.

C11(B): FIRST ORDER LOGIC WITH IDENTITY: LANGUAGE AND INTERPRETATION

Suppose, then, that we ***do*** decide that identity should be regarded as a logical notion, and hence should be incorporated into logic (thus creating a system called ***First Order Predicate Logic with Identity***).

This extension of our logic is easily achieved. On the semantic side (remember, semantics concern ***interpretations*** and their use to establish invalidity, consistency, *etc.*), we just regard '=' as a logical and therefore non-reinterpretable term, as we already indicated.

However, this is also the place to make one further addition to the **expressive power of our language** (we could have made this addition earlier but it would then havecomplicated matters without any real pay-off).

First order logic with identity is very suitable as a basis for mathematical theories and for scientific theories based on mathematics. In mathematics much use is made of ***functions.*** These are 'mappings' or 'rules of association' which take one individualand ***map it onto*** or associate it with another.

One example is the **doubling function,** usually written f(x) = 2x. It takes any natural number, say, and maps it onto its double (1 onto 2, 2 onto 4, 3 onto 6, *etc.*). Similarly, the **squaring** function, g(x) = x^2, takes any number and maps it onto its square (1 onto 1, 2 onto 4, 3 onto 9, *etc.*)

We can have functions of any (finite) number of arguments. For example, we can characterise a two–place **summing function** h(x,y) = z that maps any *pair* of numbers onto another number, *viz.* their sum. So h(1,2) = 3, h(7,9) = 16 *etc*.

(We could also characterise functions in domains of a non-mathematical kind. For example, nothing prevents us from characterising a "father function" in the domain of humans, namely the function *f* that associates any human with his or her father or takes any human and maps her/him into her/his father. So f(Cain) = Adam, f(Harry) = Charles, *etc.* It's just that it's not usual to talk in this way, except in mathematics.

Notice however that there is, for example, no 'brother function' – because (a) unlike the father case, not everyone has a brother so it is not defined for all elements of the intended domain (humans), and, (b) more importantly some people have more than one brother, so for example Brother(Charles) does not pick out one definite entity – both Andrew and Edward are his brothers – and so is not representable as a function.)

The introduction of functions greatly enlarges the class of **terms** (see above). Individual constants, individual variables, and "ambiguous names" are, remember, all terms. But now we add all functions applied to the correct number of terms: that is, we stipulate that:

If f is any n-place function symbol and t_1 ... t_n are all terms, then soalso is f(t_1 ... t_n) a term.

So, for example, if f(x,y) = z is the two-place summing function and *a* and *b* are individual numbers then f(a,b)[=a+b] also names a number and hence is a term; so is f(x,y) for variables x and y. Notice also that f(f(a,b),c) is well-defined, being the sum of the sum of a and b, and c. The definition of terms allows for iteration (indeed *any* (finite) number of iterations).

It was with the eventual introduction of functions in mind that I gave the general definition of a term being *free for a variable* earlier. (Remember this notion was necessary for an accurate description of the rule of Universal Specification.) Not only is y *not* free for x in ∃yRx,y, nor is any other term – like f(y) or g(x,y) or h(y,z,w) that *involves y.* Any such term when substituted for the free variable x in∃yRx,y gets 'captured by a pre-existing quantifier'.

Since f(x), g(x,y) *etc.* are terms, the rule of Universal Specification (US) allows us to infer from, say, ∀xPx, not just Px or Pa or Pα but also Pf(x) or Pg(x,y), *etc.* The rule about the substituted term being free for the variable still applies, though, and so, since f(y) is *not* free for x in ∃yRx,y, we could not infer from ∀x∃yRx,y to, for example, ∃yR(f(y),y) by US. Similarly, the rule of Existential Specification allows us to go from ∃xPf(x), for example, to Pf(α).

Naturally all the conventions about flagging variables and subscripting ambiguous names apply equally well when we introduce variables or ambiguous names within the context of some function. You should go back to the statement of the rules of proof (full form) and think about how they apply when we allow Specification – either Universal or Existential – using appropriate terms in the new extended sense: extended to take functions into account.

C11(c): Rules of Identity

So far as *proofs* go, as just indicated, the rules we already have carry over to the new system and in exactly the same form (they just become more powerful because applied to an extended language involving a wider notion of terms). But **two further rules,** specifically about identity, must be added to the already existing stock to extend first order logic to first order logic with identity. These are:

First rule of identity (I1):

The formula **t** = **t** for any term **t** may be derived from theempty set of premises (i.e. may be written down at any stage in a proof).

Second rule of identity (I2):

Let t_1…$t_n$and s_1…$s_n$all be terms. Then if, for alli ε {1, … ,n}, s_i is free for $t_i$in formula **F**, the formula **F'** obtained from **F** by substituting some or all of the s_i for some or all occurrences of the corresponding t_i may be inferred from **F** and the formulas $s_1=t_1$,…, $s_n=t_n$.

The first rule is obvious and Rule **I2** is actually less fearsome than it looks, as will become clear from a few examples. (It is sometimes called the principle of the '**indiscernibility of identicals**' because it basically says that if 'two' objects are identical (this basically means we have two different *names* for the *same* object) then they have all the same properties ('A rose by any other name would smell as sweet'!)

Examples of the use of I2:

1. Pb may be inferred from Pa and a = b (so, 'Cassius Clay was a great boxer' can be inferred from 'Muhammad Ali was a great boxer' and 'Muhammad Ali = Cassius Clay').
2. ∀yQx,y may be inferred from ∀yQz,y and z = x.
3. (Note that ∀yQy,y cannot be inferred from ∀yQz,y and z=y, since y is **not free for z** in ∀yQz,y.)
4. R(f(x_1), … (f(x_n)) may be inferred from R(x_1 … x_n) and x_1=f(x_1) …. x_n=f(x_n)

C11(D): Derivations in First Order Logic with Identity

Example 1:

1. Featherstonehaugh (a) killed the Russian spy (Px).
2. Anyone who killed the Russian spy was in Paris at the time (Qx)
3. Anyone who was in Paris at the time was not in Berlin at the time (Rx)
4. 006 (b) was in Berlin at the time

So, Featherstonehaugh is not 006

Proof:

1.	Pa	Premise
2.	$\forall x(Px \to Qx)$	Premise
3.	$\forall x(Qx \to \neg Rx)$	Premise
4.	Rb	Premise
5.	$Qb \to \neg Rb$	US, 3
6.	$\neg Qb$	TI, 4,5
7.	$Pb \to Qb$	US, 2
8.	$\neg Pb$	TI, 6,7
9.	$a = b$	A
10.	$\neg Pa$	I2, 8,9
11.	$Pa \,\&\, \neg Pa$	TI, 1,10
12.	$(a = b) \to (Pa \,\&\, \neg Pa)$	CP, 9-11
13.	$\neg(a = b)$	TI, 12

Example 2:

Show that $\forall x(Px \leftrightarrow \exists y(x = y \,\&\, Py))$ is a logical truth of First Order Logicwith Identity.

Proof:

1.	Px	A x
2.	$\forall y \neg(x = y \,\&\, Py)$	A x
3.	$\neg(x = x \,\&\, Px)$	US, 2 x

4. $\neg(x = x)$ v $\neg Px$ TI, 3 x
5. $x = x$ I1 x
6. $\neg Px$ TI, 4,5 x
7. $Px \& \neg Px$ TI, 1,6 x
8. $\forall y\neg(x = y \& Py) \rightarrow (Px \& \neg Px)$ CP, 2-7 x

Comment 1: Notice that x remains flagged here since although we have discharged one assumption, it was not the assumption (line 1) in which x was introduced free.

9. $\neg\forall y\neg(x = y \& Py)$ TI, 8 x
10. $\neg\exists y(x = y \& Py)$ A x
11. $(x = y \& Py)$ A x,y
12. $\exists y(x = y \& Py)$ EG, 11, x,y
13. $(x = y \& Py) \rightarrow \exists y(x = y \& Py)$ CP, 11-12 x

Comment 2: x still remains flagged since although we have discharged one assumption (line 10) in which x was introduced free, the assumption at line 1 remains undischarged.

14. $\neg(x = y \& Py)$ TI, 10,13 x
15. $\forall y\neg(x = y \& Py)$ UG, 14 x

Comment 3: Here we universally generalize, but on the unflagged variable y (we could not legitimately generalise on x since it remains flagged).

16. $\forall y\neg(x = y \& Py) \& \neg\forall y\neg(x = y \& Py)$ TI, 9,15 x
17. $\neg\exists y(x = y \& Py) \rightarrow (\forall y\neg(x = y \& Py) \& \neg\forall y\neg(x = y \& Py))$ CP, 10-16 x
18. $\exists y(x = y \& Py)$ TI, 17 x
19. $Px \rightarrow \exists y(x = y \& Py)$ CP, 1-18

Comment 4: Here at last the flagging on x stops (though we start to flag it again in the next line, which begins the 'other half' of the proof.)

20. $\exists y(x = y \& Py)$ A x
21. $x = \alpha_x\& P\alpha_x$ ES, 20 x
22. $x = \alpha_x$ TI, 21 x
23. $P\alpha_x$ TI, 21 x

24. Px	I2, 22,23 x
25. $\exists y(x = y \& Py) \to Px$	CP, 20-24
26. $Px \leftrightarrow \exists y(x = y \& Py)$	TI, 19,25
27. $\forall x(Px \leftrightarrow \exists y(x = y \& Py))$	UG, 26

Example 3:

1. There are even numbers.
2. There are odd numbers.
3. No number that is odd is even.

So, there are at least two numbers.

Proof:

1. $\exists x(Nx \& Ex)$	Premise
2. $\exists x(Nx \& Ox)$	Premise
3. $\forall x((Nx \& Ox) \to \neg Ex)$	Premise
4. $N\alpha \& E\alpha$	ES, 1
5. $N\beta \& O\beta$	ES, 2
6. $(N\beta \& O\beta) \to \neg E\beta$	US, 3
7. $\neg E\beta$	TI, 5,6
8. $\alpha = \beta$	A
9. $\neg E\alpha$	I2, 7,8
10. $E\alpha \& \neg E\alpha$	TI, 4,9
11. $(\alpha = \beta) \to E\alpha \& \neg E\alpha$	CP, 8-10
12. $\neg(\alpha = \beta)$	TI, 11
13. $N\alpha \& N\beta \& \neg(\alpha = \beta)$	TI, 4,5,12
14. $\exists x(Nx \& N\beta \& \neg(x = \beta))$	EG, 13
15. $\exists y \exists x(Nx \& Ny \& \neg(x = y))$	EG, 14

Example 4:

$\exists x \exists y(x = y \& \neg(y = x))$ is a logical falsehood in Predicate Logic with Identity.

Proof:

1.	$\exists x \exists y(x = y \,\&\, \neg(y = x))$	A
2.	$\exists y(\alpha = y \,\&\, \neg(y = \alpha))$	ES, 1
3.	$\alpha = \beta\& \neg(\beta = \alpha)$	ES, 2
4.	$\alpha = \beta$	TI, 3
5.	$\neg(\beta = \alpha)$	TI, 3
6.	$\alpha = \alpha$	I1
7.	$\beta = \alpha$	I2, 4,6
8.	$(\beta = \alpha) \,\&\, \neg(\beta = \alpha)$	TI, 5,7
9.	$\exists x \exists y(x = y \,\&\, \neg(y = x)) \rightarrow (\beta = \alpha) \,\&\, \neg(\beta = \alpha)$	CP, 1-8
10.	$\neg\exists x \exists y(x = y \,\&\, \neg(y = x))$	TI, 9

D: INFORMAL REASONING: PREDICATE LOGIC

As suggested earlier, in Section **B**, the fact that deductive logic is the logic behind ordinary informal reasoning is obscured by the fact that we seldom spell out our arguments in the detail required for a full demonstration of deductive validity. Instead, we leave various premises implicit because we assume they will be presupposed by those who hear our arguments. We saw some examples of this – and of the value of spelling out the initially hidden premises – in section **B** with informal arguments whose validity can be captured in truth functional logic. In this brief section we look at arguments where predicate logic can usefully be involved.

Example 1:

An Israeli officer, during one of the wars with Egypt in the 60s or 70s, was reported as arguing: "The man who parachuted out of the Egyptian plane had blond hair. So he must have been Russian." This clearly *is* an argument. The "so" signals the conclusion: *viz.* that the man was Russian. The only *explicit* premise is "The man who parachuted out of the Egyptian plane had blond hair". The maninvolved was obviously a particular individual – let's therefore introduce the individual constant "a" for him. Then introducing Px: "x parachuted out of the Egyptian plane", Qx: "x has blond hair", and Rx: "x is Russian", we get the following formalisation for the inference *as it stands:*

1. Pa *&* Qa

Therefore, Ra

This is of course ***invalid.*** (*Exercise:* Supply a counterexample.) Yet we can imagine that *in the circumstances* the argument was quite convincing. How can this be if the logic of ordinary argument is the deductive logic we have investigated?

Well, the answer, as in the cases in Section **B**, is that the Israeli officer was *taking for granted* certain assumptions that he did not bother to articulate. These are so-called *implicit* or *hidden* assumptions.

There can, of course, be no rules for identifying these – but only more or less plausible conjectures. It does seem fairly clear that in this case the officer was making two assumptions: (a) that only two types of people were involved on the side of his enemy – *viz.* Egyptians (openly) and Russians (covertly) and so anyone who parachuted out of the Egyptian plane was either Russian or Egyptian; and (b) that no Egyptian has blond hair. Introducing the predicate Sx for 'x is Egyptian', then (a) formalises as:

$\forall x(Px \rightarrow (Rx \lor Sx))$

and (b) formalises as:

$\forall x(Sx \rightarrow \neg Qx)$ (or, equivalently, $\neg\exists x(Sx \,\&\, Qx)$).

If we now add these initially hidden premises as *explicit* premises, then we get the following inference:

1. $Pa \,\&\, Qa$
2. $\forall x(Px \rightarrow (Rx \lor Sx))$
3. $\forall x(Sx \rightarrow \neg Qx)$

Therefore, Ra

This is, of course*, valid* in first order logic (*Exercise: supply the easy proof*).

Example 2:

As the Boston Celtics were winning the 1976 NBA championship by beating the Phoenix Suns in Arizona, a television pundit said that the Celtics "are proving that they are a great basketball team, because you can't claim to be a great team if you can't win on the road."

Here's one reconstruction:

Let Px: x is a basketball team, Rx,y: x wins at y, Qx: x is the home team, Sx: x is great, a is Boston Celtics, b Phoenix Suns, THEN the pundit's argument is:

1. $\forall x((Px \rightarrow (\neg(\exists y(Rx,y \,\&\, Qy) \rightarrow \neg Sx)$ (explicit)
2. $Pa \,\&\, Pb$ (implicit)
3. $Ra,b \,\&\, Qb$ (implicit in pundit's remark, but explicit in my introduction to it)

So, Sa

This is invalid as it stands (*Exercise:* supply a counterexample).

To make it valid you would have to add the extra assumption that if a team can win on the road then it is great; or equivalently to turn that final '→' in the first premise into a biconditional.

(*Exercise*: make sure you understand the formalisation and all these comments.)

Example 3:

Here is a 'theological' example: Consider the following passage:

> *Some fundamentalists maintain that while Adam possessed original sin and every descendant of someone with original sin himself has original sin, nonetheless Jesus did not possess original sin. This was because of the virgin birth. But surely those fundamentalists are wrong? After all, Jesus was still a descendant of Mary.*

Again some arguing is going on here, but again we need to think a bit to bring the exact structure out. The main conclusion is that Jesus did possess original sin; a sort of subsidiary conclusion is that fundamentalists are wrong in thinking that he did not possess original sin. The main argument is:

1. Jesus was a descendant of Mary (explicit)
2. Adam possessed original sin (or actually in his case brought original sin upon himself and hence the whole human race by flouting God's instruction not to eat the apple!) and every descendant of someone with original sin has original sin. (explicit, more or less)
3. Mary was a descendant of Adam (implicit – but of course, given the Genesis account, she could not be anything else)

Therefore, Jesus had original sin.

Taking Rx,y to mean" x is a descendant of y", Px: " x has original sin"; and a, m, and j to be the obvious choices for individual constants for Adam, Mary and Jesus we have:

1. Rj,m

2. $Pa \;\&\forall x\forall\; y((Rx,y \;\&\; Py) \rightarrow Px)$
3. Rm,a

So, Pj

This argument is valid. (*Exercise:* perform the relevant proof.)

As always, if someone (in this case a fundamentalist) wanted to challenge the conclusion of a valid argument, s/he must also challenge at least one premise. Here s/he would presumably want to reject the general claim in the second premise as it stands and adopt instead 'Everyone who is descended from two parents each of whom has original sin himself has original sin'. And to add that someone descended from two parents one of whom did not have original sin did not himself have original sin. This last – highly contentious – premise together with the rest of the premises would allow us, contrary to the initial case, validly to infer the fundamentalist's initial conclusion that Jesus did not have original sin. (Of course this would also require a further, so far unstated premise about Jesus – which premise? *Exercise:* formalise the second version of the premise about descendants and original sin and show that you can now validly infer that Jesus did not possess original sin.)

Example 4:

Raymond Smullyan's entertaining book called "What is the name of this book?" contains a lot of interesting anthropological detail about two islands: the island of Knights and Knaves (which I introduced you to briefly in Section **A1**) and the island of Knights, Knaves and Normals. (Knights always tell the truth; Knaves always lie and Normals sometimes lie and sometimes tell the truth.) There was a famous court case on the second (tri-partite) island. Three inhabitants (A, B and C) were charged with murder. Inspector Knacker of the Yard (flown in from England and therefore an honorary knight!) discovered for sure (that is, you are to take these as premises) that only one was guilty, that the guilty one was a Knight and that the guilty one was the only Knight among the three. The court case did not last long (to the great chagrin, of course, of the lawyers involved), since all that happened was that the three defendants made one statement each:

A: I am innocent.

B: That is true.

C: B is not normal.

Fortunately, all the jurors had previously attended this course and were quickly able, via an informal argument, to identify the guilty party. ***(Try to work it our yourself before reading on.)***

The reasoning is this: A can't be the Knight, since if he were he would be guilty, (we know that whoever is the Knight is guilty) but then would have lied in his statement and that's impossible for a Knight. A also can't be a Knave, since if he were he would be innocent (the guilty party is known to be a Knight – by 'premise 1') and hence what he said would be true, which is impossible for a Knave. So A is normal (and innocent – it is given, remember, that the only knight among the three is the guilty one). Since A's statement is therefore true, so is B's. Hence B cannot be a Knave. He is either a Knight or a Normal. If he were a Normal then C's statement would be false and so C would have to be either a Knave or a Normal. This would mean there was no Knight among the three, but we know from the premises that there is just one Knight. Therefore, B is not a Normal. Since we already decided he is not a Knave, he must be a Knight and so the guilty one.

This reasoning consists of a whole *series* of valid inferences. (This is typical of much reasoning which is sequential – like a proof.) Some steps are *reductio ad absurdums*.

You could have some "fun" capturing some of these inferences in, say, first-order logic. (In that logic we could give the following characterisation of a Knight: Knight(x) = $\forall y((Sy \,\&\, Ax,y) \rightarrow Ty)$.Here Sx: x is a statement; Tx: x is true; and Ax,y: x asserts y.)

(*Exercise:* Do the same for Knave(x); and prove (informally) that no inhabitant of the island of Knights and Knaves ever said 'I am a Knave'.)

E: Some Problems in the Foundation of Logic

An inference is deductively valid, we have found, if all models of its premises are also models of its conclusion (no counterexample). Our basic idea of a *model* of a set of sentences appears trouble-free. But it involves two notions: that of a *set* (the domain of the interpretation must be a set) and that of truth (the sentences must all be *true* for the interpretation to be a model) each of which has interesting difficulties associated with it. In the rest of the course, I shall indicate these difficulties, as examples of the interesting problems that occur in the foundations of logic (and which are studied in other courses in the Philosophy Department).

E1: TRUTH

E1(A): THE CORRESPONDENCE THEORY OF TRUTH, AND A DIFFICULTY:

The notion of truth at first appears trouble-free. So far as interpretations go, it is (declarative) *sentences* or assertions whose truth or falsity we are interested in. (Of course, we use the term 'true' in other contexts, such as true *feelings* – but for logical purposes, we are interested only in sentences.) What does it take for a sentence to be true? And what does it take for a sentence to be false? This seems a dumb question because it has an entirely obvious answer. A declarative sentence asserts that some state of affairs holds. For example, the sentence '(Pure) water freezes at 0°' asserts that water has a certain property and it is true because water indeed has that property. 'Electrons are positively charged' similarly makes a claim about the world being a certain way and it is false because it is not the case that electrons are positively charged (in fact, they are negatively charged). It seems obvious then that a declarative sentence or assertion is true just in case what it says is the case *is indeed* the case. Or as Aristotle put it:

> *"To say of what is that it is, or of what of is not that it is not is true; while to say of what is not that it is or of what is that it is not is false."*

This is the classical ***correspondence theory*** **of truth**: a sentence is true iff it corresponds with the 'facts' (in a wide sense of 'facts'). The correspondence theory implies the truth of all instances of the following schema:

X is true iff p

where p is a declarative sentence and X is a *name* of p – often formed by putting quotation marks around the sentence. So, the schema has the following instances, among many (infinitely many) others:

"Snow is white" is true iff snow is white.

"Libya is a peaceful place" is true iff Libya is a peaceful place.

Both are, of course, true (the first bi-conditional being true on both sides, the second false on both sides). Such bi-conditionals are not logically true – to be logically true they must say the same thing (in different words) about the same entities; but these bi-conditionals have an assertion about a sentence on one side and about things in the world on the other. But those bi-conditionals do seem trivial nonetheless: they express the obvious connection between a true sentence and the state of affairs (in the simple case states of affairs in the world) the sentence asserts to hold.

However, interestingly things turn out not to be quite so straightforward. To see this, first note that there is no special problem in applying the schema to sentences which themselves happen to be about some particular sentence instead of about snow or Libya or whatever. Just the same condition surely applies. Consider the sentence:

> "The first sentence in today's *Times* is true" is true iff the first sentence in today's *Times* is true.

Or the sentence:

> "All the sentences in *Genesis* are false" is true iff all the sentences in *Genesis* are false.

(Notice again that these bi-conditionals are not logically true: the left hand side of the first one is a sentence about a sentence (it makes an assertion about a sentence) whereas the right hand side is a name of a sentence – it might, for example, name the sentence 'There was a major disagreement at yesterday's meeting of the Cabinet.')

But now consider the sentence:

> The only sentence in red in these notes is false.

Apply the schema to it. We get: "The only in red in these notes is false" is true iff the only sentence in red in these notes is false. (Call this bi-conditional *)

But, in view of the fact that the only sentence in red in these notes is "The only sentence in red in these notes is false", the sentence * is logically contradictory. For assume that its LHS is TRUE, then it is indeed true that the only sentence in red in these notes is false, i.e. "The only sentence in red in these notes is false" is false. And this contradicts the LHS.

On the other hand, if the LHS is FALSE, then it is false that "The only sentence in red in these notes is false". So, assuming the sentence makes sense, which it certainly seems to, that sentence must be true but that sentence *is* "The only sentence...". So we have now derived that the sentence must be true from the assumption that it is false *plus* the truth schema. Thus the truth-schema implies that this sentence is true iff it is false.

So this instance of the truth schema is logically contradictory. The correspondence theory of truth, which seemed so obvious, implies a logical falsehood **andso is itself logically false!**

The reasoning we just went through is a rather precise version of the so-called ***Paradox of the Liar***. This originates with Epimenides the Cretan who allegedly said "All Cretans are liars" (a fact which St. Paul reported without apparently noticing anything funny about it). Hence the paradox is also sometimes called the *Epimenides.* (However, the Epimenides statement is not actually contradictory (*Exercise*: why not?). We need something more direct like "I am now lying" or "This present sentence is false" or the more precise version given above. Suppose we take the direct Liar version: 'I am now lying, *i.e.* what I am now saying is false. This is true if what it states to be the case is the case; but what it states to be the case is that it is false!)

The term 'paradox' may suggest that it is an engaging puzzle rather than a deep problem. But this is not correct. As we have seen it *refutes* the straightforward version of the apparently obviously correct account of what it means for a sentence to be true.

E1(b): Are "self-referential" Statements Meaningless?

Many logicians/philosophers argued that what the Liar Paradox shows us is simply that self-referring sentences are really ***meaningless*** (although they appear grammatically correct). A sentence is, of course, self-referring if it ascribes some property to itself. So "This sentence is false" is clearly self-referring and so was: "The only sentence in red in these notes is false" – that sentence too said *of itself* that it is false. The suggestion is that such sentences don't really mean anything – no wonder then that they appear to lead to inconsistency; but if we restrict our theory of truth to meaningful sentences (as we clearly should) then no problem arises.

But is this suggestion tenable? There are all sorts of statements that are self-referring and yet which seem to make perfect sense and indeed seem to be true. A favourite example is the car-sticker on the rear window that reads "If you can read this you are too close". What is "this" here? Well of course, it's the sentence "If you can read this you are too close". In appropriate situations, this sentence, far from being meaningless, seems to be true.

Moreover, the suggestion would *not,* even if accepted, solve the problem. For we can easily construct sentences that singly do not refer to themselves but which together give a contradiction similar to the liar.

The story goes that someone pushed a visiting card under Bertrand Russell's door, the two sides of which read as follows:

The sentence on the other side of this card is true.
The sentence on the other side of this card is false.

Neither sentence taken alone refers to itself, so both would be counted meaningful even if we barred all self-referential statements as meaningless. And indeed if one of them, say the second, were replaced by pretty well any other sentence – say 'Liverpool FC will soon become once again the best football team in the country' – then the visiting card would just amount to an elaborate way of saying that Liverpool FC will soon become once again the best football team in the country and the remaining first sentence (‘The sentence on the other side of this card is true’) would present no conceivable difficulty. Similarly, for the second sentence together with any ‘normal’ sentence – this would just be anelaborate way of saying that the ‘normal’ sentence is false.

So the sentence ‘The sentence on the other side of this card is true’, and the sentence ‘The sentence on the other side of this card is false’ taken separately do not self-refer and taken separately present no problem. But together of course they become 'paradoxical'. Assume that the sentence on the first side is true then it truly states that the sentence on the other side is true so the sentence on the other side *is* true; but then it is true that 'The sentence on the other [i.e. first side, which we started from] of this card is false', i.e. the sentence on the first side of the card is false, contrary to our assumption. So we must assume that the sentence on the first side is false. If so, it falsely states that the sentence on the other side is true, which can only mean (the sentence surely makes sense) that the sentence on the other side is false. But then it is false that the sentence on the other side is false, again contrary to assumption. We have an inconsistency. This is usually called ***The Visiting-Card Paradox.***

A still more important consideration is that self-referential statements are perfectly consistently, and indeed usefully, dealt with in various branches of mathematics: notably set theory and mathematical logic. Indeed, reasoning analogous to that underlying the Liar paradox is put to constructive effect in proving Gödel’s incompleteness theorems – central results in mathematical logic.

E1(c): Tarski’s Language-Hierarchy

The rehabilitation of the correspondence theory of truth in the face of the above 'paradoxes' is due to the Polish-American logician Alfred Tarski, who died in 1982. He pointed out that there is a natural distinction between an **'object-language'** statement like 'Mary had a little lamb' and a **'meta- language'** statement like "'Mary" has four letters' or “’Mary had a little lamb’ is the first line of a famous nursery rhyme.’ Statements of the first kind are about *physical* objects (albeit in this case *pretend* physical objects), while statements of the second kind are about *linguistic* objects like names of individuals (rather than individuals themselves) or like sentences.

'Snow is white' is an object language assertion about the physical entity 'snow'. But the statement "'Snow is white" is true' is a meta-language statement about an object language entity, *viz.* a sentence. In the meta-language, we can refer *both* to sentences *and* to objects. So, for example, the instance of our truth schema:

"Snow is white" is true iff snow is white

is a meta-language assertion which talks about *both* a sentence (on the left hand side) *and* an object – namely, snow (on its right hand side).

Since we can discuss meta-linguistic assertions in turn this points to the existence of meta-meta-languages in which we can for example make the assertion that a particular meta-language sentence is, say, false. And then there are meta-meta-meta-languages and so on. Although in natural languages, like English, we switch without noticing it between levels – English grammar is taught in English – Tarski's suggestion was that formal correctness requires us to differentiate linguistic levels and in particular to recognise that whenever we assert that a particular sentence *S* is true (or false) we do so in the meta-language of whatever language *S* happens to be in. More formally, the predicate "is true" is to be regarded as incomplete: it must always be taken as meaning "is true in (or "is a true sentence of") particular language L". This predicate cannot be a predicate of that same language L but must instead be a predicate of L's meta-language (or some language higher up the hierarchy).

There can be within language L no predicate equivalent to L's truth predicate – on pain of the Liar Paradox being derivable. But, so long as this condition is met, the Liar Paradox is *not* derivable.

When I said 'The only sentence in red in these notes is false' this was, if we accept Tarski's analysis, an incomplete assertion. To complete it we must specify which language that sentence is in. Let us try to derive the paradox again – this time specifying the language L in which it is expressed. We have:

The only sentence of language L in green in these notes is false.

According to Tarski this sentence itself must be at least in the meta-language of L. Hence *there is no sentence of language L* in green in these notes. Instead of, as before, a sentence which "asserts its own negation", we now have a sentence of the meta-language of L which asserts that some *non-existent* object-language sentence is false.

This is no longer contradictory. Its status depends on how we decide to treat the separate problem of ascriptions of truth values to statements about non-existent entities. Is, for example, the statement "The present King of France is bald" true, false or neither? Following Bertrand Russell's 'Theory of Descriptions' it is usual to analyse such statements as asserting that **there is at least one thing which is the present king of France***and* there is no more than one such thing *and* that thing is bald. This makes such a statement unambiguously false (the first of the three conjuncts is false). This would mean that the above sentence in the meta-language of L is not paradoxical but simply false.

For more information about the Liar Paradox, Tarski's language-hierarchy solution, and other solutions which have been suggested, consult the Stanford Encyclopedia of Philosophy's Liar Paradox article.

E2: THE PARADOXES OF SET THEORY

E2(A): THE PRINCIPLE OF COMPREHENSION

Aside from the notion of truth, the other notion involved in the key idea of an interpretation/model is that of *set.* It may have occurred to some of you to wonder why it is required that a particular set be specified as the 'domain of the interpretation'. Why not take for the domain of any interpretation *the* universe – the set of all things, concrete and abstract, physical and mathematical? So that when, in Predicate logic, we say 'for all' we *really* mean for *all* – the whole universe of things.

The answer, to put it dramatically if rather obscurely, is that the universe probably does not exist. (Bertrand Russell once said that he was proud that one could actually prove that there are *fewer* things in heaven and earth than are dreamt of in *his* philosophy.)

The inventor of set theory, the great mathematician Georg Cantor, took it as obvious that for any well-defined property there is a corresponding set – the set of entities that have that property. So corresponding to the property 'is red' there is a set, *viz.* the set of all red things; corresponding to the property 'is a natural number' there is a set, *viz.* the set of all natural numbers. This is indeed the basis of the idea that we can characterize predicates either *intensionally* (in terms of their meanings) or *extensionally* (as characterised by the set of all entities that satisfy the predicate). We used this idea inthe finite interpretation or finite model technique. Given *any* predicate Px, it is usual to write its extension as {x | Px} – read as: 'the set of all x such that Px'. Cantor's assumption is nowadays called:

The Naive Principle of Comprehension (sometimes of Abstraction):

Every property determines a set – or, more formally, for anypredicate Px, there is a set y, such that ∀x(x ε y ↔ Px). (Here x ε y means, as before, x is an element (or member) of the set y.)

(As we saw when dealing with finite models, the principle is extendable to predicates with more than one free variable – e.g. corresponding to the predicate Rx,y is a set of *ordered pairs* (x,y) such that Rx,y holds. So, corresponding to the predicate x>y in the natural numbers is the set of all ordered pairs of natural numbers such that the first element of the ordered pair is (strictly) greater than the second element of the pair.)

This principle seems no more than common sense. There are some funny properties – like being a natural number less than 0 or being a round square – which are satisfied by nothing; but these do not challenge Cantor's principle because one (important) set is the **empty set, ∅,** which contains no members. Hence these two properties and others like them *do* have a set as their extension in accordance with Cantor's principle – their extension being the empty set ∅. But, despite the fact that it seems so obviously true as to be trivial, Cantor's principle is not called 'naive' for nothing. It turns out to be wrong – indeed to be logically false – despite its intuitive appeal.

The fact that contradictions can be derived from the naive principle of comprehension is of wider significance than might at first be thought. Bertrand Russell, and slightly earlier, the German logician Gottlob Frege, both believed that the whole of mathematics could be *reduced* to *logic*– that mathematics consists in the end of nothing but logical truths. They included Cantor's set theory as part of logic (after all, first-order logic is the general study of predicates, and following the principle of abstraction, predicates and sets go hand in hand). It was an enormous blow to this *logicist programme* in the foundations of mathematics to discover that set theory, as it stood, was as far from being logically true as could possibly be – it was **logically contradictory**. The discovery put the logicist programme into a turmoil from which, according to most thinkers, it has never fully recovered. Why exactly is the naive principle of comprehension logically inconsistent?

E2(B): RUSSELL'S PARADOX

Given that *every* property determines a set, then any property that *applies to sets* also determines a set – one whose members themselves happen to be sets. This is certainly not problematic in itself. For example, the property of being a set with an even number of members determines a set – *viz*. the set of all sets which have an even number of members. The property of being a set of natural numbers determines the set of all sets of natural numbers.

This means that we can ask whether one set X is, or is not, a member of another set Y. And we can sensibly ask whether a set is a member of itself. Most sets are in fact not members of themselves – in order to be members of themselves they would have to satisfy their own defining characteristic. And most sets don't. *E.g.,* the set of all England cricketers is not itself an England cricketer (it's a set, after all, not a person!) and so is not a member of itself. The set of all physical objects in our galaxy is not itself a physical object in our galaxy (it's an abstract mathematical entity) and so not a member of itself.

Because most sets, indeed all of those we might normally think of, are not members of themselves – because they don't satisfy the predicate that has them as its extension – such sets are called **NORMAL sets**. A few sets are **abnormal** – though it takes a bit of ingenuity to think of examples. The easiest way is to think of 'negative properties' – like the property of *not* being an England cricketer. The set of all things which are not England cricketers is not itself an England cricketer and so *is* a member of itself. The set of all non-marijuana smokers does not itself smoke marijuana and so is a member of itself. There are also a few non-negative examples of abnormal sets, like the set of all abstract entities (itself an abstract entity and therefore a member of itself), and, more importantly, **the set of all sets** – itself a set and therefore a member of itself.

Bertrand Russell showed that we can derive a contradiction from the naive principle of comprehension by considering the property 'is a normal set'. This is a reasonable property – as we just saw, some sets (intuitively the *vast majority* of sets) satisfy the property, while a few sets, and of course all individuals (non-sets) do not satisfy it.

According to the principle, this property determines a set: the set of all normal sets, call it N. We can now ask whether N is itself normal – normality is a property of sets and N is a set, so this is a question we must be able to ask.

Well N either is normal or it isn't. Assume first that N *is* normal, then N is a member of the set of *all* normal sets, but that is N and so N *is* a member of itself. This is the defining characteristic of abnormality. So *if we assume N is normal we can derive that it isn't*.

We must conclude that N is *not* normal; but sets that are not normal are by definition members of themselves, so N is a member of the set of all normal sets, which means of course that it *is* normal. So *if we assume N is not normal we can derive that it is*.

Hence N is normal iff it isn't. This is “***Russell’s Paradox”.***

(Exercise: In view of the close connection between predicates and sets (intension and extension) it is not surprising that a paradox closely related to Russell's can be derived for properties. Define a monadic property as '**heterological**' if it fails to apply to itself. 'Long' for example is not long and so is heterological; 'in German' is in English *not* German and so is heterological. On the other hand, 'short' is itself short, and 'in English' is itself in English, and so both of these predicates are **homological** (sometimes 'autological', but in any event not-heterological). You should be able to derive a contradiction or 'paradox' by asking "Is the property of being 'heterological' itself heterological?" (Do it carefully!)—For more on this paradox, which is sometimes called the Grelling Paradox, or the Grelling-Nelson Paradox after its original authors, click here. There are a great many similarly-structured paradoxes and pseudo-paradoxes which you may also enjoy—Wikipedia has a good list.

Russell's paradox is derivable more formally as follows:

1. For any predicate Px, $\exists y \forall x(x \varepsilon y \leftrightarrow Px)$ (Naïve Principle of Comprehension)

So, substituting $\neg x \varepsilon x$ (i.e. ‘x is normal’ for P, we have:

2. $\exists y \forall x(x \varepsilon y \leftrightarrow \neg x \varepsilon x)$
3. $\forall x(x \varepsilon \alpha \leftrightarrow \neg x \varepsilon x)$ ES, 2
4. $\alpha \varepsilon \alpha \leftrightarrow \neg \alpha \varepsilon \alpha$ US, 3

(4) is, of course, a truth-functional contradiction.

This derivation, by the way, is in first order logic from line 2 onwards. We cannot fully express line 1 in first-order logic; to do so we would need to quantify not just over individuals (this includes sets considered as individuals) but also over predicates. This is not possible in first-order logic. This is why that logic is called 'first-order'. In **‘second-order’ logic** we *do* quantify over predicates as well as individuals, and the step from 1 to 2 becomes a simple instance of the rule of universal specification in that wider (and, as it turns out, interestingly problematic) system. Although nice and neat this formal derivation does not really capture the ‘paradoxicality’ of the paradox!

E2(c): Cantor’s Paradox

It turned out that Cantor himself was already aware (before Russell's demonstration) that his set theory is strictly inconsistent. Various other 'paradoxes' are in fact derivable within the theory – foremost amongst these is the one named after, and discovered by, Cantor himself. This involves the property 'x is a set' – this seems like an entirely OK property, it is satisfied by all sets and not satisfied by non-sets (individuals). According to the naive principle of abstraction, that property determines a set – *viz.* the set of all sets: the 'universal' set **U**. (So the ‘whole universe’ would consist of U together with all individuals.)

The assertion that U exists however can be shown to be contradictory – though unlike the straightforward Russell case, here a little work in set theory is required in order to exhibit the inconsistency.

(a) Elements, Subsets and Power Sets

So, first we need a little set-theoretical terminology. We already have $X \varepsilon Y$, for X is an *element of,* or a member of, Y. For example, if Y is $\{1,2,3\}$ then $1 \varepsilon Y$, $2 \varepsilon Y$, *etc.*; ifZ is $\{\{1\},\{2\}\}$ then $\{1\} \varepsilon Z$ and so is $\{2\}$, but $\neg(1 \varepsilon Z)$.

A set X is a *subset* of a set Y, written $X \subseteq Y$, if (and only if) every member of X is a member of Y. That is:

Subsets:

$$X \subseteq Y \leftrightarrow \forall x(x \varepsilon X \rightarrow x \varepsilon Y)$$

So, e.g., $\{1,2\} \subseteq \{1,2,3\}$ (but $\neg(1 \subseteq \{1,2,3\})$ – it’s a member, not a subset; and $\{\{1\}\} \subseteq \{\{1\}, \{2\}, \{3\}\}$ (but $\neg(\{1\} \subseteq \{\{1\}, \{2\}, (3\}\}$ – again, it’s a member not a subset).

(*Exercise:* make sure that you understand these claims.)

X is a *proper* subset of Y, written $X \subset Y$, iff $X \subseteq Y$ and $\exists x(x \varepsilon Y \ \& \ \neg(x \varepsilon X))$ – that is,**every element of X is in Y, but at least one element of Y is left out of X.**In other words, X is a proper subset of Y if and only if Y contains everything in X, *and something more*.

One slightly odd fact is that the *empty set,*written ∅, the set with no members, **is a subset of *every* set**, since of course it is true for any set X that $\forall x(x \varepsilon \emptyset \rightarrow x \varepsilon X)$. (*Exercise*: explain carefully why.)

Another slightly odd fact is that **every set is a subset of itself**(but of course **nota proper subset**). (*Same exercise*.)

The *power set* of a set X, written **P**(X), **is the set of *all subsets of* X**. So, if X is the set {1,2,3} then **P**(X) is:

{{1,2,3}, {1,2}, {1,3}, {2,3}, {1}, {2}, (3}, ∅}.

The use of the term 'powerset' stems from the fact that if the initial set X has n members then **P**(X) has 2^n members. So in this case X has 3 members and **P**X has $2^3 = 8$. If X is {{1,2},3} then **P**(X) is {{{1,2},3}, {(1,2}}, {3}, ∅ }.

Exercise: What is **P**(X) if X is:

(i) {{1,2}}
(ii) {{1}, (2}, (3}}
(iii) {{{1,2}} ,3}

(b) One-to-one Correspondences (or "bijections")

We need just a couple more ideas from set theory, the first of which is that of a ***one-to-one correspondence*.**There are apparently societies that do not have the naturalnumber system (*e.g.* according to the anthropologist Benjamin Lee Whorf this was true of the Hopi Indians who had only the idea of one, two, and many). A member of such a society could nonetheless decide whether the number of, say, chairs in a given room was *the same as* the number of people in that room. Without counting either set, he or she could attempt to affect a one-to-one correspondence between the two sets– that is, try to associate each chair with **one and only one** person. If this attempt succeeded, he could infer that there are as many chairs as people in the room – however many that happens to be. If there are on the contrary always some chairs left over after any attempted pairing, then there are more chairs than people, and if there are always people left standing without a chair, then the set of people is bigger than the set of chairs.

All this applies to any two sets. This suggests that the notions ‘same number as’ and ‘bigger (or smaller) number than’ are logically prior to the notion of number itself. More formally, two sets X and Y are said to have the same number (or to have ***same cardinality*** or to be ***equinumerous)*** if there is a **one-to-one correspondence *f*** between X and Y. The cardinality of set X may be written |X| and so X and Y have the same cardinality, or same size, written |X| = |Y| iff there is a 1-1 correspondence f between X and Y.

If there is a 1-1 correspondence between X and some subset Y' of Y, then $|X| \leq |Y|$; and $|X| < |Y|$ just in case $|X| \leq |Y|$ and $\neg(|X| = |Y|)$. (This last definition might seem to be unnecessarily complicated. If there is a one-to-one correspondence between the whole of X and some proper subset of Y (this would correspond to the situation in which we attempted a one-to-one correspondence between the set of seats in some lecture room and the set of students attending a lecture and there were seats left over) then surely there are strictly more members in Y than in X (so in the case of seats left over, strictly more seats than students). This is indeed true in the case of finite sets, but, fascinatingly, not so in the case of infinite sets. Indeed, it turns out to be an invariable trait of infinite sets that they always contain proper subsets which have the same cardinality as the whole set! Hence the slightly complicated looking definition of $|X| < |Y|$.)

The straightforward, and indeed seemingly obvious, idea that two sets have the same number of elements just in case there is a one-to-one correspondence between them has some surprising consequences in the case of ***infinite sets*** (for finite sets it yields only completely unsurprising consequences – that, e.g. |X| = |Y| iff they have the same number n of elements).

One example is ***'Galileo'sParadox'*** – that **there are as many even natural numbers as there are natural numbers**.This is because f(x) = 2x is a one-one correspondence between the whole set of natural numbers N and the set of even natural numbers E. This is only a 'paradox' in the sense that it is rather odd (“paradox” means “outside or beyond orthodoxy”); but no formal inconsistency is involved: the result that Galileo proved is indeed the*correct* result. We just, in general, need to get used to the idea that X may be a *proper subset* of Y, i.e.$\forall x(x\varepsilon X \rightarrow x\varepsilon Y)$ and$\exists x(x\varepsilon Y \ \& \ \neg \ x\varepsilon X)$, and yet X and Y haveexactly the same number of elements, i.e. $|X| = |Y|$. (In fact, as I already noted, it turns out to be true of *every* infinite set that it has proper subsets of the same cardinality as itself.)

Cantor could also straightforwardly prove that the cardinality of the set of **rational numbers** (natural numbers plus ratios of natural numbers) is the same as that of the set of natural numbers, *despite* the fact that there are infinitely many rationals between any two natural numbers (the rational numbers are 'dense'). You can find a simple interactive demonstration of the result here.

The suggestion arises that all infinite sets may just have the same cardinality. This would make set theory relatively boring.Cantor in fact proved that this was **not true** when he proved that the set of all **real numbers** is of a **higher infinity** than the infinity of the (counting) natural numbers. (The real numbers, which can be presented in terms of their decimal expansions, are all the points on the real line, and include natural numbers and rational numbers and lots more besides:$\sqrt{2}$, for example, although certainly a real number – it's a point on the real line – is not a rational number). Cantor showed this by showing that there is **no one-to-one correspondence** between the set of the reals and the set of natural numbers (and so since the set of natural numbers forms a proper subset of the reals, there must be strictly more reals than there are naturals). But both sets are of course infinite, so the result shows that there are orders of infinity – *some infinities are greater than others.*

In fact, Cantor proved the stronger result that there is no one-to-one correspondence between {Naturals} and {reals between 0 and 1}! The proof, like many deep results in mathematics, is by *reductio ad absurdum*. We assume that there is a one-to-one correspondence between those two sets, deduce a contradiction, and so infer that there can be no such correspondence. Here is how it goes, in outline:

Suppose that there was a one-to-one correspondence *f* between {Naturals} and {reals between 0 and 1}. Assuming such a correspondence between any set X and {Naturals} amounts to the assumption that X can be *enumerated* or *counted* – that is, written as an infinite list, without any member of X being left out: the first element of X in the list is the element associated by f with the number 1, the second is the element associated by f with the number 2, and so on.

So, if the set {real numbers between 0 and 1} can be placed in one-to-one correspondence with {Naturals} then that set of reals can be written as a list. So think of all the reals between 0 and 1 (specified by their decimal expansion 0.13579865… or whatever) written as an infinite list in any order that you like. Suppose our list is:

f(1) = 0.	0	1	4	5	3	2	1	3	**…**
f(2) = 0.	1	3	4	5	1	1	2	3	…
f(3) = 0.	9	6	5	3	4	2	9	9	…
f(4) = 0.	0	0	0	0	0	0	8	7	…
f(5) = 0.	0	0	1	2	3	5	6	1	…
f(6) = 0.	7	7	7	8	7	5	4	3	…
f(7) = 0.	1	8	6	3	6	8	4	1	…
…	…	…	…	…	…	…	…	…	…

You can then form the ***diagonalelement*** out of that list: that is, form the number $0.a_{11}a_{22}a_{33}$….a_{nn}…. where $a_{11}$is the first number in the decimal expansion of the real number, whatever it may be, that is first in the list (of course this will be a digit between 0 and 9 inclusive), $a_{22}$is the *second number* in the decimal expansion of the *second number in the list*, and so on. So, in our list, the diagonal element is highlighted:

f(1) = 0.	**0**	1	4	5	3	2	1	3	**…**
f(2) = 0.	1	**3**	4	5	1	1	2	3	…
f(3) = 0.	9	6	**5**	3	4	2	9	9	…
f(4) = 0.	0	0	0	**0**	0	0	8	7	…
f(5) = 0.	0	0	1	2	**6**	5	6	1	…
f(6) = 0.	7	7	7	8	7	**2**	4	3	…
f(7) = 0.	1	8	6	3	6	8	**4**	1	…
…	…	…	…	…	…	…	…	…	…

The diagonal element for f here begins: 0.0350624…

Now we can form the "**anti-diagonal element**" by taking each number in the diagonal element and producing a different number – of course between 0 and 9 – say by adding 1 to the original (and taking 9+1 as 0.) This number is, then, $0.\bar{a}_{11}\bar{a}_{22}\bar{a}_{33}....\bar{a}_{nn}...$

For our list, the anti-diagonal element begins: 0.1461735…

Call the anti-diagonal element **d.** d is a real number between 0 and 1 – but it **cannot be on the list**. (Try to work out why before reading on).

If it were on the list, then it would have to appear at some finite point on it. (It is a deep fact about the list of natural numbers that although the list is infinite, every element on it appears at some finite place – all the infinitely many natural numbers are finite!) So, there must be some natural number *m* such that d appears at the mth place. But that *can't be true* since d, by construction, differs from whatever number it is that appears at the mth place in the mth place of their decimal expansions. So, the assumption that we can produce a one-to-one correspondence between {Naturals} and {reals between 0 and 1}leads to contradiction. Hence, there is no such one-to-one correspondence and so the infinity of the reals, even the reals between 0 and 1, is a **higher infinity** than the infinity of the natural numbers.

This is the easiest case of Cantor's '**diagonal method**'. There is a slight wrinkle involving ensuring that you do not have infinite lists of 9s in the antidiagonal d –that some of you at least might like to think through. (The problem is easily overcome.)

What has all this, fascinating as it may be, to do with the set of all sets leading to a paradox? Well Cantor generalised his diagonal result as follows:

Cantor's Theorem: For any set X, |**P**(X)|>|X|

The cardinality of the power set of a set X is strictly greater than the cardinality of X itself. (So for example the cardinality of the set of all sets of natural numbers, i.e. **P**(N), is greater than that of the set of natural numbers N itself. So, put dramatically, the infinity of the set of all sets of natural numbers is even more infinite than the set of natural numbers itself.)

Proof:

To prove that $|X|\leq|\mathbf{P}(X)|$ we need only show that there is a one-to-onecorrespondence between the whole of the set X and some subset of $\mathbf{P}(X)$. The function $f(a) = \{a\}$, i.e. the function that maps any element *a* of X onto the set whose only member is that element, clearly is such a correspondence: the set of all singletons (all sets which just contain one natural number) is clearly a subset, indeed a proper subset, of the set of *all* sets of natural numbers.

It only remains to be demonstrated that $\neg(|\mathbf{P}(X)| = |X|)$; and this requires a demonstration that there can be no one-to-one correspondence between X and the ***whole*** of $\mathbf{P}(X)$. The proof of this, just as in the real number case, is by ***reductio ad absurdum.*** Assume that there *is* such a one-one correspondence *f*. *f* associates elements of X with subsets of X (i.e. sets of elements of X). We can therefore ask of any element $a \varepsilon X$ whether or not it is in the subset of X associated with it by *f*, i.e. **is $a\ \varepsilon\ f(a)$**? Form the set of all elements X for which the answer is negative and call this set X' i.e. $X' = \{a\ \varepsilon\ X|\ \neg(a\ \varepsilon\ f(a))\}$. X' is ofcourse, a subset of X (there is nothing in X' which is not already in X) and so $X'\ \varepsilon\ \mathbf{P}(X)$. So, given that we have supposed that *f* is a one-one correspondence between X and $\mathbf{P}(X)$, somemember of X must be associated with X' by *f*, i.e. $\exists a'\ \varepsilon\ X$ such that $f(a') = X'$.

But now, trivially, either $a'\ \varepsilon\ X'$, or $\neg(a'\ \varepsilon\ X')$. Assume $a'\ \varepsilon\ X'$ *i.e.* $a'\ \varepsilon\ \{a\ \varepsilon\ X|\neg a\ \varepsilon\ f(a)\}$ and so, by the definition of X', $\neg(a'\ \varepsilon\ f(a'))$. But $f(a') = X'$, so $\neg(a'\ \varepsilon\ X')$. Hence the assumption that $a'\ \varepsilon\ X'$ proves untenable. Therefore $\neg(a'\ \varepsilon\ X')$. But, since $X' = f(a')$ this means that $\neg(a'\ \varepsilon\ f(a'))$; hence a' satisfies the defining characteristic of the set X' and so $a'\ \varepsilon\ X'$. This is a **contradiction.**

The assumption that there is a one-one correspondence between X and the whole of $\mathbf{P}(X)$ entails a contradiction and so must be false. This means that $\neg(|\mathbf{P}(X)| = |X|)$;and since the first part of the proof easily yielded that $|X| \leq |\mathbf{P}(X)|$ we finally have Cantor's theorem that $|\mathbf{P}(X)| > |X|$.

This apparently technical result has the most mindblowing consequences: it demonstrates that, *rather than there being just one infinity, there exists a whole hierarchy (in fact an infinite hierarchy!) of distinct infinite numbers*: $|N|$ (where N as usual is the set of all natural numbers), |**P**(N)|, |**P**((**P**(N))|, |**P**(**P**(**P**(N))| and so on*ad infinitum*.

(It is easy to show that the 'continuum' – the set of all real numbers, all points on the real line – can be put in one-to-one correspondence with the set of all subsets of the naturals, *i.e.***P**(N).)

So far we have a theorem, not a 'paradox'. The 'paradox' arises by considering the ***setof all sets***.The naive principle of comprehension entails that this set exists since it isthe extension of the property 'x is a set'. Call this the 'universal' set U.

By Cantor's Theorem, |**P**(X)| > |X| for *any* X, and so in particular |**P**(U)| > |U|. **P**(U) is of course the set of all subsets of the set of all sets. This means it is certainly a set of sets and so *must itself be a subset of U*. ($\forall$x(x ε**P**(U) $\rightarrow$ x ε U) is true since every element of **P**(U) is a set and *all* sets are in U.) But it is easy to see that for *any* two sets X and Y, if X $\subseteq$ Y then |X| ≤ |Y| . This is because |X| ≤ |Y| requires only that there be aone-to-one correspondence between X and a subset of Y, and if X itself is a subset of Y then the ***identity mapping*** (which associates any element with itself) is such a one-to-one correspondence.

Hence since **P**(U) $\subseteq$ U we have |U| ≥ |**P**(U)| and this contradicts the consequence of Cantor’s theorem that |**P**(U)| > |U|.

E2(D): SOLUTIONS OF THE PARADOXES

Once the paradoxes had been spotted it proved possible to revise set theory in various ways so as to avoid them. Two ***axiomatic systems*** in particular were produced: Zermelo-Fraenkel Set Theory and von Neumann-Bernays-Gödel set theory. Neither system of course contains the full (naïve) principle of comprehension since if they did they would be inconsistent. In Z-F, *e.g.,* it is replaced by the 'Axiom of Subsets', which states that, *given a set,* any property determines a subset of it. Although it cannot be proved that either system is consistent (and so absolutely free from any 'paradoxical' derivation), it *can* be shown that the usual paradoxical reasoning (e.g. in the Russell and Cantor cases) is definitely blocked in either system.

These axiomatic systems are satisfactory from the mathematical point of view. Set theory was, however, intended to play an additional, ***foundational*** role. In particular, as I mentioned, the logicists Frege and Russell who set out to show that mathematics 'reduces' to logic, regarded set theory as a legitimate part of logic itself. The problem with the axiomatic set theories from this point of view is that the restrictions they impose on set existence seem rather *ad hoc*– aimed simply at avoiding the paradoxes – and *not* themselves 'self-evident' as one would hope (at anyrate on reflection) any truly logical principle would be. Russell himself adopted a different approach:

Type Theory. Russell suggested that the universe of sets should be regarded as stratifiedor hierarchical in structure. Every element is of a definite *type.* At typelevel 0 are *individuals* (non-sets). (These, it turns out, can be eliminated but we need not worryabout this.) At type level 1 are sets of individuals; at type level 2, sets of sets of individuals; and so on.

Each object in the universe of sets has a type indicated by a subscript, variables vary only over objects of a given type and so they too have type subscripts. The fundamental rule of type theory is that any formula of the form $x_i\varepsilon\ y_j$ (where i and j are the subscripts indicating the type level) is *well formed* (meaningful) only if $j = i + 1$. In other words, it can only be sensibly asserted that one set is, or is not, an element of a set of *next higher type.* Any other membership assertion is meaningless. In particular, the assertion $x_i \varepsilon\ x_i$is not well formed – that is, one cannot meaningfully assert in type theory that a set is a member of itself (and nor, therefore, that a set is *not* a member of itself). Hence the reasoning that led to the Russell paradox cannoteven get started. It can also be shown that, while in type theory there is a *set of allsets* of *type level i* (that set itself being of level i+1) for any i, there is no trulyuniversal set – i.e. set of *all* sets of whatever type level. This blocks Cantor’s Paradox.

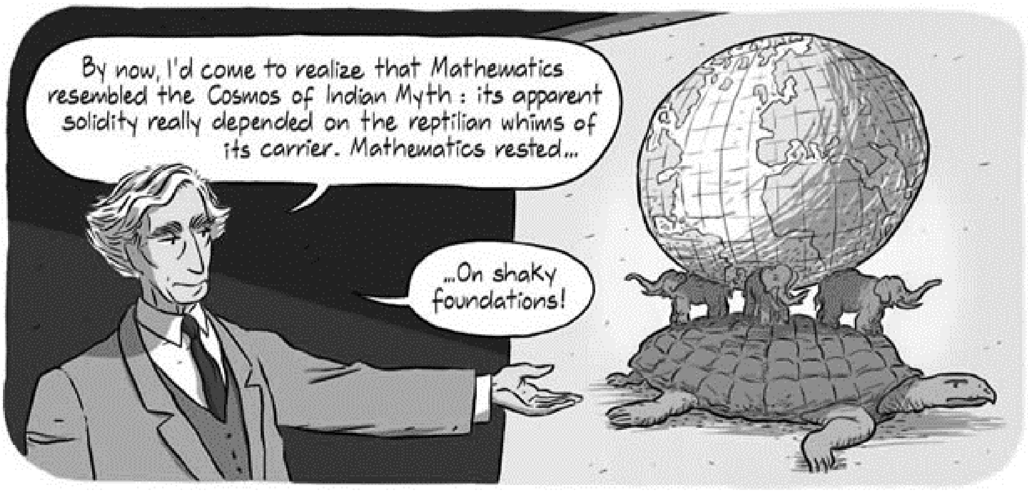
The problem again from the foundational point of view is to say why this type-level stratification is 'natural' or 'obvious', once we have cleaned our logical spectacles. Otherwise the theory appears like another *ad hoc* manoeuvre simply designed to avoid the paradoxes. Russell tried to justify the stratification using his 'vicious circle principle'.The exact import and effect of this principle is still a matter of some dispute. Historically, however, Russell's justification was not accepted and it was generally felt that the logicist programme came to grief over the paradoxes. Whatever the reason for the truth of mathematics, it was not that mathematics consists simply of logical truths.

Some further reading:

Several articles from the Stanford Encyclopedia of Philosophy provide more detail, and the further reading and references in each contain more information than one could conceivably need:

- Paradoxes and Contemporary Logic
- Self-Reference
- Russell's Paradox
- Type Theory
- Logicism

Entertaining, if lengthy, popularisations of these results can be found in **Logicomix**, a graphic novelization of the search for the foundations of logic and mathematics, as well as Douglas Hofstadter's books '*Godel, Escher, Bach'*, and '*I am a Strange Loop'*.


By now, I'd come to realize that Mathematics resembled the Cosmos of Indian Myth: its apparent solidity really depended on the reptilian whims of its carrier. Mathematics rested...
...On shaky foundations!

Excerpt from Logicomix