Model Theory and Contingent Existence

Boris Kment

Timothy Williamson's predictably impressive book *Modal Logic as Metaphysics* (*MLM*) contains an extended defense of a thesis he calls "necessitism": it is necessary that everything exists necessarily. I will focus on just one of his many interesting and thought-provoking arguments: his attempt to show that necessitists are in a better position than their "contingentist" opponents to explain the evident utility and success of model-theoretic semantics for modal first-order logic (MFOL). After a brief sketch of the necessitism dispute (section 1) and of Williamson's arguments (section 2), I will offer a contingentist interpretation of the model-theoretic framework (section 3) that I take to provide effective replies to Williamson's reasoning (section 4).

1. Necessitism and Contingentism

Modal logic in the sense that interests Williamson is a branch of modal metaphysics that studies the "sufficiently general truths" about metaphysical modality (92). To characterize the relevant standard of generality, Williamson introduces the notion of metaphysical universality (92–93). To understand this concept, note that the non-logical constants of the language of MFOL (with identity) have no fixed meanings—it hasn't been settled which properties are expressed by the predicates other than =, or which individuals are picked out by the individual constants. A formula A in the language of MFOL is *metaphysically universal* iff A's universal closure (the sentence obtained from A by prefixing universal quantifiers to bind all free variables) is true on every way of settling the meanings to the

non-logical constants.[1] For example, $\ulcorner \Diamond Fx \vee \Diamond{\sim}Fx \urcorner$ is metaphysically universal, since $\ulcorner \forall x\,(\Diamond Fx \vee \Diamond{\sim}Fx) \urcorner$ is true no matter which property we take F to express. A system of MFOL is correct iff its theorems are just the metaphysically universal formulas of the language of MFOL (95). Necessitism can be stated in the language of MFOL by the formula $\Box\,\forall x\,\Box\exists y\;x{=}y$. Since this is a closed sentence containing no non-logical constants, it is metaphysically universal (and consequently a theorem of the correct MFOL) iff it is true. Necessitism can therefore be viewed as a thesis about which MFOL is correct.

One point of contention between necessitists and contingentists is the validity of the schemata known as the "Barcan Formula" and "Converse Barcan Formula" respectively. Their unnecessitated variants run as follows:

BF $\quad \Diamond \exists x\,A \rightarrow \exists x\,\Diamond A$

CBF $\quad \exists x\,\Diamond A \rightarrow \Diamond \exists x\,A$

Contingentists will almost invariably reject both schemata. Contrary to BF, they might hold that there could have been something that was Wittgenstein's child even though there is (actually) nothing that could have been Wittgenstein's child. And contrary to CBF, they will likely say that something could have failed to exist but that there couldn't have been something that fails to exist. Necessitists will accept both schemata.

2. Model Theory

The vocabulary of the language of MFOL consists of a countable infinity of i-place predicates $F_1^i, F_2^i, \ldots$ for every natural number i, a countable infinity of individual constants $c_1, c_2, \ldots$, a countable infinity of individual variables $x, y, \ldots$, quantifiers, sentence connectives, modal operators, the identity predicate, and parentheses. An *(inhabited) model structure M* for MFOL is a mathematical structure $<W^M, R^M, w_0^M, D^M>$. W^M is informally understood as a set of "worlds" and R^M as a dyadic relation of "relative possibility" or "accessibility" between members of W^M. w_0^M is a member of W^M understood as "the actual world." D^M is the "domain function" that maps each world w to a set (understood as

---

[1] This isn't quite Williamson's official definition of metaphysical universality (92–93) but it's equivalent to it.

containing the individuals that exist at w). We obtain a *(pointed) model* on a model structure by adding an "interpretation function" V^M. V^M assigns to every individual constant an object in $\bigcup_{u \in W^M} D^M(u)$, and to every non-logical i-place predicate F_j^i a function mapping each world w to a subset of $(D^M(w))^i$ (understood as the extension of F_j^i at w). (I will simplify by dropping the superscript "M" from now on, except when it is needed to avoid confusion.) Formulas are evaluated as true or false in a model M, at a world w, on an assignment a (of objects in $\bigcup_{u \in W} D(u)$ to variables). It's customary to write "$M, w, a \vDash A$" for "A is true in M at w on a." A formula A is *true in a model M* simpliciter ("$M \vDash A$," for short) iff, for every assignment a: $M, w_0, a \vDash A$. (If A is true in model M, I will say that M "validates" A.) A is *valid on a model structure* iff A is true in every model on that model structure.

A model is an abstract mathematical structure that needs to have no special connection to the modal realm—the worlds of a model and the individuals in their domains can be absolutely anything (apples and oranges, for example). To explain how model-theoretical reasoning can nevertheless establish metaphysical conclusions about modality, we need to give an account of the connection between model theory and the subject matter of modal discourse. On Williamson's approach, model theory connects object-language formulas to a range of structures (the model structures) that include the part of reality (the "intended model structure") that is the subject matter of modal discourse. (As Williamson notes (145, 236–40), this idea needs some qualification to accommodate the cardinality constraints arising from the set-theoretic nature of models.) In the intended model structure $<W, R, w_0, D>$, W is "the set of genuine worlds," R the relation of "genuine relative possibility," and w_0 "the genuine actual world" (95). The intended model structure validates all and only the metaphysically universal formulas. *MLM* gives an account of which model structure is intended (139–47).

Williamson argues that contingentists cannot connect model theory to modal metaphysics by the same strategy, since on the definition of validity outlined above,

(1)   Any model structure (and any model) that validates all metaphysically universal formulas also validates BF. (134–5)

Since contingentists reject BF, they are forced to deny that any model structure validates just the metaphysically universal formulas, and consequently that there is an intended model structure.

Although contingentists cannot describe a specific a model structure as the subject matter of modal discourse, they can try to single out one or more model structures or models as particularly faithful *representations* of that subject matter. They can distinguish the components of a model structure that are called "worlds" (the members of W) from the genuine possible worlds that modal discourse is about—the "Possible Worlds" (capitalized), as I will call them. Possible Worlds are complex states of affairs or classes of propositions or something like that. By contrast, the worlds of a model structure are merely part of the apparatus used to represent the realm of Worlds. Williamson claims that such a purely representational connection between the models and modal reality would be "too indirect to make the model theory much more than a complicated digression." (139) Moreover, he argues that it's unclear what an adequate representational account would look like (137–8). By contingentist lights, a model structure faithfully represents the modal realm only if it invalidates BF, which requires that the domain of some world accessible from w_0 contain alien entities (i.e., entities that are not in $D(w_0)$). But what are these entities supposed to represent? They must likely be taken to stand for individuals distinct from those represented by the members of $D(w_0)$. But that seems impossible, since the entities in $D(w_0)$ presumably represent *all* (actual) individuals. Williamson concludes that contingentists have to take a more "instrumentalist" attitude towards model theory than necessitists. (195) While they "can use possible worlds model theory as a technical device for various purposes, they cannot use it to give the intended truth conditions of quantified modal sentences …." (359)

3. A Contingentist Interpretation of Model Theory

My goal is to present a contingentist construal of model theory for MFOL that circumvents these difficulties. However, for expository reasons it will be best to start with the simpler case of model theory for non-modal first-order logic (FOL). For simplicity, I will drop quotation marks and corner quotes when referring to object-language expressions, except when they're needed to avoid confusion.

4

We can think of an assignment of meanings to the non-logical constants of the language of FOL as a function that assigns a property to each predicate and an individual to each individual constant.[2] A sufficiently powerful metalanguage would allow us to formulate, for any given meaning assignment, a sentence that completely describes this assignment and that also provides information about all non-semantic matters that are relevant (given the meanings assigned) to the truth-values of object-language sentences. I'll call such a sentence a "specification." With the help of a suitable semantic background theory we could, for every specification S and (closed) object-language sentence A, derive from S either that A is true or that A is not true. If these inferences could be formalized, then that would give us a precise criterion for metaphysical universality roughly along the following lines:

(2)   An object-language formula A is metaphysically universal iff the ascription of truth to A's universal closure can be derived from every specification that provides correct information about all relevant non-semantic facts.

We would be able to use this criterion to establish results about metaphysical universality. I will argue that model theory is simply a developed and revised version of this approach.

To develop a suitable criterion of metaphysical universality, it will be best to use a metalanguage whose vocabulary includes all the logical constants (sentence connectives, quantifiers, identity predicate), individual variables, and parentheses of the object language. In addition, the metalanguage has infinitely many individual variables that aren't part of the object language and, for every natural number i, a countable infinity of i-place predicate variables X_1^i, X_2^i, …. It also needs to include the infinitary conjunction and disjunction operators, $\bigwedge$ and $\bigvee$, that can (respectively) conjoin and disjoin up to κ different formulas, for a suitable infinite cardinal number κ. (I will write $\bigwedge_C \varphi$, or $\bigwedge_C \varphi$, to abbreviate

---

[2] Some philosophers identify properties with their Intensions (where the Intension of a property is the function that maps each Possible World to the set of individuals that instantiate the property at that World). Contingentists don't have this option. By their lights, many properties don't have Intensions. Consider redness. At some Possible Worlds this property is instantiated by objects that don't actually exist. Since no function can have sets containing non-existing objects as values, there is no function that assigns to each Possible World the set of individuals that are red at that World. Fortunately, there are various alternative accounts of properties, and contingentists can adopt one of these alternatives or take the notion of a property as primitive.

the conjunction of all formulas of the form displayed by φ for which condition C is satisfied. For example, where V is a set of variables, $\bigwedge_{\substack{u,v \in V \\ u \neq v}} \sim u = v$ abbreviates the conjunction of all formulas $\ulcorner \sim u = v \urcorner$ such that u and v are distinct variables in V. If there are no formulas of the form displayed by φ for which condition C is satisfied, then $\bigwedge_C \varphi$ will be taken to express a tautology. (A conjunction is true iff all of its conjuncts are. Hence, the conjunction of the members of the empty set of claims is automatically true.) Moreover, I will use expressions of the form $\bigwedge_i \varphi_i$ to abbreviate the conjunction of all formulas φ_i where i is a natural number. $\bigvee$ will be used in the analogous ways, except that when there are no formulas of the form displayed by φ for which condition C is satisfied, then $\bigvee_C \varphi$ will be taken to express a contradiction. (A disjunction is true iff some of its disjuncts are. Hence, the disjunction of the members of the empty set of claims is false.) I'll adopt the convention that formulas containing $\bigwedge$ and $\bigvee$ are to be read so as to give these operators the smallest possible scope. For example, $\bigwedge_C \varphi \rightarrow P$ is to be read as $(\bigwedge_C \varphi) \rightarrow P$, not as $\bigwedge_C (\varphi \rightarrow P)$.) The metalanguage also needs to allow for strings of up to λ quantifiers binding different variables (where λ is a suitable infinite cardinal number). Where S is a set of no more than λ variables, $\exists S$ ($\forall S$) will abbreviate a string of existential (universal) quantifiers, one for each variable in S. (If S is the empty set, then a formula A that contains $\exists S \varphi$ ($\forall S \varphi$) is logically equivalent to the formula obtainable from A by uniformly replacing $\exists S \varphi$ ($\forall S \varphi$) with φ.) Furthermore, the metalanguage needs to have quote names for all non-logical constants and closed sentences of the object language, a truth predicate T, a dyadic first-order predicate R interpreted to mean "refers to," and a dyadic predicate E applicable to one individual and one predicate symbol and interpreted to mean "expresses." (For any object-language individual constant c_k and metalanguage individual variable x_k, R"c_k"x_k is to be read as: "c_k" refers to x_k. And for any object-language predicate F_j^i and any metalanguage predicate variable X_j^i of the same adicity, E"F_j^i"X_j^i is to be read as: "F_j^i" expresses X_j^i.) Finally, I will assume that the metalanguage has a monadic first-order

6

existence predicate $E!$ (definable in terms of $\exists$ and $=$ and therefore redundant, but convenient for our purposes).

To be able to describe every meaning assignment, we would need to enrich the metalanguage just described (henceforth called the 'basic metalanguage') to turn it into a language that is universal, in the sense that it has a predicate for every property and an individual constant for every individual. We could then define a specification as a consistent metalanguage sentence of the following form:

$$S: \quad \bigwedge_k R\text{``}c_k\text{''}o_k \ \& \ \bigwedge_{ij} E\text{``}F_j^{i}\text{''}P_j^i \ \& \ \exists D\Big[\forall y \bigvee_{x\in D} y=x \ \& \ \bigwedge_{x,y\in D,\, x\neq y} {\sim}x=y \ \& \ \bigwedge_k x_k=o_k \ \&$$

$$\bigwedge_{ij}\big(\bigwedge_{<y_1,y_2,\ldots,y_i>\in V_{ij}} P_j^i y_1 y_2\ldots y_i \ \& \ \bigwedge_{<y_1,y_2,\ldots,y_i>\in D^i-V_{ij}} {\sim}P_j^i y_1 y_2\ldots y_i\big)\Big]$$

The F_j^i are all the predicates and the c_k are all the individual constants of the object language. Each P_j^i is an i-place metalanguage predicate and each o_k a metalanguage individual constant. (The individual constants o_j and o_k might be identical even if $j\neq k$, and the same holds for the predicates P_j^i and P_k^i.) D is a non-empty set of metalanguage individual variables that aren't part of the object language, and each x_k is a variable in D. (The variables x_j and x_k might be identical even if $j\neq k$. Moreover, x_j and x_k are identical whenever the individual constants o_j and o_k are identical, and might be identical even if $o_j \neq o_k$.) $V_{ij}\subseteq D^i$ for all i and j, and $V_{ij}=V_{ik}$ whenever $P_j^i=P_k^i$.

To derive an assignment of truth-values to object-language sentences from a specification, we need a semantic background theory. A suitable theory must be based on an infinitary higher-order logic that allows open formulas to be theorems. The theory must also include the following non-logical axiom schema:

$$(\text{T}_{\text{FOL}}) \quad T\text{``}A\text{''} \leftrightarrow \exists \boldsymbol{x_k} \exists \boldsymbol{\mathcal{X}_j^i}\Big(\bigwedge_k R\text{``}c_k\text{''}x_k \ \& \ \bigwedge_{ij} E\text{``}F_j^i\text{''}X_j^i \ \& \ A[X_j^i/F_j^i, x_k/c_k]\Big)$$

A can be any (closed) object-language sentence. The X_j^i are the predicate variables of the metalanguage. Each x_k is a metalanguage individual variable that isn't part of the object language. If $j\neq k$, then the variables x_j and x_k are distinct (and of course the variables X_j^i and X_k^i are distinct as well). $\boldsymbol{x_k}$ is the set of all the variables x_k, and $\boldsymbol{\mathcal{X}_j^i}$ is the set of all the variables X_j^i. $A[X_j^i/F_j^i, x_k/c_k]$ is the metalanguage formula obtained from the object-language sentence A by replacing each individual constant c_k with x_k and each predicate F_j^i with X_j^i. Given (T_{FOL}), a specification S determines the truth-value of every object-language

7

sentence *A*, in the sense that either $\vdash S \rightarrow T\text{``}A\text{''}$ or $\vdash S \rightarrow \sim T\text{``}A\text{''}$ (where $\vdash$ symbolizes theoremhood in our semantic theory).

We can give a revised and more precise statement of our criterion for metaphysical universality, (2). Note that the non-semantic information provided by a specification is never wholly correct. For example, every specification entails that there are exactly μ individuals for some cardinal number μ, whereas there are in fact more than set-many (and hence more than μ) individuals. However, given the expressive limitations of the finitary first-order object language, this specific misinformation might not affect the distribution of truth-values over object-language sentences that is determined by *S*, provided *S* tells us that the number of individuals is infinite. Let's call a specification *S* "accurate" iff *S* represents the non-semantic facts correctly in all ways that matter to the truth-values that the object-language sentences have on the meaning assignment described by *S*. We can then amend (2) as follows:

(2*)   An object-language formula *A* is metaphysically universal iff the ascription of truth to *A*'s universal closure can be derived from every accurate specification.

If we modify this criterion for metaphysical universality slightly, then we don't need a universal metalanguage but can make do with the basic metalanguage I initially described. Note that a specification provides much more information than is needed to determine the truth-values of all object-language sentences. Call two specifications *S* and *S** "permutation variants" iff, for some permutation π of the individual constants of the universal metalanguage and some adicity-preserving permutation $\pi*$ of its non-logical predicates, *S** is obtainable from *S* by uniformly replacing every individual constant *o* with $\pi(o)$ and every non-logical predicate *P* with $\pi*(P)$. Permutation variants determine the same truth-values for all object-language sentences. Hence, for the purpose of determining which object-language formulas are metaphysically universal, there is no need to distinguish between specifications that are permutation variants. We can instead focus on what it is in common between their contents. Everything relevant that their contents have

in common can be stated in the basic metalanguage by what I'll call the "Ramsey sentence" of the relevant specifications, which is of the following form:[3]

$$M_{\text{FOL}}: \ \exists D\, \exists \mathcal{X}_j^\varepsilon \left[\bigwedge_k R\text{"}c_k\text{"}x_k \ \& \ \bigwedge_{ij} E\text{"}F_j^i\text{"}X_j^i \ \& \ \bigwedge_{x\in D} E!x \ \& \ \forall y \bigvee_{x\in D} y=x \ \& \ \bigwedge_{x,y\in D,\, x\neq y} \sim x=y \ \& \right.$$

$$\left. \bigwedge_{ij} \left(\bigwedge_{<y_1,y_2,\dots,y_i>\in V_{ij}} X_j^i y_1 y_2 \dots y_i \ \& \ \bigwedge_{<y_1,y_2,\dots,y_i>\in D^i-V_{ij}} \sim X_j^i y_1 y_2 \dots y_i \right) \right],$$

where $\mathcal{X}_j^\varepsilon$ and the X_j^i are subject to the same conditions as in the case of schema T_{FOL} above, while D, the x_k, the F_j^i, the c_k, and the V_{ij} are subject to the same conditions as in the case of schema S. The conjunct $\bigwedge_{x\in D} E!x$ is logically redundant but it will be technically convenient to include it in the Ramsey sentence. I will call a Ramsey sentence "accurate" iff it represents the non-semantic facts correctly in all ways that matter to the truth-values of object-language sentences. For every Ramsey sentence M and (closed) object-language sentence A, either $\vdash M \rightarrow T\text{"}A\text{"}$ or $\vdash M \rightarrow \sim T\text{"}A\text{"}$. Let's say that Ramsey sentence M "validates" an object-language formula A ($M \Vdash A$, for short) to express that $\vdash M \rightarrow T\text{"}\forall A\text{"}$ holds, where $\forall A$ is a placeholder for A's universal closure. We can then state a new criterion for metaphysical universality:

(MU) A formula is metaphysically universal iff it is validated by all accurate Ramsey sentences.

We can formulate a syntactical criterion for deciding whether a formula A is validated by a Ramsey sentence M. Let M^- be the open formula in the scope of the existential quantifiers $\exists D\, \exists \mathcal{X}_j^\varepsilon$ at the beginning of M. (M^- is of the same form as the formula enclosed in the square brackets in the Ramsey sentence schema M_{FOL} above.) Moreover, let an "assignment" a be a function that maps each object-language individual variable to a metalanguage variable in D. Let's write "$M, a \Vdash A$" as short for $\vdash M^- \rightarrow A[X_j^i/F_j^i, x_k/c_k, a(v)/v)]$, where $A[X_j^i/F_j^i, x_k/c_k, a(v)/v)]$ is the formula obtained from

---

[3] Two permutation variants S and S^* may have shared contents not captured by their Ramsey sentence. For example, both of them may say of two object-language predicates that they express the same property, whereas their Ramsey sentence M merely makes the weaker claim that these predicates express properties that are instantiated by the same objects. However, the shared content that is not captured by M is irrelevant to the distribution of truth-values over object-language sentences.

$A[X_j^i/F_j^i, x_k/c_k]$ by replacing the free occurrences of every object-language variable v with $a(v)$. It is easy to see that the following holds:

(3$_{\text{FOL}}$) $M \Vvdash A$ iff $M, a \Vvdash A$ for every assignment a

Moreover, we can recursively specify conditions under which $M, a \Vvdash A$ is true. Let V be a function that maps every object-language individual constant c_k to the metalanguage variable x_k, and each non-logical object-language predicate F_j^i to V_{ij}. For any assignment a, let den_a be a function that maps every object-language individual constant c_k to x_k and every object-language individual variable v to $a(v)$. We can formulate the following recursive clauses (each of $t_1, t_2, \ldots, t_i$ is either an individual constant or an individual variable of the object language, and F^j is any non-logical object-language predicate):

(4$_{\text{FOL}}$) $M, a \Vvdash t_1 = t_2$ iff $\text{den}_a(t_1) = \text{den}_a(t_2)$

$M, a \Vvdash F_j^i t_1 \ldots t_i$ iff $\langle \text{den}_a(t_1), \ldots, \text{den}_a(t_i) \rangle \in V(F_j^i)$

$M, a \Vvdash {\sim}A$ iff   not $M, a \Vvdash A$

$M, a \Vvdash A \,\&\, B$ iff   both $M, a \Vvdash A$ and $M, a \Vvdash B$

$M, a \Vvdash \forall v A$ iff $M, a[v/d] \Vvdash A$ for every $d \in D$,

where $a[v/d]$ is the assignment that maps v to d and is otherwise like a.

We can single out a Ramsey sentence up to logical equivalence by specifying the set D and the function V. $\langle D, V \rangle$ can therefore be viewed as an abbreviation of the Ramsey sentence, or if you like, as a translation of the Ramsey sentence into a more efficient and perspicuous language L_{new}. Every sentence M of L_{new} is a pair $\langle D^M, V^M \rangle$, where D^M is a set of individual variables of L_{new} and V^M is a function mapping every object-language individual constant to a member of D^M and every non-logical n-place object-language predicate to a subset of $(D^M)^n$. The representational contents of the sentences of L_{new} can be explained by specifying a method for translating them into a conventional language of the same kind as the basic metalanguage described above. The translations will be sentences of the form displayed by schema M_{FOL}. (The discussion above allows us to formulate such a translation scheme.) L_{new} doesn't need to use the same individual variables as the conventional language into which we are translating. It's enough that there be a suitable mapping between the variables of the two languages. Moreover, there is no

need to describe a specific conventional metalanguage L once and for all and then explain the content of every sentence of L$_{new}$ by translating it into L. It's sufficient to describe a way of translating each sentence of L$_{new}$ into a suitable conventional language, even if the translation scheme has us translate different sentences of L$_{new}$ into different conventional languages. That means that there is no need to restrict the cardinalities of D^M and V^M in a sentence M of L$_{new}$ to fit them to the limits of a specific conventional infinitary language. D^M and V^M can have any cardinalities. To specify L$_{new}$ uniquely, all that's left to do is to say what the individual variables of the language are. The simplest option is to let the variables include every object whatsoever[4]: any entity can be a member of D^M. Once these changes have been made, the result is the familiar model theory for FOL.

To develop a similar interpretation of model theory for MFOL, we first need to add some expressions to the metalanguage: quote names for object-language sentences containing □ or ◊; a suitably large infinity of variables u, v, w, … that range over Possible Worlds; a monadic first-order predicate *Act* expressing actualization (a contingent property possessed by different Possible Worlds at different Possible Worlds); and a "truth-at operator" $<\ >$ that can combine with a World variable w and a formula P to form the formula $^w\!<P>$, interpreted as "it is true at w that P" or "at w, P."

The semantic theory couched in the metalanguage will again be based on an infinitary higher-order logic that allows open formulas to be theorems. While I won't try to specify a suitable logic in full, I will outline some of its basic features. To begin with, the logic is free: where c is an individual constant and y an individual variable, $\forall x\, A \to A[c/x]$ and $\forall x\, A \to A[y/x]$ (y free for x in A) are not theorems, although $\forall x\, A \to (E!c \to A[c/x])$ and $\forall x\, A \to (E!y \to A[y/x])$ (y free for x in A) are theorems. Similarly, while $\forall x\, E!x$ is a theorem, $E!t$ (where t is an individual variable or constant) is not. Moreover, it's a *negative free* logic, whose theorems include the instances of the schemata $X^i_j t_1 t_2 \ldots t_i \to E!t_k$ and $P^i_j t_1 t_2 \ldots t_i \to E!t_k$ ($1 \le k \le i$), where each of t_1, t_2, …, t_i is an individual variable or constant, X^i_j is a predicate variable and P^i_j a predicate. $\forall x\, x = x$ and $E!t \to t = t$ are theorems as well, but $t = t$ is not. The logic is also "inclusive," i.e. neutral on the question of whether there are any individuals—no formula that starts with an existential individual quantifier is a

---

[4] … except proper classes (if such there are), since they cannot be members of sets.

theorem. If A is any logical axiom, then so is $\forall w\ ^w\!<A>$. (I think this principle needs some restriction, but it's a close enough approximation for present purposes.) There also need to be axiom schemata that guarantee that the set of claims that are true at a given Possible World is closed under applications of the right inference rules (e.g., $\forall w\ (^w\!<A \to B> \to (^w\!<A> \to {}^w\!<B>))$) to ensure closure under modus ponens, and something like $\forall w\ (\bigwedge_{A \in S} {}^w\!<A> \to {}^w\!<\bigwedge_{A \in S} A>)$ (where S is a set of no more than κ formulas) to guarantee closure under infinitary conjunction introduction). Moreover, we need an axiom schema that (combined with the other axioms) ensures that the claims true at a World constitute a description of reality that is complete (subject to the expressive limitations of the language) and consistent—something along the lines of $\forall w\ (^w\!<\sim A> \leftrightarrow \sim\!{}^w\!<A>)$. Yet another axiom schema encapsulates the idea that an actualized World says the whole truth and nothing but the truth: $\forall w\ (Act(w) \to (^w\!<A> \leftrightarrow A))$.

The semantic theory will rely on an analysis of the modal operators in terms of truth at a Possible World. The analysis can be stated by specifying a method of translating every object-language formula A into a formula A^* of another language L*. The vocabulary of L* includes World variables and the truth-at operator, but not □ or ◊. Otherwise, its vocabulary is the same as that of the object language. In particular, it includes the same uninterpreted predicates and individual constants. The translation is intended to preserve the truth-conditions of sentences and the satisfaction conditions of open formulas under every meaning assignment. A recursive specification of the translation is given below ("*" always takes the smallest possible scope):

(5)   (i) $A^* = A$ if A is atomic   (iv) $(\forall x\ A)^* = \forall x\ A^*$

         (ii) $(\sim\!A)^* = \sim\!A^*$ (v) $(\Box A)^* = \forall w\ ^w\!<A^*>$

         (iii) $(A\ \&\ B)^* = A^*\ \&\ B^*$

To obtain our semantic theory, we add the instances of the following non-logical axiom schema to the logical axioms (A is an object-language sentence):

(T$_{\text{MFOL}}$) $T\text{``}A\text{''}\ \leftrightarrow\ \exists \boldsymbol{x_k}\, \exists \boldsymbol{\mathcal{X}^i_j}\left[\bigwedge_k R\text{``}c_k\text{''}x_k\ \&\ \bigwedge_{ij} E\text{``}F^i_j\text{''}X^i_j\ \&\ A^*[X^i_j/F^i_j, x_k/c_k]\right]$

$\boldsymbol{x_k}$, $\boldsymbol{\mathcal{X}^i_j}$, the x_k, the X^i_j, the c_k, and the F^i_j are subject to the same conditions as in the case of (T$_{\text{FOL}}$).

A Ramsey sentence for MFOL states information about Possible Worlds, in addition to the information about semantic and non-semantic matters that are also described by the Ramsey sentences for FOL. We can define a Ramsey sentence for MFOL as a consistent metalanguage sentence of the following form:

$$M_{\text{MFOL}}: \exists O_{\underline{w_0}} \exists \mathcal{X}_j \left[\bigwedge_k R``c_k"x_k \; \& \; \bigwedge_{ij} E``F_j^{i"}X_j^i \; \& \; \exists w_0 (Act(w_0) \; \& \right.$$

$$\textbf{w}_0 < \exists D_{\underline{w_0}}^1 \left[\bigwedge_{x \in D_{\underline{w_0}}^1 \cup D_{\underline{w_0}}^2} E!x \; \& \; \forall y \bigvee_{x \in D_{\underline{w_0}}^1 \cup D_{\underline{w_0}}^2} y=x \; \& \; \bigwedge_{\substack{x,y \in D_{\underline{w_0}}^1 \cup O_{\underline{w_0}} \\ x \neq y}} \sim x=y \; \& \; \bigwedge_{ij} \left(\bigwedge_{\langle y_1,y_2,\ldots,y_i \rangle \in V_{ij\underline{w_0}}} X_j^i y_1 y_2 \ldots y_i \; \& \; \bigwedge_{\langle y_1,y_2,\ldots,y_i \rangle \in (D_{\underline{w_0}}^1 \cup D_{\underline{w_0}}^2)^i - V_{ij\underline{w_0}}} \sim X_j^i y_1 y_2 \ldots y_i \right) \right.$$

$$\left. \left. \& \; \exists R_{\underline{w_0}} (\forall w \bigvee_{u \in R_{\underline{w_0}}} w=u \; \& \; \bigwedge_{u \in R_{\underline{w_0}}} {}^u \! < \! S_{\underline{u}} \!\! >) \right] \! > \right) \right]$$

For each i and j, X_j^i is an i-place metalanguage predicate variable. X_j^i and X_k^i are distinct if $j \neq k$. $O_{\underline{w_0}}$ is a non-empty set of metalanguage individual variables that aren't part of the object language. $x_k \in O_{\underline{w_0}}$ for all k. The variables x_j and x_k might be identical even if $j \neq k$. $D_{\underline{w_0}}^2$ is a subset of $O_{\underline{w_0}}$. $D_{\underline{w_0}}^1$ is a (possibly empty) set of individual variables that aren't part of the object language and which aren't in $O_{\underline{w_0}}$. For all i and j, $V_{ij\underline{w_0}} \subseteq (D_{\underline{w_0}}^1 \cup D_{\underline{w_0}}^2)^i$. $R_{\underline{w_0}}$ is a (possibly empty) set of World variables. Moreover, for all $u \in R_{\underline{w_0}}$, $S_{\underline{u}}$ is a formula of the following form:

$$(6) \qquad \exists D_{\underline{u}}^1 \left[\bigwedge_{x \in D_{\underline{u}}^1 \cup D_{\underline{u}}^2} E!x \; \& \; \forall y \bigvee_{x \in D_{\underline{u}}^1 \cup D_{\underline{u}}^2} y=x \; \& \; \bigwedge_{\substack{x,y \in D_{\underline{u}}^1 \cup O_{\underline{u}} \\ x \neq y}} \sim x=y \; \& \; \bigwedge_{ij} \left(\bigwedge_{\langle y_1,y_2,\ldots,y_i \rangle \in V_{ij\underline{u}}} X_j^i y_1 y_2 \ldots y_i \; \& \; \bigwedge_{\langle y_1,y_2,\ldots,y_i \rangle \in (D_{\underline{u}}^1 \cup D_{\underline{u}}^2)^i - V_{ij\underline{u}}} \sim X_j^i y_1 y_2 \ldots y_i \right) \right.$$

$$\left. \& \; \exists R_{\underline{u}} (\forall w \bigvee_{v \in R_{\underline{u}}} w=v \; \& \; \bigwedge_{v \in R_{\underline{u}}} {}^v \! < \! S_{\underline{v}} \!\! >) \right],$$

where (i) $O_{\underline{u}}$ is the set of individual variables that are bound by existential quantifiers scoping over $S_{\underline{u}}$ in M; (ii) $D_{\underline{u}}^1$ is a (possibly empty) set of individual variables that aren't part of the object language and which aren't in $O_{\underline{u}}$; (iii) $D_{\underline{u}}^2$ is a subset of $O_{\underline{u}}$; (iv) $V_{ij\underline{u}} \subseteq (D_{\underline{u}}^1 \cup D_{\underline{u}}^2)^i$ for all i, j; (v) $R_{\underline{u}}$ is a (possibly empty) set of World variables; (vi) for each $v \in R_{\underline{u}}$, $S_{\underline{v}}$ is another formula of the form (6) that meets conditions (i)–(vi). Thus, $O_{\underline{u}}$, $D_{\underline{u}}^1$, $D_{\underline{u}}^2$, $V_{ij\underline{u}}$, $R_{\underline{u}}$, and $S_{\underline{v}}$ in (6) satisfy the same constraints as $O_{\underline{w_0}}$, $D_{\underline{w_0}}^1$, $D_{\underline{w_0}}^2$, $V_{ij\underline{w_0}}$, $R_{\underline{w_0}}$, and $S_{\underline{u}}$ in M_{MFOL}. (The only difference is that in order for a sentence of the form M_{MFOL} to be logically consistent (and hence in order for it to count as a Ramsey sentence for MFOL), $D_{\underline{w_0}}^2$ must be identical with $O_{\underline{w_0}}$, whereas $D_{\underline{u}}^2$ needn't be identical with $O_{\underline{u}}$ and may even be empty.) We can define the notion of an accurate Ramsey sentences and the expression "$M \Vvdash A$"

("*M* validates *A*") in the same way as before, and use (MU) as a criterion of metaphysical universality for MFOL.

(The proper-constituent relation is non-well-founded on the constituent formulas of an accurate Ramsey sentence *M* for MFOL: *M* contains a formula that contains another formula that contains another formula, and so on to infinity. In most formal languages formulahood is defined recursively in a way that precludes such non-well-foundedness, so our metalanguage would have to be somewhat unusual in that regard. To avoid this, one could reformulate the account below by replacing Ramsey sentences with "Ramsey sequences": infinite sequences of well-founded sentences, each more complex and logically stronger than the preceding one, that together say the same as a single non-well-founded Ramsey sentence. (This change would necessitate other adjustments in the formulations below, but it would be fairly straightforward to implement the needed revisions.) However, for simplicity I'll state the account in terms of non-well-founded Ramsey sentences.)

We can develop a syntactic criterion for deciding whether $M \Vdash A$ holds for an arbitrary Ramsey sentence *M* for MFOL. I will let M^- be the open formula in the scope of the existential quantifiers $\exists O_{\underline{w_0}} \exists \mathcal{X}_{\overline{\gamma}}$ at the beginning of *M*. Moreover, I will use underlined letters like $\underline{u}$, $\underline{v}$ and $\underline{w}$ for *occurrences* of World variables (while continuing to use the same letters without underlining for World variables). Let's say that an occurrence $\underline{w}$ of World variable *w* in a Ramsey sentence *M superscripts* formula *S* iff $\underline{w}$ is the first symbol of an occurrence of $^w\!<S>$ in *M*. Let $\underline{W}$ be the set of occurrences of World variables in *M* that superscript formulas. For every $\underline{u} \in \underline{W}$, I'll write "$S_{\underline{u}}$" for the formula superscripted by $\underline{u}$. $S_{\underline{u}}$ is of the form (6). I will further write "$S_{\underline{u}}^-$" for the formula in the scope of the existential quantifiers at the beginning of $S_{\underline{u}}$. ($S_{\underline{u}}^-$ is of the same form as the formula in the square brackets in schema (6).) Let w_0 be the World variable to which the actualization predicate *Act* is applied in *M*. One occurrence of w_0 in *M* (set in boldface in M_{MFOL}) is attached to an occurrence of the truth-at operator that scopes over all other occurrences of that operator in *M*. Let's call that occurrence of w_0 "$\underline{w_0}$." Let $\underline{D}(\underline{w})$ be the function assigning to every $\underline{w} \in \underline{W}$ the set of variables $D_{\underline{w}}^1 \cup D_{\underline{w}}^2$. Moreover, let an *assignment a* be a function mapping each object-language individual variable to some variable in $\bigcup_{\underline{w} \in \underline{W}} \underline{D}(\underline{w})$. Note that the following biconditionals hold:

14

(7) $M \Vdash A$ iff$_{\text{def}}$ $\vdash M \to T\text{``}\forall A\text{''}$

   iff $\quad \vdash M \to \exists_{x_k} \exists x_j^i [\bigwedge_k R\text{``}c_k\text{''}x_k \ \& \ \bigwedge_{ij} E\text{``}F_j^i\text{''}X_j^i \ \& \ (\forall A)^*[X_j^i/F_j^i, x_k/c_k]]$ $\quad$ by (T$_{\text{MFOL}}$)

   iff $\quad \vdash M^- \to (\forall A)^*[X_j^i/F_j^i, x_k/c_k]$

   iff $\quad \vdash S_{\underline{w_0}}^- \to (\forall A)^*[X_j^i/F_j^i, x_k/c_k]$

   iff $\quad \vdash S_{\underline{w_0}}^- \to (\forall(A^*))[X_j^i/F_j^i, x_k/c_k]$ $\quad\quad\quad\quad\quad\quad\quad\quad\quad$ by (5)(iv)

   iff $\quad \vdash S_{\underline{w_0}}^- \to A^*[X_j^i/F_j^i, x_k/c_k, a(v)/v]$ for all a such that:

   $\quad\quad\quad\quad a(v) \in \underline{D}(\underline{w_0})$ for every variable v occurring free in A

Now suppose we introduce an additional way of using $\Vdash$:

(8) $\quad M, \underline{w}, a \Vdash A \quad$ iff$_{\text{Def.}}$ $\quad \vdash \underline{S_w^-} \to A^*[X_j^i/F_j^i, x_k/c_k, a(v)/v]$

We can then state the upshot of (7) as follows:

(3$_{\text{MFOL}}$) $\quad M \Vdash A \quad$ iff $\quad M, \underline{w_0}, a \Vdash A$ for all a such that: $a(v) \in \underline{D}(\underline{w_0})$ for every variable

   $\quad\quad\quad\quad\quad$ v occurring free in A

   We can also formulate a syntactical criterion for deciding whether $M, \underline{w}, a \Vdash A$ holds. For any $\underline{w} \in \underline{W}$, let $^{w}\diamondsuit$ be the occurrence of the truth-at operator in M to which $\underline{w}$ is attached. Moreover, let's say that an occurrence $\diamondsuit_1$ of the truth-at operator in M *scopes directly over* another occurrence $\diamondsuit_2$ iff (i) $\diamondsuit_1$ scopes over $\diamondsuit_2$, and (ii) $\diamondsuit_1$ doesn't scope over any occurrences of the truth-at operator that scope over $\diamondsuit_2$. Let $\underline{R}$ be a dyadic relation between elements of $\underline{W}$ such that: $\underline{Rww^*}$ iff $^{w}\diamondsuit$ scopes directly over $^{w^*}\diamondsuit$. (Note that if $\underline{Rww^*}$, then $\underline{w}^*$ is an occurrence of some variable in $R_{\underline{w}}$.) Let $\underline{V}$ be a function that assigns to every object-language individual constant c_k the variable x_k, and to each non-logical object-language predicate F_j^i a function mapping every $\underline{w} \in \underline{W}$ to $V_{ij\underline{w}}$. The free variables in $A^*[X_j^i/F_j^i, x_k/c_k, a(v)/v]$ also occur free in $S_{\underline{w}}^-$ if the following condition is satisfied:

(9) $\quad a(v) \in \underline{D}(\underline{w}) \cup O_{\underline{w}}$ for every variable v occurring free in A.

We can formulate recursive clauses that hold in all cases where (9) is true (each of $t_1, t_2, \ldots, t_i$ is either an individual constant or an individual variable, F^j is any non-logical object-language predicate, and den$_a$ is defined as before):

(4$_{\text{MFOL}}$) If (9) holds, then

   (i) $\quad M, \underline{w}, a \Vdash t_1 = t_2 \quad$ iff $\quad$ den$_a(t_1) = $ den$_a(t_2)$ and den$_a(t_1)$, den$_a(t_2) \in \underline{D}(\underline{w})$

   (ii) $\quad M, \underline{w}, a \Vdash F_j^i t_1 t_2 \ldots t_i \quad$ iff $\quad \langle$den$_a(t_1)$, den$_a(t_2), \ldots,$ den$_a(t_i)\rangle \in \underline{V}(F_j^i)(\underline{w})$

15

(iii) $M, \underline{w}, a \Vdash {\sim}A$ iff   not $M, \underline{w}, a \Vdash A$

(iv) $M, \underline{w}, a \Vdash A \,\&\, B$ iff   both $M, \underline{w}, a \Vdash A$ and $M, \underline{w}, a \Vdash B$

(v) $M, \underline{w}, a \Vdash \forall v\, A$ iff $M, \underline{w}, a[v/d] \Vdash A$ for all $d \in \underline{D}(\underline{w})$

(vi) $M, \underline{w}, a \Vdash \Box A$ iff $M, \underline{w^*}, a \Vdash A$ for every $\underline{w^*}$ such that

                                              $\underline{Rww^*}$.

The following can be shown by induction on the complexity of *A*:

(10)   If (9) is true, then the extensions of $\underline{R}$, $\underline{D}$, $\underline{V}$, and *a*, together with (4$_{\text{MFOL}}$), determine whether $M, \underline{w}, a \Vdash A$ holds.

We can simplify the account by two further moves. *Firstly*, we can remove the need to qualify (4$_{\text{MFOL}}$)(i)–(vi) with condition (9) by changing the definition of "$M, \underline{w}, a \Vdash A$". Instead of defining "$M, \underline{w}, a \Vdash A$" as in (8), we use the recursive clauses (4$_{\text{MFOL}}$)(i)–(vi) as our definition, so that (4$_{\text{MFOL}}$)(i)–(vi) is true whether or not condition (9) is satisfied. ("$M \Vdash A$" remains defined as "$\vdash M \to T$"$\forall A$"".) (4$_{\text{MFOL}}$) remains true, even after this change has been made. Therefore, given (10), we can conclude that whenever (9) holds, $M, \underline{w}, a \Vdash A$ is true on the new definition iff it is true on the old definition. It follows that, if $a(v) \in \underline{D}(\underline{w_0})$ for every variable *v* occurring free in *A* (so that *a*, w_0, and *A* satisfy condition (9)), then $M, \underline{w_0}, a \Vdash A$ is true on the new definition iff it is true on the old definition. So, (3$_{\text{MFOL}}$) is still true on the new definition, and we can use (3$_{\text{MFOL}}$) and (MU) as before to reason about metaphysical universality. *Secondly*, we can reformulate our test for metaphysical universality in terms of individual and World variables rather than occurrences thereof. That requires a slight change in the definition of "Ramsey sentence for MFOL." Note that nothing in our current definition rules out the possibility that different occurrences of a single World variable in a Ramsey sentence *M* superscript different formulas. For example, a Ramsey sentence *M* may contain the formula $^w\!\!<S\!>$, where *S* in turn contains the formula $\exists w \ldots {}^w\!\!<S^*\!>$ and $S \neq S^*$. The new definition of "Ramsey sentence" will continue to allow for cases where different formulas *S* and *S** are superscripted by occurrences of the same World variable, but it imposes the following restriction on the ways in which *S* and *S** can differ from each other in such cases:

(11)   For all $\underline{w}, \underline{w^*} \in \underline{W}$, if $\underline{w}$ and w^* are occurrences of the same World variable, then

(i) $\underline{D}(\underline{w}) = D^1_{\underline{w}} \cup D^2_{\underline{w}} = D^1_{\underline{w}*} \cup D^2_{\underline{w}*} = \underline{D}(\underline{w}*)$

(ii) $V_{ij\underline{w}} = V_{ij\underline{w}*}$ for all i, j

(iii) $R_{\underline{w}} = R_{\underline{w}*}$

(Note that (11) doesn't impose the condition that $O_{\underline{w}} = O_{\underline{w}*}$—the existential quantifiers scoping over $S_{\underline{w}}$ might bind different variables than those scoping over $S_{\underline{w}*}$. Accordingly, $S_{\underline{w}}$ and $S_{\underline{w}*}$ needn't be the same formula.) (11) reduces the range of sentences that count as Ramsey sentences. However, for every sentence M excluded by (11), some logically equivalent sentence (obtainable from M by relettering World variables in the right way) still counts as a Ramsey sentence by the new definition. Our modification in the definition of "Ramsey sentence" consequently makes no difference to which object-language formulas enjoy the distinction of being validated by all accurate Ramsey sentences, and therefore doesn't affect our criterion for metaphysical universality. As can be proven by induction on the complexity of A, if we use (4_{MFOL})(i)–(vi) as our definition of $M, \underline{w}, a \Vdash A$ and adopt (11), then the following holds: for all $\underline{w}, \underline{w}* \in \underline{W}$, if $\underline{w}$ and $\underline{w}*$ are occurrences of the same World variable, then for all a and A: $M, \underline{w}, a \Vdash A$ iff $M, \underline{w}*, a \Vdash A$. Let W be the set of World variables w such that $\underline{W}$ contains some occurrence of w. For any $w \in W$, we can define "$M, w, a \Vdash A$" as short for "$M, \underline{w}, a \Vdash A$ for every occurrence $\underline{w}$ of w that is a member of $\underline{W}$." And we can recursively state the conditions under which $M, w, a \Vdash A$ holds. Let D be the function that assigns to every $w \in W$ the set S such that: $S = \underline{D}(\underline{w})$ for every occurrence $\underline{w}$ of w in $\underline{W}$. Let R be the dyadic relation on W such that: $Rww*$ iff $w* \in \underline{R}_{\underline{w}}$ for every occurrence $\underline{w}$ of w in $\underline{W}$. Finally, let V be a function that maps (i) each object-language individual constant c_k to the metalanguage individual variable x_k, and (ii) each object-language predicate F^j to a function that maps each $w \in W$ to the set S such that: $S = \underline{V}(F^i_j)(\underline{w})$ for every occurrence $\underline{w}$ of w in $\underline{W}$. By removing the underlining from all symbols in (4_{MFOL})(i)–(vi), we obtain a recursive specification of the conditions under which $M, w, a \Vdash A$ holds. Furthermore, by removing the underlining from all symbols in (3_{MFOL}), we obtain a principle that states a new necessary and sufficient condition for a Ramsey sentence to validate a formula. We can combine this principle with (MU) to obtain a new criterion for metaphysical universality.

We can single out any Ramsey sentence M up to logical equivalence by specifying the quintuple $<W, R, w_0, D, V>$. The quintuple can therefore be viewed as a translation of the Ramsey sentence in a more efficient and perspicuous language, $L_{new}*$. All sentences of $L_{new}*$ are quintuples of this kind, constructed from the individual and World variables of $L_{new}*$ and the non-logical constants of the object language. Finally, we can lift all cardinality restrictions on the members of $<W, R, w_0, D, V>$ and allow all entities to serve as individual and World variables of $L_{new}*$, in a move analogous to the one we made for the language L_{new} that we used in the discussion of FOL. The outcome is a variant of model theory (with domain constraint) for MFOL.

The resulting version of model theory differs somewhat from the one that underlies Williamson's discussion. For example, it requires that all worlds in a model stand in the ancestral relation of accessibility to the actual world. Moreover, the interpretation function V of a model can only assign entities in $D(w_0)$ to the individual constants. (If there were a model that didn't meet this condition, then that model would represent, if anything, that some individual constant refers to an object that doesn't exist. That sounds like a contradiction, and in any case there is no meaning assignment that is described by such a model.) Finally, A's validity in a model requires A's truth at the actual world not on all assignments whatsoever, but merely on those that assign elements of $D(w_0)$ to the free variables in A. This weaker condition seems independently motivated if validity in all accurate models is to be our criterion for metaphysical universality, since the metaphysical universality of a formula amounts to the *actual* truth of its universal closure on all meaning assignments.
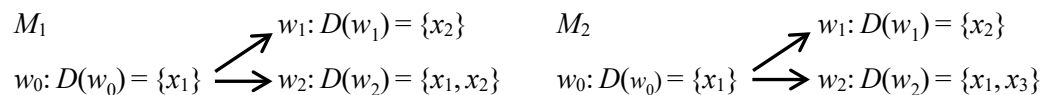
On my account, a model doesn't determine the meanings of the non-logical constants. In his discussion of Stalnaker's account (192), Williamson suggests that it is a significant weakness of an interpretation of model theory if it has that consequence. However, it's not clear to me that that is true. Model-theoretic semantics for modal logic isn't studying any specific meaning assignment but certain generalizations about *all* meaning assignments, namely truths of the form $\forall A$ *is true on all meaning assignments*. If two meaning assignments yield the same distribution of truth-values (given the relevant non-semantic facts), then the ways in which they differ are irrelevant, in which case it can only be advantageous to abstract away from these differences. What would be the benefit of using

18

models that represent inessential information? (Also, it seems clear that the models of FOL don't specify the meanings of all non-logical constants—the information they provide about the extensions of predicates don't determine which properties these predicates express. Why should matters be different for MFOL?)

4. Replies to Williamson's objections

Williamson's argument from (1) was intended to show that contingentists cannot connect model theory to the subject matter of modal discourse by appealing to an intended model structure. Since I have endorsed a different strategy for explaining the importance of model theory, my own account is not directly in the firing line of Williamson's argument. Nevertheless, I am committed to the claim that some models (the accurate ones) validate just those formulas that are metaphysically universal by contingentist lights. Therefore, if (1) were true, then that would refute my approach. However, as Williamson notes (136), the argument for (1) is blocked if the model-theoretic framework is modified in the way outlined in the penultimate paragraph of the last section. Williamson describes these modifications as unmotivated, but the discussion of the previous section seems to provide good motivation for it.

The interpretation of model theory I sketched also resolves the puzzle about what is represented by the alien entities of a model (the entities that are in the domains of worlds other than w_0 but not in the domain of w_0). The objects in the domain of a world are not names but variables, and hence don't stand for any particular individuals. We can therefore steer clear of the unpalatable idea that an alien entity represents a specific non-actual individual. Instead, an alien entity is an individual variable that is bound by an existential quantifier within the scope of a truth-at operator. Consider model M_1 represented below as an example (an arrow from one world to another represents that the latter is accessible from the former).

M_1 $w_1: D(w_1) = \{x_2\}$ M_2 $w_1: D(w_1) = \{x_2\}$

$w_0: D(w_0) = \{x_1\} \longrightarrow w_2: D(w_2) = \{x_1, x_2\}$ $w_0: D(w_0) = \{x_1\} \longrightarrow w_2: D(w_2) = \{x_1, x_3\}$

The translation of M_1 into a conventional metalanguage is a sentence of the following form:

(12)　　$\exists x_1 \exists \boldsymbol{\mathcal{x}}_j^i [\bigwedge_k R\text{``}c_k\text{''}x_1 \ \& \ \bigwedge_{ij} E\text{``}F_j^{i\text{''}}X_j^i \ \& \ \exists w_0 \left(Act(w_0) \ \& \ ^{w_0}{<}E!x_1 \ \& \ \forall y \, y{=}x_1 \ \& \ ... \ \& \right.$

　　　　$\left. \exists w_1 \exists w_2 \left(... \ \& \ ^{w_1}{<}\exists x_2[... \ \& \ {\sim}x_1{=}x_2 \ \& \ ...]{>} \ \& \ ^{w_2}{<}\exists x_2 \, [E!x_1 \ \& \ ...]{>}\right){>}\right)]$

This sentence tells us among other things that there actually exists exactly one individual x_1, such that there exists a Possible World w_1, such that an individual x_2 distinct from x_1 exists at w_1. That is a perfectly coherent interpretation of the model.

(In a conventional formal language, the fact that different variables in a sentence are occurrences of the same type variable is sometimes representationally significant and sometimes not. The first and second variables in $\exists x \, Fx \ \& \ \exists x \, Gx$ are occurrences of the same type variable as the third and fourth variable but that fact is representationally inert— $\exists x \, Fx \ \& \ \exists y \, Gy$ doesn't have that feature and yet it says the same as $\exists x \, Fx \ \& \ \exists x \, Gx$. The three variables in $\exists x \, (Fx \ \& \ Gx)$ are occurrences of the same type variable and that's representationally significant— $\exists x \, (Fx \ \& \ Gy)$ doesn't say the same as $\exists x \, (Fx \ \& \ Gx)$. It's worth noting that a similar distinction arises between two types of cases where the domains of different worlds in a model share an element e. This feature of the model is representationally significant iff

(13)　　Some world whose domain includes e stands in the ancestral relation of accessibility to some other world whose domain includes e.

In M_1, $D(w_1)$ and $D(w_2)$ have x_2 as a common element, but in this case (13) is not satisfied. Consequently, the overlap between $D(w_1)$ and $D(w_2)$ is representationally insignificant. If we replace x_2 with x_3 in $D(w_2)$ and in the extensions of all predicates at w_2, the resulting model M_2 (depicted above) has the same content as M_1. (To see this, note that we can uniformly substitute x_3 for x_2 in the scope of the truth-at operator superscripted with w_2 in (12) without changing the content of the sentence. The resulting sentence is the translation into a conventional metalanguage of M_2.) In M_1, $D(w_0)$ and $D(w_2)$ have x_1 as a common element, and in this case (13) is satisfied. Consequently, the overlap between $D(w_0)$ and $D(w_2)$ in M_1 is representationally significant. If we replace x_1 with x_3 in $D(w_2)$ and in the predicate extensions at w_2 in M_1, the resulting model M_3 (depicted below) has a different content from M_1.

$$M_3 \qquad\qquad w_1: D(w_1) = \{x_2\}$$

$$w_0: D(w_0) = \{x_1\} \quad \nearrow \atop \searrow \quad w_2: D(w_2) = \{x_3, x_2\}$$

For example, M_1 entails the claim that there actually exists an individual x_1 and a Possible World w_2 such that x_1 exists at w_2. M_3 entails the negation of that claim.)

Contingentists who endorse the account sketched above are in a good position to explain the significance of model theory. Model theory is important in part because it clarifies the truth-conditions of modal claims. For any object-language sentence A, model theory yields a necessary and sufficient condition C for a model to validate A. For A to be true, modal reality must be correctly describable by some model satisfying C . Of course, the models use the notions of a Possible World and of truth at a World to describe the modal realm, and the contingentist needs to explain these notions if they want to use model theory to illuminate the truth-conditions of sentences containing modal operators. But necessitists who endorse Williamson's interpretation of model theory are in exactly the same position: to explain the truth-conditions of sentences containing modal operators, they need to explain which model structure is intended, which requires them to explain what Possible Worlds are and what it is for a claim to be true at a Possible World. (Williamson takes on this task in *MLM*: sct. 3.7.) There is no obvious difference between contingentists and necessitists in their ability to extract from the model theory an account of the truth-conditions of modal claims.

Model theory does more than shed light on the truth-conditions of sentences containing modal operators, however. It affords a method of zooming in on the information about meanings and modal reality that is essential to questions about metaphysical universality, and a formal metalanguage in which that information can be perspicuously represented by a canonical description (a model). Moreover, it provides the formal principles (4_{MFOL}) and (3_{MFOL}), which license inferences between certain claims about Possible Worlds (the models) and claims about the truth-values of object-language sentences. Moreover, it gives us (MU) as a criterion for metaphysical universality. Now, (MU) by itself doesn't settle all questions about metaphysical universality. Even given (MU), which formulas are metaphysically universal still depends on which models are accurate, which in turn depends on the answers to some difficult questions about modal reality. However, in many cases

model theory enables us to work out the logical relationships between different claims about which formulas are metaphysically universal. With the help of (MU), (3_{MFOL}) and (4_{MFOL}), we can draw inferences from a claim about which formulas are metaphysically universal to a conclusion that tells us something about the conditions under which a model is accurate (and hence about what the space of Possible Worlds is like). From this conclusion together with (4_{MFOL}), (3_{MFOL}) and (MU), we can in turn derive further claims about which formulas are metaphysically universal. Thanks to (MU), (3_{MFOL}), and (4_{MFOL}), claims about Possible Worlds mediate inferences between different claims about which formulas are metaphysically universal.

Models are never completely correct representations of the non-semantic facts, since they misrepresent some cardinality facts. Contingentists could try to get around this result, either by modifying the model-theoretic machinery (cp. MLM: 236–40) or by giving a refined account of what is represented by a model that filters out the incorrect information. But there is no obvious need to do so. We often tolerate idealization, mere approximation, and other insignificant misrepresentation by models, maps and verbal descriptions, as long as the false information makes little or no difference for the purposes at hand. In any case, as mentioned above, necessitists trying to specify an intended model structure face similar complications arising from the cardinality constraints inherent in the model-theoretic apparatus.

In the most familiar sense of the term, you take an "instrumentalist" attitude to a theory if you regard it as a helpful device for predicting certain data but don't take it to be a correct description of facts that underlie and explain these data. The interpretation of model theory I sketched is not instrumentalist in this sense. It takes accurate models to represent certain facts about Possible Worlds that are distinct from, and explain, the facts about metaphysical universality that can be predicted by using model theory. It makes no difference to this point that the account regards accurate models as misrepresenting the facts in some irrelevant ways. What matters is that it explains the utility of the models as deriving entirely from their correct representation of certain features of reality that underlie the facts predicted.

References

Kment, B. (2014), *Modality and Explanatory Reasoning*, Oxford: Oxford UP.

Williamson, T. (2013), *Modal Logic as Metaphysics*, Oxford: Oxford UP.