The representational limits of possible worlds semantics

Jones, Nicholas

Document Version
Peer reviewed version

The representational limits of possible worlds semantics[*]

Nicholas K. Jones

This is a draft of a paper forthcoming in *Philosophical Studies*; please cite the final version.

Abstract

This paper evaluates Stalnaker's recent attempt to outline a realist interpretation of possible worlds semantics that lacks substantive metaphysical commitments. The limitations of his approach are used to draw some more general lessons about the non-representational artefacts of formal representations. Three key conclusions are drawn. (1) Stalnaker's account of possible worlds semantics' non-representational artefacts does not cohere with his metaphysics of modality. (2) Invariance-based analyses of non-representational artefacts cannot capture a certain kind of artefact. (3) Stalnaker must treat instrumentally those aspects of possible worlds formalism governing the interaction between quantification and modality, under any analysis whatsoever of non-representational artefacts.

1   Introduction

Does possible worlds semantics harbour substantive metaphysical commitments? The short answer is: no; for a purely instrumentalist interpretation of the semantics is available. On this interpretation, the semantics provides only a mathematical tool for investigating derivability in the various deductive systems relative to which versions of it are known to be sound and complete. Instrumentalist possible worlds semantics itself — as opposed to some particular variant thereof — brings only those metaphysical commitments already implicit in every such deductive system. Those commitments are clearly minimal.

A more interesting question concerns only realist interpretations of possible worlds semantics. These interpretations take the formalism to capture the underlying truth-conditional structure of modal discourse. Does a realist interpretation of possible worlds semantics harbour substantive metaphysical commitments? It is natural to think so.

The semantics makes 'There could have been something that doesn't actually exist' true just in case there's a world, and there's an object in the domain of that world, and that object doesn't actually exist. Yet if there is such a thing, it surely must actually exist; for my actual existential quantifier just ranged over it. The most straightforward realist interpretation of the semantics thus appears committed to the actual existence of everything there could possibly be, and hence to the impossibility of there being things that don't actually exist. That is a deeply non-trivial metaphysical commitment.

Robert Stalnaker's recent book *Mere Possibilities* is perhaps the most systematic and sophisticated response yet to this kind of problem. He describes his project thus:

> How, on an actualist interpretation of possible worlds as ways a world might be, is one to account for the possibility that there be individuals other than those that actually exist? That is the main focus of this book. (Stalnaker, 2012, pix)

> There are some who want to take modality seriously, and seek a theoretical account of modal discourse, but who think we cannot take possible-worlds semantics, as an account of modality, seriously, without making extravagant metaphysical commitments....But I want to defend the metaphysical innocence not only of modal concepts but also of a theoretical account of them in terms of possible worlds. (Stalnaker, 2012, p2)

> [M]y aim will be to vindicate the possible-worlds theory while making minimal commitments about substantive metaphysical questions, for example, about whether there are things or properties that exist only contingently. (Stalnaker, 2012, p8)

Stalnaker's goal is a realist interpretation of the semantics that's compatible with the possible existence of things that don't actually exist. This paper evaluates Stalnaker's approach. Problems for Stalanaker will be used to draw some more general lessons about formal representation, and especially about the nature of the non-representational artefacts of a formalism.

Stalnaker's approach combines two components. One is a view about the metaphysics of modality. The other is an account of the relationship between this metaphysics and possible worlds formal semantics. Stalnaker carefully disentangles aspects of the formalism that represent aspects of modal reality from those that don't. The result is an account of the representational role of the inhabitants of non-actual world-domains in the semantics that is supposed to both (a) cohere with Stalnaker's metaphysical view, and (b) undermine the argument above for the actual existence of everything there could possibly be. I will argue that Stalnaker's approach is flawed: his characterisation of the representational/non-representational distinction does not cohere with his metaphysical view. The problem is not merely a deficiency in Stalnaker's formulation, but a manifestation of an underlying malaise. Stalnaker's approach rests on an invariance-based characterisation of non-representational artefacts inspired by measurement theory. His problem reflects an intrinsic limitation of that approach. Invariance-based analyses of non-representational artefacts are too coarse-grained to capture a certain kind of artefact: the representational/non-representational distinction is more fine-grained than can be captured by any invariance-

based condition on whatever formal structures one uses to represent one's target system. The problems that Stalnaker faces thus illuminate the nature and limits of formal representation more generally.

Given the inadequacy of invariance-based analyses of non-representational artefacts, a question arises: could an adequate analysis be used to make the possible worlds formalism cohere with Stalnaker's metaphysical view? I will argue that any such approach brings a significant cost. A mismatch between the structures of the formal semantics and modal reality (as Stalnaker conceives it) ensures that Stalnaker cannot accept a realistic interpretation of those aspects of the semantics that bear on the present metaphysical controversy. Those aspects of the formalism governing the interaction between quantification and modality must be treated instrumentally. That is an unsatisfing way of attaining neutrality about existential contingency.

The paper proceeds as follows. §2 introduces the metaphysical controversy on which I will focus: existential contingency. The formal semantics is outlined in §3, and Stalnaker's philosophical interpretation thereof in §4. That interpretation is refined in §5 to accommodate two problems involving existential contingency. The refinement relies on a distinction between representational features of the formalism, and its non-representational artefacts. Stalnaker's invariance-based account of that distinction is introduced in §6. §7 responds to a recent criticism of that account due to Timothy Williamson. §8 uses a different objection to Stalnaker to draw some general lessons about invariance-based analyses of the representational/non-representational distinction. §9 closes by arguing that no alternative to the invariance-based approach can allow Stalnaker to retain a realistic interpretation of those aspects of the formalism that bear on existential contingency. The Appendix considers a variant semantics Stalnaker claims to be fully realistic; I show that it suffers just the same problems.

2   Existential contingency

There are many ways in which possible worlds semantics might potentially be thought to harbour substantive metaphysical commitments. This section introduces the one that will concern me, and which is central to Stalnaker.

By *existential contingency*, I mean the possibility of there being something that doesn't actually exist. This is regimented using an 'Actually' operator @ as:

$$(\exists \text{con}) \qquad\qquad \Diamond \exists x @ \forall y (x \neq y)$$

A modal semantics admits existential contingency just in case it renders ($\exists$con) consistent. Three comments follow.

First comment. I use 'exists' as synonymous with 'is identical to something': x exists iff $\exists y(x = y)$. The quantifier here is unrestricted. Other notions of existence are merely restrictions of the underlying unrestricted quantifier. The next comment justifies the focus on unrestricted quantification.

Second comment. ($\exists$con) can fail to capture an interesting notion of existential contingency. That can happen when its quantifiers are interpreted as restricted quantifiers.

Then (∃con) amounts to:

$$(\exists con^*) \qquad\qquad \Diamond \exists x \big(Fx \wedge @\forall y(Fy \rightarrow x \neq y)\big)$$

Actually, I'm not French. Suppose I could have been French. Then there could have been a French thing (i.e. me) such that, actually, it's distinct from every French thing. So (∃con*) is true when F is interpreted as synonymous with 'is French'. So (∃con) is true when its quantifiers are interpreted as restricted to French things. Existential contingency in this sense is merely classificatory or predicational contingency. A more robust and demanding notion requires contingency in the domain of things available for classification. Robust existential contingency will be my concern henceforth.

How to ensure that (∃con)'s truth requires robust existential contingecy? One straightforward option is to interpret its quantifiers as unrestricted. Others are also available. Suppose the sortal *human* is essential to its bearers in the following sense: necessarily, for any object x, if x is a human, it's impossible for x to exist without being a human. Restricting (∃con)'s quantifiers to humans, its truth requires robust existential contingency.[1] Robust existential contingency in non-humans won't register in the truth of (∃con) on this reading. Equivalence between robust existential contingency and (∃con)'s truth requires an unrestricted reading of the quantifiers. What the example shows is that the intelligibility of robust existential contingency is not dependent on the intelligibility of unrestricted quantification. So we needn't engage with disputes about the intelligibility of unrestricted quantification here. We may assume an unrestricted interpretation of the quantifiers in (∃con) without thereby prejudicing the debate. Unrestricted interpretations of all object-language quantifiers will be presupposed henceforth.

Third comment. An alternative notion of existential contingency concerns the possibility of actual things failing to exist:

$$\exists x \Diamond \forall y(x \neq y)$$

I focus on (∃con) because it's harder to accommodate. If there could be something that doesn't actually exist, the question arises: "Which non-actual thing could there be?" No answer to this question is correct. Given our unrestricted reading of the quantifiers, any putative answer concerns either an actual thing, or no particular thing at all; either way, it's incorrect. No similar difficulty afflicts the alternative notion of existential contingency.

Stalnaker's goal is a philosophically satisfying, realist interpretation of possible worlds formal semantics that can accommodate (robust) existential contingency. This will have two components. (1) A formal semantics that renders (∃con) consistent. (2) A realist (= non-instrumental) philosophical interpretation of the formalism. The next section outlines (1); §§4–5 present (2).

3 Possible worlds model theory

Stalnaker's goal is "a vindication of orthodox possible worlds semantics." (Stalnaker, 2012, p32) Although he does not explicitly state his preferred semantic theory, it is clear from

---

[1] Assumption: what's possibly necessary is actual; $\Diamond \Box A \rightarrow @A$.

context that he intends a version of Kripkean model theory, e.g. (Stalnaker, 2012, §5 of ch2.). This section introduces a version of that theory with as good a claim to orthodoxy as any. It is almost a notational variant of a theory from Stalnaker's earlier work on existential contingency. (Stalnaker, 1994) The only substantive departures are: (a) I omit constant terms and the predicate-abstraction operator for simplicity; (b) I include an 'Actually' operator with the standard recursive clause to allow formulation of (∃con). Although the details are familiar, laying them out now will facilitate discussion later.[2]

Our object-language is an ordinary first-order modal language with identity, the unary operator @, and without constant terms. A *model* is a quadruple $\langle W, D, w_@, V \rangle$. W is a set, the *points* of the model. I call them points rather than worlds to cleanly differentiate the formalism from its philosophical interpretation. D is a *domain function* taking each point $w \in W$ to a set $D(w)$. $w_@$ is a member of W, the model's *privileged point*. V is a *valuation function* from object-language predicates to their values in the model, subject to the constraints:

- For each n-place predicate Φ : $V(\Phi)$ is an n-place intension on the model.
  (i.e. $V(\Phi)$ is a function that takes each $w \in W$ to a set of n-tuples of members of $D(w)$.)

- For all $w \in W$: $V(\text{'='})(w) = \{\langle x, x \rangle : x \in D(w)\}$.

An accessibility relation is omitted for simplicity. The relevant model will often be left tacit.

A model's *outer domain* is the union of the domains of its points: $\cup_{w \in W} D(w)$. A *variable assignment* over a model is a function α from variables v to members of the model's outer domain. For any variable v and object d, $\alpha^{[d/v]}$ is an assignment just like α except at most in that $\alpha^{[d/v]}(v) = d$.

A four-place satisfaction relation ⊩ between models m, points w, assignments α and formulae A — written $m, w, \alpha \Vdash A$ — is recursively defined in the standard Tarskian/Kripkean way. For our purposes, the key clauses are:

- $m, w, \alpha \Vdash \Phi v_1, \ldots, v_n$ iff $\langle \alpha(v_1), \ldots, \alpha(v_n) \rangle \in V(\Phi)(w)$.
  (Φ is any n-place predicate; $v_1, \ldots, v_n$ are any variables.)

- $m, w, \alpha \Vdash \exists v A$ iff, for some $d \in D(w)$: $m, w, \alpha^{[d/v]} \Vdash A$.
  (v is any variable.)

- $m, w, \alpha \Vdash \Diamond A$ iff, for some $w' \in W$: $m, w', \alpha \Vdash A$.

- $m, w, \alpha \Vdash @A$ iff $m, w_@, \alpha \Vdash A$.

Satisfaction is de-relativised to a triadic and then dyadic relation thus:

- $m, w \Vdash A$ iff, for all assignments α : $m, w, \alpha \Vdash A$.

- $m \Vdash A$ iff $m, w_@ \Vdash A$.

---

[2] I will be careful about use/mention only when necessary to avoid misunderstanding. 'A' and 'B' will usually be metalinguistic variables ranging over sentences, though I'll sometimes use them schematically.

Putting these pieces together, the clause for (∃con) is:

- $m \Vdash \Diamond \exists x @ \forall y (y \neq x)$ iff, for some $w \in W, d \in D(w) : d \notin D(w_@)$.
  (i.e. iff, for some $w \in W : D(w) \nsubseteq D(w_@)$.)

A model satisfies (∃con) just in case one of its point-domains contains something not in its privileged point-domain. Call any such thing one of the model's *mere possibilia*. Note that a model's mere possibilia are not merely possible objects; they are ordinary, actually existing objects, in the range of our actual metalinguistic quantifiers. A model's mere possibilia are simply entities that play a certain role in the formal semantics. I reserve 'mere possibilia' for occupants of this theoretical role throughout. A model satisfies (∃con) iff it has mere possibilia.

Under Stalnaker's interpretation of the formalism, satisfaction by a model will represent (sentential) truth. So the interpretation must be able to accommodate models with mere possibilia; otherwise, (∃con)'s truth won't be representable. The next section introduces this interpretation; §5 refines it in response to two problems of existential contingency.

4   Stalnaker's interpretation: first pass

The theory just presented is a form of pure mathematics, without intrinsic import for modal discourse. Points and their domains may be pure sets, as far as the previous section is concerned. Various relations definable within the framework turn out to be coextensive with derivability relations in various deductive systems. We can therefore use the model theory to study those relations; instrumentalist interpretations of the formalism assign it no further role. Insight into the structure of modal discourse requires a different interpretation of the formalism, an assignment of representational content to it. This section outlines the core of Stalnaker's interpretation; the next refines it in response to two problems involving existential contingency.[3]

Properties are ways things could be; they are "to be understood in terms of what it would be for them to be exemplified." (Stalnaker, 2012, p11) A possible world is a certain kind of property of the entire universe. World-properties are maximally specific: their exemplification by the universe decides every proposition. That is, for any world u and proposition that-p, exactly one of the following strict conditionals holds:

- Necessarily, if u is exemplified, p.

- Necessarily, if u is exemplified, not-p.

The range of propositions decided by worlds is determined by theoretical context, and the questions of interest within that context:

> What matters for the applications of possible-worlds semantics is that the possible states of the world be maximal with respect to all questions that are of concern in the application at hand. I prefer to think of the worlds not as the

---

[3] Both sections draw on and elaborate (Stalnaker, 2012, chs1–2).

points in logical space, but as the cells of a relatively fine-grained partition—a partition that makes all of the distinctions we need. (Stalnaker, 2012, p13)

Properties that count as worlds in one theoretical context needn't do so in others, where other distinctions are relevant. Some useful terminology: say that *p at u* when, necessarily, if *u* is exemplified, *p*.

There is a circularity here. Modal notions were used to explain properties: what it *would* be for them to be exemplified. Possible worlds were identified with a certain kind of property. The appropriate kind of property was characterised using necessity. And necessity is truth at all world-properties. Stalnaker embraces the circularity. He does not seek a reductive analysis of modality, worlds, propositions or properties:

> It is not reduction but regimentation that the possible-worlds framework provides—a procedure for representing modal discourse, using primitive modal notions, in a way that helps reveal its structure. (Stalnaker, 2012, p11)

Stalnaker uses models to represent possible worlds, and their relation to modal discourse.

Points w represent possible worlds. The elements of w's domain $D(w)$ represent the things that would exist, were the universe the way represented by w. One-place intensions I represent monadic properties. The elements of I's w-extension $I(w)$ represent the things that would be the way represented by I, were the universe the way represented by w. Similar remarks apply to relations. Each model represents modal reality as containing exactly the worlds its points represent: W represents the extent of modal space. A point represents a world at which it's possible that p iff some point represents a world at which p. The privileged point represents the actual world, the way the universe actually is. So a model represents modal reality as such that p iff its privileged point represents a world at which p.

Only the points, domain function and privileged point are needed for this representation of modal reality. The valuation function is inert. Valuations connect representations of modal reality to the object-language. This enables representation of linguistically expressible truth-evaluable claims, and the relationships between them.

Valuation functions assign representatives of semantic values to object-language predicates. Satisfaction is then taken to represent truth. More precisely, for any model $m = \langle W, D, w_@, V \rangle$, point $w \in W$, sentence A:

- $m, w \Vdash A$ iff m represents A as true at the world represented by w, under the interpretation represented by V, given that modal reality is as $\langle W, D, w_@ \rangle$ represents it as being.

The recursive definition of satisfaction thus becomes a recursive assignment of representations of truth-conditions to sentences. The interpretations of logical and modal operators are built into the definition of satisfaction. Because $w_@$ represents the actual world, $m \Vdash A$ iff m represents A as true.

What is this notion of world-relative sentential truth? Think of it as governed by the schema:

- $\ulcorner A \urcorner$ is true at world u iff, necessarily, if u is exemplified, A.

Instances of the schema mention A on the left and use it on the right. Thus A can be true at u without existing at u. To see why, consider a world u much like actuality, except that no language users ever exist. 'Grass is green' is true at u because: necessarily, if u is exemplified, grass is green. But 'Grass is green' doesn't exist at u because: necessarily, if u is exemplified, 'grass is green' doesn't exist. So 'grass is green' is true at u without existing at u.

One final complication. Although we are using a formal object-language, the target is really (a fragment of) a natural, hence interpreted, language. The formal language represents the natural language. So valuation functions don't represent arbitrary assignments of semantic values to lexical items; they represent the actual intended assignment of truth-conditional content to the target language. Relative to a decision about what a model's points and intensions represent, at most one such assignment is correct, worries about indeterminacy etc. notwithstanding.

5   The first-pass interpretation refined

Existential contingency creates two connected difficulties for the view described above. This section outlines the difficulties and Stalnaker's solution to them, drawing primarily on (Stalnaker, 2012, §3 of ch1). A word of caution, however. Stalnaker mostly considers not existential contingency itself, but some consequences thereof: contingently existing propositions and merely possible distinctions between worlds. Since those consequences are motivated (only?) by existential contingency, a solution to the problems discussed here is required. The view I describe is a reconstruction of what I take to be Stalnaker's view.

Suppose ($\exists$con) is true: there could be something that doesn't actually exist. Which such non-actual thing could there be? Not any of the things there actually are. But on an unrestricted reading of the quantifier, there are no other things: actual unrestricted quantifiers range over all the candidates there are. A singular — alternatively: *de re* or object-involving — fact or proposition exists only if the individual it's about exists.[4] So if it's possible for there to be something that doesn't actually exist, there's no witnessing singular modal fact or truth of the form: possibly, x doesn't actually exist. That is:

No Singular Witness (NSW): If the proposition expressed by ($\exists$con) is true, it lacks a true singular witnessing proposition of the form: *possibly, x doesn't actually exist.*[5]

Equivalently: if the existential proposition *something doesn't actually exist* is possibly true, it lacks a possibly true singular witness *x doesn't actually exist.*

Note that **NSW** is consistent with: were the proposition expressed by the quantified claim $\exists x @ \forall y (x \neq y)$ embedded within ($\exists$con)'s initial modal operator true, a true singular proposition would witness it. As things actually stand, that possibly true quantified proposition lacks possibly true singular witness; for the proposition that would witness it doesn't actually exist. Were there things that don't actually exist, there would be singular propositions about them; but those propositions don't exist because there are no things that don't

---

[4] (Stalnaker, 2012, p42–43) endorses this thesis, arguing that it's required by the actualist thesis that actuality exhausts reality.

[5] I extract commitment to **NSW** from (Stalnaker, 2012, pp13–20, 28).

actually exist. So existential contingency for individuals implies existential contingency for singular propositions.

Two related problems now arise.

First problem. A model represents (∃con) as true iff it satisfies (∃con); iff one of its point-domains contains mere possibilia. A point-domain's elements represent the things that would exist, were the universe the way the point represents. So which individuals do a model's mere possibilia represent? If they represent actual individuals, the model is inaccurate: it misrepresents actual objects as not actually existing. On our unrestricted reading of the quantifier, however, there are no other individuals to represent. So representing (∃con) as true suffices for misrepresentation.

Second problem. Suppose the domain of point w contains mere possibilium d. Then d represents a thing that would exist, were the universe the way w represents, but which doesn't actually exist. Suppose d represents o. Then the model represents (∃con) as having a true singular witnessing proposition: *possibly, o doesn't actually exist*. By **NSW**: the model misrepresents the structure of modal space. Since w, d, and o were arbitrary: representing (∃con) as true suffices for misrepresentation.

The first problem involves misrepresenting actual individuals as non-actual. The second involves misrepresenting the structure of modal space as violating **NSW**. Both problems appear to show that representing (∃con) as true suffices for misrepresentation. If so, then we cannot consistently maintain that both (a) some model accurately represents modal reality and discourse, and (b) (∃con) is true.

Stalnaker's solution to both problems is the same: don't use mere possibilia to represent particular individuals; use them only to represent the existence of something or other that doesn't actually exist (Stalnaker, 2012, e.g. p41).

According to **NSW**, possibly true existentials can lack possibly true singular witnesses. Representations of modal reality should reflect this. Mere possibilia shouldn't *de re* represent particular merely possible individuals; they should represent only the existence of non-actual individuals, without representing any such individuals in particular. So mere possibilia and elements of $D(w_@)$ should have different representational import. Whereas elements of $D(w_@)$ *de re* represent particular things, mere possibilia represent only the (possible) existence of things related in various ways to other things. A particular individual can be used for this representational work without representing a particular individual; it may represent only the existence of some individual or other.

The first problem is resolved because models can represent (∃con) as true without representing any particular thing as not actually existing. The second problem is resolved because models can represent (∃con) as true without representing modal reality as containing true singular witnessing propositions of the form *possibly, x doesn't actually exist*. Satisfaction of (∃con) no longer suffices for inaccuracy. So models can represent (∃con) as true without thereby misrepresenting modal space. This completes the presentation of what I take to be Stalnaker's view. The next section introduces Stalnaker's invariance-based strategy for factoring out non-representational artefacts of the formalism, in light of this approach.

6 Representational equivalence

On Stalnaker's view, there is a structural mismatch between models and modal space. The existence of mere possibilia represents the possible existence of non-actual things.[6] So by **NSW**: the facts represented by the existence of mere possibilia lack true singular witnesses.[7] Yet facts about the existence of mere possibilia always have true singular witnesses. So if singular witnesses for existentials about mere possibilia represent singular witnesses for what those existentials represent, then models that represent ($\exists$con) as true are inaccurate.

The previous section avoids this inaccuracy by taking singular facts about mere possibilia to lack representational import. Although they're singular witnesses for existential facts that represent the possible existence of non-actual things, they don't represent singular witnesses for those represented modal-existential facts. The presence of these singular facts in the model is a non-representational byproduct of the style of representation employed: assignments of objects to variables are used to represent the truth-conditions of sentences featuring quantifiers, irrespective of whether the quantifiers occur within the scope of modal operators. This differentiates mere possibilia from elements of $D(w_@)$; for singular witnesses for existentials about the latter, but not the former, represent singular witnesses for what those existentials represent.

This involves a distinction between features of models that possess representational import, and those that don't. Call the former *representors* and the latter *artefacts*. What does this distinction amount to? How exactly should it be applied in this case? This section outlines Stalnaker's approach to these questions, drawing primarily on (Stalnaker, 2012, §§3–5 of ch2).

Stalnaker describes his approach to artefacts thus:

> One "factors out" the artifacts of the model—separates them from the realistic claims of the theory—by adding to the theory an equivalence relation between...models. Equivalent models are those that differ in artificial ways but that agree in the realistic claims they make. (Stalnaker, 2012, pp33–34)

Here's one way to understand Stalnaker's approach. Say that models x, y are *representationally equivalent* iff they have the same accuracy conditions; that is, iff x represents modal reality as F exactly when y does too (for every F whatsoever). We can think of Stalnaker as endorsing:

Invariance: Representors are the features that cannot vary between representationally equivalent models, the invariants under representational equivalence.

Artefacts are the features that can vary between representationally equivalent models.

---

[6] Recall that mere possibilia are simply actual individuals that occupy a certain formal semantic role: they're elements of one of a model's point-domains not in its privileged point-domain. They are not merely possible objects.

[7] I use fact-talk in a lightweight sense: for there to exist a fact that p is just for it to be that p. An ontology of facts is not what's at issue. Talk of facts is merely an expository convenience.

Models here are the set-theoretic quadruples introduced in §3, which we are using to represent modal reality and discourse about it. **Invariance** may be applied in other theoretical contexts by identifying models with whatever entities are used in those contexts to represent the appropriate target system. An application to the representation of space is discussed in §8.

Invariance can be motivated as follows. Artefacts lack representational import. So varying artefacts alone won't affect representational content, hence won't affect representational equivalence. Conversely, varying representationally significant features will affect representational content, thereby undermining representational equivalence. So the representors are the invariants under representational equivalence; everything else is artefactual. Although both arguments are flawed, I'll grant **Invariance** for now; its problems will emerge in §8.

With **Invariance** in place, we need an account of representational equivalence that meshes with Stalnaker's view. He doesn't consider this explicitly in *Mere Possibilities*. However, remarks in a slightly earlier paper suggest a promising approach:

> Any permutation of the 'possible individuals' that preserved identity and difference, as well as the qualitative character of the individuals, and that mapped all *actual* individuals onto themselves would be an equivalent representation — a representation of the same facts, including the modal facts. The domains of other possible worlds (or those members of the domain that are not actually existing individuals) represent the generic possibility of there being individuals of a certain kind, though if individuals of the kinds that might exist did exist, they would be individuals with modal properties and whatever concrete individuality that actual individuals have. (Stalnaker, 2010, p24)

Stalnaker is discussing representations of merely possible distinctions between worlds. Such distinctions arise when non-actual things exist at a world; they depend upon the existence of those particular non-actual individuals. Stalnaker represents these distinctions using equivalence classes of points that differ only in representationally irrelevant ways. The irrelevant differences are those deriving solely from the identities of the mere possibilia in the relevant domains. It is therefore not unreasonable to apply the account of representational equivalence suggested by this passage to the present problem of representationally irrelevant differences between models, rather than between points within a model.[8] Note, however, that the problems I raise for Stalnaker in §§8–9 are independent of any particular account of representational equivalence.

Recall that a model's outer domain U is the union $\cup_{w \in W} D(w)$ of its point-domains. A permutation of the outer domain is a total function σ from U to itself that satisfies:

- For all $x \in U$, there is a $y \in U$ such that: $\sigma(y) = x$.
  (σ is onto: it covers everything in U.)

- For all $x, y \in U$: $\sigma(x) = \sigma(y)$ iff $x = y$.
  (σ preserves identity and difference in U.)

---

[8] This application draws heavily on (Williamson, 2013, 189–190).

To ensure that (representatives of) of actual individuals are mapped to themselves we require:

- For all $x \in D(w_@) : \sigma(x) = x$.

Given a model $m = \langle W, D, w_@, V \rangle$ and permutation σ of its outer domain, σ is extended to a mapping on domain functions, intensions, and valuations. The extension is governed by:

- For each $w \in W$: $\sigma(D)(w) = \{\sigma(x) : x \in D(w)\}$.

- For each n-place intension I: $\sigma(I)(w) = \{\langle \sigma(x_1), \ldots, \sigma(x_n) \rangle : \langle x_1, \ldots, x_n \rangle \in I(w)\}$.

- For each predicate Φ: $\sigma(V)(\Phi) = \sigma(V(\Phi))$.

Finally, we set $\sigma(m) = \langle W, \sigma(D), w_@, \sigma(V) \rangle$. The clauses above vary point-domains, intensions and valuations in a way that compensates for permuting the outer domain. This ensures that $\sigma(m)$ is isomorphic to m. Hence:

- For any formula A and point $w \in W$: $m, w \Vdash A$ iff $\sigma(m), w \Vdash A$.

Call any function that satisfies these constraints a *σ-function*. The representational equivalence relation $\approx$ is characterised by quantifying over σ-functions:

- For any models x, y : $x \approx y$ iff, for some σ-function $\sigma : y = \sigma(x)$.

A complication arises. The quote above says that permutations generate equivalent representations only if they preserve qualitative character. What does this mean when the objects and intensions permuted are mathematical constructs, devoid of intrinsic qualitative character? Recall that the object-language is a formal counterpart for an interpreted natural language (§4). Its predicates may thus be taken to denote properties independently of the various valuations. This generates a (model-relative, partial) assignment of qualitative characters to intensions.[9] Suppose predicate Φ denotes property φ. Then the qualitative character of an intension I is φ iff $V(\Phi) = I$. When V maps no predicate to I, that intension lacks qualitative character.[10] By the characterisation of σ-functions: $\sigma(I)$ has the same qualitative character as I. Because $x \in I(w)$ iff $\sigma(x) \in \sigma(I)(w)$: σ preserves qualitative character. The argument generalises to relations.

This approach may appear insufficiently general. Isomorphic models with different privileged domains cannot stand in $\approx$. But surely we could use objects from the two domains to represent the same individuals. An example. Suppose one model has privileged domain $\{0, 1\}$, whereas another has $\{2, 3\}$; beyond compensating differences in their valuation functions, the models differ in no other way. Take 0 and 2 to represent Stalnaker; take 1 and 3 to represent Williamson. The models should be representationally equivalent

---

[9] A variant approach employs a separate assignment of qualitative characters to intensions, independently of any language and valuation.

[10] Alternatively: I represents a qualitative character related to those represented by other intensions in the appropriate way.

under this account of what their privileged domains represent. Yet they do not stand in $\approx$. Stalnaker must be assuming a way of assigning representational import that precludes this approach. The simplest suggestion is that elements of $D(w_@)$ represent themselves. I'll follow that suggestion for now, though I'll revisit it shortly.[11]

This account of representational equivalence appears to capture the motivating thought (though I'll cast doubt on that shortly). Elements of $D(w_@)$ represent particular individuals, namely themselves. So the elements of $D(w_@)$ and the predicates they satisfy should be invariant under $\approx$. They are. A model's mere possibilia don't represent particular individuals; each represents the existence of an individual (standing in various relations to other individuals) without representing any particular witness. So the existence of mere possibilia should be invariant under $\approx$,[12] whereas facts dependent on their being the particular individuals they are — rather than just some individuals or other — should vary under $\approx$. Plausibly, that's just how it is: the existence of mere possibilia is invariant under $\approx$, unlike the identities of the occupants of the mere-possibilia-roles.[13] As we'll see in §8, however, this approach falls short in a crucial respect. Before then, it is worth considering an objection that Timothy Williamson has recently levelled against this approach.

7 Williamson

In *Modal logic as metaphysics*, Williamson uses a sophisticated network of logical and metaphysical considerations to argue that all forms of existential contingency are impossible: necessarily, everything necessarily exists. Chapter 4 is an extended discussion of Stalnaker's view, culminating in a criticism of his approach to representational equivalence. Since this is the only published discussion of Stalnaker's approach that I know of, and since Williamson is a leading authority on the metaphysics and logic of modality, it is worth seeing why his argument fails. The next section presents a more powerful objection.

The relevant passage reads:

> Now consider a model m that [satisfies ($\exists$con)]. Thus the domain $D(w)$ of some point w in m is not a subset of the domain $D(w_@)$ of the privileged point $w_@$ of m. Some individual o in $D(w)$ is not in $D(w_@)$....One difference between m and some other models is that o belongs to the domain of the privileged point in the latter but not in m. For [defenders of ($\exists$con)], including Stalnaker, this difference should count as representationally insignificant, otherwise m would represent o as non-actual (equivalently, represent actuality as not containing o), and therefore represent o falsely, since o is ac-

---

[11] Because our object language lacks constant terms, elements of the privileged domain cannot receive representational import from the object language directly, in the manner of intensions.

[12] Better: the existence of mere possibilia and elements of $D(w_@)$ standing in a certain network of relations to one another should be invariant under $\approx$.

[13] Although models with a single mere possibilium are problematic, they threaten only the letter of the view, rather than its spirit. Employing the actuality-fixing isomorphisms of (Williamson, 2013, note 49 on p192) in place of σ-functions in the definition of $\approx$ would resolve the problem. Actuality-fixing isomorphisms are like σ-functions, except allowing the sets of mere possibilia and points to vary (with fixed cardinality).

tual. Therefore, by Stalnaker's criterion of representational significance as applied to m, o should belong to the domain $\sigma(D)(w_@)$ of the privileged point in a model $\sigma(m)$ for some [σ-function] σ. But that is impossible, for if $o \in \sigma(D)(w_@)$ then $o = \sigma(i)$ for some $i \in D(w_@)$; but $\sigma(i) = i$ [by the characterisation of σ-functions], so $o = i$, so $o \in D(w_@)$, contrary to hypothesis. (Williamson, 2013, p191; notation and terminology have been changed slightly to fit with present usage.)[14]

To see what's going on here, we can reconstruct the argument as follows:

Stage One: Suppose for *reductio* that model m accurately represents ($\exists$con) as true. Then m satisfies ($\exists$con). So for some $w \in W$ and $o \in D(w)$: $o \notin D(w_@)$.

Stage Two: Let model m^* be just like m except insofar as $o \in D^*(w_@)$. Elements of the privileged domain represent themselves. So m^* represents o as actually existing.

Stage Three: Suppose for *reductio* that the difference between m and m^* is representationally significant. Those models differ only over o: $o \notin D(w_@)$ but $o \in D^*(w_@)$. [A] So that difference must be representationally significant: m and m^* differ in how they represent o. [B] From Stage Two: m^* represents o as actual. So by [A] and [B]: [C] m represents o as non-actual. Since o is actual, m is inaccurate, contrary to Stage One. By *reductio*: the difference between m and m^* is representationally *in*significant. So m is representationally equivalent to m^*, i.e.: $m \approx m^*$.

Stage Four: From Stage Three: $m \approx m^*$. By definition of $\approx$: $o \in D(w_@)$ iff $o \in D^*(w_@^*)$. By Stage Two: $o \in D^*(w_@^*)$. So $o \in D(w)$, contrary to supposition at Stage One. So by *reductio* and since m was arbitrary: no accurate model represents ($\exists$con) as true.

This argument purports to show that, by Stalnaker's own lights, any model that represents ($\exists$con) as true is inaccurate. However, the key inference at Stage Three is invalid: [A] and [B] do not entail [C].

The only relevant difference between m and m^* is that $o \in D^*(w_@)$ and $o \notin D(w)$. So if m and m^* differ representationally, they must differ in how they represent o. Since m^* represents o as actual: m doesn't represent o as actual. It doesn't follow that m represents o as non-actual; for m might not represent o at all. An argument of the form

- Not: x represents y as F.
  Therefore: x represents y as not-F.

is valid only when x's representational content concerns whether y is F. When x's content is silent about y — or even just silent about whether y is F — the conclusion is false and

---

[14] Williamson's primary concern here lies not with ($\exists$con) but with the Barcan Formula $\Diamond \exists x A \rightarrow \exists x \Diamond A$. The issues are closely connected. Every countermodel to the Barcan Formula satisfies ($\exists$con). And every model that satisfies ($\exists$con) is either a countermodel to the Barcan Formula, or differs from one only over the valuation function.

the premiss true. So the argument requires an additional assumption: m's representational content (at least partially) concerns whether o actually exists, hence isn't silent about o.

Stalnaker's view involves rejection of this missing assumption. On that view, mere possibilia don't represent particular individuals. So none of m's mere possibilia represent o. Since the elements of m's privileged domain represent themselves, they also don't represent o. So m is silent about o: it doesn't represent o as actually existing or as not actually existing, because m doesn't represent o at all. Because m^* represents o as actual, it isn't representationally equivalent to m. So we should have: $m \approx m^*$. And on our definition of $\approx$, we do. The representational difference between m and m^* is thus consistent with Stalnaker's account of $\approx$, with m's silence about o,[15] and hence also with the invalidity of the argument as presented above.

A response is available. Although models that satisfy ($\exists$con) needn't be inaccurate by virtue of *de re* representing their mere possibilia as not actually existing, each such model's accuracy is nonetheless incompatible with its mere possibilia actually existing. Each model represents what there is as exhausted by its privileged domain. So models that satisfy ($\exists$con) represent what there is as exhausted by things other than their mere possibilia. Since each model's mere possibilia actually exist, models that satisfy ($\exists$con) misrepresent what there is as exhausted by only some of what there really is.

To see why this objection fails, note that a similar problem afflicts even those who reject existential contingency. Privileged domains are sets. But there is no universal set. So every model whatsoever misrepresents actuality as exhausted by the members of some set. I now explain how two different solutions to this problem undermine the objection above.

First solution: accept this in-principle limitation of set-theoretic semantics and adopt an account of representational adequacy that coheres with it. Interesting deficiencies in Stalnaker's approach should flow from specifically modal considerations, not from failings in model theoretic semantics quite generally. Moreover, one might think that model theoretic semantics gets something deeply right, despite no model's domain exhausting what there is. Our account of representational adequacy should reflect this. The best candidate I see is: rather than assessing models for accuracy *tout court*, assess them for accuracy under the hypothesis that some set exhausts what there is. Must all models that satisfy ($\exists$con) be inaccurate under that hypothesis? No.

Suppose some set s exhausts what there is. Let m^s be a model with privileged domain s and mere possibilia from outside s. Although m^s misrepresents actuality as exhausted by s, that doesn't render m^s inaccurate under the hypothesis that s exhausts what there is. Since m^s satisfies ($\exists$con), the hypothesis that some set exhausts what there is doesn't render every satisfier of ($\exists$con) inaccurate.

The above reconstruction of Williamson's argument now breaks down at Stage Three. The appropriate initial supposition at Stage One is that, on the hypothesis that some set exhausts what there is, m accurately represents ($\exists$con) as true; for unless we're to count every model whatsoever as inaccurate, that's the standard for accuracy that counts. At Stage

---

[15] Objection: o is a mere possibilium of all models bearing $\approx$ to m; so thats representational; so m's representational content isn't silent about o. Response: modify the definition of $\approx$ by replacing σ-functions with actuality-fixing isomorphisms. See note 13.

Three, it still follows that m represents what there is as exhausted by a set s that excludes o. Supposing that s exhausts what there is, however, that doesn't render m inaccurate. So Stage Three's closing *reductio* step fails: it doesn't follow that the difference between m and m^* is representationally insignificant, or that $m \approx m^*$ contrary to Stalnaker's account.

Second solution: modify the formalism to avoid the problem. One prominent option involves switching to a higher-order metalanguage, in which higher-order analogues of models, valuations, and point-domains can be specified and quantified over.[16] Relative to each (higher-order analogue of a) model, point-domains will be specified by a dyadic predicate 'DOM', taking point-variables in its first argument position and monadic predicates of individuals in its second: 'DOM(w, F)' says, in effect, that the w-domain comprises exactly the Fs. Since F may be a universally applicable predicate of individuals, a (higher-order analogue of a) model's (analogue of) a privileged domain may, on this higher-order approach, comprise everything there actually is. For simplicity, I use 'model' and 'domain' for higher-order analogues of models and domains throughout the rest of this section.

This alone doesn't solve the problem. Suppose the elements of each model's privileged domain represent themselves. Then a model accurately represents which things exist only if it lacks mere possibilia; for the candidate mere possibilia will have all been used up in the model's privileged domain. However, other approaches are available. We can use any individual to represent whatever we like. We needn't use a model whose privileged domain comprises, say, 0 and 1 to represent actuality as containing only 0 and 1. We might instead take 0 to represent Stalnaker and 1 to represent Williamson. Then the model represents actuality as containing only Stalnaker and Williamson. To apply **Invariance** in this more flexible setting, think of representational equivalence as defined relative to a given choice of what each individual is to represent when in a model's privileged domain.

Assume that, for some F, (a) the Fs are in one-correspondence with everything, even though (b) some things are not F.[17] Set F as some model's privileged domain, taking the Fs to represent their images under some one-one correspondence from (a). This model accurately represents actuality as exhausted by exactly what there really is. Yet since the Fs don't exhaust what there is, the model may also contain mere possibilia. So possessing mere possibilia is compatible with accurately representing what there is. Higher-order models may therefore represent ($\exists$con) as true without thereby misrepresenting which things actually exist.

The above reconstruction of Williamson's argument now breaks down at Stage Two. Take m as the higher-order model just described. A new model is obtained by adding one of m's mere possibilia o to the privileged domain. Since elements of the privileged domain don't represent themselves, however, the new model needn't represent o as actually existing, contrary to Stage Two. And since the elements of m's privileged domain represent everything there is to represent, o represents no thing: o lacks representational content. How does this affect the variant model's representational content? This is, in effect, a version of

---

[16] For details, see (Williamson, 2013, §7 of ch5). Williamson's semantics interprets object-language quantifiers as having the same (unrestricted higher-order analogue of a) domain at each point. This needs complicating to accommodate the proposal in the text. Each point needs associating with a higher-order analogue of a domain, and the point-relative satisfaction-conditions — captured using Williamson's predicate 'TRUE' — for quantified sentences restricting accordingly. I won't go into details here.

[17] The assumption is reasonable given reality's infinitude.

the problem of empty names. Whatever the correct solution to that problem, the variant model with non-representing inhabitant of its privileged domain, should differ representationally from m,[18] in line with Stalnaker's account of representational equivalence.

8   The limits of representational equivalence

This section argues that Stalnaker's conception of non-representational artefacts cannot accommodate his interpretation of the possible worlds formalism. The problem is an instance of a more general difficulty facing **Invariance**, which is independent of any particular account of representational equivalence. It reflects a deficiency in invariance-based approaches more generally: they're in principle too coarse-grained to capture certain kinds of non-representational artefact. The next section argues that no alternative approach to representors and artefacts can subserve Stalnaker's view.

As we have seen, facts about Kripke models outstrip the facts about modal space. Existential facts about a model's mere possibilia always have singular witnesses. Those existential facts represent facts about the possible existence of non-actual things. By **NSW**, those represented (modal-existential) facts lack singular witnesses. So to avoid misrepresentation, singular facts witnessing existentials about a model's mere possibilia had better not represent singular facts witnessing what those existentials represent, namely facts about the possible existence of non-actual things. Singular facts about a model's mere possibilia are non-representational byproducts of using the (actual) existence of mere possibilia to represent the possible existence of non-actual things.

An adequate characterisation of representational equivalence will factor out the ways in which facts about Kripke models outstrip facts about modal space. The surplus facts won't be invariant under representational equivalence. So the presence of singular witnesses for true existentials about a model's mere possibilia shouldn't be invariant under representational equivalence. And on Stalnaker's account, there's a sense in which they aren't. Which individuals occupy which of a model's mere-possibilia-roles varies between representationally equivalent models. So no particular singular witness for an existential about the model's mere possibilia is representational. Models don't represent modal reality as containing any particular singular fact about which non-actual things there could be.[19]

Unfortunately for Stalnaker, this does not go far enough. Every true existential about a model's mere possibilia has a true singular witness. Different models contain different singular witnesses. But *the presence of some singular witness or other* is invariant throughout all models, hence invariant under representational equivalence. So by **Invariance**, the existence of a singular witness for each true existential about a model's mere possibilia is

---

[18] Example: sentences featuring empty names plausibly lack truth-conditions. On this approach, the variant model lacks accuracy conditions, unlike the original model m.

[19] Objection: the set of mere possibilia is invariant under actuality-fixing permutations of the outer domain; so the identities of that set's members is representational; so models represent modal reality as fixing which non-actual things there could be, leaving it open which qualitative roles they could occupy. Reply: liberalise Stalnaker's definition of $\approx$ by employing actuality-fixing isomorphisms in place of σ-functions, as suggested in note 13. Note that the argument in the text survives this modification.

representational; modal reality is represented as containing singular witnesses for what existentials about mere possibilia represent. So models represent modal reality as containing singular witnesses for all facts about the possible existence of non-actual things. That is, models represent modal reality as containing singular facts about which non-actual things there could be. Because the identities of a model's mere possibilia aren't invariant under ≈, models don't represent modal reality as containing any one such singular fact rather than another; they're silent about exactly which non-actual things there could be. But they do represent modal reality as containing facts that settle which particular non-actual things there could be. So by **NSW**: models that represent (∃con) as true are inaccurate.

The problem cannot be solved by a different account of representational equivalence. In any model whatsoever, every true existential about its mere possibilia possesses a true singular witness. The presence of such witnesses is invariant throughout all Kripke models, hence invariant under any account of representational equivalence. This mismatch between the formal representations and modal reality is intrinsic to the style of representation employed.

We can see Stalnaker's approach as an application of a more general strategy, inspired by techniques drawn from measurement theory and the semantic conception of theories.[20] I'll focus on measurement theory. Measurement theorists study numerical representations of empirical properties and relations. They use variation under a specified kind of transformation to characterise those truths that turn on the choice of one representation, rather than an alternative equivalent one. The equivalent representations are those that can be obtained from one another by application of the relevant transformations. Exactly the invariants under all such transformations are independent of one's choice of numerical representation.

Whatever one's view of this as a strategy for isolating the distinguishing features of particular representations, it is an inadequate account of the representor/artefact distinction in full generality. That distinction is more fine-grained than the measurement-theoretic approach was designed to capture. Non-representational artefacts intrinsic to the style of representation employed will be invariant under all transformations, hence independent of one's choice of particular representation. That's what going on here. On Stalnaker's approach, the existence of singular witnesses for all existentials about a model's mere possibilia is a non-representational artefact intrinsic to the style of formal representation employed, i.e. to the use of Kripke models to represent modal space. Another example of the same phenomenon emerges from the discussion of relationalism about space that Stalnaker uses to motivate his approach.

Stalnaker's relationalist denies that spatial locations exist: "there are really no such things as spatial locations—there are just spatial relations between things." (Stalnaker, 2012, p33) Nevertheless, this relationalist uses relations on an abstract space (= collection of points) to represent spatial relations. A worry arises: doesn't this approach represent spatial relations as relations on locations, contrary to relationalism? Stalnaker responds by supplementing spatial models with an equivalence relation. In line with **Invariance**, the representors are exactly what cannot vary between equivalent spatial models. The over-

---

[20] On measurement theory, see (Krantz et al., 1971, esp. 9–12); on the semantic conception of theories, see (Suppes, 2002, esp. ch.4).

all structure of spatial relations is invariant between equivalent models, and is therefore representational. The identities of the points standing in spatial relations varies between equivalent models, and is therefore not representational. However, this doesn't quite capture the relationalist thought. In every spatial model, spatial relations are represented by relations on the constituents of an abstract space; that's invariant under any account of representational equivalence, hence representational given **Invariance**. So spatial models represent spatial relations as relations on the constituents of physical space; that is, as relations on spatial locations. Because Stalnaker's relationalist rejects the existence of spatial locations, she therefore faces a choice: (a) reject every spatial model as inaccurate; or (b) reject the **Invariance**-based analysis of non-representational artefacts.

Invariance-based analyses of artefacts should be rejected; for they are too coarse-grained to capture non-representational byproducts of the material from which representations are made. I've illustrated this with examples involving controversial philosophical doctrines: relationalism about space and Stalnaker's view of modal reality. One might respond that the problem lies with those doctrines, not with **Invariance**-based analyses of artefactuality. To allay that suspicion, one more informal (and slightly fictionalised) example follows to illustrate the point.[21] It's relatively uncontroversial that some invariants under representational equivalence are non-representational in this example.

Imagine building model boats from exceptionally flimsy balsa. Representors include the number of masts on the deck, their relative positions, the size of the deck, the dimensions of the hull, and so forth. Because the balsa is so flimsy, every boat requires supports within its hull; otherwise, the deck would collapse under the weight of the masts. These supports are not representational; they serve only to ensure structural integrity, given the exceptionally flimsy material from which the models are made. The models don't represent boats as having supports within their hulls. But because every model contains supports within its hull, no equivalence relation can factor them out. So by **Invariance**, the presence of supports is representational. If supports can occupy different locations within the model hulls, their presence in any particular location won't count as representational; but their presence in some location or other will. **Invariance** counts too much as representational; for invariance under representational equivalence might be a non-representational byproduct of the style of representation together with the material from which models are made. A more fine-grained conception of representors and artefacts is required.

In §6, **Invariance** was motivated thus: because artefacts lack representational import, varying them alone won't affect representational content, hence won't affect representational equivalence. We now see why the argument fails: variation in purely artefactual respects might be impossible, or might entail variation in representational respects. One will then be unable to factor out the artefacts by varying them whilst holding the representors fixed. At most, representors are invariant under representational equivalence, though not all such invariants are representors.[22]

Stalnaker's equivalence relation on Kripke models was supposed to make his view about their non-representational artefacts precise. On the resulting view, however, models

---

[21] The example is inspired by a similar one in (Cook, 2002).

[22] Certain forms of representational redundancy, or overdetermination of content, may even allow representors to vary between representationally equivalent models. I won't examine this in detail here.

that represent (∃con) as true also represent **NSW** as false. Four responses are available:

(1) Reject the model-theoretic representation of modal reality and language.

(2) Reject (∃con).

(3) Reject **NSW**.

(4) Reject **Invariance**.

Responses (1)–(3) involve giving up on Stalnaker's project: a realist interpretation of the possible worlds formalism on which actuality and its inhabitants exhaust reality, and which can admit existential contingency. The discussion of model boats and spatial relationalism provide independent motivation for (4). The next section argues that no alternative account of non-representational artefacts can cohere with Stalnaker's view. The precise nature of the mismatch between models and modal reality precludes an interpretation of the formalism on which it (accurately) represents the source of the metaphysical controversy at issue: the interaction between quantification and modality.

9   The representational limit of realistic model theory

Given **Invariance**, the existence of singular witnesses to existentials about mere possibilia isn't artefactual. Can a different conception of non-representational artefacts save Stalnaker's view? Not while maintaining a satisfying form of realism about the formal semantics, or so I'll now argue.

Truth-conditions are represented by satisfaction-conditions. The satisfaction-condition for $\exists xA$ is, in effect, the existence of an individual that witnesses A (given the model's valuation of A). The semantic effect of placing the quantifier within the scope of a modal operator is simply to allow witnesses from outside $D(w_@)$. In the case of (∃con), the witness cannot come from $D(w_@)$.

On Stalnaker's view, requiring the existence of a witness to A is a non-representational artefact of the satisfaction-condition for $\Diamond \exists xA$. But the existence of such an individual just is the satisfaction-condition for $\Diamond \exists xA$. So that satisfaction-condition must itself be a non-representational artefact, a technical device of merely instrumental value that doesn't capture the underlying truth-conditional structure of $\Diamond \exists xA$. This is a direct consequence of representing the intensional truth-condition for $\Diamond \exists xA$ with an extensional satisfaction-condition. Yet that style of representation is supposed to be one of the key theoretical benefits of the possible worlds formalism: it provides an extensional tool with which to investigate the intensional.

The point applies whenever existential quantifiers occur within the scope of modal operators, including in (∃con). There's a structural mismatch between the truth-conditions and satisfaction-conditions of such sentences. Whereas the satisfaction-condition is the existence of a witness for the sentence within the scope of the quantifier, the truth-condition does not require such a witness. To avoid misrepresentation, Stalnaker must treat the satisfaction-condition instrumentally. This renders the formalism silent about

the truth-conditions of these sentences. So the truth-condition of (∃con) isn't represented.[23] On Stalnaker's interpretation of the formalism, it cannot represent the semantic interaction between quantification and modality, and the truth-conditions of (∃con).

Here is another way of making the point. To avoid misrepresentation, Stalnaker needs to regard true existentials about a model's mere possibilia as representational, and the existence of witnesses for them as artefactual. But as things actually stand, there is no distinction here. The actual truth of an existential just is the existence of a witness. Either both are representational, or neither is. The former brings misrepresentation via conflict with **NSW**. The latter brings representational silence about the truth-conditions of (∃con).

Note that a similar line of argument doesn't undermine **NSW**. $∃xA$ is possibly true iff, possibly, its truth-condition is satisfied. The truth-condition of $∃xA$ is the existence of a witness for A. So $∃xA$ is possibly true iff, possibly, there is a witness for A. By **NSW**: $∃xA$ can be possibly true when no thing possibly witnesses A. Putting these together, **NSW** requires the consistency of:

- Possibly, something witnesses A.

- Nothing possibly witnesses A.

This forces a distinction between the possible truth of an existential and the actual existence of a possible witness for it.[24] That is consistent with there being no distinction between (i) the actual truth of an existential and the actual existence of a witness, or between (ii) the possible truth of an existential and the possible existence of a witness.

On the extent of realism under his interpretation of the semantics, Stalnaker writes:

> I have argued that our overall semantic theory is, in the sense that matters, a realistic one, while acknowledging that there is a precise sense in which we are not realists about the possible worlds, and possible individuals of the Kripke models. (Stalnaker, 2012, p39)

Is the view realistic in the sense that matters? It depends on our interests. If our interest lies in the interaction between quantification and modality, and in the semantics of quantifiers within the scope of modal operators, the answer is: no; for Stalnaker's approach treats as non-representational those aspects of the formalism that bear on these issues. As we saw in §1, he wants to combine a realist interpretation of the formalism with metaphysical neutrality. We now see that this neutrality is achieved only by an instrumentalist formal treatment of the metaphysical controversy. That, I submit, is an unsatisfying way to obtain metaphysical neutrality.

References

Cook, R. T. (2002). Vagueness and mathematical precision. *Mind*, 111:225–247.

---

[23] More cautiously: the nature and intrinsic structure of that truth-condition aren't represented; only its relationships to other truth-conditions are represented.

[24] Rejection of the Barcan Formula $◇∃xA → ∃x◇A$ is also required. Defenders of (∃con) must reject the Barcan Formula anyway; otherwise (∃con) entails the actually inconsistent: $∃x◇@∀y(x ≠ y)$. Note 14 outlines the intimate model-theoretic connection between (∃con) and the Barcan Formula.

Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A. (1971). *Foundations of measurement volume 1: additive and polynomial representations*. New York: Academic Press.

Stalnaker, R. (1994). The interaction of quantification with modality and identity. In W. Sinnott Armostrong, D. Raffman, N. A., editor, *Modality, Morality and Belief: Essays in honor of Ruth Barcan Marcus*. Cambridge University Press. Reprinted with a Postscript in (Stalnaker, 2003, ch8).

Stalnaker, R. (2003). *Ways a World Might Be: Metaphysical and Anti-Metaphysical Essays*. OUP.

Stalnaker, R. (2010). Merely possible propositions. In Hale, B. and Hoffman, A., editors, *Modality: Metaphysics, Logic, and Epistemology*, chapter 1, pages 21–33. OUP.

Stalnaker, R. (2012). *Mere Possibilities: Metaphysical Foundations of Modal Semantics*. Princeton University Press.

Suppes, P. (2002). *Representation and invariance*. CSLI Publications.

Williamson, T. (2013). *Modal Logic as Metaphysics*. OUP.

Appendix: A fully realistic semantics?

An appendix to *Mere Possibilities* presents a variant semantics. Stalnaker describes it thus:

> It is technically feasible to do the compositional semantics in a way that mixes reference to real things with artefacts of the model and then use the equivalence relations to filter out the artifacts at the end, but one might hope to do the semantics more directly, where all the values of a complex expression…are entities that exist in the actual world. Our Tarskian semantics makes reference to infinite sequences of possible individuals, sequences that may include "merely possible individuals" that are artifacts of the model. But this was just a notational convenience. Using the methods sketched in appendix B, we could do the compositional semantics directly, where all intermediate values in the composition are actual things. (Stalnaker, 2012, pp124–5)

However, the semantics described in appendix B addresses not the problems above, but a slightly different issue. To avoid confusion, I now explain what Stalnaker's appendix B does and doesn't address.

The semantic theories of appendix B and §3 differ in two relatively trivial ways.

First way. Stalnaker does not relativise satisfaction to variable assignments. Assignments are needed to handle quantification. So Stalnaker takes an alternative approach to quantification. Rather than quantifying over assignments, he quantifies over extensions of the valuation function to new constant terms. The new clause for $\exists$ is:

- $m, w \Vdash \exists v A$ iff, for some $d \in D(w) : m^{[d/a]}, w \Vdash A^{[a/v]}$.

a is a new constant, not present in the original language. $A^{[a/v]}$ is the sentence obtained by substituting a for all occurrences of v bound by the initial quantifier in $\exists vA$. $m^{[d/a]}$ is a model just like m, except for its valuation $V^{[d/a]}$. $V^{[d/a]}$ is just like m's valuation V, except that $V^{[d/a]}(a) = d$. The resulting satisfaction-condition for, say, $\Diamond\exists xA$ involves quantification over all elements of the outer domain and consideration of all the associated extensions of the valuation function. So this is almost a notational variant of the assignment-based approach in §3. The only real difference is that Stalnaker's assignments are defined only for variables appearing in the sentence, rather than for all variables in the language. That could be mimicked with partially defined assignments.

Second way. Stalnaker's valuations of constant-terms are point-relative. For a constant term a, $V_w(a)$ is the same object d at every point whose domain contains d, and undefined otherwise. We can mimic this with point-relative assignments subject to the same constraint. For example, the new clause for atomic predication is:

- $m, w, \alpha \Vdash \Phi v_1, \ldots, v_n$ iff $\langle \alpha_w(v_1), \ldots, \alpha_w(v_n) \rangle \in V(\Phi)(w)$.

What are the gains of this approach? Think of the semantic value of an open sentence (of one free variable) as a function that takes, at a world u, each individual x to a singular proposition about x. This function should be defined at u only for individuals that exist at u: since there are no other individuals at u, there are no other arguments to consider at u, and no singular propositions about them to return as value. Functions defined on objects that don't exist at u themselves don't exist at u. Treating assignments as insensitive to worlds forces the (formal representatives of) semantic values of open sentences to be defined relative to whatever there could possibly be (since they're defined relative to all mere possibilia). Relativising assignments to worlds/points avoids this.

The quote above contrasts Stalnaker's semantics with one that uses infinite sequences (plus a global ordering of object-language variables) to determine assignments. On that approach, the semantic value of an open sentence at a world is defined relative to all such sequences. So it's defined at some worlds relative to sequences containing things that don't exist at that world. Interpreting this realistically will result in inaccuracy. Stalnaker's focus on valuations for extensions of the language by a single constant, together with his world-relativisation of valuations, avoids this (Stalnaker, 2012, pp143–147). The semantic theory in §3 does not involve infinite sequences; it quantifies over variable assignments directly. Relativising those assignments to points therefore avoids the problems arising from satisfaction by infinite sequences.

None of this bears on the issues discussed above. The satisfaction-condition for $\Diamond\exists xA$ still requires a witness for the satisfaction-condition of A. Since its truth-condition doesn't require a witness for the truth-condition of A, the mismatch between satisfaction-conditions and truth-conditions remains. Stalnaker's appendix B does not resolve and (as I read Stalnaker) is not intended to resolve, the problems with existential contingency that I have been discussing.