5 Irving Terrace
Cambridge
Mass.
02138
6 March 1966

Professor Dagfinn Føllesdal
Department of Philosophy
Stanford University
Stanford
California

Dear Professor Føllesdal,

I have been spending some of my time recently working on a way of treating the modal operators, and I would be very interested and grateful to hear any reactions you might have. The idea was suggested to me by some informal discussion in a paper Saul Kripke wrote for a seminar his senior year; but it turns out that I took a somewhat different turn from Kripke (at least from his "Semantical Considerations on Modal Logic" in _Acta Philosophica Fennica_ fasc. 16 (1963), which is the only one of his papers I've read carefully). My treatment is of the same family as yours, Kripke's, Hintikka's, and Kanger's, but I hope it may differ from them somewhat.

Suppose we have a formalized language Lc, based on standard quantification theory with identity and without ineliminable singular terms, in which the variables are taken to range over a universe including many possible worlds and the things in them. We need not say that everything is in some world--perhaps God, numbers, universals, sets with elements in several worlds, wholes with parts in several worlds, or worlds themselves might best be taken as not in any world. However--_contra_ Kripke--things are in at most one world. And there is a world which contains just those things which are in the actual world. So Lc contains these three predicates--

Wx........x is a possible world
x in y....x is something in world y; world y contains x
Ax........x is something in the actual world

governed by at least these postulates--

$x \text{ in } y \supset Wy$
$x \text{ in } y \,\&\, x \text{ in } z . \supset y = z$
$(\exists x)[Wx \,\&\, (\forall y)(y \text{ in } x \equiv Ay)]$.

Nothing is in more than one world. But in place of saying that something in one world is strictly identical to something in another world, we may say that one is a _counterpart_ of the other. This is to say, roughly, that the two are similar in content and context to a high degree in important respects. So Lc contains another predicate--

x C y.....x is a counterpart in x's world of y in y's world

governed by at least these postulates--

$x \,C\, y \supset . (\exists z)(x \text{ in } z) \,\&\, (\exists w)(y \text{ in } w)$
$x \text{ in } z \,\&\, y \text{ in } z . \supset . x \,C\, y \equiv x = y$

by which only things in worlds are or have counterparts and anything in a world is its own unique counterpart in that world.

For the sake of generality, let us not postulate that the counterpart relation is symmetric or transitive; nor that everything in a world has a unique counterpart in any other world; nor that everything in a world is the unique counterpart in that world of something in any other world. We may give the name <u>counterpart theory</u> to the apparatus just described; Lc is thus an arbitrary formalized language containing counterpart theory.

Before going on, let us define some abbreviations--

$\not{c}xy$.......$(\iota z)(z \text{ in } x \ \& \ z \ C \ y)$; the unique counterpart of y in world x

$E\not{c}xy$......$(\exists w)(w = \not{c}xy)$; y has a unique counterpart in world x

@.........$(\iota x)[Wx \ \& \ (\forall y)(y \text{ in } x \equiv Ay)]$; the actual world (existence and uniqueness of @ are guaranteed)

$(\forall x{:}\phi x)$...$(\forall x)(\phi x \supset \_\_)$; for every x such that ϕx, $\_\_$

$(\exists x{:}\phi x)$...$(\exists x)(\phi x \ \& \_\_)$; for some x such that ϕx, $\_\_$.

Now suppose we have a second formalized language Lm which contains the modal operators $\Box$ and $\Diamond$. Let variables in Lm range only over things in the actual world; and let Lm not contain the four predicates of counterpart theory. Otherwise, let Lm be like Lc. I claim that the translation scheme from Lm to Lc which I am about to describe preserves our intended interpretation of the modal operators in Lm.

Given a modal-free sentence ϕ (closed) or $\phi y_1 \ldots y_n$ (open), the sentence obtained from it by replacing all unrestricted quantifiers by corresponding restricted quantifiers of the form $(\forall x{:}x \text{ in } w)$ or $(\exists x{:}x \text{ in } w)$--its <u>restriction to world w</u>--may be written as ϕ^w or $\phi^w y_1 \ldots y_n$.

To translate from Lm to Lc, start by repeatedly applying the following substitutions, which remove innermost modal operators--i. e. modal operators whose scopes do not contain other modal operators. At every step, let w be a variable which does not yet occur in the sentence under translation.

If ϕ is a modal-free closed sentence--

$\_\_\Box\phi\_\_ \rightarrow \_\_(\forall w{:}Ww)(\phi^w)\_\_$

$\_\_\Diamond\phi\_\_ \rightarrow \_\_(\exists w{:}Ww)(\phi^w)\_\_$.

And if $\phi y_1 \ldots y_n$ is a modal-free open sentence with free $y_1 \ldots y_n$--

$\_\_\Box\phi y_1 \ldots y_n\_\_ \rightarrow \_\_(\forall w{:}E\not{c}wy_1 \ \& \ldots \& \ E\not{c}wy_n)(\phi^w \not{c}wy_1 \ldots \not{c}wy_n)\_\_$

$\_\_\Diamond\phi y_1 \ldots y_n\_\_ \rightarrow \_\_(\exists w{:}E\not{c}wy_1 \ \& \ldots \& \ E\not{c}wy_n)(\phi^w \not{c}wy_1 \ldots \not{c}wy_n)\_\_$.

(Since $E\not{c}wy \supset Ww$, it is unnecessary to include Ww explicitly in the restriction on the quantifier.) After all modal operators have been removed in this way, restrict any remaining quantifiers to things in the actual world by the substitution--

$\phi \rightarrow \phi^{@}$.

(Notice that restricted quantifiers may not be expanded by the definition above until translation of a sentence is complete; otherwise they will be restricted over again.)

For instance, take the sentence "There must be somebody who might have been mayor but is not." Writing "P" for "is a person" and "M" for "is mayor" we would presumably render this sentence in Lm as--

$$\Box(\exists x)(Px \;\&\; \Diamond Mx \;\&\; -Mx).$$

That would be translated into Lc as--

$$(\forall u{:}Wu)(\exists x{:}x \text{ in } u)(Px \;\&\; (\exists v{:}E{\not c}vx)(M{\not c}vx) \;\&\; -Mx)$$

which we can read as "In every possible world, there is somebody whose unique counterpart in some possible world is mayor, but who is not himself mayor." I claim that this sentence in Lc is a correct analysis of the original English sentence, and <u>a fortiori</u> of the corresponding sentence in Lm.

It is nothing new, of course, to think of the modal operators as quantifiers over possible worlds. Kripke also suggests that they are restricted quantifiers, but for him they are restricted in a different way: to worlds "possible relative to" a world under discussion in a context immediately surrounding the modal operator, rather than to worlds containing unique counterparts of the things denoted by variables free within the scope of the modal operator. (We might combine the two approaches by taking a modal operator as a quantifier over possible worlds which is restricted in both ways.) The difference is shown, for instance, in the fact that Kripke gets a counterexample to the converse Barcan formula--

$$\Box(\forall x)(\varnothing x) \supset (\forall x)\Box(\varnothing x)$$

and, as will be seen, I do not.

Another difference between my treatment and Kripke's is that Kripke is providing a truth-definition for known modal logics; I am providing an alternative apparatus which, I claim, can take the place of modal logic. Therefore it is not essential for me that the laws for modal operators which are available as translations of theorems of counterpart theory should be those of a known modal logic--though it is, of course, an interesting question whether they are. If they are not, so much the worse for that system of modal logic; not, so much the worse for my translation-scheme. Unfortunately, I do not have enough familiarity with systems of modal logic to make the comparisons.

If we translate the controversial sentence of Lm--

$$(\forall x)(\forall y)(x = y \supset \Box x = y)$$

we get this sentence of Lc--

$$(\forall x{:}x \text{ in } @)(\forall y{:}y \text{ in } @)(x = y \supset (\forall w{:}E{\not c}wx \;\&\; E{\not c}wy)({\not c}wx = {\not c}wy))$$

which can be shown to be true. Hence the sentence of Lm ought to be an axiom or theorem of modal logic.

If we translate these four sentences of Lm--

$\Box(\forall x)(\varnothing x)$......Necessarily everything is $\varnothing$*

$(\forall x)\Box(\varnothing x)$......Everything is necessarily $\varnothing$

$\Box(\exists x)(\varnothing x)$......Necessarily something is $\varnothing$

$(\exists x)\Box(\varnothing x)$......Something is necessarily $\varnothing$

we get these four sentences of Lc--

---

* In this paragraph and the next, $\varnothing x$ is to contain no modal operators, no quantifiers, and no free variables except x; however it may have truth-functional structure.

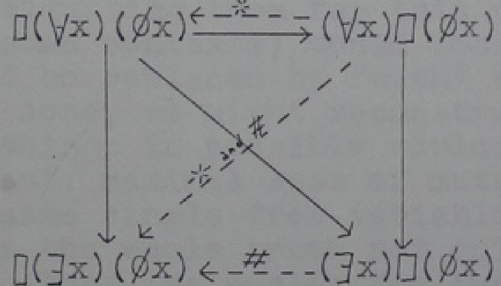
(∀w:Ww)(∀x: x in w)(∅x).......Everything, in every possible world, is ∅

(∀x:x in @)(∀w:Eøwx)(∅øwx)....Every unique counterpart of any actual thing is ∅

(∀w:Ww)(∃x:x in w)(∅x)........Every possible world contains something which is ∅

(∃x:x in @)(∀w:Eøwx)(∅øwx)....There is some actual thing, every unique counterpart of which is ∅.

By considering the implications between these translations, we get the following table of implications which ought to hold in modal logic between the original four sentences in Lm--



*...holds if everything, in any world, is the unique counterpart in its world of some actual thing.
#...holds if every actual thing has a unique counterpart in every world.

If we translate this sentence of Lm--

◇□□(∀x)(∅x)

we get this sentence of Lc--

(∃u:Wu)(∀v:Wv)(∀w:Ww)(∀x:x in w)(∅x)

in which all the modal operators except the rightmost have been translated as vacuous quantifiers. If, on the other hand, we translate this sentence of Lm--

(∀x)◇□□(∅x)

we get this sentence of Lc--

(∀x:x in @)(∃u:Eøux)(∀v:Eøvøux)(∀w:Eøwøvøux)(∅øwøvøux)

in which the quantifiers are not vacuous. (A free variable which occurs in a sentence as an argument of a ø-term or within the restriction-clause of a restricted quantifier is to be treated just like any other free variable in that sentence. A ø-term may be the second argument of another ø-term.) So a string of iterated modal operators should not collapse in general, but should collapse when it governs a closed sentence.

It is clear that this treatment of the modal operators involves essentialism. The essential properties of something are just those of its properties which it shares with all of its unique counterparts, in all the possible worlds in which it has a unique counterpart. I do not claim that essentialism is entirely clear; however, on my treatment the unclarity of essentialism is resolved into two distinct components. There is unclarity over the question of what worlds are possible; this is the question of what purported descriptions succeed in consistently describing worlds, and thus it runs into the usual difficulty about analyticity etc. (Plus difficulty about the distinction between empirical laws and other truths, if by possible worlds we mean physically possible worlds.) There is also an independent unclarity in our notion of the counterpart relation. We certainly have some such notion, but it is very vague. We should expect that if it can be made precise at all, it can be made precise in several incompatible and equally artificial ways (perhaps leading to different additional postulates for counterpart theory).

We might tamper with counterpart theory in order to cut down the basis of primitive predicates and the ontology. If we had another predicate--

x with y..x and y are in the same possible world

we could use it to eliminate "in" and "W". If Lc contains set theory (or the calculus of individuals) we can say that a world is a maximal set (or whole) whose elements (or parts) are related pairwise by "with"; if so, "in" would be replaced by "ε" (or "<"). Or we could get the effect of quantifying over worlds by quantifying over things in worlds (i. e. things with themselves), and individuating them for this purpose by "with". Thus "in" and "=" (in some contexts) would both be replaced by "with", and "W" would be replaced by "with" with the same argument in both places. That done, we might reconstrue the theory as dealing not with the things in possible worlds, but with complete descriptions thereof: maximal sets of mutually consistent open sentences with the same single free variable. Observe that such a description gives the whole truth not only about something in a world but also about its world; for instance, because it includes a maximal consistent set of open sentences of the form x = x & Ø, Ø being a closed sentence. If counterpart theory thus reconstrued were embedded in an adequate semantical metalanguage, we would presumably be able to define at least "with" and "A", if not "C".

* * *

Enough. A little news, to maintain a pretense that this is a letter--I am married, since six months ago; my wife Steffi is a senior in philosophy at Radcliffe (perhaps you remember her; she was Stephanie Robinson). I hope to finish my thesis ~~over the summer~~ (on conventions of language, with Quine as adviser) this summer. Next fall we will be at UCLA; Steffi will be a graduate student in philosophy, and I will be an assistant professor.

I hope we will have a chance to see you while you are in the country. We will be here until late summer--we don't know just when, it depends how my thesis goes--and will arrive at Los Angeles by at least the end of September, and perhaps much sooner. Or if we miss you this trip, we hope you will be back soon; or we may get a chance to spend a summer in Europe.

Best wishes,

David Lewis