Natural deduction really can be natural

Ian C. Stirk

Introduction

The notation I use in this paper is that of Whitehead and Russell (1910), including the dot notation as revised by W.V. Quine in earlier editions of his *Methods of Logic* (for instance Quine, 1974). It is both elegant and practical.

In that same work, Quine uses a method of proof in symbolic logic which he calls "the Main Method". It is a kind of *reductio ad absurdum*. RaA is a convenient abbreviation for this. Observe the following three examples:

i			ii			iii		
1.	(∃x)Fx		1.	(x)Fx		1.	(∃x)Fx	
2.	(∃x)-Fx		2.	(x)-Fx		2.	(x)-Fx	
3.	Fa	1	3.	Fa	1	3.	Fa	1
4.	-Fb	2	4.	-Fa	2	4.	-Fa	2
			5.	⊗	3,4	5.	⊗	3,4

In example (i), we assume that the first two lines, the premises, are both true together. The first line states that at least one individual is F, so we can pick some name for that individual, or one of those individuals. We choose the name "a", and write "Fa" as line 3. The number 1 to the right shows that line 3 was derived from line 1. This kind of derivation is called "existential instantiation", or EI for short. Line 2 states that at least one individual is not F. Clearly lines 1 and 2 are not inconsistent, for the individual or individuals which are F and –F may be different ones. We show this by selecting a different name, "b", to instantiate the second line. We end up with Fa and –Fb, and there is nothing *absurdum* about those.

In example (ii), however, the premises state that all individuals are F and that all individuals are –F, which is obviously inconsistent. Line 3 here is a case of "universal instantiation", or UI. Clearly if all individuals are F, we could pick any one of them to receive the name "a", as in line 3. We can also use the name "a" for UI on line 2, since all individuals are involved, and one of them is already named "a". Thus we obtain the *absurdum* lines 3 and 4. Line 5 contains a convenient symbol for inconsistency, and the numbers to the right of it show just which lines were inconsistent.

Example (iii) is a mixed case. We choose the name "a" for EI on line 1. Going from line 2 to line 4 is a case of UI, so we can choose any individual, including the one already named "a". An inconsistency again arises.

Even such simple examples indicate the power and simplicity of RaA. The main rule of it is just that each existential quantifier we come across must be instantiated with a new letter, because a different individual with a different name may be involved. Universal quantifiers can clearly be instantiated with whatever letter we like, including those that are already the names of some individuals. It is best to follow those rules strictly, even though in example (iii) we might have used UI on line 2 first, thus using the name "a" for any individual, and then used EI on line 1 using the same name for some particular one. But strict application of the rules just means that existential quantifiers should be instantiated before universal ones whenever possible. It is better to be strict rather than compromise and fall into error.

There is an alternative method of proof called "natural deduction", ND for short. The concept seems pretty natural at first sight. It is to begin with some premise or premises and see what can be deduced from it or them. To start with a trivial example, given the premise (x)Fx we should be able to deduce (∃x)Fx from it, since if something is true for all individuals it must also be true for some of them. A deduction like the following seems plausible:

i

1.	(x)Fx	
2.	Fa	1
3.	(∃x)Fx	2

The premise, line 1, states that all individuals are F, so we can pick any one and give it the name "a", as in line 2. Line 2 is fairly deduced from line 1. If line 2 is true, then clearly we can deduce line 3: if F is true of the individual named "a", it has to be true of something.

The example above was labelled (i), and there are three other basic possible deductions to go with it:

ii			iii			iv		
1.	(∃x)Fx		1.	(x)Fx		1.	(∃x)Fx	
2.	Fa	1	2.	Fa	1	2.	Fa	1
3.	(∃x)Fx	2	3.	(x)Fx	2	3.	(x)Fx	2

Cases (ii) and (iii) are clearly all right, as they just go there and back again, as it were. But case (iv) cannot be right. Just because some individuals are F, we cannot deduce that all of them are.

In ND, we are clearly adding quantifiers as well as removing them. We could call the process of adding an existential quantifier "existential generalisation", EG, and that of adding a universal quantifier "universal generalisation", UG. Using those terms, the case we want to avoid in ND is EI followed by UG of the same letter, which happens in case (iv). A simple way to avoid that case is to bear in mind, or "flag", to use Quine's term, the letters involved in EI and UG, like this:

ii			iii			iv		
1.	(∃x)Fx		1.	(x)Fx		1.	(∃x)Fx	
2.	Fa	1 a	2.	Fa	1	2.	Fa	1 a
3.	(∃x)Fx	2	3.	(x)Fx	2 a	3.	(x)Fx	2 a

This is because an instance of EI is rather special. It may be a possible name of only some individuals, or only one. Similarly in UG we need to be sure that the name involved could be the name of any individual. All will be well as long as a letter is not flagged twice in a deduction, as "a" is in example (iv). That would mean trying to go from "at least one" to "all", which of course is not proper. Examples (ii) and (iii) are all right, as is example (i), where no flag is needed.

So ND is already more complicated than RaA, despite its "naturalness". Quine used ND exclusively in his earlier work, but later abandoned it for RaA, not so surprisingly.

In fact there is yet another complication in ND, which will be investigated in the next section.

RaA and ND compared

A classic problem that illustrates the advantages and disadvantages of RaA and ND is to prove that while

$$(\exists x)(y)Rxy \supset (y)(\exists x)Rxy$$

is logically true,

$$(y)(\exists x)\ Rxy \supset (\exists x)(y)\ Rxy$$

is not.

The two symbolic sentences can receive a straightforward interpretation in ordinary language. The first could translate "If someone loves everyone, then everyone is loved by someone", which rings true. The second sentence, in that case, would be "If everyone is loved by someone, then someone loves everyone". That doesn't sound right: "Someone loves everyone" suggests just one person doing all the loving, while "Everyone is loved by someone" could mean that each person has a different admirer.

Let us try to deal with these using RaA. There will be two premises in each case, one being the left hand side of the conditional, the other the negation of the right hand side. In the first case, the negation will be $-(y)(\exists x)Rxy$, or $(\exists y)(x)$ $-Rxy$ if we move the negation to the right of the quantifiers. The RaA might start

1.	(∃x) (y) Rxy	
2.	(∃y) (x) -Rxy	
3.	(y) Ray	1
4.	(x) -Rxb	2

The existential quantifiers have been instantiated with two different letters, according to the rule, and now only universal quantifiers are left, which can be instantiated as we please. We want to reach an inconsistency if we can, and instantiating the (y) of line 3 with "b", and the (x) of line 4 with "a", will get us one:

5.	Rab	3
6.	-Rab	4
7.	⊗	5,6

Now for the case of "If everyone is loved by someone, then someone loves everyone". The RaA could start as follows:

1.	(y) (∃x) Rxy	
2.	(x) (∃y) -Rxy	
3.	(∃x) Rxa	1
4.	Rba	3
5.	(∃y) -Rby	2

It is clear, though, that we cannot reach any inconsistency. A universal quantifier has to be instantiated first, and the two existential quantifiers must later be instantiated with different letters, so there will be three different letters around. In the working above, the (∃y) of line 5 cannot be instantiated with "a", which would be the only way to get an inconsistency with line 4.

The example illustrates a potential difficulty with RaA. In this simple case, it is easy to show that there cannot be an inconsistency, but in more complicated cases it might be much more difficult. We might wonder whether there really was no inconsistency, or whether we had missed some possible instantiation. Also in RaA we have to guess some conclusion to start with in order to find out whether or not its negation is inconsistent with the premises. It would be nice to have some way of just exploring: beginning with some premises, and playing around with them to see where they lead.

ND promises to let us do just that, so let us try the deductions above with that method. "If someone loves everyone then everyone is loved by someone" is straightforward:

1.	(∃x) (y) Rxy	
2.	(y) Ray	1 a
3.	Rab	2
4.	(∃x) Rxb	
5.	(y) (∃x) Rxy	4 b

That is encouraging: the quantifiers were reversed in the working, and the rule about only flagging a letter once was followed. The difficulty arrives with the following:

1.	(y) (∃x) Rxy	
2.	(∃x) Rxa	1
3.	Rba	2 b
4.	(y) Rby	3 a
5.	(∃x) (y) Rxy	4

Again the flagging rule has been followed, but the result has to be wrong! It needs a bit of subtle thought to work out where that deduction fails. In going from line 1 to line 2, "a" could be the name of any individual, as it came from the instantiation of a universal quantifier. In going from line 2 to line 3, however, "b" is the name of some particular individual or individuals, not of any. Line 3 states that "b" loves "a", to use the ordinary language illustration. Now, however, "a" may not be the name of any individual, but only of those that are loved by "b". The existential instantiation that provided line 3 had an effect on the reference of "a" also. There is a trick to avoid this problem, which can be called the "alphabet rule:

The flagged letter of any line must be alphabetically later than any free letter in the same line.

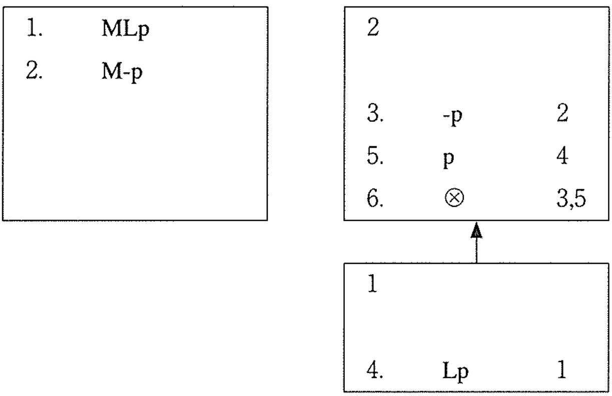
Here "free" means that the letter is not in the scope of some quantifier using the same letter for its variable name. Thus in line 2 above, "a" is free but "x" is not, while in line 3 both "a" and "b" are free. Clearly in line 4 the alphabet rule has been broken: "a" is flagged, but an alphabetically later letter, "b", is free in line 4. Following the alphabet rule would prevent us from deducing anything from line 3, apart from replacing the quantifiers in the same order and getting back to line 1, a pointless exercise, or proceeding with an existential quantifier as follows:

1.	(y) (∃x) Rxy	
2.	(∃x) Rxa	1
3.	Rba	2 b
4.	(∃y) Rby	3
5.	(∃x) (∃y) Rxy	4

In that case, "a" does not need to be flagged.

A bit of thought shows us that the alphabet rule prevents a universal quantifier being moved to the right of an existential quantifier. But the whole procedure is rather unnatural, even if it is easy enough to follow in practice.

Although Quine would have hated the idea, it is quite straightforward to adapt RaA as a method of proof in modal logic, using the analogy between "M" and "(∃)" and "L" and "(x)". (I use "M" to mean "true in at least one accessible possible world" and "L" to mean "true in every accessible possible world", following Hughes and Cresswell in their standard text, Hughes and Cresswell, 1996.) As an example, I show that MLp ⊃ Lp is logically true in S5:



Possible worlds are indicated by rectangles. The one on the left contains the two premises, the left hand side of the conditional and the negation of the right. The next step is to "instantiate" the "M" of line 2 with another possible world on the right. In this world, "-p" is true. The world is also labelled "2" at the top left, to show its origin. Next the "M" of line 1 is instantiated with world 1, where Lp is true. But since in S5 all worlds are accessible to each other, if "Lp" is true in one world, "p" has to be true in every one. Hence line 5 in world 2, which provides the inconsistency. An arrow connects worlds 1 and 2, to show accessibility – not really necessary in S5.

RaA is thus easy to apply to modal logic. But what about ND? What could possibly be the analogue of flagging, or the alphabet rule? A few years ago I struggled to find some way of using ND with modal propositional logic, and succeeded to some extent with S5. The method was clumsy and I soon gave up on it. It is described in Stirk (2005). Recently, however, I stumbled across a way of making ND in general a whole lot more natural, as described in the next section.

More natural natural deduction

Let us go back to what we can deduce from $(\exists x)(y)Rxy$, "Someone loves everyone". If we name the individuals in question a1, a2 and so on, we could make a deduction as follows:

1. $(\exists x)(y)Rxy$

2. $(y)Ra_1y$ 1   3. $(y)Ra_2y$ 1   ....

The branches are meant to indicate alternatives: in the first branch, it is a_1 who loves everybody, in the second branch it is a_2, and there may be other possibilities, as shown by the third branch and the dots. At least one of the branches must be true. This use of branching to indicate alternation is often convenient in RaA:

1.	(x)-Fx				
2.	(x)-Gx				
3.	(∃x)(Fx ∨ Gx)				
4.	Fa ∨ Ga		3		
5.	Fa		8.	Ga	
6.	-Fa	1	9.	-Ga	2
7.	⊗	5,6	10.	⊗	8,9

Here both branches end in inconsistency, so the three premises are inconsistent. Now the previous ND branching could continue:

1.	(∃x)(y)Rxy					
2.	$(y)Ra_1y$	1	3.	$(y)Ra_2y$	1	....
4.	Ra_1b	2	5.	Ra_2b	3	
6.	(∃x)Rxb	4	7.	(∃x)Rxb	5	
8.	(y)(∃x)Rxy	6	9.	(y)(∃x)Rxy	7	
10.	(y)(∃x)Rxy					

All the branches end up the same, so they can be collapsed into one, the conclusion. Steps like the one from line 4 to line 6, an existential generalisation, are always possible because Fa ⊃ (∃x)Fx is logically true. It is impossible for the left hand side of the conditional to be true while the right is false. Steps like that from line 6 to line 8, a universal generalisation, are justified because "b" itself arose from a universal instantiation in the same branch.

That all seems reasonable, so let us have a look at the more problematic case, what we can deduce from (y) (∃x) Rxy:

1. (y) (∃x) Rxy
2. (∃x) Rxa 1

3. Rb_1a 2 4. Rb_2a 2 ....

Now we see right away that universal generalisation of "a" is impossible in any particular branch, since they may not all be true. One way to complete a deduction would be

1. (y) (∃x) Rxy
2. (∃x) Rxa 1

3. Rb_1a 2 4. Rb_2a 2 ....
5. (∃x) Rxa 3 6. (∃x) Rxa 4

7. (∃x) Rxa
8. (y) (∃x) Rxy 7

This is the pointless deduction which takes us right back to the starting point. The end points of the branches become identical in line 5 and line 6, so the collapse into one in line 7 is justified. Now that there is again only a single branch, the final universal generalisation can be allowed, since "a" came from a universal instantiation further back on the same branch. Another possible deduction is the slightly more interesting

1. $(y)(\exists x)Rxy$
2. $(\exists x)Rxa$ 1

3. Rb_1a 2 4. Rb_2a 2 ....
5. $(\exists y)Rb_1y$ 3 6. $(\exists y)Rb_2y$ 4
7. $(\exists x)(\exists y)Rxy$ 5 8. $(\exists x)(\exists y)Rxy$ 6

9. $(\exists x)(\exists y)Rxy$

where existential generalisation is used on each separate "a" in each separate branch. Finally the branches become identical and can be collapsed.

The difficulty of the alphabet rule has been solved! Branching is much clearer than flagging in making the possibilities for generalisation apparent. The rules for the new style of ND can be summed up as follows:

Universal instantiation (UI)	A universal quantifier may be instantiated by any letter in the next line of the same branch.
Existential instantiation (EI)	An existential quantifier may be instantiated by any letter, but must also be instantiated by other letters

in other branches.

Existential generalisation (EG) Any free letter in any branch may be existentially generalised.

Universal generalisation (UG) A letter that came from a previous UI in the same branch may be universally generalised.

Branch collapsing (BC) Branches that end with identical lines may be collapsed into one branch.

All that seems natural enough. Let us see if there are analogues in modal logic to this new style of ND. What can be deduced from MLp in S5?

1.	MLp

So there we are, a world in which MLp is assumed to be true. The "M" can be instantiated with another world, but allowing also for the possibility of other worlds, in the spirit of the EI rule:

1.	MLp

1		
2.	Lp	1

1

There is at least one accessible world in which Lp is true, and other worlds which we so far know nothing about. Both the worlds illustrated have "1" at the top left, since both represent the instantiation of the "M" in line 1. Now we can continue

1.	MLp	
3.	p	2
6.	Lp	3,4,5

1		
2.	Lp	1
4.	p	2

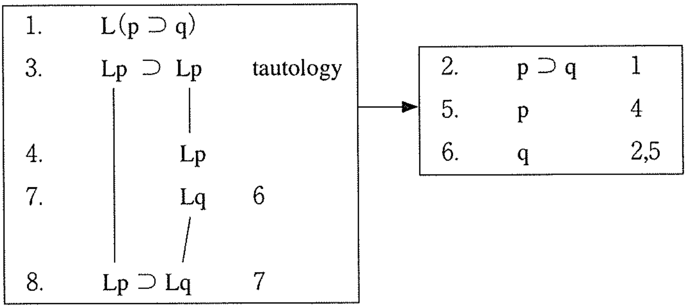
1		
5.	p	2

All worlds are accessible to one another, so we can of course deduce lines 3 and 4 from line 2. Also we can deduce line 5 from line 2, in another world. In fact we see that "p" is going to be true in all worlds, just because of the Lp in line 2. That means we can go back to our original world, and deduce line 5, Lp, there, since in all accessible worlds p is true. Natural enough!

Let us try a more complicated example. We can show that

$$L(p \supset q) \supset . Lp \supset Lq$$

is logically true. Here is the proof:



The world on the right can be any world in this case, so there is no complication there. Notice also that nothing is assumed about the nature of the accessibility relation. The same proof will go through in any modal system, including K. It is often useful to introduce a tautology as a line in a proof, as in line 3 above. Tautologies clearly have to be true in any possible world. Branching is a way of dealing with an alternation, and line 3 is an alternation in disguise, since Lp ⊃ Lp is equivalent to –Lp ∨ Lp. But nothing is deduced from the left hand side of the alternation in the proof, so there is no point in changing the conditional to its alternation form.

Introducing tautologies can be very useful in ND, and in fact enables us to prove one or two things which cannot be proved by RaA, for instance the logical truth of (x)Fx ≡ -(∃x)-Fx. This is best done in two parts, firstly showing that (x)Fx ⊃ -(∃x)-Fx:

1. $(x)Fx$

2. $-(\exists x)-Fx \quad \vee \quad (\exists x)-Fx$ tautology

3.	$-Fa$		4.	$-Fb$	...
5.	Fa	1	6.	Fb	1
7.	$\otimes$	3,5	8.	$\otimes$	4,6

9. $-(\exists x)-Fx$

On the right hand side of the alternation, EI takes place allowing for various possibilities. However, every one of these branches has to end in inconsistency, because of line 1. Thus only the left hand branch remains. The proof is certainly ND, of course, but it has an odd admixture of RaA!

Now to show that $-(\exists x)-Fx \supset (x)Fx$:

1. $-(\exists x)-Fx$

2. $-Fa \quad \vee \quad Fa$ tautology

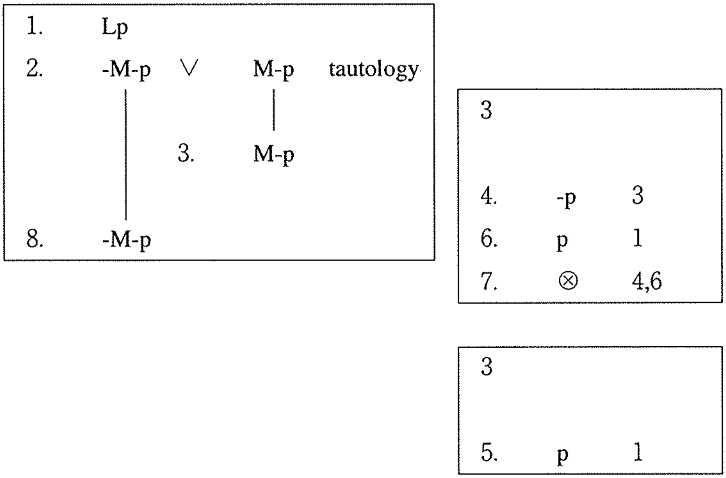
3. $(\exists x)-Fx$ 2

4. $\otimes$ 1,3

5. $(x)\ Fx$ 2

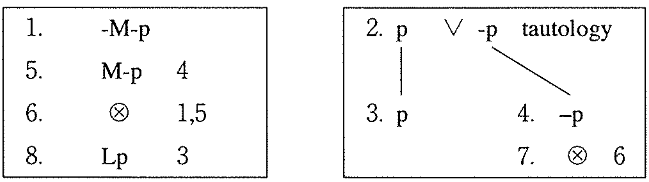
Another cunning trick is illustrated here. Because line 2 is a tautology, the name "a" could refer to any individual. When the left hand branch turns out to be inconsistent, only the right remains, so UG is justified.

The analogous formulae in modal logic can also be proved, starting with $Lp\supset-M-p$:



Here the M-p of line 3 is "instantiated" with a couple of worlds, one where –p is true, and another where we are not sure. But clearly, given line 1, any world where –p is true would be inconsistent, so the right hand branch of the tautology is inconsistent too.

Now for –M-p ⊃ Lp:



I'm not sure I yet have the best notation for this, but the meaning is clear, I think. Line 2 in the world on the right is a tautology, so that world could be any world. However, when we consider the branches separately, we may have something true in one or more worlds or none at all. If the right hand branch, line 4, is true in some possible accessible world then line 5 will be true, but it is inconsistent with line 1. Thus there can be no world where line 4 is true – only the left hand branch of the alternation remains. Thus p must be true in every accessible world, hence line 8.

The new style of ND seems to work very well with modal propositional logic. We already know it works with first order predicate calculus, but we must make sure it will work also with higher order calculi. Here is a very straightforward example:

1. $Fa \vee -Fa$ tautology

2. $(\exists X)Xa$ 1 3. $(\exists X)Xa$ 1

4. $(\exists X)Xa$ 2,3

5. $(x)(\exists X)Xx$ 4

Here we begin again with a tautology, so any individual might be the one called "a", and any predicate the one called "F". In lines 2 and 3 EG is applied to the predicate of the particular branch: F in the left hand, -F in the right. The branches then collapse to give line 4, and UG can apply to the one instance of "a". The proof shows that $(x)(\exists X)Xx$ is logically true, and of course it is plausible. For any individual, there should be at least one predicate which is true of it.

A word about notation here. I am using X, Y, Z as predicate variables and F and G as predicate names. In previous work I have used F and G ambiguously as both variables and names, as do many writers. But this seems undesirable.

It should also be the case that $(\exists X)(x)Xx$, that is, there is at least one predicate which is true of everything. It is difficult to see what kind of tautology it could be deduced from, however. For if we had the name "a" in both branches, the branches would need to collapse before we could use UG on it, and get $(x)...x$, whatever the dots might represent. But we can prove $(\exists X)(x)Xx$ like this:

1.	$Fa \vee -Fa$	tautology
2.	$(\lambda x)(Fx \vee -Fx)a$	1
3.	$(x)(\lambda x)(Fx \vee -Fx)x$	2
4.	$(\exists X)(x)Xx$	3

This time, in going from line 1 to line 2, the alternation is collapsed by forming a λ-expression. UG can then give line 3. The λ-expression is not a general case that UG could be applied to, but we can use EG on it to give line 4, the result. It seems then that the λ notation represents not just a convenient abbreviation of more complex expressions, but a genuine increase in power.

ND seems to work with higher order calculi, then, so let us use it on a more complicated example. Identity is an important relation, and it seems to be adequately described by the two axioms $(x)(x = x)$ and $(x)(y)(Fx . x = y . \supset Fy)$. The second of those is generally known as Leibniz' Law.

If we can quantify over predicates, it looks plausible that identity could actually be defined as follows:

$$(x)(y)[x = y . \equiv (X)(Xx \equiv Xy)]$$

It turns out that this definition is actually equivalent to the conjunction of the two axioms. With quantification over predicates, Leibniz' Law can be written as

$$(X)\,(x)\,(y)\,(Xx\,.\,x = y\,.\supset Xy)$$

and using ND, it is easy enough to show that the axioms imply the definition:

1.	$(x)\,(x = x)$	
2.	$(X)\,(x)\,(y)\,(x = y\,.\,Xx\,.\supset Xy)$	
3.	$a = b\,.\supset.\,a = b$	tautology
4.	$a = b$	3
5.	$(X)\,(x)\,(y)\,(x = y\,.\,Xx\,.\supset Xy)$	2
6.	$a = b\,.\,Fa\,.\supset Fb$	5
7.	$Fa \supset Fb$	4,6
8.	$b = a$	4, symmetry
9.	$b = a\,.\,Fb\,.\supset Fa$	5
10.	$Fb \supset Fa$	8,9
11.	$Fa \equiv Fb$	7,10
12.	$(X)\,(Xa \equiv Xb)$	10
13.	$a = b\,.\supset (X)\,(Xa \equiv Xb)$	3,12
14.	$(X)\,(Xa \equiv Xb) \supset (X)\,(Xa \equiv Xb)$	tautology
15.	$(X)\,(Xa \equiv Xb)$	14
16.	$(\lambda x)\,(a = x)\;a \equiv (\lambda x)\,(a = x)\;b$	15
17.	$a = a\,.\equiv.\,a = b$	16
18.	$a = a$	1
19.	$a = b$	17,18
20.	$(X)\,(Xa \equiv Xb) \supset.\,a = b$	14,19

21.	$a = b \,.\, \equiv\, (X)(Xa \equiv Xb)$	13,20
22.	$(x)(y)[x = y \,.\, \equiv\, (X)(Xx \equiv Xy)]$	21

The proof has been abbreviated slightly. In line 8, symmetry has been invoked to deduce b = a from a = b. But the symmetry of identity can easily be deduced from the axioms. In line 22, UG has been applied to both "b" and "a" simultaneously, clearly not a problem.

It is also easy enough to show that the definition implies the axioms. It is most straightforward to deal with the axioms separately. First it is shown that Leibniz' Law follows from the definition:

1.	$(x)(y)\ [x = y \,.\, \equiv\, (X)(Xx \equiv Xy)]$	
2.	$Fa \,.\, a = b \,.\, \supset \,.\, Fa \,.\, a = b$	tautology
3.	$\quad\vert\quad Fa \,.\, a = b$	2
4.	$\quad\vert\quad a = b \,.\, \equiv\, (X)(Xa \equiv Xb)$	1
5.	$\quad\vert\quad (X)(Xa \equiv Xb)$	3,4
6.	$\quad\vert\quad Fa \equiv Fb$	5
7.	$\quad\vert\quad Fb$	3,6
8.	$Fa \,.\, a = b \,.\, \supset Fb$	2,7
9.	$(X)(x)(y)(x = y \,.\, Xx \,.\, \supset Xy)$	8

Going from line 8 to line 9, there are three instances of UG compressed into one, but there is clearly no problem with this.

It is even easier to show that the definition of identity implies $(x)(x = x)$:

1.	$(x)(y)\ [x = y\ .\equiv (X)(Xx \equiv Xy)]$	
2.	$a = a\ .\equiv (X)(Xa \equiv Xa)$	1
3.	$Fa \equiv Fa$	tautology
4.	$(X)(Xa \equiv Xa)$	3
5.	$a = a$	2,4
6.	$(x)(x = x)$	5

No special comment needed there.

Now everything can be brought together in the following rather unfortunate result:

1.	$(x)(y)\ [x = y\ .\equiv (X)(Xx \equiv Xy)]$	definition
2.	$a = b\ .\equiv (X)(Xa \equiv Xb)$	1
3.	$a = b\ .\supset .\ a = b$	tautology
4.	$a = b$	
5.	$(X)(Xa \equiv Xb)$	2,4
6.	$(\lambda x)L(a = x)a \equiv (\lambda x)L(a = x)\ b$	5
7.	$L(a = a) \equiv L(a = b)$	6
8.	$Fa \equiv Fa$	tautology
9.	$(X)(Xa \equiv Xa)$	8
10.	$a = a\ .\equiv (X)(Xa \equiv Xa)$	1
11.	$a = a$	8,10
12.	$L(a = a)$	11
13.	$L(a = b)$	7,12
14.	$a = b\ .\supset L(a = b)$	3,13
15.	$(x)(y)[x = y\ .\supset L(x = y)]$	14

That demonstrates that if two names belong to the same individual, then they necessarily belong to that individual. Notice here that no rectangle has been placed around the main world, an obvious saving of effort. Also the other world cited could be any world, as only a tautology is assumed to be true in it.

There is nothing wrong with the proof, but the result certainly seems strange. “George Orwell” and “Eric Blair” are two names for the same person in this world, but why should that mean that they refer to the same person in every possible world? Surely in the worlds of our imaginations at least they could be different.

There could be even worse. If identity of individuals is defined by

$$(x)(y)[x = y\ .\equiv (X)(Xx \equiv Xy)]$$

then maybe we could use a similar idea for the identity of predicates:

$$(X)(Y)[X = Y\ .\equiv (\mathfrak{X})(\mathfrak{X}X \equiv \mathfrak{X}Y)]$$

where Gothic letters are used for predicates of predicates. Distinguishing a hierarchy of predicates is one way of avoiding the deadly problem known as *Russell’s Paradox*. If there were no such hierarchy then we could imagine sentences such as FF, meaning “F is F”, or –FF, meaning “F is not F” and so on. Then we would be led to this:

1.	$(\exists X)(Y)(XY \equiv -YY)$		$\vee$ $(X)(\exists Y)(XY \equiv YY)$		tautology
2.	$(Y)(FY \equiv -YY)$	1	5. $(\exists Y)((\lambda Z)(-ZZ)Y \equiv YY)$	1	
3.	$FF \equiv -FF$	2	6. $(\lambda Z)(-ZZ)F \equiv FF)$	5	

4.	$\otimes$	3	7.	$-FF \equiv FF$	6
			8.	$\otimes$	7

Here we have a tautological alternation either branch of which leads to an inconsistency. A hierarchy of predicates, predicates of predicates and so on is the most obvious way to avoid this paradox, and it is adopted by writers like Richard Montague in their work on language. Unfortunately it is not a very satisfactory solution from the mathematical point of view, since it leads to a hierarchy of arithmetics: we need one set of numbers for counting individuals, another for counting predicates and so on. The problem still has no clear solution.

Anyway, if we do decide to define identity of predicates by $(X)(Y)[X = Y. \equiv (\text{Ӿ})(\text{Ӿ}X \equiv \text{Ӿ}Y)]$, similar definitions would obtain for predicates on all levels. The difficulty is that just as we proved $(x)(y)[x = y. \supset L\ (x = y)]$ above, we would obviously be able to prove $(X)(Y)[X = Y. \supset L\ (X = Y)]$, $(\text{Ӿ})(Y)[\text{Ӿ} = Y. \supset L(\text{Ӿ} = Y)]$ and so on ad infinitum. Every possible statement of identity would be necessarily true, and every possible world would be exactly the same as every other, a weird state of affairs!

Luckily it is possible to define the identity of predicates in a different way:

$$(X)(Y)[X = Y. \equiv (x)(Xx \equiv Yx)]$$

that is, predicates are identical if and only if they are true of just the same individuals. This is treating predicates as sets of individuals, clearly. This definition could also be extended up the hierarchy: $(\text{Ӿ})(Y)\ [\text{Ӿ} = Y. \equiv (X)(\text{Ӿ}X \equiv YX)]$ and so on.

Anyway, how are these alternative definitions related? It would be bad news indeed if they were equivalent: we would be stuck with the problem of only one possible world. Observe the following proof in ND:

1.	$(\mathfrak{X})(\mathfrak{X}F \equiv \mathfrak{X}G) \supset (\mathfrak{X})(\mathfrak{X}F \equiv \mathfrak{X}G)$	tautology
2.	$(\lambda Z)(x)(Fx \equiv Zx)F \equiv (\lambda Z)(x)(Fx \equiv Zx)G$	1
3.	$(x)(Fx \equiv Fx) \equiv (x)(Fx \equiv Gx)$	2
4.	$Fa \equiv Fa$	tautology
5.	$(x)(Fx \equiv Fx)$	4
6.	$(x)(Fx \equiv Gx)$	3,5
7.	$(\mathfrak{X})(\mathfrak{X}F \equiv \mathfrak{X}G) \supset (x)(Fx \equiv Gx)$	1,6
8.	$(Y)[(\mathfrak{X})(\mathfrak{X}F \equiv \mathfrak{X}Y) \supset (x)(Fx \equiv Yx)]$	7
9.	$(X)(Y)\ [(\mathfrak{X})(\mathfrak{X}X \equiv \mathfrak{X}Y) \supset (x)(Xx \equiv Yx)]$	8

The proof begins with a tautology, so line 7 is logically true, which justifies the two uses of UG at the end. Clearly the proof shows that the definition $(X)(Y)[X = Y \,.\equiv (\mathfrak{X})\ (\mathfrak{X}X \equiv \mathfrak{X}Y)]$ implies the other definition $(X)(Y)[X = Y \,.\equiv (x)\ (Xx \equiv Yx)]$. This is not very surprising, because clearly if identical predicates were necessarily identical, they would have to be true of exactly the same sets of individuals, since identical individuals are necessarily identical too.

Given that, we try the following rather anxiously:

1. $(x)(Fx \equiv Gx) \quad \supset \quad (x)(Fx \equiv Gx)$

But unlike in the previous proof, there is nowhere to go with the right hand side of this. We can construct λ -expressions at a certain level of the hierarchy out of entities at the next lower level: for instance, $(\lambda x)(..x..)$ is a predicate formed with an individual variable, $(\lambda X)(..X..)$ is a predicate of predicates, and so on. But here we have only an individual variable to instantiate, so no λ -expression is possible, and thus no way of reaching the other definition.

That is quite a relief. The fact that identical predicates are not necessarily identical gives us one way round the problem of necessarily identical individuals. Instead of individual names, we might think of a unique conjunction of predicates which are true of an individual. The conjunction could thus pick out some particular individual in one world, but a different individual in another. Such an approach is considered in chapter 17 of Hughes and Cresswell (1996).

Another possibility is to bring in functions from possible worlds to individuals. The function could then pick out different individuals in different worlds. This is the approach taken by Richard Montague in his (1973).

Conclusion

I think I have been able to show that a more natural way of regarding natural deduction can provide a useful tool for students of higher order predicate calculi and modal logic, a tool which nicely accompanies reductio ad absurdum.

Bibliography

G.E. Hughes and M.J. Cresswell (1996) *A New Introduction to Modal Logic* (Routledge)

Richard Montague (1973) *The proper treatment of quantification in ordinary English* reprinted in Thomason, ed (1974)

W.V. Quine (1974) *Methods of Logic* (Routledge and Kegan Paul, 3rd edition)

Ian C. Stirk (2005) *Restoring the naturalness of deduction* (大阪外国語大学英米研究 vol 29)

Richmond H. Thomason, ed (1974) *Selected Papers of Richard Montague* (Yale University Press)

A.N. Whitehead and B.A.W. Russell (1910) *Principia Mathematica vol 1* (Cambridge University Press)