THE MODAL ÆTHER

THOMAS FORSTER

This essay owes its appearance here to the good offices of Reinhard Kahle who, at short notice, allowed me the opportunity to turn into something publishable a rat's nest of clandestine fragments hitherto available only to the author's friends and students. I had always assumed that they were in any case so heretical that nobody would publish them even if I tidied them up and I am grateful to Kahle for giving me the chance to be burnt rather than merely ignored. The one major regret I now have about the delay caused by my timidity is that in the dozen-or-so years that have passed since the first draughts of this essay were circulated David Lewis has died. Readers familiar with the literature will immediately recognise Lewis as the most important single creator of the Augean stables I am reporting on, and as the writer most likely to wish to maintain them intact as a World Heritage Site. It is true that I thought his ideas terrible, but he defended them with personal integrity and without malice: I enjoyed his company, and am sorry that he is not around to reply.

The three essays out of which it grew were entitled 'The modal æther', 'Indexicality' and 'The closest possible world'. In the first essay I argue that if possible worlds are to be used at all to explain necessary truth, then at least some truths (those concerning relations between worlds) are necessary in virtue of something other than truth in all those worlds. In the second I argue that the idea that actuality as indexical is in need of a lot of explanation. In the third I argue that there is no logical notion of closest possible world. The essays get progressively more technical, but I preface them with an introductory essay whose general drift can be caught even by those with no formal background, and insert between the first and second essays a brief sketch of the topological ideas on which ideas of indexicality and proximity (of the two final essays) presumably ultimately rely.

§1. How did it all start? I once heard a philosophical colleague telling a lecture-threatre full of first year students that an inference was valid if "there was no possible world in which" the antecedent was true and the conclusion was false. Perhaps my readers are not as shocked by this as I am. They should be: the idea of a possible world is not needed for an explanation of the concept of valid argument, since we knew what valid arguments were before we had possible world semantics.

However the point is not that this is a shocking story, and that people teaching first-year logic ought to know better. The point is not even that this represents an unwarranted intrusion of possible world jargon into an area where it has nothing to offer. The worry is that there are nowadays many other applications of possible world imagery in more complex settings than this (counterfactual, fictions etc etc) and this story raises the possibility that they all may be fully as misconceived as this one was. I fear they are: in fact over the years I have been driven reluctantly to the conclusion that all the uses known to me of possible world semantics beyond formal logic are misconceived, and that the light that possible world imagery appears to have shed on various philosophical problems in recent years will be seen in the years to come to have been entirely spurious.

I want to emphasise that my concern is a philosophical concern not a mathematical concern. I am not a mathematician annoyed at someone else playing with my toys: I am a philosopher who is concerned that bad use of mathematics makes for bad philosophy.

Anyone with a naturalistic philosophy who wishes to say that a research programme is misconceived had better have not only evidence of the misconception, but a theory about how the mistake came to be made: the fact that people spin the theories that they do is itself a part of nature and deserves an explanation just as much as do the phenomena those theories are theories of. I shall be offering such a theory.

By the middle of the last century the soil out of which possible world analyses were to grow had been prepared and mulched. By then the concept of a mathematical structure was more-or-less settled. A structure was a set ("carrier set" or "domain") with some associated relations or functions. For example a *group* was a set of elements with a multiplication operation defined on it satisfying certain rules and so on. This picture is useful and unproblematic and is serving mathematics well, but it was never going to give any way of describing a logic internal to a structure that is different from that being used by the researcher externally to erect it.

In possible world semantics there is not merely one single carrier set, but a whole host, and the host is equipped with relation(s) which hold *between* the sets (worlds) on top of the old relations that hold *within* the worlds. The most important of these relations is a binary relation of *accessibility*. This relation of accessibility can be used to give more complicated definitions of what it is for a formula to be true in a world than had been possible before. For example we are no longer compelled to say that $W \models \neg p$ iff $W \not\models p$: we can say $W \models \neg p$ iff no world accessible from $W \models p$. This can give us a semantics for a logic in which the law of double negation fails. Suppose there is a world W that does not believe[1] p, but that there is, from every world accessible from W, a world that *does* believe p then W believes $\neg\neg p$—even though it doesn't believe p. A particularly pleasing illustration can be given where the accessibility relation is transitive, and can be thought of loosely as succession in time. Then $\neg\neg(\exists x)(S(x))$ is "One day there will be a saviour"; $(\exists x)\neg\neg(S(x))$ is "the one who will become the saviour is already among us" and $(\exists x)(S(x))$ is "The saviour is here!". A transitive accessibility relation is associated with intuitionistic logic, and this illustration reminds us that we should think of intuitionistic double negation as *prophecy*.

Since the several worlds in the family of possible worlds can disagree about the truth values of particular formulæ, if we wish to ascribe a truth-value to a formula

[1] Slang for "$W \neg \models p$".

on the basis of its behaviour in such a family, it is customary to designate one of the worlds to speak up when verdicts on truth-values are required. Accordingly the world so designated is called the *designated* world.

The expression 'laboratory toy' is information technology slang for a physical device that is understood from a theoretical point of view, and which will behave according to specification when handled in a laboratory by an expert, but which cannot be let loose on the public because the conditions for its safe or reliable use cannot be guaranteed. Those who use laboratory toys outside the laboratory do so not only at their own risk, but to a certain extent at the risk of others too. One is reminded not only of Goethe's *Zauberlehrlinge* but also of Hergé's delightful possible world *Les Bijoux de la Castafiore* wherein the jewels of the title are stolen by magpies who are entirely unaware of their significance. Possible world semantics is a laboratory toy: safe in the hands of logicians perhaps, but not in the beaks of magpies, nor in the beaks of their students.

The basic assumptions that make possible-world analysis work in the logician's laboratory are roughly as follows: (i) The language for which the semantics is being provided is agreed and defined in advance; (ii) There is a completely unproblematic transworld identity relation (between inhabitants of different worlds); (iii) The identity critereia for possible worlds is unproblematic, and (iv) The web of worlds is being created by fiat in some well-controlled non-circular process, and not happening inside any of the component worlds.

All these assumptions are too basic ever to have been spelled out by the logicians who use this style of semantics. The result is that the health warnings never got written. Let us consider these assumptions in turn: what they mean and how they do not hold outside the laboratory. The discussions will overlap, since the conditions fail in related ways.

I can well imagine that many philosophers reading this will irritably exclaim that these assumptions are distorting oversimplifications which they do not make when applying possible world semantics to their concerns. However these are assumptions that have to be made if the techniques of possible world semantics are to be applied, and if these assumptions are wrong then the point is being conceded that possible world semantics are not applicable to their concerns. If this is so, then the uses of possible world discourse outside logic are rhetorical not substantial.

1.1. An agreed language. It is a matter of record that there is no agreement on what might be the formal language in use during any of these applications of possible world semantics to philosophy. Nor indeed is there any attempt to secure agreement. This looks defensible: can it really matter what the language is? Can we not use this gadgetry whatever canonical language we finally decide on?

The important point about the canonical language is that every possibility must be a consistent theory in it, and *vice versa*. Thus there is a problem about what the constant symbols of this language are to be. Is there to be one for each inhabitant of the union of all the worlds? There are numerous reasons for the answer to be 'yes'. If not all objects have names then those that do are specially privileged and we have thereby built in a *de re* modality from the outset. We can permute unnamed members of a carrier set for a possible world to obtain a new possible world. Or is it new? If not all objects have names then we can construct elementarily equivalent worlds which are nevertheless palpably distinct: they are non-isomorphic for example. Is this one possibility or two?[2]

But if every object in any carrier set for any possible world has a name in the canonical language it is no longer possible to invent new constants.[3] Inventing new constants is an essential technique in the standard construction of saturated models of theories[4] by means of the completeness theorem. For any finitely satisfiable set of formulæ $\{\Phi_i(x) : i \in I\}$ of formulæ we invent a new constant symbol a and an axiom $\Phi_i(a)$ for every i. We would find ourselves in a situation where although saturated models of appropriate theories remain possible, we cannot appeal to the usual devices of the completeness theorem to create them. Thus it would no longer be the case that possibilities are all reified into worlds.

Most philosophers would probably want for the canonical language something a great deal richer than the spavined first-order languages which are essentially the only languages which admit the completeness theorem, and would feel that a higher-order language is called for. But the completeness theorem is what tells us that existence of a model is the same as freedom from contradiction, and if it is not available there is no longer any reason to believe that to every possibility there corresponds a world.

1.2. Unproblematic Transworld Identity. The mere fact that transworld identity excites debate at all shows that this condition for the application of possible world semantics is not met. But even if the philosophical community were agreed on the nature of transworld identity it could still have got it wrong. Philosophy has a way of throwing up new problems which have interesting parallels with familiar old ones.[5] There are obvious parallels between the problem of ascertaining who I am in other possible worlds and the familiar problem of who I was in a previous life in *this* possible world. It might be an idea to use the familiar problem as a point of departure. After all, if I can find out who I am in some *other* possible world, finding out who I was in a previous life in *this* possible world should be comparatively straightforward. Even this appears to be quite hard: at any given time there are plenty of people who think they were Napoleon in a previous life (I saw many when I worked in a psychiatric hospital); at any time at most one of them can be right, and there doesn't seem to be any obvious way to ascertain which.[6]

[2] Two structures are elementarily equivalent for a language if the language cannot express the difference between them.

[3] 'The set of all constants' sounds like a denoting term from the Age of Nightmares of 100 years ago. It is true that consistent set theories with universal objects like the set of all sets and the set of all ordinals can be found, but they disallow certain manipulations of such sets which would certainly have to be allowed in a theory-of-everything. However, I shall not pursue the matter, as I believe that the incoherence of philosophical possible world talk has nothing to do with the logical or semantical paradoxes.

[4] A structure is *saturated* if whenever $\{\Phi_i(x) : i \in I\}$ is a set of formulæ with 'x' free such that for any finite $J \subseteq I$ there is an object a such that $\bigwedge_{i \in J} \Phi_i(a)$ (we say $\{\Phi_i(x) : i \in I\}$ is *finitely satisfiable*) then there is an object a satisfying *all* the Φ_i simultaneously ($\{\Phi_i(x) : i \in I\}$ is *satisfiable*).

[5] See Cresswell, *op. cit.*

[6] And now we'll never know which. The psychiatric hospital in question has been long since closed down and all the possible-Napoleons untraceably dispersed into care-in-the-community.

If the canonical language is one of a sort for which we can prove the completeness theorem then we can indeed reify possibilities into models. But the completeness theorem does not enable us to reify *de re* possibilities into worlds. In situations where we are wondering if *some given object* might be Φ the completeness theorem has nothing to say at all. Consider theory of groups with a constant 'a' denoting an involution, and theory of groups with a constant 'a' denoting an object of order 4. These are both possibilities, and according to the completeness theorem there are models (which of course are groups) manifesting these possibilities. But, as every user of the completeness theorem knows, this doesn't mean that there is an object that might be an involution and also might be an element of order four. That's not to say that we cannot put these two groups into a possible world structure and identify the two designata of 'a' by fiat, but it is precisely *by fiat* that that identification arises, and not from the two applications of the completeness theorem. This difference between the obscurity of transworld identity in the philosophical applications of possible world semantics and the utter straightforwardness of transworld identity by fiat in the original setting is very striking.

1.3. Noncircularity and stipulation. Consider the formula $\Diamond\exists x\Phi(x) \to \exists x\Diamond\Phi(x)$ Suppose we wish to falsify it. We want a possible world model in which it is possible that there is a thing which is Φ, but there is no thing which is possibly Φ. It is easy to arrange this: merely put in one of the worlds something which is Φ-in-the-sense-of-that-world, and ensure that no such object inhabits the designated world. This is feasible because all the worlds in a model—the designated world included—are constructs into which we can put whatever we please: the world we inhabit is a toolkit that has things just lying around and we can choose to put them into any given world W or not, as the spirit moves us. In particular, we can choose what to put into the designated world, and indeed which world to designate. However the *actual* world is actual because it just *is* actual, not because we have designated it as actual. We are most certainly *not* free to put whatever we choose into the actual world, still less are we free to choose to leave things out of it. And the situation is rather worse than that, for on any sensible account of what possibility is, we are not going to be able to stipulate what any other possible worlds contain either. We do not control possible worlds by stipulation in the way we control the components of a model. Possible worlds are not brought into existence by our stipulations, but are brute facts of life.

This difference from the original setting doesn't of course mean that possible world semantics cannot be applied elsewhere, but it does serve as a warning that the setting is very different.

Thinking you might have been Napoleon in a previous life is apparently no longer grounds for being judged a menace to yourself or others and confined at public expense in a psychiatric hospital. On the other hand thinking that you are (were? might be?) Napoleon in another world is apparently still grounds for being maintained in a philosophy department. Perhaps this will change once the *Nouveau Right* realise that philosophers, too, can be rehabilitated by care-in-the-community at great savings to higher-rate taxpayers.

The question of what happens if the possible worlds can individually describe the relations between them is complex enough to deserve an entire section to itself.

§2. The Modal Æther. There is an argument due to Aristotle that is designed to show that there is only one universe: if there were two, they would just be two parts of one. This does not apply straightforwardly to the modern theory of possible worlds, for these are conceived expressly in response to the compulsion we are under (given the administrative complexity of life in a world with possibilities) of submitting universes in multuplicate. In this context, Aristotle's point is the slightly more subtle one that nothing happens *outside* possible worlds—*there is no modal æther*. If a possible world is a way the world might be, then it must furnish answers to all questions about the way the world might be. In particular it must have answers to all questions about possible worlds, since there are such things.

The invention of multiple possible worlds (over the old-style models I wrote about above where there is only one carrier set) creates an enormous amount of new paperwork. There are various accessibility relations between possible worlds; relations of satisfaction between propositions and worlds; there are relations of *habitation* between objects and worlds; relations of correspondence or identity across worlds; the property of being a possible world and so on. The machinery that I am trying to point to can be recognised by its two salient features: (i) none of it was needed under the old dispensation (ii) it doesn't appear to be located inside any one possible world, and nothing seems to be gained by attempting to locate discussions of it inside possible worlds at all. Indeed it could be characterised as that part of our theory of nature that remains when all information internal to possible worlds is ignored altogether, rather in the way in which the geometry of space-time is what remains once we expunge events. Let us call it the *machinery*.

What is the metaphysical status of the machinery? Let us say that a property $\Phi(\vec{x})$ is noncontingent iff $\forall\vec{x}(\Phi(\vec{x}) \longleftrightarrow \Box\Phi(\vec{x})))$. The list of free variables may be empty, so we can talk of noncontingent *propositions* as well as noncontingent *properties*. It's pretty clear that all the properties the machinery is concerned with are going to have to be noncontingent, and this is the received view: the property of being a possible world; the binary property of being an inhabitant of a possible world; all relations in the style "$\Psi(\vec{x})$ holds in W" where Ψ is atomic and quite possibly even when it isn't, are one and all noncontingent. If we do not make the assumption that the property of being a possible world is noncontingent, it would seem that one of the questions which will occupy us below, "does the truth value of '$\Psi(\vec{x})$ holds in W' depend for its answer on the world in which it was being asked?" could not itself even be asked, for the W we are asking about might not even be a world in the other world we are asking the question! And what is this other world anyway? Who says it is a world? It is easy to feel that failure to make the assumption that "W is a world" is noncontingent lets loose an infinite regress. It is indeed fortunate that the possible worlds tradition, by choosing the horn of the dilemma that represents truth-in-a-possible-world as neccessary, spares us the need to explore this regress.

Can the machinery be properly described inside all possible worlds, despite appearances? Or does it go on outside them, in a modal *aether*? In what follows I shall go along with the received view that facts about the machinery are noncontingent, and explore what happens if one attributes their noncontingency not to the æther but to their having the same truth-value in all possible worlds. I extract consequences that seem to me absurd, and conclude that at least some necessary truths are true not in virtue of what happens in possible worlds, but true in virtue of what happens in the æther. In other words, possible world semantics doesn't provide a uniform account of necessary truth.

2.1. Truth in a Possible World. It is a simple matter to devise a paradox along the lines of Tarski's theorem on the undefinability of truth, by requiring that the actual world contain truth-definitions for the possible worlds, and *a fortiori* for itself. One might be tempted to claim that this demonstrates the incoherence of the enterprise. This would be unfair, for the whole point of Tarski's theorem is that *any* attempt to construct truth-definitions in this style for an entire language (with certain minimal closure conditions) will result in paradox, and this will happen even if there is no plurality of possible worlds. Mostly this problem is simply ignored. I shall follow general practice, and will try to reason in such a way as to ensure that any absurdities that do crop up do not arise merely because of Tarski's theorem.

If semantics is to go on inside possible worlds rather than the *ae*ther then we have to make sense primarily not of "ϕ is true in W_1" but rather "It is true in W_2 that ϕ is true in W_1". There is an immediate and obvious possibility of an infinite regress here: the same argument will establish that we have to make sense primarily not of "It is true in W_2 that ϕ is true in W_1", but rather "It is true in W_3 that it is true in W_2 that ϕ is true in W_1".

How vicious is this regress? The ωth stage presents us with a situation in which each original n-place predicate has become an $\omega^* + n$-place predicate[7] where all argument places except the last n are occupied by variables ranging over possible worlds. A language with predicate letters that have infinitely many argument places is an alarming prospect to non-logicians, but even more alarming is the thought that the regress might not halt there. After all, the same step of asking whether or not this (by now infinitary) atomic relation between individuals and worlds holds can be done in any possible world, and the regress resumes and will continue transfinitely.

This has to be nipped in the bud before the Burali-Forti paradox comes up over the horizon. We absolutely must have a proof that for all W_1 and W_2, and for all ϕ, ϕ is true in W_1 iff it is true in W_2 that ϕ is true in W_1. The obvious way to do this is by structural induction on formulæ.

Presumably noncontingency will tell us that, for atomic ϕ at least, W_1 thinks that W_2 thinks that ϕ iff W_2 does indeed think that ϕ. But even this comes at a price.

THEOREM 1. *All worlds and all individuals inhabit all worlds.*

[7] ω^* is the length of the negative integers in their natural order. Thus $\omega^* + n$ is in fact equal to ω^*, but I write it this way to keep in mind the fact that the last n places are different from the first ω^* of them.

Proof:

W_1–believing–$\phi(x)$ is a relation between W_1 and x. Noncontingency tells us that every W_2 must believe W_1 and x to be related in this way. But for W_2 to have any beliefs about W_1 and x they must both inhabit W_2.

■

So far so good. Let us now consider the induction step for quantifiers and connectives. By way of illustration consider the induction step for →. Suppose W_1 thinks that W_2 thinks that $A \to B$. So W_1 thinks that every possible world accessible from W_2 that thinks A also thinks B. But by noncontingency of possible-worldhood and of accessibility (both from the machinery) something is a possible world accessible from W_2 iff W_1 thinks it is. And by induction hypothesis such a world will satisfy A iff W_1 thinks it does. So if for all W_1 and W_2, W_1 thinks that W_2 thinks that A iff W_2 does indeed think that A, and the same holds for B, then it holds for $A \to B$.

The other propositional connectives are straightforward too: let us just check the existential quantifier. Let us consider the inductive step for the existential quantifier. Suppose W_2 thinks that W_1 thinks that $(\exists x)(\phi(x))$. The semantics for the existential quantifier are that W thinks that $(\exists x)(\phi(x))$ iff there is an x such that W thinks that $\phi(x)$. Applying this twice we arrive at: there is an x such that W_2 thinks that W_1 thinks that $\phi(x)$ and by induction hypothesis that will be W_1 thinks that $(\exists x)(\phi(x))$ as desired. The universal quantifier is similar.

2.2. Interworld Relations.

We have just seen an argument to the effect that every world and every individual inhabits every world. This relies on the assumption that the truth-value of "W_1 believes ϕ" for ϕ atomic, doesn't depend on the world it is evaluated in. If we weaken the assumption we can obtain a weakened conclusion by a different argument.

Theorem 2. *Some objects belong to more than one world.*

Proof: If x (in W_1 but not W_2) is a counterpart of y (in W_2 but not W_1) then this assertion has to be made in some possible world, W_3, say, which is different from W_1 and W_2. So $W_3 \models x$ is the W_1 counterpart of the W_2 object y. But if objects belong to only one possible world, then $W_1 = W_2 = W_3$, and there is only one world.

■

Theorem 3. *$K5$ is true*

Proof:

If I assert that p is (logically) necessary I am claiming that p is true in all possible worlds. *All* possible worlds, note. This is in sharp contrast to the standard situation where I am asserting that, say, all dodos have died, where I don't actually mean *all* dodos, merely all dodos in this world. This difference is very important! The standard situation is captured by the recursive definition above: $W \models \forall x \Phi$ iff $\forall x \in W$ $W \models \Phi$. So if we are to stick to the recursive definition and successfully produce the intended effect of the universal quantification implicit in "p is necessary" we conclude that these two accounts have to give the same result. Therefore (at least in the case where we are quantifying over worlds) it cannot make any difference whether our domain of quantification is the world we inhabit or the union of all possible worlds. That is to say

$$\forall W[((\forall W')(W' \models \Phi)) \longleftrightarrow ((\forall W' \in W)\ W' \models \Phi)]$$

or equivalently

$$\forall W[((\exists W')(W' \models \Phi)) \longleftrightarrow ((\exists W' \in W)\ W' \models \Phi)]$$

In other words, if there is a W in which Φ holds, then in any possible world there is such a W. Thus, if Φ is possible, every world thinks it is possible. This is the $K5$ principle $\Diamond\Phi \to \Box\Diamond\Phi$.

■

Thus we have shown that logical necessity obeys $K5$. Notice that none of this depends on Φ being *closed*: the same proof works for Φ with $\vec{x}$ free if for "all possible worlds" we read "all possible worlds inhabited by $\vec{x}$".

2.2.1. *Do all possible worlds inhabit themselves?* A possible world is merely a world that could be actual, so for every W there is W' that thinks W is actual. This W' is presumably W itself, since no world can simultaneously believe that two distinct worlds are both actual, and the idea that we can have two worlds W_1 and W_2 each of which thinks the other is the actual world is unpalatable. All this does is tell us that all questions about actuality can be easily solved: $\forall W \forall W'(W \models \text{Actual}(W') \longleftrightarrow W = W')$.

What can we say about relations between objects that belong to different possible worlds? Do objects ever belong only to different worlds, or for any x and y is there some possible world they both inhabit? The problem is that, in any possible world W in which the question is asked, the answer is trivial. For any two objects there is a world they both inhabit, namely W, since inside W we cannot quantify over anything outside W. *And that is the only answer we are going to get.* This is a totally unsatisfactory state of affairs. The only way of getting any other answer would be to admit that there is a modal æther.

There is a strong temptation to believe that these questions have answers, that is to say that the answers do not depend on which possible world we ask them in. That is to say, the answers are logically necessary. But it does not seem to be *logically* necessary that any two objects inhabit a common possible world, nor that there should be two objects that don't. We seem forced to the conclusion that these questions do not have answers.

2.3. An *amuse-gueule*, and a conclusion. I close with a simple argument discovered only in the process of final revision of this essay.

Let T be the theory consisting of all facts about the machinery. Suppose T had an axiomatisation with an axiom ψ and remaining axioms T' not entailing ψ. Then $T' \cup \{\neg\psi\}$ would be a consistent theory, and therefore true in one of the possible worlds.[8] So there would be a possible world that thought that the truth about relations between possible worlds was represented by $T' \cup \{\neg\psi\}$ rather than T. But then ψ would be a truth about relations between possible worlds that was not neccessary. But ψ was necessary. So there is no axiomatisation of

[8] I hope nobody will object to this in the grounds that $T' \cup \{\neg\psi\}$ can't be consistent because ψ is neccessary. Such an objector would have us believe that there are no independent axiomatisations of any theory consisting of necessary truths.

T with ψ as an independent axiom. But it is standard that for any theory T and any theorem ψ of T the set $\{\psi\} \cup \{\psi \to \sigma : \sigma \in T\}$ is an axiomatisation wherein ψ is independent. We will need this last fact again, so we may as well have a proof now. Suppose *per impossibile* that ψ followed from $\{\psi \to \sigma : \sigma \in T\}$. Then it would follow from finitely many $\psi \to \sigma_i$ and we would have $(\bigwedge_{i \in I}(\psi \to \sigma_i)) \to \psi$ for some finite set I. But this last is equivalent to $(\psi \to \bigwedge_{i \in I} \sigma_i) \to \psi$. Now $(((\psi \to \bigwedge_{i \in I} \sigma_i) \to \psi) \to \psi)$ is a truth-table tautology (Peirce's law) so we can deduce ψ outright.

(I suspect that a reply to this last argument can be given along the following lines. By all means $T' \cup \{\neg\psi\}$ would be a consistent theory, and therefore true in one of the possible worlds: the point is that no problem arises because in those worlds in which it is true it is not a theory of the machinery, but a pointless theory describing only some artifice wished on us by the completeness theorem.)

To summarise: if we want possible world semantics we are stuck with the machinery; we have to give an account of truths about the machinery; there seems to be no sensible alternative to the received view that makes propositions about the machinery noncontingent; there doesn't seem to be any way that this necessity can arise from truth in all possible worlds. It must come from the æther.

This is messy rather than catastrophic, but accepting the reality of the modal æther commits one to a philosophy in which the mediæval distinction between knowledge from reason (of the innards of the possible worlds) and knowledge from faith (about the æther) play a rather larger rôle than one might think desirable. Surely we have made more progress since the renaissance?

§3. Topology and Possible Worlds. The two remaining essays treat issues in possible world theory that are normally approached topologically when encountered elsewhere. In spacetime consideration of indexicality and proximity involves topology. What can we say about topologies on possible worlds? A brief sketch of some background topology may be in order.

A topology on a set X is a family O of "open" subsets of X. Any set that is a union of open sets is open; the empty set is open, and any intersection of finitely many open sets is open. A function from X to X is *continuous* iff $\{f^{-1}(x) : x \in X'\}$ is open whenever $X' \subseteq X$ is open. In euclidean space an open set is either a *ball* (for every point p and real number α, the set of all points within α of p form a ball) or obtained from such balls by (arbitrary) union and (finite) intersection as above. The balls are said to be a *basis*. A continuous map from the space onto itself, whose inverse map is also continuous is said to be an *autohomeomorphism*.

If, for any two points x and y in the space, there is a way of splitting the space into two open subsets, one of which contains x and the other y, then the space is said to be *totally disconnected*.

These ideas evolved first in connection with attempts to describe transformations of space, but there is a natural topology on spaces of models too. An *ultrafilter* in a boolean algebra is family F of elements satisfying the three conditions (i) if a and b are both in F, so is $a \wedge b$. (ii) if $a \geq b$ and b is in F, so is a. Finally (iii) either a or $\neg a$ is in F. The stone space of a boolean algebra B

is the space whose points are the ultrafilters of B, and where a basic open set (corresponding to the balls in euclidean space) arise from elements of B. For each b in B, the set of ultrafilters containing b form a basic open set.

If one is looking to topology to provide a weapon for the analysis of relations between possible worlds the natural thing to reach for is the stone space of all complete theories extending the canonical theory T containing all necessary truths.

This extension of topological ideas from physical space to Logic is due to Tarski, though curiously spaces like this are usually known as 'Stone Spaces', after M. H. Stone.

Stone spaces tend to be totally disconnected: the set of extensions of T that prove ϕ and the set of extensions of T that prove $\neg\phi$ are complementary and both open. Euclidean space is not totally disconnected: indeed it is not possible to split euclidean space into two disjoint open sets at all: the complement of an open set is never open.

Lindenbaum algebras—at least of first-order theories—tend to be *homogenous*: for any two points of the algebra other than 0 and 1 there is an automorphism sending one of them to the other. This will ensure that the Stone space of complete theories extending T is homeogeneous in an analogous sense: given any two points of the space, there is an autohomeomorphism sending one of them to the other. Euclidean space, too, is homogeneous in this sense. The question of the homogeneity of the Stone topology on possible worlds is one that hasn't been addressed, and I suspect the answer will depend sensitively on the canonical language. Believing, as I do, that there is very little sense to be made of the idea of a space of all possible worlds in the sense envisaged by some philosophers, any attempt on my part to examine the question of the homogeneity of its space would be an exercise in *mauvaise foi*.

In the two following essays I examine two ideas for applications of possible worlds that turn out to pull in opposite directions. To argue that actuality is indexical one wants the stone space on possible worlds to be homogeneous. On the other hand if it is homogeneous then there is no nontrivial logical structure on it and one cannot look to topology to provide a notion of closeness of possible worlds.

§4. Actuality and Indexicality. The thesis that actuality is indexical is the thesis that the difference between a possible world and the actual world is just like the difference between here and there, or between now and then. The thoughtful reader will immediately suspect there might be a parallel with McTaggart's arguments about the unreality of time. Such a reader should cast an eye over Cresswell [1990].

Part of the difference between there and here (or rather the relation between there and here) is that one can *get from* there to here. Or from then to now. These ideas are developed in point-set topology and have given rise over the years to a well-understood notion of a *connected space*. In layman's terms a space is connected iff any two points in it can be connected by a continuous line lying entirely within the space. If one believes that the relation between possible worlds is indexical, it is natural to seek a topology on the family of possible worlds which will give us a similar analysis. There is a natural topology on the family of possible worlds, but unfortunately it is not connected. However it turns out that connectedness is only part of the story anyway.

The picture below shows a standard illustration from point-set topology of a connected space. Although this topology is connected (one can get from any point in it to any other point in it by following a line lying within the space) the difference between pairs of points is not always what we would consider indexical. The difference between any two points in the disk is indexical: not only is there a path between them, but there is an autohomeomorphism of the space that sends one of them to the other. Likewise the difference between any two points on the excrescence: there is a path between them and there is an autohomeomorphism of the space sending one to the other. The relation between a point on the excrescence and a point in the disk is not indexical: there is a path between them all right, but no autohomeomorphism that moves one to the other. The point where the excrescence meets the disk is not related indexically to any other point. There are four bundles of points: (i) the points in the disk, (ii) the points on the excrescence, and (iii) the single point at the junction, and (iv) the remaining points on the boundary of the disk. The relations between points within each bundle is indexical: the relation between points in different bundles is not.

Figure 1. A connected space

This reminds us that another important ingredient of indexicality is the idea of *indistinguishability* between indexically-related points so beguilingly alluded to by David Lewis who would always happily tell all comers that "Possible worlds are *just like* this one, only they're possible not actual" (my italics). The idea is that there is no way of distinguishing between addresses in spacetime (or between worlds) on the basis solely of information that refrains from mentioning events at those addresses.

This idea of indistinguishability can be captured in various ways: for example by the idea of an automorphisms, where we say that two things are indistinguishable if there is an automorphism sending one to the other. One could also approach it logically by saying that two things are indistinguishable if in the appropriate language there is no formula that tells one from the other. If there is no connectedness analysis available to us (remember that the usual "Stone" space on possible worlds is not connected) we will have to look to a logical analysis like this to explain what indexicality of worlds is.

On this account, claiming that the relation between any two worlds is indexical is to say that the machinery has no way of telling two worlds apart: any one-place predicate from the machinery holds of all worlds or of none. But then we can ask: can the machinery distinguish *tuples* of worlds? There is some point to this question, as the following example will make clear. Consider a universe consisting of two causally disconnected spacetimes, as in the illustration below:

Figure 2. Two causally disconnected spacetimes

The relation between the two members of a pair both of which are in one of the two halves is indexical (like the pair $\{a, b\}$) and the relation between the two members of a pair which belong to different components is not ($\{a, c\}$). When (as is the case here) we have a connectedness analysis available, we can say that the relation between the components of the first pair is indexical (we can travel from a to b) but the second is not. However, even if we do not have a connectedness analysis available, we can detect the difference between pairs whose two components come from the same half of the space and those whose two components come from different halves becuase there is no automorphism sending a pair of the first flavour to a pair of the second flavour. So if we have no connectedness analysis available, we will want to explain the indexical relation between all points not by saying merely that any two *points* are indistinguishable, but that any two *pairs of points* are indistinguishable.

There is a temptation at this point to say that the machinery must be logically trivial. If the machinery cannot distinguish pairs of worlds then if even one world can see one other world then every world can see every other world! This would indeed make for a trivial machinery. However this would be a bit hasty: it is probably only a part of the machinery that has to be logically degenerate in this way. After all, relations between dates are indexical even though there is an order relation on them: when thinking about indexicality one discards time and surveys matters *sub specie æternitatis*. Defenders of the idea that relations between possible worlds are indexical must be given the chance to offer a reasoned explanation of which details can be discarded and why: after all, the process of judicious discarding that took us from cartesian geometry to point-set topology took several hundred years and a great deal of ingenuity. We cannot expect the analogous task for possible world theory to be done overnight, but a start would be nice.

§5. There is no Logical Proximity Relation. The most baroque of the extravagant claims made by possible world theorists was the claim that possible world theory had something useful to say about counterfactuals. The plan was that was that a counterfactual would be true at a world W if the consequent was true at the closest world to W in which the antecedent held. Sadly this project had no logical underpinning.

For these purposes we may take possible worlds to be complete theories. Let T be an arbitrary theory and ψ an arbitrary theorem of T. A "closest world to T in which ψ is false" will be a theory which is like T in all possible respects except that ψ is false in it instead of true. We will explore the possibility of a logical basis for this notion.

There are many conceptions of logic, but we need distinguish here only between a *narrow* one (logic as deduction) and a *broad* one (logic as the centre of the web). The question of whether or not there is a logical (in the broad sense) notion of closeness is too extensive and technical to be treated here. However it does seem to be fairly straightforward to establish that there is no notion of closeness which is logical in the narrow sense.

If we construe logic narrowly, then the only thing we could mean by a (logically) closest possible (ersatz) world to T in which $\neg\psi$ holds is a theory which is like T in all possible respects except that it proves $\neg\psi$ instead of ψ. We can show that in no interesting cases is there a unique such theory.

First we show that any theory T which proves ψ has an independent axiomatisation in which ψ is an axiom, and this we have already done. In such an axiomatisation of T we can replace ψ by $\neg\psi$. The result is an axiomatisation of a new theory which is a desired closest possible (ersatz) world. Finally we show that this can be done in (infinitely) many ways, with no (narrowly) logical grounds for preferring any one to the others. The construction is very elementary, uses only manipulations of the propositional calculus and accordingly is unaffected by enrichment through quantifiers, higher-order variables, predicate modifiers etc.

There will remain the question of whether or not there is a notion of proximity which behaves in the way we want and is logical in some more broadly construed way. Of course if Logic is merely the stuff at the centre of the web we can choose where to draw the line between the core and the periphery, and simply rule that the answer is yes, by assigning the necessary proximity relations to Logic. A more interesting question is whether there is some natural pre-existing notion of Broad Logic for which the answer is yes.

It is a standard result that every theory has an axiomatisation in which none of the axioms can be derived from the remaining axioms. Sadly there is no space here to provide a proof. Now let $\{B_i : i \in A\}$ be an independent axiomatisation of T, let ψ be an arbitrary theorem of T and let $\{B_i : i \in I\}$ be a maximal subset which does not prove ψ. (We are not making any assumptions about the size of A). Let us call these *maximal subsets.* For the moment we will duck the question of how many maximal subsets there are, since this depends in complicated but fundamentally uninteresting ways on the language in which T is expressed.

Now let I be such a maximal set, and consider the new axiomatisation of T devised as follows: let ψ be an axiom, B_i is an axiom if $i \in I$, and $\psi \to B_i$ is an axiom for $i \notin I$.

It is clear that this is an axiomatisation of T. We wish ψ to be an independent axiom in this presentation. Suppose ψ is *not* independent.

Then there is a finite J disjoint from I and a finite $I' \subset I$ so that

$$\{B_i : i \in I'\} \cup \{\psi \to B_j : j \in J\} \vdash \psi$$

which is to say

$$\{B_i : i \in I'\}, (\psi \to \bigwedge_{j \in J} B_j) \ \vdash \psi$$

$$\bigwedge_{i \in I'} B_i \ \vdash \ (\psi \to \bigwedge_{j \in J} B_j) \to \psi$$

But '$((\psi \to \bigwedge_{j \in J} B_j) \to \psi)) \to \psi$' is an instance of Peirce's Law so we can infer

$$\bigwedge_{i \in I'} B_i \ \vdash \psi$$

contradicting the assumption that $\{B_i : i \in I\}$ is a subset which does not prove ψ, so ψ is independent as desired. Accordingly we can form an axiomatisation of a new theory $T*$ by replacing ψ by $\neg\psi$ in this axiomatisation.

Now fix an independent axiomatisation of T, and for each I a maximal subset as above let T_I be the theory $T*$ with axiomatisation derived as above from the given axiomatisation of T. Evidently for $I \neq J$ we have $I \not\subseteq J \not\subseteq I$. We can also prove $T_I \not\subseteq T_J \not\subseteq T_I$ as follows.

Since $I \not\subseteq J$ there is some axiom B in I which is not in J. We shall show that $T_J \not\vdash B$. Suppose $T_J \vdash B$, then

$$\langle B_i : i \in J\rangle, \neg\psi, \langle \psi \to B_i : i \notin J\rangle \vdash B$$

Now $\neg\psi \to (\psi \to B_i)$ so we don't need $\langle \psi \to B_i : i \notin J\rangle$ anyway. This simplifies the last assertion to

$$\langle B_i : i \in J\rangle, \neg\psi \vdash B$$

Now J is *maximal* not proving ψ so $\langle B_i : i \in J\rangle \cup \{B\} \vdash \psi$. Therefore

$$\langle B_i : i \in J\rangle, \neg\psi \vdash \psi$$

whence

$$\langle B_i : i \in I\rangle \vdash \psi$$

contradicting assumption on J.

Therefore any two distinct maximal subsets of the axiomatisation will give rise to distinct theories.

There is of course no guarantee that T_I is a complete theory when I is a maximal set. If it isn't we will not even need the fact that $I \neq J \to T_J \neq T_J$ to show that there are many "closest" theories to T in which $\neg\psi$, for any incomplete theory (of this sort, at least) has many complete extensions. For theories with countably many constants there will be uncountably many such completions.

Model theory and logic have more to tell us than that there is no sensible logical notion of proximity. A *logical* notion pertaining to a structure is one preserved by all automorphisms of that structure. Thus a logical relation is one which—considered as a set of n-tuples—is fixed setwise by all automorphisms of the structure. This is of course the Erlanger programm view of Logic, and since I wrote the earlier versions of this draught I have learned of some interesting literature on the subject: Tarski [1986], Vann McGee [1996] and Keenan [2001]. Frobenius in 1897 gave the definition of a *characteristic* subgroup of a group G as one fixed setwise by all automorphisms of G. There is also a discussion in Rogers [1967] chapter 4.

What does this show? Well, only that there is no *logical* basis (given a world W) for designating any particular world in which $\neg\psi$ to be that world which is closest (for example) to W. We need something else. This something else is presumably a metric arising from a relative proximity relation, or perhaps merely a relative proximity relation itself. Each way of Gödel numbering the canonical language gives rise to a metric on the stone space: let the distance between two theories be the sum of all 2^{-n} where n is the Gödel number of a sentence in the symmetric difference. All these metrics give rise to the same

Stone topology, but the metrics themselves can be quite different and give rise to different truth-values for counterfactuals.

The point is the simple one that possible worlds do not come naturally equipped with the relations necessary to explain the phenomena for whose supposed explanation it was that possible worlds were invoked. We have to supply the metrics and proximity relations ourselves, since there is no *logical* reason for preferring one to the other. Batteries not included. This will not surprise thoughtful users of this machinery, but it should serve as a warning to others that possible world semantics is not a candidate for an *explanation* of (for example) counterfactuals, but is merely a procedural device for presenting the theoretical entities (the metrics) that are. We could equally well try providing a description of the problem in iambic pentameters. In short, it is question-begging. There is nothing wrong with this as long as we retain a sense of the difference between *expressing* something and *explaining* it. Over the years far too many people have been seduced by the beguiling imagery of possible world semantics into thinking that once a problem has been expressed in that formalism then progress has automatically been made. The virtue of this simple illustration is that it reminds us that this is not so.

§6. Coda. I acknowledged at the outset of this essay a responsibility on me as a naturalistic philosopher to provide an explanation of how the errors I allege are being made come to be made. If possible world semantics outside logic is such a huge mistake, why do so many clever people make it? There's no mystery about that. It was obvious from the outset that possible world semantics was going to be very useful in formal logic, and it didn't take long for philosophers with an interest in logic to recognise an affinity between its ideas and a strand of thought in philosophy going back to Leibniz and beyond and jump to the conclusion that it would be useful to them.[9] As well as looking like a useful technique, possible world semantics came with a lot of appealing imagery—worlds *seeing* one another, vicariously enjoying the wicked adventures of our counterparts in less humdrum worlds and so on; by reifying possibilities into worlds it collaborates insidiously with our natural inclination to realism, and finally in its allusions to formal logic it attracts not only those who value logic and have high expectations of it, but also those who—while having no exalted view of the rôle Logic can play in Philosophy—are happy to appear inclusive. Finally, it's fun without guilt: it allows us to combine the delight of reading Dunsany or Tolkien while pretending that we are doing Philosophy. An unbeatable combination!

Bibliography.

1990 Cresswell, M.J. Modality and Mellor's McTaggart. Studia Logica Vol 49, 1990, pp.163-170

[9] This is a very human thing to do: fault-tolerant pattern-matching has long been useful for spotting lions in undergrowth or edible fruit in thick canopy, but we pay for the speed by sacrificing reliability. Normally this is not a problem, as one does not lose one's life by jumpily mistaking something for a lion, and a hastily misidentified fruit can be spat out.

2001 Keenan, E. Logical Objects *in* Logic, Meaning and Computation, Essays in memory of Alonzo Church, Anderson and Zelëny, eds. Kluwer 2001. pp. 149-180.

1996 McGee, V. Logical Operations *Journal of Philosophical Logic* **25** (1996) pp 567-580.

1967 Rogers, H. Theory of recursive functions and effective computation McGraw-Hill

1986 Tarski, A. What are logical notions? History and Philosophy of Logic v 7 1986 143-154

Depatment of Pure Mathematics and Mathematical Statistics
Centre for Mathematical Sciences
Wilberforce Road
CB3 0WB
United Kingdom