Non-Aristotelian Logic in Practice, or How to be much cleverer than all your friends (so they *really* hate you).

by

Mike Alder

Long, long ago, before **Philosophy *Now*** was even a gleam in its editor's eye, there were bright and lively minded people around, just like you. People who liked new ideas, liked a bit of intellectual stimulation, enjoyed debate and discussion, people who liked to use their brains. And what did they do when there was no **Philosophy *Now***? Well, some of them used to read Science Fiction.

In these latter, degenerate days when Science Fiction mainly means movies, and furthermore movies which are of the *leave-your-brain-where-you-buy-your-pop-corn-and-pick-it-up-on-the-way-out* category, this may seem unlikely; but all us old wrinklies who were around before mobile phones know it's true. It used to come in things called *books*, and they were genuine science fiction, not just tales of dragons and elves, and bold warriors with IQs lower than that of their swords. The stories, it has to be admitted, were every bit as silly as Star Wars, and the characterisation even worse. But in the good old days now gone, there were ideas that knocked your socks off. For an example, one I've mentioned before when discussing the sins of Miss Blackmore, take A E van Vogt's *World of Null-A*. Two ideas come up in the book: one is that the hero has two brains and the second is that he can think in a non-Aristotelian logic. The combination makes the hero ever so clever, able to out-think lesser mortals stuck with ordinary logic, and to defeat the bad guys with relative ease. Only relative ease, or the book would be rather short, but he defeats an evil galactic overlord and gets the girl, and no hero can reasonably expect to do better than that. All done, moreover, by superior thought rather than bulgier muscles, bigger zap-guns, or some kind of moral superiority associated with being sentimental as in the case of Captain Kirk. To the small adolescent lad I used to be, not conspicuous for either bulgy muscles, sentiment or moral superiority, this had definite appeal.

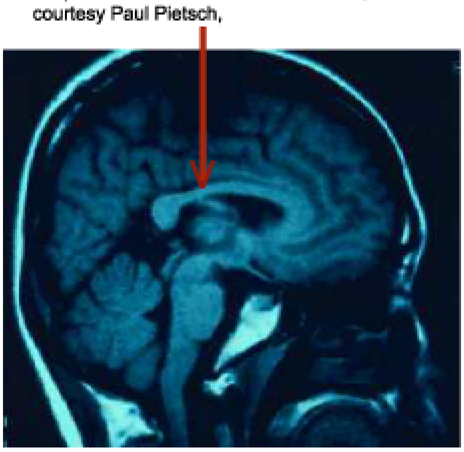
Some ideas turn out, after close and critical thought, to be pure tosh. So much so that whole organisations are devoted to stamping out close and critical thought or at least pointing it somewhere else. The principal idea in the Star Wars movies is that the Universe allows what used to be known as magic, which is mastered by listening to Master Yoda passing on his profound insights. These sound the sort of thing that come in fortune cookies or christmas crackers, and one would have to regretfully conclude that the central idea in Star Wars is complete tosh. What is rather striking about both the ideas in *World of Null-A* is that they aren't. In fact both are applicable to you. You do have two brains. And you can, in principle, learn to think in a non-Aristotelian logic. And if you master the skill, it really does make you more intelligent in an objective sense, namely, better able to predict the world, than those who can't. So read on and I shall explain how to think in a null-A logic, and explain what needs to be done in order to get good at it and develop a superior intellect. You probably won't be able to teleport as van Vogt's hero could, that part we have to admit was pure fiction. But there *is* a language based on a non-Aristotelian Logic. And you can learn to think in it.

I have to admit now, to avoid disappointment later, that there is a down side to this. For starters, no evil galactic overlords will be provided for thwarting purposes, nor any girls for any purposes whatever. Giving you a superior mind will certainly make it hard for you to put up with the lesser abilities of your friends, and will also make it even harder for them to put up with you. But it has to be confessed that these are not the main drawbacks. After all, if I offered to make you ten centimetres taller I could do it: there's a device called a rack which will accomplish it quite quickly. The trouble is, it's rather painful. And given the choice of being stretched on the rack or learning to think in a non-Aristotelian logic, level-headed people might well consider the rack the softer option. So be prepared for a tough time in the remainder of this article, and an even tougher time if you decide to go the whole hog and become a super-being.

Touching on the two brains aspect first; Roger Sperry got half a Nobel Prize (maybe it should have been two) for showing that people who had the corpus callosum severed showed signs of distinct personalities, one in each hemisphere of the brain. The corpus callosum is a big nerve trunk which connects the two hemispheres, and cutting it was done in cases of extreme epilepsy in order to prevent the brain from going wholly into the kind of spasm that manifests itself as an epileptic attack. Presenting pictures of half-naked ladies to one hemisphere of a male subject got pursed lips and a disapproving frown, the other hemisphere of the same subject giggled. This doesn't happen with most people, presumably because one hemisphere or the other wins some sort of internal debate about the best response, the argument being conducted via the corpus callosum. Unless you have had some currently unfashionable surgery, you are probably well equipped in this regard, which is just as well, or you would be like the two headed pushmi-pullyu, and in serious trouble if you had a fundamental disagreement of values.

Here is a picture of a brain showing the corpus callosum.

Corpus callosum seen on an MRI scan,
courtesy Paul Pietsch,



(See it at the website http://www.indiana.edu/~pietsch/callosum.html#mri)

Disagreement between the two hemispheres may manifest itself as adolescent angst or chronic indecisiveness, who knows? But that you have two hemispheres, and that each is capable of some sort of intellectual activity, is not in serious doubt. The usual story that the 'dominant' hemisphere, the left in right handed people, is responsible for linguistic skills and the 'subdominant' hemisphere for spatial skills, is a crude simplification, but not wholly devoid of truth. The reality is much richer than the simple tales told to the credulous by the optimistic, and worthy of some investigation, but for the present it suffices to make the claim that you indeed have two brains. You are not alone in this. We all do.

And yes, you will need both of them if you decide to master thinking in a non-Aristotelian logic.

For those who came in late, or even for those familiar with the sins of Miss Blackmore but with less than perfect memories, first some words of wisdom about the ordinary, common or garden, Aristotelian logic.

Classical Logic, first codified by Aristotle (hence van Vogt's habit of attaching his name to it), is all about the rules for correct argument. Two things should be noticed: first, everyone over the age of ten has some grasp of logic. We work it out by inferring the rules of what words mean, in particular how *all* and *and* and *not* and *or* are used. There is a remarkable ability of human brains to extract rules from data, and we do it a great deal in childhood, somewhat less frequently in later life. As a result of listening to logical arguments, those who did have some idea of how to follow and produce a logical argument themselves, having extracted the rules from samples. It has to be faced that some of us are better at it than others. The second thing to notice is that having extracted rules, these rules can be written down and listed. A rat could, learn that when the bell rings it is safe to reach for the food and when the buzzer sounds it is a good idea to get off the metal plate before you get zapped. It has extracted a rule from data. But it takes a human being to write the rule down and pass it on to its young. And some of the human beings who have worked out what the rules of valid argument are, have obligingly written them down for us. This allows even those of us who aren't smart enough to work out all of them to have access to the complete list.

Aristotle classified the valid syllogisms; the first (called AAA) is:

All A's are B's,
All B's are C's
Therefore all A's are C's.

The second (EAE) is

Some A's are B's,
All B's are C's
Therefore some A's are C's.

There are lots more of them.

You should have no difficulty checking that these are legitimate arguments whenever you replace the capital letters by anything you want: thus if all human beings are mammals and all mammals are animals, then it must be true that all human beings are animals whether you like this conclusion or not. On the other hand, the equivalent looking

Some A's are B's,
Some B's are C's
Therefore some A's are C's

is not a valid syllogism. There are some choices of A, B and C which make it fail.
And good old Aristotle provided a comprehensive list of all the valid syllogisms, and for over a thousand years, schoolchildren were forced to learn Latin rhymes in order to memorise them.

The psychologist Piaget concluded that the final stage of intellectual development of the child is this ability to extract the rules of logic. By this he seems to mean that in their later teens, most people can extract the rules from hearing other people use them without being able to say what the rules are. They know them in the sense of being able to *follow* them, but not in the sense of being able to *state* them. Obviously the second level of knowing the rules, being able to say what they actually are, is more difficult if you have to work them out for yourself, but they should be easy to recognise when someone has already done the hard part of writing them down in a rule book. Piaget doesn't say at what age the child figures out how to write down the rules of Logic, but it must be on the high side since most people never get there.

Why should we care about the rule book? Well, if arguments get complicated, you can either fall back on guessing what the right rules are, or you can use the rule book if you happen to know it. So for many centuries, part of every Western educated person's background consisted of learning the rules as laid down by Aristotle. It was supposed to make them better at producing and following arguments.

As with most educational theories, nobody bothered to find out if the theories actually worked. After all, the only people qualified to judge had already learnt the laws of logic, and being educated meant being like them.

Don't laugh, your own education has been based on even sillier ideas.

Other parts of classical logic also centred on when one could be sure of the truth of propositions if one was sure of the truth of other propositions. For any proposition, A, there was a negation, written ~A, which was the denial of A. Then one can write out some of the laws of logic in a very algebraic looking way:

(1) $A + \sim A = 1$
(2) $A \,.\, \sim A = 0$
where + is short for 'or' and . is short for 'and'. George Boole did this in the nineteenth century. Maths strikes again.

The first law (the law of the excluded middle) says that any proposition is either true or false. The second law says that to assert something and it's negation is to say something false. We use 1 for truth and 0 for falsity, so a proposition is defined to have a value, either true or false, 1 or 0. Logic, please note, is not generally concerned with telling us which: that's a matter of experience. Logic is concerned not with the truth, but with whether arguments are sound, although admittedly this came to be formulated in some propositions being true for logical reasons, for example: 'A or not A'. This is sometimes said to be a *necessary truth*, or a *tautology*. Tautologies aren't statements about the world, they only look a bit like them. Their truth is not a matter of the world being some way when it might have been another. They are true in a thoroughly boring way: they are statements about how language is correctly used. They are the rules that have been extracted from how language was in fact used. Later they became the rules of how you'd better use language if you didn't want to be laughed at as too dumb to figure out the rules from the samples.

All this worked for a long time; although the algebraic notation came more recently; all it did was say in shorter symbol strings what Aristotle knew two and a half thousand years ago.

Time for something new. It was the philosopher Leibnitz who is credited (by John Maynard Keynes, the economist) with the observation that people also pick up a different and more powerful form of reasoning. If you see a fire engine going flat out and flashing its lights and making a lot of noise, and see a column of smoke rising in the distance, you can't be certain that something is on fire, but it seems likely. And if A is very likely to be the case, and whenever A is true then B is almost always true too, then you are inclined to think B is likely to be true. You cannot be sure, and Aristotelian logic is no help, but people make a living betting on less likely things than that. And we all reason this way, we follow some set of rules, what one might call rules of plausible reasoning, but most of us don't know what the rules are. So it's a bit like logic before Aristotle; most of us are smart enough to have worked out rules of plausible reasoning from seeing other people do it, but are not quite smart enough to articulate those rules. We can *follow* them but not *state* them. So when we see the fire engine and the column of smoke, or the man wearing a Mickey Mouse mask climbing through the jewellers window with a bag over his shoulder, we frame some plausible hypotheses rather quickly, despite the claim of fire or burglarious intent not being warranted by strict logic. And again, just as with Aristotelian logic, there is a case for being taught what the rules actually are by someone who has articulated them, because knowing the rules, or at least where to find the rule book, helps when the situation is complicated. People who can do this can reason better and are less likely to make blunders in their thinking. In other words, they are, in operational terms, smarter.

Leibnitz wrote enthusiastically about the prospects for an extension of classical logic to the case of plausible reasoning, but never got down to the nitty-gritty of saying what the rules actually were. Many philosophers while excellent at seeing the big picture are, alas, regrettably fluffy about the fine detail. And some things depend rather a lot on the fine detail.

Jump forward about two and a half centuries to the early twentieth century, when logicians were wondering about generalising logic, just for fun. An obvious possibility is to have more than two possible values for the truth value. Instead of just False and True, how about another third option? It is perfectly possible to have a new possibility and build a system of rules for working in this new three valued logic. The question is, what would it mean? The answer is it could mean anything you wanted it to, so the next question is, what meaning could you give to it that would make it useful? When an innocent student, I once found a moderately useful meaning for a five valued logic in electronic circuitry, but it seemed a bit restricted and not as exciting as one might like. Can one find a generalisation of logic, where you have more than two values, that would be really applicable and useful?

John Maynard Keynes, in 1920, argued that you could. He reasoned that you could have a continuous logic, where you had not only the numbers 0 and 1, but all the other numbers in between as well. And if you had a proposition B, you could assign a number between 0 and 1 to B to represent the extent to which you believed B, the credibility of B. A credibility value of 1 meant that you thought B was true, 0 meant you thought it was false, and the other numbers in between were for representing different degrees of belief. This, he thought, was what we were doing when we assigned probabilities to things. Mostly, people had assigned probabilities to events up to this point. Keynes thought we should assign probabilities to propositions. If you take a coin and toss it and no funny business is going on, then Keynes said that the statement ‘ the coin will land with head up’ had probability one half. The probability isn’t a property of the coin, it’s a property of a statement about what the coin will do, made by someone with a certain amount of information. Change the relevant information that the person has, and you change the probability. If a close investigation of the coin showed a head on both sides, it would change a lot.

Many people heartily disliked this approach, arguing that it made everything subjective. They wanted to believe that the probability was something the coin had, and moreover that you could measure it by tossing it lots of times and counting the outcomes. They said that the probability was one half when the ratio in the limit, as you did more and more tosses, of the number of heads to the total number of throws was one half. Since you cannot in fact toss a coin an infinite number of times, it follows that you cannot ever actually know the probability exactly, but you can get closer and closer estimates. This is what is known as the frequentist interpretation of probability.

There are a number of serious problems with this. One of them is that mostly it isn’t practicable to perform the equivalent of tossing a coin more than once. If you see a race between four spavined, three-legged donkeys and the horse that just won the grand national, you are not entitled to say the probability is high that the horse will win, until the race has been held and repeated a few dozen times. If you see someone collapse, purple in the face, you are not allowed to say that he probably had a heart attack unless he has done it several times and it was mostly a heart attack before. Probability of the frequentist sort depends on replications of experiments, although frequentists are conveniently vague as to what counts as a repetition. Many events couldn’t be repeated at all, but we still feel inclined to think that some are more likely than others.

Another objection is that the tossing of a coin is, if you happen to believe Newton got things pretty much right, a deterministic process. So how does the probability come in? An obvious answer is that the final state depends in a perfectly deterministic way on the initial orientation of the coin and the angular and linear momentum it got when it left your hand – but we don't know what those initial conditions were, and for every initial state that finishes up with a head showing, there is another close to it where a tail shows. So to say that the probability is one half, is saying something about our ignorance of the details of the tossing. A very precisely made coin tossing machine which always gave almost exactly the same initial oomph to the coin, and where the coin was placed in the machine in the same way every time, would, according to Newton, always give the same outcome. The symmetry of the coin has something to do with it: tossing a hat might not give equally frequencies for the crown winding up on top. But the way we toss the thing makes a difference too. Knowing the exact details of the initial state (and maybe the wind) would, in principle, allow us to calculate the result. So the probability must surely be a property of our ignorance of the details, not a property of the coin. Or hat.

As to the subjectivity, Keynes pointed out that if two people have exactly the same prior knowledge and beliefs, then they will, if we assume they are both rational, assign the same number to a probability of an event. And if they assign different numbers because they have different prior knowledge, then they they damn well *ought* to assign different numbers. And finally, the subjectivity of the assignment of values to propositions is in any case irrelevant. The crucial things are the rules saying what follows from what, not the assignments. People differ over whether propositions are true or false, but that doesn't invalidate logic, which isn't actually concerned with whether a given proposition is true or false, just with what follows if it is. Or isn't. The actual truth value is a separate matter determined by other quite different procedures and quite frequently disputed, hence, presumably, subjective.

Armed with these reflections, we can now set about the business of generalising logic to take values in the infinite continuum consisting of all the real (decimal) numbers between 0 and 1. This will provide us with a genuine, functional non-Aristotelian Logic, and one with some practical value. It will enable us to put our informal feelings about, say, the intent of people crawling through jeweller's windows while wearing masks and carrying sacks, on a formal footing. We shall be able to make logical decision on whether or not to do the football pools or the lottery or to buy insurance. And we shall be able to make better decisions than people who have not studied the correct rules of plausible inference, providing us with a better chance of defeating the evil overlords, galactic or otherwise, and getting the girl (or whatever) of our choice. (And anybody who doesn't believe in the existence of evil overlords simply hasn't been looking at out politicians, so don't say this isn't useful stuff.) We shall also, of course, be completing the Leibnitz programme and providing a vindication of that kindly, if fluffy, philosopher.

As you can see, the promise of making you smarter is close to being kept. No short changing going on, no deception practised. Master the next part and you will be effectively smarter than Leibnitz. For he knew there were rules for plausible reasoning, and even followed them, but he wasn't able to say what they were. I'm shortly going to tell you.

We have, for any proposition B some value P(B) which is going to mean the extent to which we believe B. You can call P(B) the *plausibility* of the proposition B. And you can think of it as being a measure of the truth of B. It is true that you might assign one number to P(B) and I might assign a different one, but that might have been true even if we were only allowed to assign zero or one for false or true. Life is full of conflict. P(B) is required to be some number between 0 and 1 and to have the usual meaning of falsity and truth at the end points.

We have certain constraints. One is that in the cases where the values are in fact 0 or 1, we want to recover the usual rules of logic. We are trying to generalise Aristotelian logic, not make something totally different.

Another is a continuity assumption. Sometimes it makes sense to say that we can change a proposition continuously, for example 'the length of this bar is x centimetres' is a family of propositions depending on x, and changing continuously as x does. It is desirable that if for any x in some range, B_x is the above proposition, then $P(B_x)$ changes continuously with x too. There should be no sudden jumps in the value of $P(B_x)$ with x. When you cross the road, you don't suddenly vanish from this side and reappear on the other, and degrees of belief shouldn't change that way either. Of course, when some new information comes in that is substantial, then your value of P may change dramatically and quickly. But if the new information is almost identical to what you already knew, sudden jumps in P are unreasonable.

There are other reasonable properties we want our new logic to have; for example if we take a coin and toss it a thousand times and get five hundred Heads and an equal number of Tails, then we would expect than any sane assessment of P(B) where B Is the statement 'next time I toss this coin it will come down Heads' ought to be, in the light of the data, in the vicinity of one half. Happily, all the reasonable properties anyone could want are forced by a very small number indeed. We don't need to write down all the requirements we need. Given a modest few, the rest are guaranteed and follow by ordinary logic.

One thing we can do is to argue that if we have a proposition B, and we have a value P(B), then if A is some proposition that is logically equivalent to B, and we have P(A) then we ought to have

RULE 1: If B and A are logically equivalent using classical logic, then P(B) = P(A).

Another thing we can argue is that if we have B and P(B) then we ought to be able to say what P(~B) is. This should not depend on B: if B and A have the same truth value, that is if P(B) = P(A), then P(~B) = P(~A). So I shall give

RULE 2: P(~B) is a continuous function of P(B)

You should feel free to brood over this and decide if some sort of confidence trick is being pulled here. I hope you will conclude that RULE 2 is reasonable. You might feel that we ought to come right out and put P(~B) = 1–P(B), and this is certainly one possible continuous function, but let's go with the weaker assumption.

Can we do the same thing with AND and OR? If I know the value of P(B) and the value of P(C) can I say what the value of P(B.C) is? The answer to this is no, not if we want to have the interpretation we want. The example I give is from Ed Jaynes recent (and vastly entertaining) book *Probability Theory, The Logic of Science*, which has caused a certain stir in some quarters. Following Jaynes, we suppose there are fifty people in a room who have blue eyes and fifty who have brown eyes, and someone sends us one of them, picked by what rule we do not know. Then we might reasonably say that the proposition B, 'The person's right eye is blue' has credibility value around one half. And the proposition C, ' The person's left eye is blue' also has value about one half. Now the proposition B.C says that both eyes are blue, which also has credibility value about one half. On the other hand if D is the proposition 'The persons left eye is brown', then this also has value about one half, while B.D, the statement that the persons right eye is blue and his left eye is brown has credibility close to zero. So if we are to use our generalised logic to have the meaning of credibility, we conclude that the value of P(B.C) depends on what B and C actually are, not just on their credibility values. Incidentally, Fuzzy Logic tries to actually force a value on B.C which depends only on the value of B and the value of C, which tells us immediately that whatever Fuzzy Logic is about, credibility isn't it.

We are similarly stopped if we try to work out P(B+C) as a function of P(B) and P(C) only. We persevere however. We go to Modus Ponens, which is the law of inference

$$\begin{array}{l} B \\ B \Rightarrow C \\ \hline C \end{array}$$

which says that if B is true and B implies C then we can safely deduce C. This can be turned into formal logic:

$(B \,.\, (B\Rightarrow C)\,) \Rightarrow C$

It can be made more symmetric by writing it in the equivalent form:

$B.\,(B\Rightarrow C) \;=\; B.C$

This says that to assert that B is true and that B implies C, is equivalent to asserting that both B and C are true. It is a tautology, a theorem of classical Aristotelian logic. The equals sign doesn't mean that both sides are identical, it means that whenever the left side is true the right side is true, and vice versa.

Now the truth value of the left hand side should be equal to the truth value of the right hand side whenever both sides are logically equivalent by RULE 1. It is not too far fetched to believe that the truth value of B.C is some continuous function of the truth value of B and the truth value of $B \Rightarrow C$. So I postulate

RULE 3 P(B.C) is a continuous function of P(B) and $P(B \Rightarrow C)$

And it might occur to you that just multiplying the values would work nicely, giving the right answer in the extreme cases where the P values are either 0 or 1.

The above three rules are called the Cox axioms, after the physicist Richard Cox who wrote *The Algebra of Probable Inference* in 1960. It can be shown by somewhat messy algebra that if you accept these rules, then there is only one possibility for each of the functions, we must have P(~C) = 1-P(C) and we must also have

$P(C.B) = P(B \Rightarrow C).P(B)$, where the dot in the right hand side means ordinary multiplication of the numbers.

One of the consequences of this is that we can define P for an implication:

$P(B\Rightarrow C) = P(B.C)/P(B)$

Now this technically breaks one of our rules, the one that says that in the limiting case of probabilities being 0 or 1 we should reduce to classical logic. We run into the vexing problem of what happens to the truth of $B \Rightarrow C$ when B is false.

This worries a lot of people when they first meet it. Philosophy students at university who embark on a course on Logic and are told that when B is false, $B\Rightarrow C$ is true, are frequently baffled. One such, when assured that this was so by Bertand Russell, challenged Russell to show that if 0 was equal to 1 then Bertrand Russell was the Pope. Russell proved it on the spot. I give a variant of his argument:

If $0 = 1$ then, by adding 1 to both sides we get $1 = 2$, and by properties of $=$ we deduce that $2 = 1$.
The set of people consisting of me and the Pope has two elements.
Since $2 = 1$ the set of people consisting of me and the Pope has one element.
This can only happen if 'me' and 'the Pope' are different names for the same thing.
So I am the Pope.

This is a valid argument, but many people feel unhappy about it. They feel even more unhappy about the claim that if you, the reader, are a tree then Bill Gates is a pauper and Microsoft is bankrupt. Nevertheless, this is a true statement according to the rules of Logic, although it is doubtful if even Bertand Russell could have provided a convincing proof. There is a strong feeling shared by many that $A\Rightarrow B$ ought to mean something about the truth of A having something to do with the truth of B and in order to placate the unhappy, Russell chose to call this $\Rightarrow$ *material* implication, with the suggestion that the complainer was thinking of some subtly different kind of implication. Well, maybe he was thinking of the new sort of implication given by

$$P(B\Rightarrow C) = P(B.C)/P(B)$$

This is simply not defined when B is false. And it behaves in a rather reasonable manner when B is not false, ranging from when it is rather unlikely to when it is absolutely certain. Try putting in some values for B and C as statements and choose reasonable looking values for P(B) and P(C), and verify that your belief in $P(B\Rightarrow C)$, defined as $P(B.C)/P(B)$, behaves sensibly. Many pleasant hours can be passed doing this for a variety of Bs and Cs. You will find that $P(B\Rightarrow C)$ when defined this way does indeed behave in a reasonable manner, reflecting your feelings about your faith in $P(B\Rightarrow C)$ in simple cases. Since it would be misleading to pretend that the two sorts of implication are the same when they aren't, I shall use the modern notation and write $P(C|B)$, read as 'P of C given B' in place of $P(B\Rightarrow C)$. My position on implication is that this was the implication we ought to have had, because we shouldn't have used Aristotelian Logic in the first place. Had Aristotle been a bit smarter, we could have saved a few thousand years of muddle by doing logic the proper way with a continuum of values from the beginning. All classical logic was any good for was the simplest kinds of arguments, anyway. And building computers.

We might as well go the whole hog and use the probability theory formalism for everything else: we get for the modus ponens law:

$P(BC)= P(C|B) P(B)$ which reads:

"P of both B and C is equal to P of C given B times P of B."

In frequentist theory this is a definition of what is called 'conditional probability' but to us logicians it is just good old modus ponens in a generalised logic.

With the new improved implication, the three axioms given are sufficient to deduce all of Bayesian probability Theory; you have to throw in Logic as well of course to nail down the extreme case. The version of probability theory you get if you follow this line of thought is called Bayesian Probability. It has to be said that there is a sort of religious war going on in Universities between Bayesians and Frequentists, and the religion of Bayesianity has been steadily making more adherents. My own view is that it is perfectly respectable to choose whichever interpretation seems convenient, depending on the problem, and that one ought not to get dogmatic about these choices. I suspect that any problem that can be solved in one interpretation can be solved in the other, but for any problem, one is usually easier than the other (and sometimes a lot easier) to work in. The drawback of my approach is that it is considered vile heresy by both religions, but I'd rather be an apostate than a nong.

Some people find the Bayesian perspective more natural and easier to defend. Jaynes' book mentioned above is a lovely, polemical defence of Bayesian thinking and one of the more interesting books of the millennium. Get your local library to order a copy. No, I don't get a commission, I just think it's a great book and an exciting read, and bashing through it is going to give you some power-thinking skills that will beat the hell out of anything that master Yoda ever came out with.

So I have come to the crux of the case. If you want to, you can learn Bayesian probability theory. Start with Jaynes, it shouldn't take more than about ten years to finish the book, assuming you don't waste time on anything else, making it excellent value for money. If you do, you will be acquiring the skill of thinking in a non-Aristotelian Logic, just as advertised. This will make it possible for you to solve problems that are currently beyond your powers to even state let alone solve. People who can reason in such a way about the world are readily employable and useful members of society: we call them statisticians.

I claimed that mastering a non-Aristotelian logic makes you smarter and able to see things lesser mortals cannot. An example would help at this point; you can see a small problem though: if you are still a lesser mortal, how will *you* see it? Still, I shall give one anyway; it deals with the expected lifetime of the human species. Papers have been written explaining that it is very likely that the Human race will be extinct within a few thousand years. The argument is one which the simple minded non-Bayesian might find convincing, but which the Bayesian super-mind can penetrate easily and dispose of as a pile of dingo-droppings. Naturally, since you are not, as yet, a Bayesian super-mind, you won't follow this– but you may get the flavour of it.

Imagine that you are given a box which is fixed on a desk top and has a button on top. You are told that the box may contain either ten balls or a thousand balls. All the balls are the same except that one and only one has your name printed on it. You are asked to decide which box you have here, the thousand ball box or the ten ball box. All you can do is to press the button, and you are told that when you do, a ball will fall out of the box.

You reason that you have to press the button eleven times. If the eleventh button press produces a ball, then it must have been the thousand ball box, since the ten ball box wouldn't have anything to produce. So far we have conventional Aristotelian type reasoning.

You press the button once and a ball comes out. You press it again and another ball comes out. You press it again and a third ball comes out- and this one has your name on it.

You can now make a pretty good guess as to which box you have. It is one hundred times as likely to be the ten ball box as the thousand ball box. This result should agree with your intuitions if you have any. The Bayesian can provide a justification for this very quickly– but this is easy and understandable only for superbeings and you aren't one yet. You should, however, be able to see that getting your name up in the first three goes is not too improbable if there are only ten balls in the box but is awfully unlikely if there are a thousand. And if it is a hundred times as unlikely, then the ten ball explanation ought to be about a hundred times as believable. This is the intuitive, common sense approach. To a Bayesian, it is not just plausible it is blindingly obvious– although it requires some additional assumptions, which he or she can state precisely and you can't. This is because as a result of using a powerful non-Aristotelian Logic, they are smarter than you. Annoying, isn't it?

Now we come to the life of species. We accept the opinion of anthropologists that the human race has been in existence for less than a million years, and more than one hundred thousand. Just how long depends on how exactly you define human, so there is some unavoidable fluffiness about this time, but it does not affect the argument materially. Now consider the two possibilities: first that the human race will last another million years, and the second that humanity will be extinct within five thousand years.

In the first case, the total number of human beings who will ever have lived, a number which grows exponentially with time, becomes something colossal. The number in the first hundred thousand (or million) years is approximately twice the number currently on the planet, around six billion. If the present population level continues for a million years, a very modest assumption indeed, the total number who will ever have lived at the end of that time will be about ten to the power fourteen.

In the second case, with a life time of the species of five thousand years, the total number of human being who will ever have existed is much smaller, only about fifty times as many as at present.

Now given these two possibilities, and *given that you are alive at present*, your name is on the ball, the lifetime for the species of five thousand years is much more likely than the lifetime of a million years. The probability that you would be here, right at the beginning, in the first fraction of a percent of all people, is obviously very small. The analogy with the boxes and balls is obvious and the same kind of reasoning gets you to the result. From which we deduce that the human race is likely to become extinct quite soon in historical time.

You may find this conclusion utterly convincing or totally unconvincing. Some people may be found who take it very earnestly indeed, others think the argument sucks. It has provoked a lot of debate, and many pages of sometimes heated writings can be found. The point I wish to make is that amateurish argument of the 'It seems to me …' sort is a waste of time. A Bayesian can dispose of it quite quickly and make the underlying assumptions explicit in both the case of the two boxes and the two lifetimes. A simple example of a problem that can lead the ordinary muddled human being into endless hours of debate with no clear end in sight, but where the properly trained Bayesian thinker can cut through it immediately. If you imagine an evil galactic overlord wishing to cause alarm and despondency by throwing the expected lifetime of the species at us, ('Har har, terran scum, you will be extinct soon anyway!!!') and the beautiful girl falling into the arms of the man who can solve the problem in short order, you know what to do to collect the beautiful girl. If you don't want a beautiful girl, preferring perhaps a handsome man or a few bottles of plonk, make the appropriate changes.

No, I won't tell you the answer, unless you are a beautiful girl or a few bottles of plonk. If you want to dispose of the matter in a clean and compelling way, learn Bayesian probability theory and apply both your brains. You will also be able to solve a good many other more important problems.

Two final matters which might trouble the sceptic. First, is it possible that nobody can learn to be more intelligent, more competent by these means, but only that you have to be much cleverer than average to master the damned stuff in the first place? In other words that I am cheating you, the causality is the other way around. It is not that a training in probabilistic logic makes you smarter than average, it is that you have to be smarter than average in order to survive the training.

I am able to assure the sceptical reader that a close investigation of some of my colleagues who are professional statisticians has revealed no signs of innate intellectual superiority whatever. One cannot rule out the possibility that they are merely concealing superior minds, possibly in the hope of making more friends, but if so they are doing a very fine job of it.

The second worry is altogether a graver matter. Are there any side effects? I understand that being streched ten centimetres on the rack does indeed make you taller, or at least longer, but gives a certain languor to the personality. Ex-rackees are said to spend a lot of the time lying down and are slow off the mark when pursued by bears or vampires. Are there similar undesirable side effects of being put on the mental rack, being made to learn Bayesian probability theory?

It is hard to say. A comparison between those who have learnt orthodox probability theory and statistics and those who have done the Bayesian theory would seem to indicate that the former does indeed have much the same effect as being stretched on the rack. The victims are frequently pallid and harrassed looking, and low on humour and vivacity. They appear to have been trained beyond their natural intelligence. Bayesians on the other hand seem to be of a sunnier disposition, wittier and altogether better company. But this was a small sample and neither orthodox nor Bayesian statisticians would be inclined to build much on the data. You will just have to take your chances.