Logique & Analyse **135-136** (1991), 271-282

IN DEFENCE OF THE BARCAN FORMULA

M J CRESSWELL

The principle of modal logic known as the Barcan Formula

BF $\forall xL\phi x \supset L\forall x\phi x$

has tended to have a rough passage in the literature of quantified modal logic. A typical negative reaction may be found in Garson 1984, p.257, and a preference for a semantics which fails to validate BF is expressed on p.127 of Fine 1978. In chapter 7 of Cresswell 1990 I gave a philosophical defence of the formula, but my present concern, although related, is more internal to quantified modal logic, and argues that it is systems *with* the Barcan Formula which should be taken as basic, with the systems without BF being taken as systems of restricted quantification.

The semantics for quantified modal logic is of course an extension of the familiar possible-worlds semantics for modal propositional logics. It is by no means a trivial question to examine what happens when various different propositional logics are extended to predicate logics. For instance a complete propositional logic may well lose this property when extended to a predicate logic. Thus, the propositional system S4.2 is characterized by models in which the accessibility relation is reflexive, transitive and satisfies the convergence condition, that if a world can see two worlds, those two can between them see another world. S4.2 + BF is not characterized by this or any other condition on an accessibility relation - though the extension of S4.2 which *lacks* BF *is* so characterized (Cresswell, forthcoming). Fascinating as all this is the primary interest of modal predicate logic is by and large independent of which propositional system is taken as basic.

In addition to a set of possible worlds modal predicate logic must assume a domain D of individuals. It is with respect to such a domain that the quantifiers are interpreted, and it is here that the Barcan Formula enters the picture. For modal systems with BF the case is simple. Where w is any world

[V∀] $\forall x\phi x$ is true in w iff ϕx is true in w whatever member of D is assigned to x.

(This rule normally requires a more complex formulation since, first, a value to a wff with a (free) variable x is relative to an assignment to the variables and, second, it must be so stated as to give a value to all wff of the form $\forall x\alpha$, however complex a wff α is. Such complexity is not required for my purposes.) I hope it is not difficult to see why [V∀] satisfies BF. For suppose the consequent $L\forall x\phi x$ is false in some world w. Then there is a world w' accessible from w such that $\forall x\phi x$ is false in w'. So there must be some u in D such that when x is assigned u, ϕx is false in w'. But then $L\phi x$ will be false in w when x is assigned u, and since $u \in$ D, $\forall xL\phi x$ will be false in w.

I've said that the rule is simple. And indeed its simplicity is one of the things I'm going to hold in its favour. To get a semantics for systems without the Barcan formula what happens is that instead of a single domain D there is for each world a domain D_w which is the things which exist in that world. The relations between the different D_w's are often supposed to be a metaphysical matter. At one extreme is David Lewis's view (beginning in Lewis 1968) that nothing exists in more than one world. At the other extreme would be the view that exactly the same things exist in all possible worlds. The quantifier is then supposed to range only over things which exist in the world of evaluation. More precisely

[V∀′] $\forall x\phi x$ is true in w iff ϕx is true in w whatever member of D_w is assigned to x.

To get the Barcan formula on this view one supposes that exactly the same things exist in all possible worlds, so that D_w in [V∀′] can be replaced by D, since D no longer depends on w, and we get back to [V∀]. That is the sense in which it is supposed that the validity of the Barcan Formula reflects the view that the same things exist in all possible worlds. Now this looks like the view that everything is a necessary existent, and that makes the Barcan Formula look not only like a special case of a more general semantics, but in fact like an implausible special case. One of the earliest philosophers to get worried about BF was Arthur Prior in *Time and Modality* (Prior 1957). Prior was concerned with a temporal interpretation of the necessity operator. He read $L\alpha$ as 'it is and always will be that α' and he read BF as saying that if everything will always be ϕ then always everything will be ϕ. And he thought this was false because even if everything *now existing* will always be ϕ it does not follow that always it will be that everything *then* existing is ϕ.

But you don't have to interpret BF that way. (See Cresswell 1990, p.96) You can interpret $\forall$ as ranging over all past, present or future individuals, and if every one of *them* will always be ϕ then it will always be that everything is ϕ. The point is simple. Even if each world w has its own domain D_w of the things which exist in w there is no reason why all these D_w's can't be collected into one single domain D. Perhaps the culprit here was isolated by Ruth Marcus herself in Marcus 1962. It is easier to see in the case of the existential quantifier $\exists$. For $\exists$ the rule is

[V$\exists$] $\exists x\phi x$ is true in w iff ϕx is true in w for some assignment to x of a member of D.

Marcus was concerned to protest against reading $\exists x\phi x$ as 'there *is* an x such that ϕx' on the ground that that implied existence. (That article has subsequently been considered one of the originating articles of the 'substitutional' interpretation of the quantifiers. That is the view that $\exists x\phi x$ is true if there is a name a such that ϕa is true. However as I read her Marcus's 1962 article is neutral on whether to read it that way or as in [V$\exists$], which is not a substitutional reading.) The problem with reading $\exists$ as 'there is a' is that, in a modal context it seems to claim actual existence and in a temporal context it seems to claim present existence. Prior at any rate so read it, and thus rejected the Barcan Formula.

But, as Marcus says, why must one read it that way? One could of course read it substitutionally as subsequent defences of BF by Marcus and others have done, but even if you read it referentially why must you interpret the quantifier only in the domain of things which exist in the world in question? Why not interpret it in the whole domain which is the union of the domains of each world? It might be thought there are metaphysical reasons against doing this. It might be thought that it is contradictory to say that there are things which don't exist. Arthur Prior certainly believed that only presently existing things existed and was perplexed about how you could speak of anything as coming into existence. For before it existed there was no 'it' to come into existence. I don't want this paper to be concerned with metaphysics, though I will mention one confusion which does I think cloud the issue here. If the causal theory of reference is correct, and if something in one world can never be part of a causal chain leading to an utterance in another world then no one can refer to any particular non-existing thing. But when we say 'there might have been more things than there are' we are not refering to any particular one of them, even though our statement cannot be true without quantifying over non-existent things.

Now those who eschew modality altogether may refuse to quantify over the non-existent, as also may those who use modal languages but do not take seriously a possible-worlds semantics. My concern is with authors who offer [V∀′] as a semantics for the universal quantifier in a modal language. For such a semantics already assumes a set of possible worlds and already assumes a domain of individuals for each one of them. Perhaps that is what drove Marcus to the substitutional interpretation of the quantifiers. Once you have chosen a semantics involving worlds and domains the metaphysical plunge is taken, and it is then a purely semantical matter whether you use [V∀] or [V∀′].

Could there be semantical reasons for [V∀′]? One might be the following. You might hold that the only reason for studying intensional languages *at all* is that the semantics of natural language is intensional. And you might also hold that the quantifiers of natural language only refer to actually existing things, and so should be modelled on [V∀′] rather than [V∀]. I don't in fact believe that natural language quantifiers are actualist and have argued so in chapter 7 of Cresswell 1990 and chapter 4 of Cresswell 1994. However I shall not pursue these claims here since I am at present concerned with reasons for BF internal to modal logic. And in the choice between various modal systems I am concerned to defend the view that taking systems with BF as basic gives a better account of systems without BF then does proceeding in the other direction.

If systems with BF are taken as basic, and [V∀] rather than [V∀′] is used then the quantifiers in systems without BF are best regarded as restricted quantifiers. Restricted to the domain of the world in question. But how can this restriction be expressed? Well, that will depend on whether the system has an existence predicate either primitive or defined. I shall begin by looking at systems which *do* have an existence predicate, and then proceed to those which do not. But before I can look at systems without BF I will have to look at the following question. In a wff like ϕx in which x is free what happens in a world w in which x is assigned as a value an individual which is not in D_w? One way of avoiding this question would be to prohibit such assignments altogether and I shall look later at what kind of semantics you get when you do. But for now I shall take it that such cases can arise. Another way, which actually turns out to be equivalent, is to say that ϕx lacks a truth value in such worlds. That was the way we proceeded in chapter 10 of Hughes and Cresswell 1968 (IML). A third way, the way chosen in Kripke 1963,p.85f, is to say that it is either true or false, just as when x is assigned something which *is* in D_w. Which of these values it has of course is up to the model, since the value of ϕ will deliver in each world the set of things which satisfy ϕ. (it might be tempting to require that if $\langle u,w\rangle \in V(\phi)$ then $u \in D_w$, but although this would make ϕx false for every atomic wff when x has a value not in D_w it would make every $\sim\phi x$ true for every such value, and if we don't want to say that ϕx is true when x doesn't exist, does it mean that we do want to say that $\sim\phi x$ is always true when x doesn't exist. As Kripke points out on p.86n it would require certain axioms to be stated for atomic wff which would not hold of all wff.)

Look first at languages which do have an existence predicate. I.e., assume then that there is a predicate E which has the very simple semantics:

[VE] Ex is true in w iff x is assigned a member of D_w.

It is now easy to define a quantifier satisfying [V$\forall'$] in terms of a quantifier satisfying [V$\forall$]. To avoid confusion I shall temporarily use $\forall$ for the 'possibilist' quantifier, that is the quantifier satisfying [V$\forall$] and ranging over the whole of D, and will follow Prior in using Łukasiewicz's universal quantifier symbol Π for the 'actualist' quantifier, that is the quantifier satisfying [V$\forall'$] and ranging over the domain of the world in question. It is trivial to note that $\Pi x\phi x$ can be expressed as

[Def Π] $\forall x(Ex \supset \phi x)$

This means that systems without BF but which have an existence predicate emerge as subsystems of systems with BF. One reason for treating Π as defined is that it does not obey all the laws of standard quantificational logic. One such law, which I'll state for $\forall$ is

[$\forall$1] $\forall x\phi x \supset \phi y$

This law is sometimes called universal instantiation and is a law of standard non-modal logic. Now although in modal logic the schematic version of this law

[∀1′] $\forall x\alpha \supset \alpha[y/x]$

can have modal instances, as when α is $L\phi x$, the simple [∀1] is not a modal wff at all, and it would therefore be reasonable to expect that its analogue

[Π1] $\Pi x\phi x \supset \phi y$

would also be a law. However it is easy to see that this is not so, for [Π1] becomes

[∀1″] $\forall x(Ex \supset \phi x) \supset \phi y$

and if we consider an interpretation in which every member of D_w satisfies ϕ, but y is assigned something which is not in D_w and does not satisfy ϕ then [∀1″] will be false. Of course [Π1] can be turned into a truth by replacing it with

[Π1E] $(\Pi x\phi x \wedge Ey) \supset \phi y$

or schematically

[Π1E′] $(\Pi x\alpha \wedge Ey) \supset \alpha[y/x]$

[Π1E] is one of the forms universal instantiation can have in what is called *free* logic (meaning, as I understand it, logic 'free' of existential assumptions) and free logic is often considered the appropriate way to deal with quantified modal logic (see Garson 1984.) However care is needed in taking free logic as our model. In a non-modal free logic it is tempting to think of the variable y, in those cases when [Π1] fails, as a 'non-denoting' term. This is because in non-modal logic we don't normally have a class of things which don't happen to exist but might have. But in a modal semantics as I have been presenting it so far there are no non-denoting terms. The y in [Π1] denotes all right. But the thing it denotes does not exist in the world in which the sentence is being evaluated.

If we are to make a choice between ∀ and Π then the natural one seems to me to take the quantifier ∀ as basic and treat Π as definable. The issue here is one of expressibility; for in a language with ∀ one can easily express Π as a quantifier restricted by E but not vice-versa. And indeed, a language with Π and E has no simple way to express $\forall x\phi x$ at all. (One might think that $\forall x\phi x$ could be expressed as $L\Pi x\phi x$ but that would be a mistake. $\forall x\phi x$ says that every possible object is ϕ *in this world*. $L\Pi x\phi x$ says that in every world the things that exist in that world are ϕ *in that world*.)

In logics with an existence predicate we may easily express Π in terms of $\forall$. I now want to consider what happens in a language with Π rather than $\forall$ as primitive, but no existence predicate. If we add identity to modal LPC the issue is changed since Ex can be defined as $\Sigma x(x = y)$ where Σ is the actualist existential quantifier defined as $\sim\Pi x\sim$. This assumes that identity is given its standard meaning that $x = y$ is true iff x and y are both assigned the same value. However, identity raises issues beyond the scope of this article and I will now proceed to consider what happens in a language with actualist quantifiers but without identity and without E. Such a language does not have E and [Π1] is not valid. What can we do? What Kripke did was to replace [Π1] by its universal closure.

[UΠ1] $\Pi y(\Pi x\phi x \supset \phi y)$

[UΠ1] is indeed valid, for the initial Π restricts the values of y to those in domain of the evaluation world. Kripke follows Quine 1940 in maintaining that the validity of wff containing free variables is really the validity of their universal closures and so replacing [Π1] by [UΠ1] is not in fact changing anything. The problem is that this procedure only changes nothing if the quantifiers range over the whole domain from which the values of the free variables are chosen, and that is just what we don't have here since in evaluating [Π1] in a world w the variable y may be assigned something not in D_w. In fact we don't need to go to modal logic to distinguish between [Π1] and [UΠ1]. For imagine a non-modal predicate logic in which there is a domain D and a subdomain $Q \subseteq D$. Assume that Π is interpreted so that

[VΠ] $\Pi x\phi x$ is true iff ϕx is true whenever a member of Q is assigned to x.

With this semantics [UΠ1] will be valid but [Π1] will not. Kripke's technique gives us the logic of restricted quantification. (For a discussion of issues connected with Kripke's axiomatization see Fine 1983.)

Looking at the non-modal case is helpful since it would I think be granted that the more natural way to express restricted quantification is with the use of a restricting predicate. In discussing languages with an existence predicate I pointed out that a wff using only Π and Σ quantifiers can be translated (with the aid of E) into a wff with the quantifiers $\forall$ and $\exists$. In such languages the notion of translation is quite a strong one since an interpretation for a language with E stipulates that $V(E)$ must satisfy [VE], which stipulates that Ex is true in w iff x is assigned a member of D_w. This will ensure that every interpretation will require that in any world a Π/Σ wff receives the same value, under any assignment to its variables, as its $\forall/\exists/E$ translation. Without E we have no way of forcing any particular predicate to match up with the system of domains in the way E does. However a slightly weaker result can be proved, and that is that there is a simple translation which preserves logical truth. Consider a wff α in the Π/Σ language. Choose some predicate ϕ which does not occur in α and let the translation function τ into the $\forall/\exists$ language be obtained as follows, for any subformulae β and γ of α:

If β is atomic then $\tau(\beta) = \beta$
$\tau(\sim\beta) = \sim\tau(\beta)$
$\tau(\beta \supset \gamma) = (\tau(\beta) \supset \tau(\gamma))$
$\tau(\Pi x\beta) = \forall x(\phi x \supset \tau(\beta))$

If α is not logically valid, then take the model in which it fails and define a model for the new language just like it except that $\langle u,w\rangle \in V(\phi)$ iff $u \in D_w$. This model will falsify $\tau(\alpha)$ so that it will not be valid either. And if $\tau(\alpha)$ is not valid define a model for the original language in which $u \in D_w$ iff $\langle u,w\rangle \in V(\phi)$ in the model which falsifies $\tau(\alpha)$. $\forall x\phi x$ has no comparable translation in terms of Π or Σ.

Kripke's approach consists in weakening principles of standard non-modal logic, specifically in replacing [Π1] by [UΠ1]. It might therefore be profitable to consider whether there is any way of saving modal systems without BF which do not sacrifice principles of standard non-modal logic. One suggestion might be to modify the definition of validity. For restricted quantification in a non-modal language, one can simply define validity by saying that α is valid in an interpretation iff it is true in that interpretation for all assignments to its variables which give values in the restricted domain (the domain called Q in [VΠ]). In a non-modal language this means that members of D not in Q play no role at all in the determination of validity, and we get just the same result as by taking Q to be the whole domain. As might be expected, in the modal case things are a bit more complicated. For suppose we say simply that α is valid in an interpretation iff α is true in that interpretation in every world w when its variables are assigned members of D_w. This certainly validates [Π1] but consider its necessitated version

[LΠ1] $L(\Pi x\phi x \supset \phi y)$

Suppose that y is assigned a member u of D_w and suppose that wRw'. Suppose that everything in $D_{w'}$ satisfies ϕ. So $\Pi x\phi x$ is true in w'. But suppose $u \notin D_{w'}$ and suppose that u does not satisfy ϕ in w'. Then ϕy will be false in w' thus making $\Pi x\phi x \supset \phi y$ false at w' and so making [LΠ1] false at w. So perhaps we should insist that the variables of α should be assigned values from the domains of the evaluation world and all the worlds in the *posterity* of w, where this is defined to be the smallest set POS_w such that

(i) $w \in POS_w$
(ii) If $w' \in POS_w$ and $w'Rw''$ then $w'' \in POS_w$.

But that will not do either since, although such a semantics validates

[L2Π1] $L\Pi x\phi x \supset L\phi y$

(because if y's value is in the domain of every world accessible from w and if $\Pi x\phi x$ is true in every such world then $L\phi y$ must be also) is valid its universalized version

[UL2Π1] $\Pi y(L\Pi x\phi x \supset L\phi y)$

is not valid, since despite [L2Π1]'s validity it can be false when y is assigned something in D_w which is not in some accessible w' and which does not satisfy ϕ in w' even though everything in $D_{w'}$ is ϕ, and that is enough to shew the invalidity of [UL2Π1]. Now, one can save the situation by changing the evaluation rule for the quantifier so that it reads

[VΠ′] $\Pi x\alpha$ is true at a world w iff α is true at w for all assignments to x which assign it some u which is in the domain of every world in the posterity of w.

The reason this works is the following. Let us define D_w^* to be the set such that $u \in D_w{}^*$ iff u is in the domain of every world in the posterity of w. Then [VII′] is just the usual rule except that $D_w{}^*$ is used as the domain instead of D_w. So although it may *look* as if D_w is the domain of things existing in w the real domain is $D_w{}^*$, and when we look at $D_w{}^*$ we notice an interesting fact. Suppose wRw'. Then if u is in the domain of every world in the posterity of w it is certainly in the domain of every world in the posterity of w'. In other words if wRw' then $D_w{}^* \subseteq D_{w'}{}^*$. And that means that the * domains satisfy what in IML p.171 is called the *inclusion requirement*. The inclusion requirement says that when you move from one world to another which is accessible from it you are allowed to add things which don't already exist, but you are not allowed to drop anything which does already exist. It is known that imposing the inclusion requirement gives the semantics for the systems you get when you simply combine the axioms and rules of standard propositional modal logic with those of standard non-modal predicate logic. The inclusion requirement also makes it clear why any interpretation which has a symmetrical accessibility relation will interpret the quantifiers in the same domain for each world. And indeed BF is a theorem of the predicate logics standardly based on extensions of the Brouwerian system, B. (B is the extension of T defined by the axiom schema

B $\sim\alpha \supset L{\sim}L\alpha$.)

Although this result may please those who want to provide a semantics for standardly axiomatized systems of modal predicate logic the semantics it provides can hardly be considered philosophically plausible. For while it allows you to say that there might have been more things than there are it does not allow you to say that there might have been fewer. And that seems quite the wrong direction, for one can surely point to something and say '*that* might not have existed.' And finally it prohibits you from using a logic like S5 for necessity on pain of re-introducing the Barcan Formula.

The semantics I have been discussing have all assumed that every wff has a truth value at every world even if its variables are assigned values outside the domain of that world. In chapter 10 of IML we presented a semantics in which a wff was said to be *undefined* in a world w if its variables were given values outside D_w. I want to make a few brief remarks to the effect that doing this doesn't give you any new results, and doesn't add anything to the issue of whether or not to accept the Barcan Formula. The first thing to notice is that the semantics offered on p.171f of IML assumes the inclusion requirement. With the inclusion requirement the rules for V presented on p.172 entail that for any given world w the truth of a wff α at w whose free variables are all assigned members of D_w never depends on undefined subformulae, and so would remain the same in a model which does not permit undefined wff, so long as the new model agrees with the original model on all defined atomic wff.

When the inclusion requirement is dropped it is unclear just what to do in models which allow undefined wff. For consider a wff like $\Pi x L\phi x$. If rule 6 on p.172 of IML is used this will be true in w iff $L\phi x$ is *true* in w (not just never false) for all values for x taken from D_w. Now consider a value u for x where $u \in D_w$, but where wRw' and $u \notin D_{w'}$. What are we to say here? From rules 3 and 7 on p.172 of IML we say that since ϕx is undefined in w' (by rule 3) then $L\phi x$ is undefined in w. But then $L\phi x$ will not be true for every $u \in D_w$, since it will be undefined for some. If the model does not satisfy the inclusion requirement for some wRw' then there will have to be some $u \in D_w$ not in $D_{w'}$, and so *all* wff of the form $\forall x\alpha$ where x is free in α in the scope of a modal operator will be undefined at w irrespective of the meanings of the predicates in α. So $\Pi x L\phi x$ will be undefined at w. And notice that it will be undefined so long as there is some value u for x which makes $L\phi x$ undefined, i.e. so long as there is some $u \notin D_{w'}$, which is precisely what a failure of the inclusion requirement amounts to. In short the effect will be that the only defined cases of quantification into modal contexts will arise in worlds which respect the inclusion requirement in the sense that everything in their domain is also in the domains of all the worlds they can see.

One might be tempted to propose an alternative evaluation rule for $\forall$ to cover the undefined cases. One might say that $\Pi x L\phi x$ is true provided $L\phi x$ is *true where defined* for every value u of x taken from D_w. However that would have the effect of invalidating

[$\wedge\Pi$] $\Pi x(\phi x \wedge L\phi x) \supset \phi x$

This will be false in a model with wRw' and x assigned some $u \in D_w$ but $u \notin D_{w'}$. For suppose u does not satisfy ϕ in w, but every other individual satisfies ϕ both in w and in every other world. This will mean that ϕx will be false at w. But $\phi x \wedge L\phi x$ will be undefined at w when x is assigned u, since ϕx is undefined at w' so by rule 7, $L\phi x$ is undefined at w, so by rules 4 and 5, $\phi x \wedge L\phi x$ is also undefined at w. But otherwise it is true, so it is true where defined for every member of D_w. So $\Pi x(\phi x \wedge L\phi x)$ is true at

w, and so [$\wedge\Pi$] is false at w and so not valid. Further (for what it is worth) $\Pi x(\phi x \wedge L\phi x) \supset \Pi x\phi x$ also fails, so the problem cannot be blamed on the presence of free variables. Admitting truth value gaps is not a way of avoiding the problems which arise when you drop the Barcan Formula.

The upshot of this discussion is to suggest that since the possible worlds semantics for quantified modal logic already assumes that we have a domain of possible objects, not all of which need actually exist, there is no metaphysical bar to quantifying over them, and that considerations of naturalness, expressibility, and axiomatizability as extensions of standard axiomatic predicate logic, all strongly indicate that we should. In consequence it is systems possessing the Barcan Formula which should be taken as basic in modal predicate logic.

Victoria University of Wellington

REFERENCES

Cresswell, M.J., 1990, *Entities and Indices*, Dordrecht, Kluwer

- 1994, *Language in the World*, Cambridge, Cambridge University Press
- (forthcoming) Incompleteness and the Barcan Formula. *Journal of Philosophical Logic*

Fine, K, 1978, Model theory for modal logic. *Journal of Philosophical Logic*, Vol 7, pp.125-156

- 1983, The permutation principle in quantificational logic. *Journal of Philosophical Logic*, Vol 12, pp.33-37

Garson, J.W, 1984, Quantification in modal logic. *Handbook of Philosophical Logic, Vol II*, Dordrecht, Reidel, pp.249-307

Hughes, G.E., and M.J.Cresswell, 1968, *An Introduction to Modal Logic*, London, Methuen

Kripke, S.A., 1963, Semantical considerations on modal logics. *Acta Philosophica Fennica (1963) Modal and Many-valued Logics*, pp.83-94

Lewis, D.K., 1968, Counterpart theory and quantified modal logic. *The Journal of Philosophy*, Vol 65, pp.113-126

Marcus, R.B., 1962, Interpreting quantification, *Inquiry*, Vol 5, pp.252-259

Prior, A.N., 1957, *Time and Modality*, Oxford, Oxford University Press

Quine, W.V., 1940, *Mathematical Logic*,Cambridge Mass., Harvard University Press (2nd edition, revised, 1951.)