To appear in a festschrift for Robert Stalnaker, edited by Judith Thomson (Oxford University Press)

Stalnaker on the Interaction of Modality with Quantification and Identity

Timothy Williamson

0. Logic is sometimes conceived as metaphysically neutral, so that nothing controversial in metaphysics is logically valid. That conception devastates logic. Just about every putative principle of logic has been contested on metaphysical grounds. According to some, future contingencies violate the law of excluded middle; according to others, the set of all sets that are not members of themselves makes a contradiction true. Even the structural principle that chaining together valid arguments yields a valid argument has been rejected in response to sorites paradoxes. In each case, a deviant metaphysics corresponds to the deviant logic. Of course, if one is trying to persuade deviant metaphysicians of the error of their ways, one is unlikely to get far by relying on logical principles that they reject. But that obvious dialectical exigency stably marks out no realm of logic. Each logical principle has persuasive force in some dialectical contexts and not in others. We do better to admit that logic has metaphysically contentious implications, and embrace them — if we know what they are.

Logic and metaphysics are not mutually exclusive. They overlap in the logic and metaphysics of existence, identity and possibility, for instance. The exploration (but not total conquest) of that area was one of the great achievements of twentieth century philosophy. Here, as in so many other areas, Bob Stalnaker has played an exemplary role, as a voice for metaphysical sobriety and the careful archaeology of logical structure. His intervention clarifies and deepens every debate in which he participates. In this essay I will examine the innovative argument of his 1994 paper 'The Interaction of Modality with Quantification and Identity'.

1. Stalnaker begins by considering two languages. One is a first-order language with quantification, predication and an identity sign but no modal operators. The other is a propositional language with modal operators but no quantification, predication or identity sign. For each language, he provides a sound and complete axiomatization with respect to what we may call its *standard semantics*; this semantics differs slightly from the usual Kripke semantics in ways noted below, but for many purposes we can ignore the differences. Combining the two languages gives us a language for quantified modal logic with identity. We might hope that combining the axiomatizations for the two languages would give us a correspondingly sound and complete axiomatization with respect to the standard semantics for the combined language. However, Stalnaker proves that the hope is vain, by providing non-standard semantic theories that validate the combined axiomatization but falsify some formulas that are valid on the standard semantics: consequently, those formulas are underivable in that axiomatization. Thus the combined

axiomatization, although sound, is not complete with respect to the standard semantics for the combined language.

What is less clear is the significance that Stalnaker attaches to his result. In one case, he says that the non-standard semantics allows for a ‘nonstandard conception of individuals and their modal properties’ (a form of counterpart theory); whether or not that conception is defensible, ‘the issue is a philosophical one that cannot be settled by logical theory’ (p. 154).[1] That might lead one to interpret Stalnaker as denying that the formulas invalidated by the non-standard semantics are genuine logical truths. Presumably, the standard semantics would be at fault, for employing too narrow a range of models and thereby validating formulas that are in some sense too substantive to deserve the status of logical truth. But Stalnaker later describes the formulas at issue as ‘logical principles that are valid’ (p. 157). If he is employing a notion of validity only relative to a semantic theory, then of course every formula of the language is valid relative to some semantic theories and invalid relative to others. In a postscript added in 2002, he describes a formula valid on the standard but not the non-standard semantic theory as ‘less central’ to the logic in question than is a formula valid on both semantic theories (p. 161). That informal idea of comparative centrality may come closest to what he has in mind.

Let us postpone these questions of philosophical significance, and examine Stalnaker’s argument in detail. We start with a description of the syntax and semantics for the combined language, from which those for the two original languages can easily be derived by deletion of inapplicable features. The syntax is slightly unusual, because Stalnaker parses quantification in terms of a predicate abstraction device $\frown$.

The simple expressions of the language are as follows. What Stalnaker calls the 'descriptive' (non-logical) vocabulary consists of denumerably many atomic sentences (sentence letters), denumerably many n-place atomic predicates for each $n > 0$, and denumerably many individual constants. There are also denumerably many individual variables and parentheses. The logical vocabulary consists of the two-place predicate '=' and $\wedge$, $\sim$, $\Box$, $\forall$ and $\frown$. In the usual way, sentences involving other logical symbols, such as $\vee$, $\rightarrow$, $\leftrightarrow$, $\Diamond$ and $\exists$, are read as meta-linguistic abbreviations for sentences made up of the primitive vocabulary. The singular terms are the variables and the individual constants.

The complex expressions are constructed thus. If F is an n-place predicate and t_1, …, t_n are singular terms then $Ft_1 \ldots t_n$ is a sentence. If φ and ψ are sentences then $\sim\varphi$, $\Box\varphi$ and $(\varphi \wedge \psi)$ are sentences. If F is a one-place predicate then $\forall F$ is a sentence. If φ is a sentence and x is a variable then $\hat{x}\varphi$ is a one-place predicate. Informally, if we can read φ as '… x …' then we can read $\hat{x}\varphi$ as 'is such that … it …' and therefore $\hat{x}\varphi t$ as 't is such that … it …' and $\forall \hat{x}\varphi$ as 'Everything is such that … it …'.

Here is Stalnaker's standard model-theoretic semantics for the language. A model is a quadruple <W, R, D, *v*>, where: W is a nonempty set; R is a binary relation on W; D is a function from members w of W to sets D_w (which may be empty); v is a function that takes each individual constant c to a partial function from members w of W to members $v_w(c)$ of D_w; v takes each sentence letter A to a total function from members w of W to members $v_w(A)$ of $\{0, 1\}$; v takes each n-place atomic predicate F to a total function from members w of W to sets $v_w(F)$ of n-tuples of members of D_w. Informally, we can think of W as a set of possible worlds, R as a relation of accessibility between worlds, D_w as the domain of the world w (the set of individuals that exist at that world), $v_w(c)$ as the denotation of c at w (if any), $v_w(A)$ as the truth-value of A at w (1 if A is true, 0 otherwise), and $v_w(F)$ as the extension of F at w; in particular, $v_w(=)$ is the set of pairs $<d, d>$ for all d in D_w. Thus v maps descriptive expressions to intensions of the appropriate type; the identity sign is mapped to its intended intension. The semantic rules assign values to simple and complex expressions relative to a model and an assignment, which is a partial function from individual variables to members of the union of the sets D_w for all w in W. If s is an assignment and x is a variable, $s[d/x]$ is the assignment that maps x to d but otherwise is like s. Given a model <W, R, D, v>, semantic values are assigned thus (where s is an assignment and w is in W):

If x is an individual variable, $v^s{}_w(x) = s(x)$.

If φ is an atomic descriptive expression, $v^s{}_w(\varphi) = v_w(\varphi)$.

If F is an n-place predicate and $t_1, \ldots, t_n$ are singular terms,

$$v^s{}_w(Ft_1\ldots t_n) = 1 \text{ if } <v^s{}_w(t_1), \ldots, v^s{}_w(t_n)> \varepsilon\ v^s{}_w(F)\text{; otherwise } v^s{}_w(Ft_1\ldots, t_n) = 0.$$

If φ is a sentence, $v^s{}_w(\sim\varphi) = 1 - v^s{}_w(\varphi)$.

If φ and ψ are sentences, $v^s{}_w((\varphi \wedge \psi)) = \min\{v^s{}_w(\varphi), v^s{}_w(\psi)\}$.

If φ is a sentence, $v^s{}_w(\Box\varphi) = 1$ if $v^s{}_u(\varphi) = 1$ whenever wRu; otherwise $v^s{}_w(\Box\varphi) = 0$.

If F is a one-place predicate, $v^s{}_w(\forall F) = 1$ if $v^s{}_w(F) = \mathrm{D}_w$; otherwise $v^s{}_w(\forall F) = 0$.

If φ is a sentence and x is a variable, $v^s{}_w(\hat{x}\varphi) = \{d \ \varepsilon\ \mathrm{D}_w \colon v^{s[d/x]}{}_w(\varphi) = 1\}$.

To obtain the semantics for the non-modal first-order language, delete W and R in the definition of a model, the clause for □φ, and the world subscript (w) throughout, and conceive the constituents D and v of a model accordingly. To obtain the semantics for the propositional modal language, delete instead D from the original definition of a model, the clauses that involve singular terms or predicates, and the assignment superscript (s) throughout.

Stalnaker's axioms and rules for the combined language are as follows, where $\vdash$ expresses provability, which is restricted to closed sentences.

Propositional Logic	If φ is a truth-functional tautology, $\vdash \varphi$
Modus Ponens	If $\vdash \varphi \rightarrow \psi$ and $\vdash \varphi$ then $\vdash \psi$
K Schema	$\vdash \Box(\varphi \rightarrow \psi) \rightarrow (\Box\varphi \rightarrow \Box\psi)$
Necessitation	If $\vdash \varphi$ then $\vdash \Box\varphi$
Abstraction	$\vdash \forall \hat{x}(\hat{y}\varphi x \leftrightarrow \varphi^{x}/_{y})$, where $\varphi^{x}/_{y}$ is the result of substituting

x for all free occurrences of y in φ (relettering bound variables in φ where necessary to prevent clashes)

Quantification $\vdash \forall \hat{x}(\varphi \rightarrow \psi) \rightarrow (\forall \hat{x}\varphi \rightarrow \forall \hat{x}\psi)$

Redundancy[2] $\vdash \forall \hat{x}Fx \leftrightarrow \forall F$, where x is not free in F

Existence $\vdash Ft_1 \ldots t_n \rightarrow \exists \hat{x}x{=}t_i$, where F is an n-place predicate

Identity[3] $\vdash s{=}t_i \rightarrow (Ft_1 \ldots t_i \ldots t_n \rightarrow Ft_1 \ldots s \ldots t_n)$
where F is an n-place predicate

Universal Generalization If $\vdash \varphi \rightarrow \psi$ and t does not occur in φ then $\vdash \varphi \rightarrow \forall \hat{x}\psi^{x}/_{t}$

To obtain the axioms and rules for the non-modal first-order language, delete the K schema and Necessitation. To obtain the axioms and rules for the propositional modal language, delete instead Abstraction, Quantification, Redundancy, Existence, Identity and Universal Generalization.

Now consider the following schemas, where E is the existence predicate $\hat{x}\exists \hat{y}x{=}y$ (thus $v^{s}{}_{w}(E) = \mathrm{D}_w$):[4]

CBF $\Box \forall \hat{x}\varphi \rightarrow \forall \hat{x}\Box\varphi$

QCBF $\Box\forall\hat{x}\varphi \rightarrow \forall\hat{x}\Box(Ex \rightarrow \varphi)$

EI $\forall\hat{x}\forall\hat{y}(x=y \rightarrow \Box(Ex \rightarrow x=y))$

NEI $\forall\hat{x}\ \forall\hat{y}\Box(x=y \rightarrow \Box(Ex \rightarrow x=y))$

ND $\forall\hat{x}\forall\hat{y}(\sim x=y \rightarrow \Box\sim x=y)$

Each of these principles except CBF is valid ─ it (or all its instances, in the case of QCBF) is true on all assignments at all worlds in all models ─ on Stalnaker's semantics.

CBF is the converse Ibn-Sina-Barcan schema, usually known just as the converse Barcan formula.[5] Informally, it says that that if necessarily everything meets a certain condition, then everything is such that necessarily it meets that condition. CBF is highly controversial; for instance, it implies that if necessarily everything exists then everything has necessary existence. Many philosophers regard it as a trivial necessary truth that everything exists (everything is something), but an obvious falsehood that everything has necessary existence. Since Stalnaker's semantics permits contingent existence, it invalidates CBF. QCBF is a weak consequence of CBF; it says that if necessarily everything meets a certain condition, then everything is such that necessarily it exists only if it meets that condition. The existence qualification in QCBF finesses the usual objections to CBF; the relevant instance says only that if necessarily everything exists then everything is such that necessarily it exists only if it exists, which is clearly harmless. On Stalnaker's semantics, if the antecedent of QCBF is true, then in every accessible world everything in the domain satisfies $\hat{x}\varphi$, so everything in the domain of the original world is such that in every accessible world if it is in the domain of that world then it satisfies $\hat{x}\varphi$, so the consequent is true; thus QCBF is valid.

EI is the essentiality of identity. Informally, it says that if things are identical then necessarily one of them exists only if they are identical. NEI is a strengthened form of EI. It says that all things are such that necessarily if they are identical then necessarily one of them exists only if they are identical. Both EI and NEI are valid on Stalnaker's semantics because variables are rigid designators: they are assigned values absolutely, not relative to worlds. If they have the same value at any world, they have the same value at every world. The existence qualification in EI and NEI is needed on Stalnaker's semantics, for the extension of '=' at a world is restricted to the domain of that world.

ND is the necessity of distinctness. Informally, it says that if things are not identical then they could not have been identical. It too is valid on Stalnaker's semantics because variables are rigid designators. If they have different values at any world, they have different values at every world. No existence qualification is needed, for the identity claim is automatically false at any world whose domain does not contain the values of both variables.

EI is derivable in Stalnaker's axiomatization. CBF is of course underivable, since it is invalid on his semantics, for which the axiomatization is sound. However, although QCBF, NEI and ND are all valid on his standard semantics, they are all underivable in his axiomatization. Stalnaker proves their underivability by providing non-standard semantics on which the axioms are valid, the rules preserve validity but the formula in question is invalid. For QCBF and NEI, the non-standard semantics deviates by treating variables as non-rigid: although they are initially assigned a world-independent value, they are subsequently evaluated as denoting a counterpart of the initial value with respect to a given world. The way in which QCBF and NEI fail on this semantics is quite subtle; the details can be checked in Stalnaker's paper and are not crucial for the argument to come. Stalnaker shows that even if one adds QCBF as an axiom schema (which permits the derivation of NEI too), ND remains underivable. He provides another deviant semantics on which variables behave rigidly but '=' is interpreted as indiscernibility rather than identity. His original axioms and QCBF are valid on this semantics, and the rules preserve validity, but ND is invalid: discernibles could have been indiscernible.

2. What philosophical import does Stalnaker attribute to his independence results? Concerning QCBF (and NEI), he writes:

> The variant [counterpart] semantics brings to the surface an assumption about the relation between the modal properties of an individual in different possible worlds ─ an assumption implicit in the standard semantics that is not grounded in the nonmodal logic of predication, or in the modal logic of propositions, or in their combination.

(p. 153). Concerning ND, he says:

> the combination of identity theory with modality provides the resources to distinguish identity from a weaker relation [indiscernibility] that cannot be distinguished from it in a nonmodal context. Perhaps this shows that there is in some sense something modal about the concept of identity.

(p. 157). Are these claims justified?

Let us first ask what methodology is implicit in Stalnaker's form of argument. In schematic terms, he considers a language L1 with an axiom system A1 that is sound and complete with respect to a semantics S1 (all and only formulas valid on S1 are derivable in A1), and a language L2 with an axiom system A2 that is sound and complete with respect to a semantics S2. He combines L1 and L2 into a joint language L1+L2, S1 and S2 into a joint semantics S1+S2 for L1+L2 and A1 and A2 into a joint axiom system A1+A2 for L1+L2. He then treats A1+A2 as exhausting what the logics of L1 and L2 tell us, when combined, about the logic of L1+L2. But what do 'combine', 'joint' and '+' mean here?

First, it is not always clear what it would be to combine two languages. For example, what is English+Japanese? In the particular case at issue, another way of combining the first-order language of predication, quantification and identity with the propositional modal language would have permitted the application of □ only to closed formulas; such a language would have a lower grade of modal involvement (Quine 1966). However, let us simply take Stalnaker's way of combining the two languages as given. It is standard enough, and philosophically attractive.

Second, it is not always clear what it would be to combine two semantic theories for different languages. For example, in his official combined semantics, Stalnaker makes the assignment of values to variables absolute, not relative to possible worlds, on the plausible grounds that 'open sentences are devices for the formation of complex predicates that express properties of individuals, and not properties of some kind of intension' (p. 150). But one can also give a semantics for first-order modal languages in which the assignment of values to variables is world-relative; the variables may then be said to stand for individual concepts (see Hughes and Cresswell 1996: 334-42 for an introduction to such systems). Thus the form of the semantics for the two original languages underdetermines the form of the semantics for the more inclusive language. For the time being, let us simply take Stalnaker's standard form of the semantics for the first-order modal language as given.

Third, it is not always clear what it would be to combine two axiomatizations for different languages. We have to decide what counts as an instance in one language of an axiom schema or rule of inference originally formulated with respect to another language; more than one extrapolation is generally possible, since we may or may not permit the distinctive vocabulary of one language to occur in instances of an axiom schema or rule inherited from the axiomatization of the other language. In the present case, Stalnaker permits such instantiations; the combined system would otherwise be extremely weak. Given his way of presenting his axiomatizations, his way of combining them is natural enough. We can accept it too as given.

There is a more urgent question. Distinguish *logics* (consequence relations, or sets of theorems) from *axiomatizations* (sets of axioms and inference rules); many different axiomatizations can generate the same logic. But then why should we assume that Stalnaker's combined axiomatization generates all and only the formulas in the first-order modal language to which one is committed by acceptance of his combined logics for the first-order non-modal language and the propositional modal language, even granted that his logics for those languages are sound and complete with respect to their original semantics?

Let us start with the soundness of Stalnaker's combined axiomatization with respect to his combined semantics. It does *not* follow merely from the soundness of his two axiomatizations for the original languages with respect to their original semantics. For some other axiomatizations for those languages are sound with respect to their semantics even though, when one combines them in a way analogous to Stalnaker's, the resulting combined axiomatization is *un*sound with respect to Stalnaker's combined semantics.

Here is an example. Replace the axiom schema Identity above with this axiom schema, a familiar form of Leibniz's Law:

Identity* $\vdash s{=}t \rightarrow (\varphi \rightarrow \varphi^{s}/_{t})$

Every instance of Identity is an instance of Identity*, so the completeness of the resulting axiomatization for first-order non-modal logic follows from the completeness of Stalnaker's original axiomatization. But instances of Identity* in which φ is not a predication are not instances of Identity. However, as Stalnaker says, every instance of Identity* in the first-order non-modal language is derivable in his original axiomatization (p. 148). Since his axiomatization is sound with respect to his semantics for the non-modal language, every instance of Identity* in that language is valid on his semantics for that language. Consequently, the proof system in which Identity* replaces Identity is sound with respect to Stalnaker's semantics for the first-order non-modal language. Nevertheless, the axiomatization for the first-order modal language that results from combining (in Stalnaker's way) the axiomatization that uses Identity* with Stalnaker's axiomatization for the propositional modal language is not sound with respect to Stalnaker's semantics for the combined language. For, as Stalnaker says, Identity* has instances in the combined language that are invalid on his semantics (p. 148). One such instance is:

(1) $s{=}t \rightarrow (\Diamond(s{=}s \wedge \sim s{=}t) \rightarrow \Diamond(s{=}s \wedge \sim s{=}s))$

Here *s* and *t* are distinct constants. The reason is that Stalnaker classifies individual constants as descriptive terms and does not require them to be rigid designators; he allows them a flexibility that he does not allow to individual variables (in his standard semantics). Thus *s* and *t* may designate the same individual with respect to one world while designating distinct individuals with respect to another world accessible from it.

Although that particular example depends on Stalnaker's questionable treatment of individual constants, the general point does not. A less controversial example uses an extensionality principle for predicates:

Coextensiveness $\vdash \forall \hat{x}(Fx \leftrightarrow Gx) \rightarrow (\varphi \rightarrow \varphi^F/_G)$

For simplicity, *F* and *G* here are one-place atomic predicates; $\varphi^F/_G$ is the result of substituting *F* for *G* in φ. All instances of Coextensiveness in the first-order non-modal language are valid on any standard semantics, but the same schema has clearly invalid instances in the first-order modal language, such as:

(2) $\forall \hat{x}(Fx \leftrightarrow Gx) \rightarrow (\sim\Box\forall \hat{x}(Fx \leftrightarrow Gx) \rightarrow \sim\Box\forall \hat{x}(Fx \leftrightarrow Fx))$

For *F* and *G* may be accidentally coextensive.

The general point is this. The mere soundness and completeness of an axiomatization with respect to the semantics for a language does not entitle one to extrapolate the axiom schemas and rules of inference of the axiomatization in the natural way to a more inclusive language as being what the logic of the logical constants in the original language has to offer the combined logic of the logical constants in the extended language. Different axiomatizations that yield the same set of theorems in the original language may yield different sets of theorems from each other when extrapolated to the extended language. In effect, Stalnaker makes this point himself about the example above: 'The validity of the general schema [Identity*], unlike the validity of the identity axioms [Identity], depends on the expressive limitations of the extensional theory' (p. 148). In axiomatizing the restricted language, one must choose axiom schemas and rules of inference whose appropriateness does not depend on the expressive limitations of the restricted language. Doing that is no straightforward matter, since proving soundness and completeness for the restricted language does not suffice. One must somehow use a conception of how the restricted language may legitimately be extended, in both syntax and semantics.

In the present case, of course, Stalnaker has carefully chosen his axiom schemas and rules of inference for the original languages so that they remain sound when extrapolated to the combined language. But an analogous issue arises about completeness. For one might provide a sound and complete axiomatization for a restricted language that is unnecessarily weak when extrapolated to more inclusive languages.

Here is an example. Stalnaker's propositional modal logic is K. Like most of the familiar systems of propositional modal logic, K is decidable. Thus, instead of using Stalnaker's axiom schemas of Propositional logic and the K schema and rules of Modus Ponens and Necessitation, one can in principle axiomatize K simply by taking all formulas that pass some given decision procedure for K (in effect, all theorems of K) as axioms. Call that axiom family 'Cheap K'. Analogously, Stalnaker's axiom family of Propositional Logic in effect axiomatizes non-modal propositional logic simply by taking all tautologies (in effect, all theorems) as axioms. Cheap K by itself constitutes a sound and complete axiomatization of K. Now combine it in the natural way with Stalnaker's axiomatization of the logic of the first-order non-modal language (Propositional Logic, Abstraction, Quantification, Redundancy, Existence, Identity, Modus Ponens and Universal Generalization) into an axiomatization for the first-order modal language. That axiomatization is manifestly inadequate. The instances of Cheap K are just substitution instances in the first-order modal language of theorems of K. There is no rule of Necessitation. One cannot even derive:

(3) $\Box \forall \hat{x}(Fx \rightarrow Fx)$[6]

But it would not be plausible to conclude that the deviant semantics brings to the surface 'an assumption implicit in the standard semantics that is not grounded in the nonmodal logic of predication, or in the modal logic of propositions, or in their combination' in any serious sense. Rather, its underivability is a mere artefact of the specific way in which the propositional modal logic was axiomatized.

The question now naturally arises: are Stalnaker's underivability results similarly artefacts of the specific ways in which he axiomatized the logics of the propositional modal language and of the first-order non-modal language? If so, they lack the philosophical significance that he claims for them.

Let us start with QCBF. The first point to notice is that in Stalnaker's system we can derive a schema very close to QCBF:

QCBF* $\Box\forall\hat{x}\varphi \rightarrow \forall\hat{x}\Box(Ex \rightarrow \hat{x}\varphi x)$

For where F is an atomic one-place predicate and t is an individual constant, the formula $\forall F \rightarrow (Et \rightarrow Ft)$ is valid on Stalnaker's semantics for the first-order non-modal language, and therefore provable. Consequently, substituting $\hat{x}\varphi$ for F in the proof gives us in his axiomatization for the first-order modal language:

(4) $\forall\hat{x}\varphi \rightarrow (Et \rightarrow \hat{x}\varphi t)$

(Exercise: lay out the proof in full.) By Necessitation and the K schema, we have:

(5) $\Box\forall\hat{x}\varphi \rightarrow \Box(Et \rightarrow \hat{x}\varphi t)$

Universal Generalization yields QCBF* from (5). From QCBF*, we could easily prove QCBF if we had:

(6) $\forall\hat{x}\Box(\hat{x}\varphi x \rightarrow \varphi)$

Although (6) is valid on Stalnaker's semantics for the first-order modal language, it is unprovable in his axiomatization. We cannot derive it from his axiom schema of Abstraction, because the latter does not allow us to slip a modal operator between the outer and inner occurrences of the predicate abstraction operator. His Abstraction schema constrains the effect of the abstraction operator only in a limited range of contexts. But that is an artefact of Stalnaker's version of Abstraction. Consider this rule of inference, designed to constrain the effect of the abstraction operator in a wider range of contexts:

Free Abstraction — If B is the result of replacing some or all occurrences of $\hat{y}\varphi x$ in A by $(Ex \wedge \varphi^{x}/_{y})$, where $\varphi^{x}/_{y}$ is as in the original Abstraction principle, and $\vdash$ A, then $\vdash$ B.

The rule is called 'Free Abstraction' because it includes an existence qualification, which is characteristic of free logic. Since x and y are allowed to be the same variable, Free Abstraction has the special case in which occurrences of $\hat{x}\varphi x$ are replaced by $(Ex \wedge \varphi)$. Thus from QCBF*, in the system in which Free Abstraction replaces Stalnaker's Abstraction principle, the rule yields this theorem:

(7) $\Box\forall\hat{x}\varphi \rightarrow \forall\hat{x}\Box(Ex \rightarrow (Ex \wedge \varphi))$

By dropping a conjunct from (7), we easily prove QCBF.

We have still to justify Free Abstraction. To show that it preserves validity on Stalnaker's semantics for the first-order modal language, it suffices to check that in any model, $v^{s}_{w}(\hat{y}\varphi x) = v^{s}_{w}((Ex \wedge \varphi^{x}/_{y}))$, for any assignment s and world w, for then the compositional nature of the semantics ensures that the sentences A and B in the rule always have the same semantic value. The check is easily made. First, suppose that $v^{s}_{w}(\hat{y}\varphi x) = 1$. Then $s(x)\ \varepsilon\ v^{s}_{w}(\hat{y}\varphi) \subseteq D_{w}$, so $v^{s}_{w}(Ex) = 1$ and moreover, by the validity of Stalnaker's own Abstraction schema, $v^{s}_{w}(\varphi^{x}/_{y}) = v^{s}_{w}(\hat{y}\varphi x) = 1$. Thus $v^{s}_{w}((Ex \wedge \varphi^{x}/_{y})) = 1$. Conversely, if $v^{s}_{w}((Ex \wedge \varphi^{x}/_{y})) = 1$, then $v^{s}_{w}(Ex) = 1$ and $v^{s}_{w}(\varphi^{x}/_{y}) = 1$; but by the former

$s(x)$ ε D_w, so by the validity of Stalnaker's own Abstraction schema $v^s_w(\varphi^x/_y) = v^s_w(\hat{y}\varphi x)$. Thus $v^s_w(\hat{y}\varphi x) = 1$. By dropping the world subscript w throughout the argument, we can also show that Free Abstraction preserves validity on Stalnaker's semantics for the first-order non-modal language. The underlying point is that abstracting on an open sentence with respect to a variable and then applying the resultant predicate to that variable makes no difference when the value of the variable is in the relevant domain of quantification but produces falsity otherwise (since that value is excluded from the extension of the predicate), on Stalnaker's standard semantics. He allows abstraction with respect to an individual constant to have more extensive effects, since it may involve replacing a non-rigid by a rigid designator, but both Stalnaker's original Abstraction schema and Free Abstraction concern only abstraction with respect to variables.

We can also check that, in the presence of Free Abstraction, Stalnaker's own Abstraction schema becomes redundant. Using just Propositional Logic, Universal Generalization and Modus Ponens we obtain:

(8) $\forall \hat{x}(\hat{y}\varphi x \leftrightarrow \hat{y}\varphi x)$

Consequently, by Free Abstraction:

(9) $\forall \hat{x}(\hat{y}\varphi x \leftrightarrow (Ex \wedge \varphi^x/_y))$

We must eliminate the first conjunct on the right-hand side. Let T be any closed truth-functional tautology. As with (8), we obtain:

(10) $\forall \hat{y}\mathrm{T}$

By an instance of Redundancy (with $\hat{y}\mathrm{T}$ for F) we have:

(11) $\forall \hat{x}\hat{y}\mathrm{T}x$

Applying Free Abstraction to (11) gives:

(12) $\forall\, \hat{x}(Ex \wedge \mathrm{T}))$

From (9) and (12), Propositional Logic, Universal Generalization, Quantification and Modus Ponens yield Stalnaker's Abstraction schema. Thus we lose no theorems by dropping his Abstraction schema and employing Free Abstraction instead.

In the first-order non-modal logic, Free Abstraction makes no difference. Stalnaker's original axiomatization is sound and complete; as we have seen, the new axiomatization is sound and extends Stalnaker's, so is also complete; thus the set of theorems is the same. In the first-order modal logic, the new system is still sound, but properly extends Stalnaker's, because QCBF is derivable only in the new system. As Stalnaker notes, NEI is derivable given QCBF (p. 155). NEI is therefore another theorem of the axiomatization with Free Abstraction in place of Stalnaker's Abstraction schema. Thus the underivability of QCBF and NEI is no deep fact about the relation between the logic of quantification and the logic of modality. It merely reflects Stalnaker's unforced choice amongst ways of formulating the logic of quantification that are equivalent in a non-modal setting but not in a modal setting. As we have seen, such choices can leave utterly innocuous truths of quantified modal logic unprovable. Thus his result casts no metaphysical doubt on QCBF and NEI. The new Abstraction rule is both formally correct and informally plausible: (*x* is such that … it …) is equivalent to (*x* exists and … *x* …). QCBF and NEI are straightforwardly valid in quantified modal logic.

Could Stalnaker reply that the invalidity of QCBF and NEI on his deviant counterpart semantics shows that Free Abstraction is metaphysically contentious? Indeed, the derivability of QCBF and NEI from Free Abstraction implies that the latter is invalid on Stalnaker's deviant counterpart semantics. For example, that semantics allows an object *d* not in the domain of a world *w* to satisfy $\hat{y}\Diamond Gyx$ with respect to *w* (because the counterpart of *d* in *w* has a counterpart in a world *w** accessible from *w* that belongs to the extension of *G* in *w**) even though *d* does not satisfy $\Diamond Gx$ with respect to *w** (because no counterpart of *d* in any world belongs to the extension of *G* in that world). However, for Stalnaker to object to Free Abstraction on that basis would be to argue in a circle. For his original reason for taking the counterpart semantics seriously was precisely that it validated all the principles of first-order non-modal logic and of propositional modal logic. But now that turns out to be so only on an unjustifiedly narrow view of the principles of first-order non-modal logics. For any theorem of first-order modal logic, one can cook up an unintended formal semantics on which it is invalid, and a metaphysical fairy tale to add colour to the semantics. Such a methodology is a recipe for shallowness and confusion. But it was not Stalnaker's methodology. His original argument laudably relied on the constraints of first-order non-modal logic and propositional modal logic; it failed only because it made unjustified claims about the limits of those constraints. That is no reason to throw the constraints away altogether, as the direct appeal to the counterpart semantics would do.[7]

Of course, counterpart semantics in its own right still finds some defenders, who are willing to put up with its ugly complications for the sake of the freedom that it delivers from constraints to which they object on metaphysical grounds. But we have already seen that the contentiousness amongst metaphysicians of a principle is compatible with its being a valid law of logic. Stalnaker's argument promised to do something more interesting: to introduce an objective procedure for determining how far logic constrains modal metaphysics. The trouble is that the putatively objective procedure is over-sensitive to the way in which a given logic is axiomatized. Metaphysicians who start from the idea that counterpart theory must be logically coherent and then tailor their logic to suit are not even attempting to do the more interesting thing.[8]

3. Free Abstraction is not the only means by which one can argue for QCBF and NEI within something like Stalnaker's framework. In particular, suppose that we are granted the stronger identity principle Identity*. Of course, Stalnaker regards Identity* as invalid in a modal context, because he allows non-rigid individual constants. However, that decision has no obvious bearing on the logical status of QCBF. Individual constants occur neither in the schema itself nor in the instance of it that Stalnaker shows to be invalid on the counterpart semantics (p. 153). One could instead declare some or all individual constants rigid without undermining the rationale for other aspects of Stalnaker's combined system. Indeed, since he axiomatizes the logic of quantification using only closed formulas (p. 147), individual constants are pressed into playing a double role, as both descriptive terms and the analogue of free variables ('arbitrary names') in proofs. In effect, the rule of Universal Generalization exploits them in the latter capacity. By contrast, their non-rigidity is justified only by their descriptive content. For since Stalnaker's official semantics treats variables as rigid, the role of closed terms in acting like free variables in the corresponding logic of quantification is best served by rigid designators. To mark the difference between these contrasting functions for individual constants, we could divide them into two categories: rigid arbitrary names and possibly non-rigid descriptive terms. Identity* would then be valid for the arbitrary

names but not for the descriptive terms. Moreover, as already noted, Identity* has only valid instances on Stalnaker's semantics for first-order non-modal logic, so it is in any case unclear with what right he rejects Identity* as a constraint on the semantics for first-order modal logic, given his methodology elsewhere in the paper.

Consider an extension of Stalnaker's axiomatization by Identity* for arbitrary names. Thus for any formula φ in which only the variable *x* is free, and distinct arbitrary names *s* and *t* that do not occur in φ, we have the theorem:

(13) $\sim\varphi^{s}/_{x} \rightarrow (\varphi^{t}/_{x} \rightarrow \sim s{=}t)$

Then Universal Generalization gives:

(14) $\sim\varphi^{s}/_{x} \rightarrow \forall \hat{x}(\varphi \rightarrow \sim s{=}x)$

Hence by Quantification:

(15) $\sim\varphi^{s}/_{x} \rightarrow (\forall \hat{x}\varphi \rightarrow \forall \hat{x}{\sim} s{=}x)$

But in Stalnaker's system we can already prove:

(16) $Es \rightarrow \sim\forall \hat{x}{\sim} s{=}x$

Propositional reasoning from (15) and (16) yields:

(17) $\forall \hat{x}\varphi \rightarrow (Es \rightarrow \varphi^{s}/_{x})$

Hence Necessitation and the K schema give:

(18) $\Box\forall \hat{x}\varphi \rightarrow \Box(Es \rightarrow \varphi^{s}/_{x})$

But from (18) Universal Generalization yields QCBF. As before, NEI can then be derived as a corollary of QCBF.

Thus all that blocks this alternative derivation of QCBF and NEI are Stalnaker's decisions to allow non-rigid individual constants and have no separate category of rigid arbitrary names (to make the logic as free as possible, one could still permit constants in the latter category to be world-independently empty). Those decisions are in no way compelled by first-order non-modal logic. This reinforces the conclusion that the underivability of QCBF and NEI in Stalnaker's axiomatization is an artefact of its detailed workings and does not undermine their status as logical truths.

4. We turn to the case of ND, the necessity of distinctness. As Stalnaker notes, it is underivable even when QCBF is added to his axiomatization. More generally, it is underivable when the Free Abstraction rule replaces his Abstraction schema. For Free Abstraction preserves validity on Stalnaker's other deviant semantics, on which '=' is interpreted to mean indiscernibility but everything else is standard, while ND is invalid. ND would remain underivable even if we were to add the schema Identity* to the axiomatization, because it remains invalid even when we validate Identity* by requiring individual constants to be rigid designators on that deviant semantics. Is ND a principle that really cannot be settled by the combination of first-order non-modal logic with identity and propositional modal logic?

Stalnaker himself notes a reason for qualifying his claim that ND cannot be so settled (p. 156). The deviant semantics equates the extension of '=' at a world with the set of ordered pairs of members of the domain of that world that are mutually indiscernible, in the sense that they are in the extension of the same one-place predicates (open or closed, simple or complex, but not containing '=' itself) at that world. Stalnaker's counter-model to ND on the indiscernibility semantics requires two individuals *a* and *b* in the domain of a world *w* that are discernible in *w* but indiscernible in some world *w** accessible from *w*. This can happen only if *w* is not accessible from *w**, for otherwise the discernibility of *a* and *b* in *w* makes them discernible in *w** too by modal predicates: specifically, *a* but not *b* is in the extension of $\hat{x}\Box(x=x \rightarrow x=y)$ at *w** on an assignment that maps the variable *y* to *a*. Thus the counter-model depends on an underlying propositional modal logic in which non-symmetric accessibility relations are permitted. Stalnaker opts for the weakest normal propositional modal logic K, which imposes no constraints whatever on accessibility (since it can be non-reflexive, necessity does not even entail truth in K). But many philosophers take the propositional logic of metaphysical modality to be S5, the logic of the class of models in which every world is accessible from every world, and also of the wider class of models in which accessibility is an equivalence relation. In such models, accessibility is symmetric. The same holds of weaker propositional modal logics with the Brouwerian schema:

B $\vdash \Diamond\Box\varphi \rightarrow \varphi$

This schema (which is of course derivable in S5) corresponds to the symmetry of accessibility but not to its reflexivity or transitivity. Indeed, even axioms that are weaker than B in the presence of the reflexivity axiom T ($\vdash \Box\varphi \rightarrow \varphi$) will suffice to rule out Stalnaker's counter-model to ND, for it has no counter-model on the deviant interpretation of '=' in which, whenever *w** is accessible from *w*, *w* can be reached from *w** in finitely many steps of accessibility.

Syntactically, we can derive ND once we strengthen Stalnaker's axiomatization by Free Abstraction on the first-order non-modal side and by the B schema on the propositional modal side. For we can derive NEI using the Abstraction rule, from which it is routine to derive:

(19) $\forall \hat{x} \forall \hat{y}(\Diamond x=y \rightarrow \Diamond\Box(Ex \rightarrow x=y))$

But from B we can also derive:

(20) $\forall \hat{x} \forall \hat{y}(\Diamond\Box(Ex \rightarrow x=y) \rightarrow (Ex \rightarrow x=y))$

From (19) and (20) we have:

(21) $\forall \hat{x} \forall \hat{y}(\Diamond x=y \rightarrow (Ex \rightarrow x=y))$

Using (12) above (everything exists), we easily obtain ND from (21) by contraposition.[9] Thus the underivability of ND in Stalnaker's axiomatization depends on the separate weaknesses of his abstraction principle and his propositional modal logic.

Suppose, however, that we are working with an interpretation of $\Box$ for which we do not wish to impose any constraints on the accessibility relation. Nevertheless, Stalnaker's semantic framework enables us meaningfully to expand the language by introducing a second necessity-like operator $\blacksquare$ by the semantic clause:

If φ is a sentence, $v^s{}_w(\blacksquare\varphi) = 1$ if $v^s{}_u(\varphi) = 1$ for all $w \varepsilon W$; otherwise $v^s{}_w(\blacksquare\varphi) = 0$.

The dual operator $\blacklozenge$ is of course defined as $\neg\blacksquare\neg$. The semantic clause obviously validates the principles of S5 for $\blacksquare$:

K schema for $\blacksquare$ $\vdash \blacksquare(\varphi \rightarrow \psi) \rightarrow (\blacksquare\varphi \rightarrow \blacksquare\psi)$

T schema for $\blacksquare$ $\vdash \blacksquare\varphi \rightarrow \varphi$

E schema for ■ $\vdash \blacklozenge\varphi \rightarrow \blacksquare\blacklozenge\varphi$

Necessitation for ■ If $\vdash \varphi$ then $\vdash \blacksquare\varphi$

It is well known that those principles for ■ enable one to derive:

B schema for ■ $\blacklozenge\blacksquare\varphi \rightarrow \varphi$

4 schema for ■ $\blacksquare\varphi \rightarrow \blacksquare\blacksquare\varphi$

We also add a valid schema linking the two box operators:

Bridge schema $\vdash \blacksquare\varphi \rightarrow \Box\varphi$

For if φ is true in all worlds whatsoever, then *a fortiori* it is true in all accessible worlds. Of course, the Bridge schema is essentially a bimodal principle: we obviously cannot hope to derive it from the separate logics of ■ and □. But it is not clear that Stalnaker can object to that. For, analogously, he doers not attempt to derive his non-modal logic of quantification with identity from separate non-modal logics of quantification without identity and identity without quantification.

Informally, we can think of ■ and ♦ as, respectively, metaphysical necessity and metaphysical possibility, and of □ and ◊ as restricted modalities of some sort; Stalnaker himself sometimes speaks of metaphysical necessity in just such terms (2003: pp. 202-3). Thus we have a bimodal logic. In the presence of Necessitation for $\blacksquare$ and the bridge schema, Necessitation for $\Box$ is of course redundant, and so may be dropped from the axiomatization, although we still need the K schema for $\Box$ in addition to that for $\blacksquare$. We can now employ a strategy of first using the S5 principles to prove a result for $\blacksquare$ and then using the Bridge schema to deduce a corresponding result for $\Box$.

As an instance of the strategy, we start by proving ND in the system that results from replacing Stalnaker's Abstraction schema by Free Abstraction in his combined axiomatization and adding the K, T and E schemas and Necessitation for $\blacksquare$ and the Bridge schema. In the way already sketched, we use the B schema for $\blacksquare$ to derive:

ND for $\blacksquare$ $\qquad$ $\forall \hat{x} \forall \hat{y}(\sim x=y \rightarrow \blacksquare\sim x=y)$

From the Bridge schema and Universal Generalization we prove:

(22) $\quad \forall \hat{x} \forall \hat{y}(\blacksquare\sim x=y \rightarrow \Box\sim x=y)$

ND for $\blacksquare$ and (22) yield ND for $\Box$ by non-modal reasoning.

We can generalize the result to the following, for any natural numbers j and k, where $\Box^j$ is a sequence of j occurrences of $\Box$:

ND+: $\qquad$ $\Box^j \forall \hat{x} \forall \hat{y}(\sim x=y \rightarrow \Box^k\sim x=y)$

To see this, note that we can prove the following for any j:[10]

Extended Bridge schema: $\blacksquare\varphi \rightarrow \square^j\varphi$

Consequently, we can strengthen (22) to:

(23) $\forall \hat{x} \forall \hat{y}(\blacksquare{\sim}x{=}y \rightarrow \square^j{\sim}x{=}y)$

Just as ND for ■ and (22) yield ND for □, so ND for ■ and (26) yield:

(24) $\forall \hat{x} \forall \hat{y}({\sim}x{=}y \rightarrow \square^j{\sim}x{=}y)$

To complete the argument for ND+, subject (24) to Necessitation for ■ and then apply the Extended Bridge schema.[11] Many similar results are derivable by such means.

According to Stalnaker, 'the necessity (or essentiality) of identity is more central to the logic of identity than the necessity of distinctness' (p. 161). That may well be so in the sense that natural systems of first-order modal logic with identity require significantly richer resources to prove the necessity of distinctness than they require to prove the necessity (or essentiality) of identity: more axioms or rules of inference and, in some cases, greater expressive powers. But proofs of the necessity of distinctness and strengthenings of it such as ND+ need not employ principles that derive neither from first-order non-modal logic with identity nor from propositional modal (bimodal) logic. The proofs for □ above used only principles taken from first-order non-modal logic (including Free Abstraction) with identity and the bimodal logic of □ and ■. No distinctively modal principles concerning '=' were assumed. Thus Stalnaker's suggestion that there may be 'in some sense something modal about the concept of identity' is not supported when one examines the issue in a wider range of logical settings.

5. The result of the preceding discussion is that the underivability in Stalnaker's axiomatization of the principles QCBF, NEI and ND, which are valid on his semantics, casts no serious doubt on their status as logical truths. His treatment of individual constants as non-rigid was also queried in passing. The remainder of the paper raises some more radical questions about his system. Nothing in the critique of Stalnaker's argument above depends on what follows.

We may start by reflecting on Abstraction principles. Free Abstraction preserves validity on Stalnaker's semantics. It has the extra complexity of the existential conjunct *Ex*. Is that extra conjunct really wanted? If we delete it, we obtain this simpler rule:

Simple Abstraction rule: If B is the result of replacing some or all occurrences of $\hat{y}\varphi x$ in A by $\varphi^{x}/_{y}$, where $\varphi^{x}/_{y}$ is as in the original Abstraction principle, and $\vdash$ A, then $\vdash$ B

On Stalnaker's first-order non-modal semantics, Simple Abstraction preserves validity (and, indeed, truth in a model). Admittedly, in any model $v^s(\hat{y}\varphi x)$ and $v^s(\varphi^{x}/_{y})$ will differ for some assignments *s* and formulas φ, for Stalnaker treats assignments as partial functions from individual variables to members of the domain; if the domain is empty, there are no such total functions (p. 158). For example, if *s* assigns no value to *x*, then $v^s(\hat{y}{\sim}Eyx) = 0$ (because $<v^s(x)> \varepsilon\ v^s(\hat{y}{\sim}Ey)$ only if v^s is defined on *x*); but $v^s({\sim}Ex) = 1$ (because $v^s(Ex) = 0$). However, for any member *d* of D, $v^{s[d/x]}(\hat{y}\varphi x) = v^{s[d/x]}(\varphi^{x}/_{y})$. Now Stalnaker restricts his logic to closed formulas, and in his language a variable *x* is bound only by the predicate-forming operator $\hat{x}$, which forms a predicate whose semantic value relative to *s* depends only on the semantic values of the formula to which $\hat{x}$ was applied relative to assignments $s[d/x]$ that assign x a member d of D. Consequently, if B is the result of replacing some or all occurrences of $\hat{y}\varphi x$ in a closed formula A by $\varphi^{x}/_{y}$, then $v^{s}(A) = v^{s}(B)$ for all assignments *s*, including those undefined on *x*. For example, $v^{s}(\forall\ \hat{x}{\sim}\hat{y}{\sim}Eyx) = v^{s}(\forall\ \hat{x}{\sim}{\sim}Ex) = 1$. Thus Simple Abstraction preserves validity in the non-modal system.

The corresponding argument fails on Stalnaker's semantics for the first-order modal language. For example, $v^{s}_{w}(\forall\ \hat{x}\Box{\sim}\hat{y}{\sim}Eyx) = 1$ since $v^{s^*}_{w^*}(\hat{y}{\sim}Ey)$ is empty for any s^* and w^*; but $v^{s}_{w}(\forall\ \hat{x}\Box{\sim}{\sim}Ex) = 0$ whenever some member of the domain of *w* is absent from the domain of some world accessible from *w*. In general, although $v^{s}_{w}(\hat{y}\varphi x) = v^{s}_{w}(\varphi^{x}/_{y})$ whenever $s(x) \in D_{w}$, a modal operator may intervene between those formulas and the occurrence of $\hat{x}$ that binds occurrences of the variable *x* in them, so that the truth-values of A and B in Simple Abstraction can be sensitive to differences between $v^{s}_{w}(\hat{y}\varphi x)$ and $v^{s}_{w}(\varphi^{x}/_{y})$ for assignments *s* such that $s(x)$ belongs only to the domains of some worlds other than *w*. But that depends on Stalnaker's decision to relativize domains to worlds. For models in which all worlds have the same domain, the earlier argument for the non-modal case can easily be adapted to show that Simple Abstraction preserves truth. One can combine the propositional modal semantics with the first-order non-modal semantics without relativizing domains to worlds, just as Stalnaker himself deliberately refrains from relativizing assignments of values to variables to worlds.

Simple Abstraction gives a smoother account of the effect of abstraction than Free Abstraction does, since the content of $\hat{y}\varphi x$ is unpacked wholly in terms of the abstracted formula φ and the variable *x*, without the introduction of extraneous elements such as *Ex*. The smoother account *might* be unsatisfactory if the variable had been assigned no value, but, as before, since the logic is confined to closed formulas, in which the variable *x* cannot occur unbound by $\hat{x}$, the relevant assignments all assign it a value. The truth-condition of the sentence φ is equivalent to a condition on the individual assigned to *y*, and the truth-condition of the sentence $\varphi^{x}/_{y}$ is equivalent to the condition on the individual assigned to *x*, irrespective of whether it belongs to the domain of the current world, that it meets the former condition.

Formally, the rationale for Simple Abstraction is strong. But once Stalnaker's system is extended to include the rule, highly controversial theorems are forthcoming. Note first that his Existence schema yields:

(25) $\hat{y}{\sim}Eyt \rightarrow Et$

From (25), Necessitation and then Universal Generalization give:

(26) $\forall \hat{x}\Box(\hat{y}{\sim}Eyx \rightarrow Ex)$

Applying Simple Abstraction to (26), we have:

(27) $\forall \hat{x}\Box({\sim}Ex \rightarrow Ex)$

Since $({\sim}p \rightarrow p) \rightarrow p$ is a truth-functional tautology, by standard reasoning we can derive from (27):

NE $\forall \hat{x}\Box Ex$

Everything has necessary existence (the necessity of existence). Indeed, by applying Necessitation *j* times before Universal Generalization and *i* times afterwards, we can prove:

NE+ $\Box^i \forall \hat{x} \Box^j Ex$

But, it might be thought, NE is quite bad enough already. Is it not obvious that many actually existing things, including ourselves, exist only contingently? If so, NE constitutes a reductio ad absurdum of the system in which it was derived. On that basis, one might reject Simple Abstraction, or perhaps keep it and reject Stalnaker's Existence principle instead.

That reaction would be too quick. What is obvious enough is that many things that exist in space and time, such as ourselves, could have failed to exist in space and time. But we should not assume that existing in space and time is the only way of existing. For example, if pure sets exist, as they arguably do, they presumably do it without existing in space and time. To say that pure sets exist is just to say that there are pure sets. To say that the null set exists is just to say that there is one and only one set with no members. Something in space and time is a counterexample to NE only if it could have failed to exist *at all*; it is insufficient that it could have failed to exist in space and time. Of course, if we had existed without existing in space and time, we would not then have been persons, let alone sets. Rather, we would have been merely possible persons: non-persons that could have been persons. Being a merely possible person is not a way of not existing; it is a way of existing in the only sense of the term of special interest to logic, that is, of being something or other. Elsewhere, I have defended such a conception of modal metaphysics in more detail (Williamson 1990, 1998, 2000a, 2000b, 2002).

One can validate Simple Abstraction by modifying the semantics to have a single domain D, unrelativized to a world. For the general plan of combining the semantics of the propositional modal language with the semantics of the first-order non-modal language simply leaves it open whether the domain of the latter should be relativized to the worlds of the former or not.

Given Simple Abstraction, one can derive CBF, the unqualified converse Ibn-Sina-Barcan schema. For Simple Abstraction is equivalent to Free Abstraction in the presence of NE+; the existence conjunct becomes redundant. We have already seen how to derive QCBF by Free Abstraction. We can therefore derive QCBF by Simple Abstraction, and then obtain CBF from QCBF by NE. It is well known that when domains are world-relative, CBF corresponds to the semantic condition that whenever a world *x* is accessible from a world *w*, the domain of *w* is a subset of the domain of *x*. Similarly, consider the Ibn-Sina-Barcan schema itself:

BF $\forall \hat{x} \Box \varphi \rightarrow \Box \forall \hat{x} \varphi$

BF corresponds to the condition that whenever *x* is accessible from *w*, the domain of *x* is a subset of the domain of *w*. Together, the two conditions are equivalent to the constancy of the domain across chains of accessibility; variation in domain between worlds not linked by a chain of accessibility by itself makes no difference to the truth of any sentence. The semantics with a constant domain validates both CBF and BF. However, BF differs from CBF in being underivable even when Stalnaker's system is expanded by Simple Abstraction. For, as already noted, that rule is derivable from Free Abstraction together with NE+. But Stalnaker's system, Free Abstraction and NE+ are all validated by a semantics like Stalnaker's with world-relative domains but subject to the semantic condition corresponding to CBF, whereas BF remains invalid on that semantics.

One way to expand the system to permit the derivation of BF is by adding a modal operator $\Box^{-1}$ whose accessibility relation S is required to be the converse of the accessibility relation R for $\Box$ in all models: wSx if and only if xRw. Naturally, $\Diamond^{-1}$ is $\neg\Box^{-1}\neg$. Thus $\Diamond$ and $\Diamond^{-1}$ are related like past and future tense operators in tense logic.[12] These new modalities make as good sense as $\Box$ and $\Diamond$ do within Stalnaker's semantics. Their interrelationship automatically validates two axiom schemas for their propositional bimodal logic:

Converse1 $\vdash \varphi \rightarrow \Box\Diamond^{-1}\varphi$

Converse2 $\vdash \varphi \rightarrow \Box^{-1}\Diamond\varphi$

Of course, the K schema for $\Box^{-1}$ is also valid, and Necessitation for $\Box^{-1}$ preserves validity. Thus we consider the extension of Stalnaker's system with Converse1, Converse2 and the K schema for $\Box^{-1}$ as additional axiom schemas, Necessitation for $\Box^{-1}$ as an additional rule and Simple Abstraction in place of his Abstraction schema. We can now derive CBF for $\Box^{-1}$ just as we derived it for $\Box$, using the extra power of Simple

Abstraction in the expanded language. To derive BF for $\Box$ we proceed thus. As already noted, $\forall F \to (Et \to Ft)$ is derivable in Stalnaker's system, so it is derivable in this extension. By substituting $\hat{x}\Box\varphi$ for F in the proof and then using Universal Generalization, the proof that everything exists and Simple Abstraction we prove:

(28) $\forall \hat{x}(\forall \hat{x}\Box\varphi \to \Box\varphi)$

Applying Necessitation for $\Box^{-1}$ gives:

(29) $\Box^{-1}\forall \hat{x}(\forall \hat{x}\Box\varphi \to \Box\varphi)$

As noted, we can derive CBF for $\Box^{-1}$, so (29) yields:

(30) $\forall \hat{x}\Box^{-1}(\forall \hat{x}\Box\varphi \to \Box\varphi)$

By standard manipulations on (30) we obtain:

(31) $\forall \hat{x}\Diamond^{-1}\forall \hat{x}\Box\varphi \to \forall \hat{x}\Diamond^{-1}\Box\varphi$

We can remove the outer quantifier in the antecedent by Universal Generalization and the modalities in the consequent by Converse2 (as contraposed), so we have:

(32) $\Diamond^{-1}\forall \hat{x}\Box\varphi \to \forall \hat{x}\varphi$

By Necessitation and the K schema for $\Box$ we have:

(33) $\Box\Diamond^{-1}\forall \hat{x}\Box\varphi \to \Box\forall \hat{x}\varphi$

Finally, Converse1 allows us to remove the outer two modalities in the antecedent, thereby obtaining BF for $\Box$. We obtain BF for $\Box^{-1}$ in exactly parallel fashion. The proofs involve only principles from the propositional bimodal logic of $\Box$ and $\Box^{-1}$ and first-order non-modal logic; they do not require any extra assumptions about the interaction of modal operators with quantification and identity.

In the special case in which the accessibility relation for □ is assumed to be symmetric, $\Box^{-1}$ reduces to □, Converse1 and Converse2 both reduce to the B schema, and the proof of BF reduces to the usual proof for monomodal systems with BF and B.

6. How should we decide between variable domains and constant domains versions of possible worlds semantics? Given the overall approach, both versions seem to confer truth-conditions on sentences of the formal object-language in a coherent way. Of course, the model theory by itself does not completely fix the meaning of those sentences. For example, it is compatible with many different readings of □, depending on how the accessibility relation R is to be understood. In fixing the intended interpretation of the formal language, we also need to decide how any domains are to be understood: that may enable us to decide whether they should be variable or constant.

However, a prior issue arises. For whether domains are variable or constant, their role is to restrict the universal quantifier of the meta-language that is used in the semantic clause for $\forall$ to state its contribution to the truth-conditions of sentences of the object-language in which it occurs. On this view, for $\forall F$ to be true at a world, it suffices that the predicate *F* applies at that world to all members of the relevant domain, even if it fails to apply to some things outside the domain. In effect, $\forall$ is being interpreted as a restricted quantifier. The Ibn-Sina-Barcan schemaa and its converse can undoubtedly fail on a restricted reading of the quantifier, even when the modality is read as metaphysical. For example, call a number *popular* if and only if it is many people's favourite number. 'Every popular natural number is necessarily prime' does not entail 'Necessarily every popular natural number is prime', since the only actually popular ones may be prime, and therefore necessarily prime, even though some composite natural numbers could have been popular. Conversely, the obviously true ‘Necessarily every popular natural number is popular’ does not entail the obviously false ‘Every popular natural number is necessarily popular’. But for both metaphysics and logic the most interesting reading of $\forall$ is as totally unrestricted, ranging over everything whatsoever. The restricted readings can then be recovered as complex quantifiers constructed out of the unrestricted one and a restricting condition; the basic reading is the unrestricted one. On that unrestricted reading, $\forall F$ is true at a world if and only F applies to everything whatsoever at that world; domains do not come into it. Such a domain-free semantics automatically validates BF, CBF and NE (Williamson 2000a). On this view, the existence and identity of individuals (being something and being the same thing) are entirely non-contingent matters.

It might be objected that the appearance of domains in the standard semantics is no real restriction on the quantifiers, because in a given world there is nothing except what exists there to quantify over, and on the intended interpretation the domain of a world by definition contains whatever exists there. But that objection fails to take the possible worlds semantics seriously. It treats the semantic clause for $\forall$ as though it were a misleading approximate translation of a more fundamental clause in which an unrestricted universal quantifier of a more fundamental meta-language occurred within the scope of a modal operator. For present purposes, we take the standard meta-language for the semantics of quantified modal logic seriously, as Stalnaker does; the discussion of a more homophonic form of semantics for quantified modal logic must be postponed for another occasion. Once the standard semantic clauses for $\forall$ are taken at face value, the restriction to the domain of a world must be understood as genuinely imposing a restricted reading on the object-language quantifiers; so the objection fails.

The logic of absolutely unrestricted quantification is highly controversial in at least two ways. First, it is controversial whether absolutely unrestricted quantifiers even make sense, especially given the threat of set-theoretic paradox (see Cartwright 1994 and Williamson 2003 for arguments that they do). Second, even granted that they do make sense, it is controversial whether they should count as logical constants (see Williamson 2000a and Rayo and Williamson 2003 for an account on which they do). If they do so count, then the formalization of the claim that there are at least *n* things is a logical truth for each natural number *n*, in the Tarskian sense that it is true under all interpretations of the non-logical vocabulary, for it is true (after all, the formula itself contains at least *n* variables) and contains no non-logical vocabulary; it cannot be invalidated by a restricted domain of fewer than *n* things because its semantics involves no restriction to a domain. Thus logic has substantive existential commitments, just as it has on Frege's logicist conception. Those controversies about absolutely unrestricted quantification cannot be resolved here.

The case of absolutely unrestricted quantification illustrates in an extreme way the potential of logic for metaphysical controversy. Less extreme illustrations are provided by the formulas that are invalid on Stalnaker's deviant counterpart semantics but nevertheless valid on the intended semantics, and provable in an appropriate axiomatization. But no other science is bound by the constraint that its laws must be uncontroversial. Why should logic be any different? Of course, when logic is controversial it cannot easily act as arbiter of fair play in extra-logical disputes; but we have already observed that to define logic by that role would be to condemn it to extreme unsystematicity. There is no science of fair play. A better proposal is that the primary function of logic is to investigate logical consequence, that is, truth-preservation from premises to conclusion however the argument is interpreted, given its logical form. After all, we need *some* science to investigate that, and logic is by far the best candidate. Logic is not defined by its dialectical or epistemological status. But if we carry out the investigation well enough, the generality of its results can still carry an authority sufficient for our needs.

Notes

* An earlier version of this paper was presented to a meeting of the Belgian Society for Logic and Philosophy of Science and in classes at Oxford; thanks to the audiences for useful discussion. I thank Agustín Rayo and Yannis Stephanou for detailed and thoughtful written comments. Above all, I thank Bob Stalnaker himself, not just for his answers to some technical questions but for all that I have learnt from him about the geometry of conceptual space.

1 All pages references are to Stalnaker 1994 as reprinted in Stalnaker 2003 unless otherwise specified.

2 Only the left-to-right direction of Redundancy figures in the original 1994 axiomatization, supplemented with two extra axiom schemas: the universal instantiation principle $\vdash \forall \hat{x}(\forall F \rightarrow Fx)$ and the permutation principle $\vdash \forall \hat{x} \forall \hat{y}\varphi \rightarrow \forall \hat{y} \forall \hat{x}\varphi$. Stalnaker informs me (p.c.) that the change in the 2002 version was just to improve the economy of the axiomatization: universal instantiation becomes derivable once the biconditional Redundancy principle is used and permutation (which Kit Fine (1983) had shown to be independent of Kripke's (1963) axiomatization of free quantified modal logic) is derivable in Stalnaker's system with the help of his principles about identity.

3 Obviously, it makes no difference if $x=t_i$ in Existence and $s=t_i$ in Identity are replaced by $t_i=x$ and $t_i=s$ respectively. However, if the latter of these replacements is

made without the former, then the symmetry principle $s=t \rightarrow t=s$ is underivable, because a deviant semantics for identity on which $s=t$ is true if and only if either s and t denote the same thing or s fails to denote validates the axiomatization while invalidating symmetry (analogously if the former replacement is made without the latter). The problem arose for an earlier, unpublished version of the system. Delia Graff and Gabriel Uzquiano prompted the correction; Stalnaker supplied the independence argument. Thanks to all three (p.cs.) for this information.

4 Stalnaker defines E as $\exists \hat{y} x=y$ (p. 151). Since the latter is an open sentence, not a predicate, as he makes the grammatical distinctions, he intends $\hat{x} \exists \hat{y} x=y$. But Stalnaker notes (p.c.) that it is essential that the Existence principle be stated as it is, rather than with $\hat{y}(\exists \hat{x} x=y)t_i$ in the consequent, since otherwise $\hat{y}(\exists \hat{x} x=y)s \leftrightarrow \exists \hat{y} y=s$ will be underivable. One can prove this by considering a deviant semantics on which all predications and universal quantifications count as true (adding the symmetry principle from the previous footnote does not help).

5 I am grateful to Zia Movahed for the information that the first known discussion of the Barcan and converse Barcan principles was by Ibn Sina (Avicenna, 980-1037); see Movahed 2004. Of course the credit for independently rediscovering and initiating modern discussion of them remains with Ruth Barcan Marcus. I follow Movahed's obvious proposal for renaming. For clarity, I continue to use Stalnaker's acronyms.

6 Proof: Consider a deviant semantics in which worlds are divided into *sensible* worlds, in which $\forall$ is interpreted in the usual way as 'all', and *silly* worlds, in which $\forall$ is interpreted as 'not all'; the semantics is exactly like Stalnaker's in all other respects, except that validity is defined as truth in all sensible worlds in all models. On the deviant semantics, all instances of Cheap K are true in every world, sensible or silly, because their truth does not depend on the specific interpretation of $\forall$; *a fortiori*, all instances of Cheap K are valid. Since all instances of Propositional Logic are also instances of Cheap K, all instances of Propositional Logic are valid. Similarly, Modus Ponens preserves validity, because in every world it preserves truth. All instances of Abstraction, Quantification, Redundancy, Existence and Identity are valid, because true in every sensible world, since their truth at a world depends only on the interpretation of the non-modal vocabulary, which is standard in every sensible world. Similarly, Universal Generalization preserves validity. Thus the axiomatization is sound on the deviant semantics. The formula $\forall \hat{x}(Fx \rightarrow Fx)$ is derivable in the usual way; it is valid because it is true in all sensible worlds, even though it is false in all silly worlds. But since a silly world can be accessible from sensible ones, (3) is false in some sensible worlds in some models, and so is invalid on the deviant semantics. Thus (3) is underivable on this axiomatization.

7 The basic objection was formulated in Williamson 1996b.

8 Some defenders of counterpart theory (I exclude Stalnaker) think that they are not really rejecting classical logic (for example, the classical logic of identity) because

sentences of quantified modal logic do not really have the logical form that they superficially appear to have, but rather one given by a counterpart-theoretic translation. However, such claims must then be assessed by the standard methods for assessing claims about logical form in semantics. It is doubtful that they can withstand such assessment. If defenders of counterpart theory drop the claims about logical form and protest that it is just a theory about the metaphysical truthmakers for sentences of quantified modal logic, then they cannot reconcile it with classical logic in the way just envisaged. Unclarity on this point has made counterpart theory look more defensible than it really is. For a more specific critique along related lines see Fara and Williamson 2005.

9 The main idea of the proof is from Prior 1955: 206-7.

10 Proof: By mathematical induction on j. For $j = 0$, the Extended Bridge schema reduces to the T schema. Suppose that the Extended Bridge schema holds for j. By Necessitation and the K schema for ■ we prove $■■\varphi \rightarrow ■\Box^j\varphi$. Thence the 4 schema for ■ yields $■\varphi \rightarrow ■\Box^j\varphi$. An instance of the original Bridge schema is $■\Box^j\varphi \rightarrow \Box^{j+1}\varphi$. Together, these yield the Extended Bridge schema for j+1.

11 Somewhat similar results are established in Williamson 1996a using the logic of an 'actually' operator rather than ■. Stalnaker (2003: 159-61) comments on those results in a postscript added in 2002 to the reprinting of Stalnaker 1994. He objects that ND+ remains unprovable (although valid) even when the QCBF schema and a complete logic for 'actually' are added to his axiomatization, although ND itself is provable. The present

result shows that to be a specific feature of the logic of ‘actually’, not a more general phenomenon. For discussion of related issues see Karmo 1983 and Humberstone 1983.

12 See again Karmo 1983 and Humberstone 1983.

References

Cartwright, R. 1994. ‘Speaking of everything’, *Noûs* **28**: 1-20.

Fara, M., and Williamson, T. 2005. ‘Counterparts and actuality’, *Mind* **114**: 1-30.

Fine, K. 1983. ‘The permutation principle in quantificational logic’, *Journal of Philosophical Logic* **12**: 33-37.

Hughes, G.E., and Cresswell, M.J. 1996. *A New Introduction to Modal Logic*. London: Routledge.

Humberstone, I.L. 1983. ‘Karmo on contingent non-identity’, *Australasian Journal of Philosophy* **61**: 188-91.

Karmo, T. 1983. ‘Contingent non-identity’, *Australasian Journal of Philosophy* **61**: 185 -7.

Kripke, S. 1963. ‘Semantical considerations on modal logic’, *Acta Philosophica Fennica* **16**: 83-94.

Movahed, Z. 2004. ‘Ibn-Sina’s anticipation of Buridan and Barcan formulas’, to appear in A. Enayat, I. Kalantari and M. Moniri (eds.), *Proceedings of the Workshop and Conference on Logic, Algebra and Arithmetic, Tehran 2003* (in the Lecture Notes in Logic series of the Association for Symbolic Logic). Natick, Mass.: A.K. Peters.

Prior, A.N. 1955. *Formal Logic*. Oxford: Clarendon Press.

Quine, W.V. 1953. ‘Three grades of modal involvement’, in *Proceedings of the XIth International Congress of Philosophy*, Brussels, 1953, vol. 14. Amsterdam: North-Holland. Reprinted in Quine 1966.

Quine, W.V. 1966. *The Ways of Paradox and Other Essays.* New York: Random House.

Rayo, A., and Williamson, T. 2003. 'A completeness theorem for unrestricted first-order languages', in J.C Beall (ed.), *Liars and Heaps: New Essays on Paradox*. Oxford: Clarendon Press.

Sinnott-Armstrong, W., Raffman, D., and Asher, N. (eds.) 1994. *Modality, Morality and Belief: Essays in honor of Ruth Barcan Marcus*. Cambridge: Cambridge University Press.

Stalnaker, R.C. 1994. 'The interaction of modality with quantification and identity', in Sinnott-Armstrong, Raffman and Asher 1994. Reprinted with revisions and a postscript in Stalnaker 2003, to which page numbers refer.

Stalnaker, R.C. 2003. *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*. Oxford: Clarendon Press.

Williamson, T. 1990. 'Necessary identity and necessary existence', in R. Haller and J. Brandl (eds.), *Wittgenstein — Towards a Re-Evaluation: Proceedings of the 14th International Wittgenstein-Symposium*, vol. I. Vienna: Holder-Pichler-Tempsky.

Williamson, T. 1996a. 'The necessity and determinacy of distinctness', in S. Lovibond and S. Williams (eds.), *Essays for David Wiggins: Identity, Truth and Value*. Oxford: Blackwell.

Williamson, T. 1996b. Review of Sinnott-Armstrong, Raffman and Asher 1994. *Philosophy* **71**: 167-72.

Williamson, T. 1998. 'Bare possibilia', *Erkenntnis* **48**: 257-73.

Williamson, T. 2000a. 'Existence and contingency', *Aristotelian Society* sup. vol **100**: 181-203.

Williamson, T. 2000b. 'The necessary framework of objects', *Topoi* **19**: 201-8.

Williamson, T. 2002. 'Necessary existents', in A. O'Hear (ed.), *Logic, Thought and Language*. Cambridge: Cambridge University Press.

Williamson, T. 2003. 'Everything', *Philosophical Perspectives* **17**: 415-65.