Ramsey equivalence

September 29, 2016

In this paper, I critique the claim that a theory's Ramsey sentence (or something like it) is a good candidate for encoding that theory's structural content, by showing that such a claim leads to implausible criteria of theoretical equivalence.

1 Introduction

This paper is about the so-called *Ramsey-sentence approach to structural realism*.[1] As is well-known, taking the structural content of a theory to be expressed by its Ramsey sentence leads to a trivialisation problem, since the process of Ramseyfication appears to wash out all non-empirical content (save that concerning cardinality). This concern is known as the *Newman problem*.[2] The basic question I want to ask here is the following: can we, with sufficient ingenuity, modify the Ramsey-sentence approach so as to avoid the Newman problem; and can we do so in a way that delivers a plausible conception of a theory's structural content? My answer to the first question will be "yes"; my answer to the second will be a (somewhat cautious) "no".

The structure of the below is as follows. First, §2 briefly reviews the formalisms of first- and second-order model theory, and introduces the Ramsey sentence. §3 introduces the Newman problem. §4 and §5 consider whether moving to (respectively) frame or Henkin semantics for second-order logic could be the basis of a feasible response to the Newman problem; whilst §6 looks at whether employing modal re-

[1][Maxwell, 1971]
[2]For a useful overview of the Newman problem, see [Ainsworth, 2009].

sources will do so. Having concluded that none of these ideas is enough by itself, §7 looks at whether combining them will do the trick. §8 concludes.

2 The Ramsey sentence

In what follows, I will suppose that the theory with which we begin comprises a set of sentences of first-order logic. Obviously, this isn't a realistic assumption, but it's perfectly in order given the aim of this paper: if the Ramsey sentence approach cannot deliver a plausible conception of structural content in this (highly idealised) case, then it is very unlikely to be able to deal with more complex or realistic examples. I will also simplify things by considering only languages without constants or function-symbols (save for a brief discussion in §**??** below). I'll begin by briefly reviewing the relevant aspects of first-order model theory, as much to fix notation as anything else.

So, a *signature* is a set Σ of monadic and polyadic predicates. Given a signature Σ, one can define the set of well-formed *first-order Σ-formulae*, using the standard compositional rules of first-order predicate logic. I'll denote that set as $L^1(\Sigma)$. A first-order Σ-sentence is a Σ-formula containing no free variables.

The semantics for $L^1(\Sigma)$ is given by first-order *Σ-structures*. A Σ-structure $\mathcal{M}$ consists of

- A set M (which will sometimes be denoted as $|\mathcal{M}|$)
- For each n-ary $R \in \Sigma$, a set $R^{\mathcal{M}} \subseteq M^n$

I will refer to a subset of M^n as an (n-ary) *extension* over $\mathcal{M}$.

A Σ-structure $\mathcal{M}$ determines the truth or falsity of elements of $L^1(\Sigma)$, relative to a variable-assignment g for $\mathcal{M}$ (i.e. a map from the variables to M): if $\mathcal{M}$ makes a formula ϕ true relative to g, we write $\mathcal{M}[g] \models \phi$. The truth-value assigned to ϕ by $\mathcal{M}$ and g is determined by the standard recursive clauses. For example, the clause for determining the truth-value of a formula of the form $\exists x\phi$ is

- $\mathcal{M}[g] \models \exists x\phi$ iff for some $a \in M$, $\mathcal{M}[g^x_a] \models \phi$.

where g^x_a is an assignment which differs from g on at most x, and is such that $g(x) = a$. Given a variable x and an individual $a \in M$, I'll use $x \mapsto a$ to denote an arbitrary assignment which assigns a to x. If ϕ is a sentence, then the variable-assignment no longer matters, and we may write simply $M \models \phi$.

A *theory* T in the signature Σ (for short, Σ-theory) is a set of Σ-sentences.[3] A Σ-picture M is said to be a *model* of T if it satisfies each member of T; we denote the class of all models of T by $\mathrm{Mod}(T)$.

The Ramsey sentence is a sentence of second-order logic, so it will also be valuable to briefly review that second-order model theory. Syntactically, second-order logic goes beyond first-order logic by introducing new variables. For each $n \in \mathbb{N}$, we introduce a stock of *n-ary relation-variables*. Relation-variables will typically be denoted $X, X_1, X_2, \ldots, Y, Y_1, Y_2, \ldots$. We now modify the standard formation clauses for formulae, in two ways. First, atomic formulae can now be formed using relation-variables rather than predicates: if X is an n-ary relation-variable, then $Xx_1 \ldots x_n$ is an atomic sentence. Second, we can now quantify over second-order variables, not just first-order variables: that is, if ϕ is a formula and X a relation-variable, then $\exists X \phi$ is a formula. If Σ is the set of predicate-letters, we will refer to the language generated over Σ by these rules as $L^2(\Sigma)$.

The standard semantics (also known as the *full* semantics) for second-order logic goes as follows. A *full structure* $\mathcal{M}$ for $L^2(\Sigma)$ consists of the same data as a Σ-structure: that is,

- A set M
- For each n-ary $R \in \Sigma$, a set $R^{\mathcal{M}} \subseteq M^n$

Such a structure evaluates formulae of $L^2(\Sigma)$ relative to a first-order variable-assignment g, and a *second-order variable-assignment* G: an arity-respecting map from the second-order variables to extensions over $\mathcal{M}$. For the new atomic formulae, the relevant clause is

- $\mathcal{M}[g, G] \models Xx_1 \ldots x_n$ iff $\langle g(x_1), \ldots, g(x_n) \rangle \in G(X)$

whilst for formulae formed using the new quantifiers, the clause is

- $\mathcal{M}[g, G] \models \exists X \phi$ iff for some $E \in \mathcal{P}(M^n)$,$\mathcal{M}[g, G^X_E] \models \phi$

where G^X_E is a variable-assignment just like G, save that it assigns E to X.

For the purposes of forming the Ramsey sentence, suppose that the signature Σ of T is the union of two disjoint signatures, Ω and Θ: Θ contains the vocabulary that is to

[3] In accordance with standard practice in model theory, I don't require theories to be deductively closed.

be Ramseyfied, whilst Ω contains the vocabulary we are going to refrain from Ramseyfying. The labels "Θ" and "Ω" are used because the classic formulation of these issues supposes that we Ramseyfy the "theoretical" vocabulary but not the "observational" vocabulary. We will suppose that $\Theta = \{R_1, R_2, \ldots, \}$, and that $\Omega = \{S_1, S_2, \ldots, \}$.

The Ramsey sentence of a Σ-theory T is then defined as generated from T by applying the following procedure:

1. Conjoin all the sentences of T into a single (perhaps infinitely long) sentence, $\bigwedge T$.
2. Replace each n-ary predicate symbol $R_i \in \Theta$ occurring in $\bigwedge T$ by an n-ary second-order variable X_i, thereby obtaining an open second-order sentence; we will denote this T^*.
3. Prefix T^* by a (perhaps infinite) string of second-order existential quantifiers $\exists X_1 \exists X_2 \ldots$, one for each free second-order variable in T^*.

It is clear from this description that if T is an arbitrary first-order theory, the language in which the Ramsey sentence is formulated must have rather powerful logical resources. If the cardinality of T is κ, and the cardinality of the subset of Σ occurring in T is λ, then we are dealing with a second-order language that permits κ-many sentences to be conjoined, and permits the introduction of λ-many second-order quantifiers. One topic that is not usually discussed in detail concerns what signature the second-order language of the Ramsey sentence has. The most natural choice, however, is to suppose that the second-order signature is Ω; we will suppose this in what follows.

3 The Newman problem

However, taking the Ramsey sentence to encode the "structural content" of a theory has a seemingly disastrous consequence: if we Ramseyfy a theory, then we wash out the theory's non-observational content. More precisely, we have the following result.[4]

Proposition 1. Let T be a satisfiable theory. Suppose that $\mathcal{M}$ is a second-order full Ω-structure such that for some model $\mathcal{N}$ of T, $\mathcal{M}$ and $\mathcal{N}$ are Ω-isomorphic: that is,

[4] cf. [Ainsworth, 2009, Theorem 5].

that there is a bijection $f : M \to N$ such that for each $S_i \in \Omega$,

$$\langle a_1, \ldots, a_n \rangle \in (S_i)^{\mathcal{M}} \Leftrightarrow \langle f(a_1), \ldots, f(a_n) \rangle \in (S_i)^{\mathcal{N}} \tag{1}$$

Then $\mathcal{M} \models T^R$.

Proof. For any extension $E \subseteq N^n$, define the *pullback of E by f* to be

$$f^*E := \{\langle a_1, \ldots, a_n \rangle \in M^n : \langle f(a_1), \ldots, f(a_n) \rangle \in E\} \tag{2}$$

Given that for any $S_i \in \Omega$, $f^*(S_i)^{\mathcal{N}} = (S_i)^{\mathcal{M}}$, we then have that

$$\mathcal{M}[X_i \mapsto f^*(R_i)] \models T^* \tag{3}$$

and hence, that $\mathcal{M} \models T^R$. □

The standard way of explaining why this result is problematic is as follows. Let $\mathcal{W}$ be "the world". Say that a theory is *observationally adequate* if it has a model which is Ω-isomorphic to the world.[5] The above result then shows that the Ramsey sentence of T is true (of the world) if and only if T is observationally adequate. This is then a Bad Thing, if the Ramsey sentence was supposed to be part of a realist strategy, since realists (by definition) are those committed to more than just the observational adequacy of scientific theories.[6]

However, I suggest that it is more illuminating to present the problem in terms of theoretical equivalence. If the Ramsey sentence really captures the "structural content" of a theory, and if that structural content is the only content to which we ought to be committed (or to which we are entitled to be committed), then we obtain a very natural associated criterion of theoretical equivalence: two theories are equivalent just in case they have logically equivalent Ramsey sentences. What Proposition 1 shows, though, is that this criterion of equivalence is implausibly weak (at least, implausibly weak for any position that aspires to be described as realist). For suppose that T_1 and T_2 are two theories, with signatures Σ^1 and Σ^2 respectively, such that $\Omega^1 = \Omega^2$.[7] Say that T_1 and T_2 are *Ω-equivalent* if it is the case that for every model of T_1, there is an Ω-

[5]cf. [van Fraassen, 1980, chap. 3]

[6][Votsis, 2003] and [Zahar, 2004] both argue that—this result notwithstanding—the Ramsey sentence does indeed go beyond the observational content of a theory. However, see [Ainsworth, 2009] for some fairly convincing replies.

[7]Note that this condition is required if it is even to be possible that T_1^R is logically equivalent to T_2^R.

isomorphic model of T_2, and vice versa. It is straightforward to show from the above that the following corollary holds:

Proposition 2. T_1^R and T_2^R are logically equivalent (under full second-order semantics) if and only if T_1 and T_2 are Ω-equivalent.

Thus, Ramsey-sentence realism is just Ω-realism: two theories have equivalent Ramsey sentences if and only if their classes of models agree on the Ω-structure, and on the cardinality of the underlying domain. Note that if we Ramseyfy *all* the predicates of the theory—i.e., if $\Omega = \varnothing$—then the only information retained by the theory is, at best, information about cardinality.

What we have proved here is a fairly general result. From it, we can easily obtain most of the other statements of the Newman problem found in the literature. For instance, [Ketland, 2004] works in a two-sorted formalism, where there is a sort σ_Θ of theoretical objects and another sort σ_Ω of observational objects. Ω then comprises all predicates of sort $\langle \sigma_\Omega, \ldots, \sigma_\Omega \rangle$ (i.e., those ranging only over tuples of observational objects), whilst Θ contains all other predicates (i.e., those ranging over tuples of theoretical objects *or* tuples of observational and theoretical objects). Given a two-sorted structure $\mathcal{M}$, let $|\mathcal{M}|_\Omega := \{a \in |\mathcal{M}| : a \text{ is of sort } \sigma_\Omega\}$, and let $|\mathcal{M}|_\Theta := \{a \in |\mathcal{M}| : a \text{ is of sort } \sigma_\Theta\}$.

We then say that a second-order two-sorted Ω-structure $\mathcal{M}$ is isomorphic to a first-order two-sorted Σ-structure $\mathcal{N}$ if there is a *sort-preserving* bijection $f : |\mathcal{M}| \to |\mathcal{N}|$ such that for each $S_i \in \Omega$, (1) holds. It's straightforward to show that the above result then still holds, i.e., that if $\mathcal{M}$ is Ω-isomorphic to a model $\mathcal{N}$ of T, then $\mathcal{M} \models T^R$. Ketland defines a model $\mathcal{N}$ as *empirically correct* just in case $\mathcal{W}$ is Ω-isomorphic (in this sense) to $\mathcal{N}$. Putting these ingredients together, we obtain Ketland's main result: that for any theory T, T^R is true (of $\mathcal{W}$) iff T has an empirically correct model $\mathcal{M}$ such that $|\mathcal{W}|_\Theta$ is equinumerous with $|\mathcal{M}|_\Theta$.

The more general proposition, however, makes clear that the Newman problem is not a function of particular aspects of Ketland's treatment (such as the use of a two-sorted formalism), but is a direct consequence of Ramseyfication. This is important, since some of the responses in the literature consider tweaking various aspects of exactly how to cash out the observational/theoretical distinction, or else how to Ramseyfy. [Melia and Saatsi, 2006], for example, argue (contra Ketland) that the structural realist is not required to Ramseyfy all mixed predicates; they claim that the structural realist is entitled to refrain from Ramseyfying "structural" mixed predicates such as "x is a part of y", "x is located in region r", or "x has velocity v".

But even if this is granted, Proposition 1 makes clear that all this is so much deckchair-shuffling. If we Ramseyfy to any extent at all, then the Newman problem will apply to just that same extent; the only way the Newman problem could be entirely blocked is if we Ramseyfy no vocabulary at all. Indeed, Melia and Saatsi go on to observe exactly this point, i.e. that (in their words) "Even when predicates that apply to observables and unobservables alike are handled with the requisite care, there are still cases where Ramseyfied theories demand too little of the theoretical world."[8]

4 Frame semantics

Nevertheless, there is a natural response to the Newman problem. This begins by observing that Proposition 1 is a straightforward consequence of our decision to interpret the Ramsey sentence in terms of the full semantics: that is, in such a way that the second-order n-ary quantifiers range over *all* sets of n-tuples in the model. As soon as we have existentially quantified over the predicates in Θ, then it becomes far too easy to find extensions to witness them—*if*, that is, we are working with a full semantics for second-order logic. But why should we do that? That is, what is it that compels us to let the second-order quantifiers range over every extension available in the models?[9] In this section, I consider proposals for how we might restrict the range of the quantifiers, and the semantics naturally associated with such proposals.

To restrict the range of the quantifiers means supposing that, in a model, not all subsets of the domain are created equal: some of them are privileged (namely, those over which the quantifiers range). Many discussions of the Newman problem are highly sceptical that the structural realist is entitled to such a distinction between sets of objects in the domain: Psillos, for example, argues that

> in order for [(epistemic) structural realists] to distinguish between natural and non-natural classes they have to admit that some non-structural knowledge is possible, viz. that some classes are natural, while others are not.[10]

I do not find this argument compelling. As we have seen, if no distinction between different sets of n-tuples is admitted, then nothing can have more structure than that of

[8] [Melia and Saatsi, 2006, p. 571]

[9] In Ketland's version of the Newman problem, the second-order quantifiers in the Ramsey sentence only range over extensions of the right sort; nevertheless, it remains the case that the Newman problem arises because they range over *all* extensions of the right sort.

[10] [Psillos, 1999, p. 66]

a bare set. Insofar as we have any grip on the notion of structure at all, it is presumably drawn from our experience of mathematics, in which it is entirely standard for some sets of objects to be privileged over others: the structure of a group G, for example, arises exactly from the privilege accorded to the set containing those triples of the form $\langle g_1, g_2, g_1 \cdot g_2 \rangle$ (as well as the privilege given to the singleton set $\{e\}$, and the set containing the pairs of the form $\langle g, g^{-1} \rangle$). As such, the fact that certain classes are more natural than others would seem to be a precondition on the possibility of structure at all.

Resistance to this claim may stem from a certain misreading of the structuralist thesis. The structural realist maintains that we can only have structural knowledge of the world, i.e., that the only facts we can know about the world are structural facts. Let's grant, for the sake of argument, that whether a set of objects is a "natural" set (i.e. is the extension of some natural property) is not a structural fact. It doesn't follow that such a fact is inadmissible to the structural realist, since the structural realist only maintains that our knowledge of the *world* is purely structural; and a set of objects in the world is not the same thing as the world itself. Again, consider the case of the group. It is surely right that for an abstractly or algebraically described group, we know nothing more about it than its structure. Nevertheless, we also know that certain sets of (tuples of) objects in the group are privileged (natural). Our knowledge of the *group* is purely structural; but that is consistent with (indeed, is in some sense constituted by) our knowledge that certain *sets of tuples of elements of the group* are more natural than others. Having only structural knowledge of some (structured) entity does not rule out having knowledge of where the entity's "joints" lie.[11]

So, there is not a prima facie reason why the structural realist may not distinguish some extensions from others, and take the second-order quantifiers to range only over the distinguished extensions. However, that does not mean that doing so is not also subject to difficulties. In order to assess things, we need to get more precise about what the proposal looks like.

First, we introduce the concept of a *frame* for second-order logic (also known as a *pre-structure*).[12] A frame $\mathcal{F}$ (of signature Σ) consists of

- A set F
- For each $n \in \mathbb{N}$, a set $\mathcal{E}_n^{\mathcal{F}}$ of subsets of F^n (the *extension-universe* for $\mathcal{F}$); let $\mathcal{E}^{\mathcal{F}} := \cup_{n \in \mathbb{N}} \mathcal{E}_n^{\mathcal{F}}$

[11] cf. [Redhead et al., 2001]

[12] The below follows [Manzano, 1996, chap. 4].

- For each $n \in \mathbb{N}$ and n-ary $\Pi \in \Sigma$, a set $\Pi^{\mathcal{F}} \in \mathcal{E}_n^{\mathcal{F}}$

A second-order variable-assignment G for a frame $\mathcal{F}$ assigns each n-ary variable to some element of $\mathcal{E}_n^{\mathcal{F}}$. A frame provides sufficient structure to interpret the language $L^2(\Sigma)$: each set $\mathcal{E}_n^{\mathcal{F}}$ gives the range of the second-order n-ary quantifiers. (Note that we require that the extension of each relation-letter fall in that range.) More precisely, given first- and second-order variable-assignments g and G, a frame $\mathcal{F}$ determines the truth-value of formulae involving the second-order quantifier via the clause

- $\mathcal{F}(g, G) \models \exists X^{(n)}\phi$ iff for some $A \in \mathcal{F}_n$, $\mathcal{F}(g, G_X^A) \models \phi$

Consequently, one can base a second-order semantics on frames (by taking validity to be truth-in-all-frames, etc.), but the logic obtained is very weak. A correlate of this is that the Ramsey sentence of a theory is very strong. Certainly, it is strong enough to block the Newman problem as discussed above.

Example 1. Let $\Omega = \varnothing$, $\Theta = \{R\}$ (where R is a unary predicate), and consider the theory

$$T = \{\exists! x(x = x), \forall x Rx\} \tag{4}$$

Clearly,

$$T^R = \exists X\, (\exists! x(x = x) \wedge \forall x Xx) \tag{5}$$

Since $\Omega = \varnothing$, Ω-isomorphism reduces to equinumerosity. But if we use the frame semantics, then it is not the case that any pair of equinumerous frames must either both satisfy or fail to satisfy T^R. For example, the frame $\langle\{0\}, \{\{0\}\}\rangle$ is a model for T^R, whilst the equinumerous frame $\langle\{0\}, \varnothing\rangle$ is not.

The reason why the logic based on frame semantics is so weak is because there are no constraints on what the privileged extensions in a frame are like. In particular, just because a frame $\mathcal{F}$ privileges (say) a pair of unary extensions E and E', it doesn't follow that their "conjunction" $\{a \in |\mathcal{F}| : a \in E \text{ and } a \in E'\}$ is privileged in $\mathcal{F}$. This makes some sense if the privileged extensions are thought of as the extensions of perfectly natural properties:[13] such properties are fully metaphysically independent from one another, and there is no reason (in general) why logical constructs out of the extensions of two natural properties should be the extension of another natural property.

[13] In (something like) the sense of [Lewis, 1983].

However, there are good reasons to think that using frame semantics for the Ramsey sentence does not deliver a conception of theoretical content that is any more palatable to the structural realist than that based on using full semantics—but for the opposite reason. Using the full semantics meant, as we saw above, that it was too easy for two theories to be Ramsey-equivalent: they needed only to agree on matters of cardinality in order for their Ramsey sentences to be (full-)logically equivalent. Using the frame semantics, however, makes it too hard for two theories to be Ramsey-equivalent. More specifically, there are *definitionally equivalent* theories whose Ramsey sentences are not equivalent under the frame semantics.

To explain this, we need a little more apparatus. Suppose we have a single observational vocabulary Ω but disjoint theoretical vocabularies Θ_1 and Θ_2; let $\Sigma_1 := \Omega \cup \Theta_1$ and $\Sigma_2 = \Omega \cup \Theta_2$. For any n-ary $R \in \Theta_2$, an *explicit definition of R in terms of* Σ_1 is a formula δ_R of the form

$$\forall x_1 \ldots \forall x_n (Rx_1 \ldots x_n \leftrightarrow \tau_R(x_1, \ldots, x_n)) \tag{6}$$

where τ_R is an n-place Σ_1-formula. A *dictionary* for Θ_2 in terms of Σ_1 is a set Δ of explicit definitions, one for each $R \in \Theta_2$. Given a dictionary Δ for Θ_2 in terms of Σ_1, any Σ_1-structure $\mathcal{S}$ can be converted into a Σ_2-structure $\Delta(\mathcal{S})$, by first taking the unique expansion of $\mathcal{S}$ to $\Sigma_1 \cup \Sigma_2$ that satisfies Δ, and then taking the reduct to Σ_2.

Now suppose that T_1 and T_2 are a pair of first-order theories, in signatures Σ_1 and Σ_2 respectively. A dictionary Δ for Θ_2 in terms of Σ_1 is a *translation manual* for T_2 in terms of T_1 if, for every model $\mathcal{M}$ of T_1, $\Delta(\mathcal{M})$ is a model of T_2. If there is a translation manual Δ for T_2 in terms of T_1 and a translation manual Δ' for T_1 in terms of T_2, then we will say that T_1 and T_2 are *bi-interpretable*. Since Ω is held fixed, bi-interpretability implies Ω-equivalence. If the translation manuals are such that $T_1 \cup \Delta$ is logically equivalent to $T_2 \cup \Delta'$, then we say that T_1 and T_2 are *definitionally equivalent*; the theory $T_1 \cup \Delta$ (or the logically equivalent $T_2 \cup \Delta'$) is referred to as the *common definitional extension*.

Recall that Ω includes, at least, all the observational vocabulary. This provides a good reason for thinking that definitional equivalence, as defined here, should be considered sufficient for theoretical equivalence: definitionally equivalent theories are empirically equivalent (in at least some sense), and their theoretical vocabularies are entirely intertranslatable.[14] Certainly, it seems that it would be a mistake for the structural realist to insist on a criterion of equivalence more fine-grained than definitional

[14]For arguments to this effect, see [Glymour, 1970], [Glymour, 1977].

equivalence. If they are to see off the pessimistic meta-induction, then they will want to regard theories as equivalent if they define the same theoretical structures, even if they do so using different basic resources—i.e., if they have a common definitional extension.[15] This is a problem for applying frame semantics to the Ramsey sentences of theories: there are definitionally equivalent theories whose Ramsey sentences are *not* frame-equivalent.

Example 2. Let $T_1 = \{\exists x Fx\}$, $T_2 = \{\exists x \neg Gx\}$. T_1 and T_2 are definitionally equivalent, by the dictionaries

$$\begin{aligned} \Delta &= \{\forall x(Gx \leftrightarrow \neg Fx)\} \\ \Delta' &= \{\forall x(Fx \leftrightarrow \neg Gx)\} \end{aligned} \tag{7}$$

However, their Ramsey-sentences

$$\begin{aligned} T_1^R &= \exists X \exists x Xx \\ T_2^R &= \exists X \exists x \neg Xx \end{aligned} \tag{8}$$

are *not* frame-equivalent. Indeed, let $\mathcal{F}$ be any frame such that $\mathcal{E}_1^{\mathcal{F}} = \{|\mathcal{F}|\}$. Then $\mathcal{F} \models T_1^R$, but $\mathcal{F} \not\models T_2^R$.

5 Henkin semantics

This example suggests that the advocate of the Ramsey sentence should attend more closely to the notion of definability. Roughly speaking, given a frame $\mathcal{F}$, an extension E over $|\mathcal{F}|$ is *definable* if there is some n-place formula ϕ such that E contains all and only those n-tuples which satisfy ϕ. The most natural way to generate a more fruitful semantics is to limit our attention to those frames which are closed under definability: i.e., which are such that for any extension E definable over $\mathcal{F}$, $E \in \mathcal{E}^{\mathcal{F}}$. Say that such a frame is a *Henkin structure*. The challenge, however, is that as we change what language ϕ may be written in, we will get different conceptions of definability, and

[15]cf. Melia and Saatsi's critique of restricting the second-order quantifiers to range over just the extensions of "natural" properties, on the grounds that some properties thought to be natural/fundamental later turn out to be somehow "disjunctive"; this is a problem for the structural realist, they argue, since he "wants his Ramsey sentences to be preserved across theory change—they are supposed to capture something that is constant between theories, else the structural realist does little better than the full blown realist in dealing with the pessimistic meta-induction." [Melia and Saatsi, 2006, p. 576]

hence different notions of Henkin structure.[16] Moreover, we may want consider not just (mere) definability, but the broader notion of *definability with parameters*. An extension E over a frame $\mathcal{F}$ is definable with parameters if there is (a) some formula ψ with n or more free first-order variables and zero or more second-order variables, (b) some individuals from $|\mathcal{F}|$, and (c) some extensions from $\mathcal{E}^{\mathcal{F}}$, such that: E contains all and only those n-tuples which satisfy ψ when its remaining free variables are assigned to the chosen individuals and extensions. As we proceed, we will need to be somewhat careful to keep an eye on this moving part in the account.

The reason to move to Henkin structures is that we are far more restricted in which frames we can consider. For example, the frame $\mathcal{F}$ considered in Example 2 is not a Henkin structure on any plausible unpacking of definability: the set F is definable over $\mathcal{F}$ by a formula such as $x = x$, and so it at least would have to be included in the extension-universe. More generally, by choosing a notion of definability for frames designed to "mimic" definability over associated first-order structures, we can show that definitional equivalence is sufficient for equivalence of Ramsey-sentences (under the chosen Henkin semantics—and hence, under any Henkin semantics which permits a richer notion of definability). In fact, we can show that not just definitional equivalence, but bi-interpretability, is sufficient. I turn to showing this now.

First, the relevant notion of definability. A formula of second-order logic is said to be *first-order* if it contains no second-order quantifiers (note that a first-order formula is permitted to contain second-order variables). Given a frame $\mathcal{F}$, over signature Ω, we say that an extension E over $\mathcal{F}$ is *first-order definable with second-order parameters* if there is some first-order formula $\psi \in L^2(\Omega)$ with free first-order variables $x_1, \ldots, x_n$ and free second-order variables $Y_1, \ldots, Y_m$, and there are some extensions $E_1, \ldots, E_m \in \mathcal{E}^{\mathcal{F}}$, such that, for any $a_1, \ldots, a_n \in |\mathcal{F}|$,

$$\langle a_1, \ldots, a_n \rangle \in E \Leftrightarrow \mathcal{F}[Y_i \mapsto E_i, x_j \mapsto a_j] \models \psi \tag{9}$$

We then have the following result.

Proposition 3. Suppose that T_1 and T_2 are two bi-interpretable theories, in signatures $\Sigma_1 = \Omega \cup \Theta_1$ and $\Sigma_2 = \Omega \cup \Theta_2$ respectively (where $\Omega = \{S_i\}_i$, $\Theta_1 = \{R^1_j\}_j$, and $\Theta_2 = \{R^2_k\}_k$). Then T_1^R (wherein each R^1_j has been replaced by a variable X^1_j) is logically equivalent to T_2^R (wherein each R^2_k has been replaced by a variable X^2_k), on the Henkin

[16] So my usage of the term "Henkin structure" is a little non-standard, given that I have not fixed on a specific notion of definability: usually, a Henkin structure is a frame which is closed under (specifically) definability, with parameters, in the language of finitary second-order logic.

semantics generated by first-order definability with second-order parameters.

Proof. Suppose the proposition were false; then (without loss of generality) we can suppose that there is some Henkin structure $\mathcal{H}$ (of signature Ω) such that $\mathcal{H} \models T_1^R$ but $\mathcal{H} \not\models T_2^R$, where $\mathcal{H}$ is closed under first-order definability with second-order parameters.

Hence, $\mathcal{H}[G] \models T_1^*$, for some second-order variable-assignment G. So consider the Σ_1-structure $\mathcal{M}$ defined by

$$M = H \tag{10}$$

$$S_i^{\mathcal{M}} = S_i^{\mathcal{H}} \tag{11}$$

$$(R_j^1)^{\mathcal{M}} = G(X_j^1) \tag{12}$$

Clearly, $\mathcal{M} \models T_1$.

Since T_1 and T_2 are bi-interpretable, we can use the translation manual Δ for T_2 in terms of T_1 to construct a model $\Delta(\mathcal{M})$ of T_2. But now let G' be any second-order variable-assignment such that, for any second-order variable X_k^2 occurring in T_2^R,

$$G'(X_k^2) = (R_k^2)^{\Delta(\mathcal{M})} \tag{13}$$

For any other n-ary second-order variable X, let $G'(X)$ be some arbitrary element of $\mathcal{E}_n^{\mathcal{H}}$.

I now show that G' is a variable-assignment for $\mathcal{H}$. Given any $R_k^2 \in \Theta_2$, we know that the formula $\tau_{R_k^2}$ (occurring in the definition $\delta_{R_k^2}$ of R_k^2 in terms of Σ_1, as per equation (6)) defines the extension $(R_k^2)^{\Delta(\mathcal{M})}$ in $\mathcal{M}$. It cannot define the extension in $\mathcal{H}$, however, since $\mathcal{H}$ is of signature Ω and $\tau_{R_k^2}$ is a Σ_1-formula (and so will, in general, contain predicates from Θ_1). But now observe that for each j, $G(X_j^1) \in \mathcal{E}^{\mathcal{H}}$ (since G was an assignment for $\mathcal{H}$); that means, by equation (12), that $(R_j^1)^{\mathcal{M}} \in \mathcal{E}^{\mathcal{H}}$. So now consider the formula $\tau^*_{R_k^2} := \tau_{R_k^2}[Y_j/R_j^1]$, i.e., the formula obtained by uniformly substituting variables Y_j for the predicates R_j^1 in Θ_1. Note that $\tau^*_{R_k^2}$ is a first-order formula of $L^2(\Omega)$: i.e., it has second-order variables but no second-order quantifiers, and is of signature Ω. If the second-order variables Y_j are assigned to the parameters $(R_j^1)^{\mathcal{M}}$, then $\tau^*_{R_k^2}$ defines the extension $(R_k^2)^{\Delta(\mathcal{M})}$. But since all of those parameters are in $\mathcal{E}^{\mathcal{H}}$, that means that $\tau^*_{R_k^2}$ parametrically defines $(R_k^2)^{\Delta(\mathcal{M})}$ in $\mathcal{H}$. Given that $\mathcal{H}$ is closed under first-order definability with second-order parameters, it follows that $(R_k^2)^{\Delta(\mathcal{M})} \in \mathcal{E}^{\mathcal{H}}$. By equation (13), this means that $G'(X_k^2) \in \mathcal{E}^{\mathcal{H}}$. Therefore, for any second-order variable X, if X occurs in T_2^R, then $G'(X) \in \mathcal{E}_n^{\mathcal{H}}$; and by stipulation, if X does not occur in T_2^R, then $G'(X) \in \mathcal{E}_n^{\mathcal{H}}$. So G' is a variable-assignment for $\mathcal{H}$.

But now, given that $\Delta(\mathcal{M}) \models T_2$, it is clear that $\mathcal{H}[G'] \models T_2^*$. Therefore, $\mathcal{H} \models T_2^R$. This contradicts our assumption, and so the proposition follows. $\square$

So moving to Henkin semantics, with its more restricted notion of a model, enables us to avoid the problem canvassed in the previous section. However, there is a problem: with a sufficiently liberal conception of definability, perhaps it will turn out that any Henkin model is forced to include *all* extensions—bringing us back to full semantics, and hence to the Newman problem. This is how I read the following remarks of Newman:

> The only possibility of combating this objection [i.e., the Newman problem] seems to be to deny the truth of the proposition about relation-numbers [i.e. extensions] on which it depends, namely that given an aggregate A, there exists a system of relations, with any assigned structure compatible with the cardinal number of A, having A as its field. This involves abandoning or restricting Mr. Russell's own definition of a relation, namely, the class of all sets $(x_1, x_2, \ldots, x_n)$ satisfying a given propositional function $\phi(x_1, x_2 \ldots, x_n)$. If this definition is retained our assertion is clearly true. For example if a, α, β, γ are any four objects whatever, a relation which holds between a and α, a and β, and a and γ, but no other pairs is the set of all couples, x and y, satisfying the propositional function
>
> $$x \text{ is } a, \text{ and } y \text{ is } \alpha \text{ or } \beta \text{ or } \gamma \tag{14}$$

Note that it is granted in the argument that we may only consider those extensions which are definable. Newman's claim, however, is that if we are allowed to freely name elements of the domain, then restricting our attention to definable extensions is no restriction at all: every relation will be definable by some means or other.

In more precise terms, the claim would be something like the following. Suppose that $\mathcal{H}$ is a Henkin structure over signature Σ, and that for every $a \in |\mathcal{H}|$, there is some constant $\alpha \in \Sigma$ such that $\alpha^{\mathcal{H}} = a$. It then follows (says Newman) that every set of n-tuples over $|\mathcal{H}|$ is definable, and hence that $\mathcal{E}_n^{\mathcal{H}} = \mathcal{P}(|\mathcal{H}|^n)$. A more modern way of making Newman's point would appeal to the notion of parametric definability, rather than to the introduction of new constants. In those terms, the relevant claim is then that for any Henkin structure $\mathcal{H}$, every extension is parametrically definable.

It should be observed that the truth of this claim is not quite so trivial as Newman makes out. In order to prove it, we need to suppose that the cardinality of $|\mathcal{H}|$ is no greater than the cardinality of permissible disjunction in the language used to formulate definitions. If the language in which we formulate definitions permits disjunction of κ-many formulae, then if $|\mathcal{H}|$ contains no more than κ-many elements, any extension E over $|\mathcal{H}|$ can be (parametrically) defined by a formula of the form

$$\bigvee_{\lambda}(x_1 = y_1^\lambda \wedge x_2 = y_2^\lambda \wedge \cdots \wedge x_n = y_n^\lambda) \tag{15}$$

with parameters b_i^λ, each of which gets assigned to y_i^λ: here λ indexes the different elements of E, so that for every $\langle a_1, \ldots, a_n\rangle \in E$, there is some value of λ such that $b_i^\lambda = a_i$ $(1 \leq i \leq n)$. The fact that $|\mathcal{H}|$ has no more than κ elements means that E has no more than κ elements,[17] and hence that the above disjunction is well-formed. As a result, the claim is only true in full generality if we have *no* upper bound whatsoever on the formation of disjunctions in the defining language. That said, it is hard to see how one could motivate such an upper bound, at least insofar as we are doing metaphysics rather than logic.

A better response to this trivialisation objection is to argue that the notion of definability is too generous in a different way: the issue is not the expressive resources available within the defining language, but rather the use of definability with parameters. Newman considers just such a response, which he puts as follows:

> It may, however, be held that "real" relations can be distinguished from "fictitious" ones; that the example just given is a fictitious one, while the generating relation of the structure of the world is real; and that there is not always a real relation having an assigned structure and a given field. Here "fictitious" has a well defined sense; it means that the relation is one whose only property is that it holds between the objects that it does hold between; i.e., the propositional function defining it is of the type (14) above.[18]

[Melia and Saatsi, 2006] consider a similar proposal: rather than the "real/fictitious" distinction, they consider the distinction between *qualitative* properties (those "tied to what the objects are *like*, the kinds of *features* that they have, the *qualities* that they

[17]Unless κ is finite—but in that case, E will have only finitely many elements. So unless we are working in a (very strange) language with finite bounds on disjunction, this won't be a problem.
[18][Newman, 1928, p. 145]

possess")[19] and non-qualitative properties (properties such as "*being identical to a, b or c*")[20].

The most natural way of making this precise is to restrict ourselves to definability *without* first-order parameters. Of course, in structures in which everything carries a name, then this is no limitation (as the Newman quote above points out); but not all structures are like that. (It seems that we should still admit definability by second-order parameters, since we want to allow that "theoretical" properties—i.e., properties whose corresponding predicates we are seeking to Ramseyfy—are still qualitative properties.)

Even without getting any more specific about the language of definition, we can show that if definability with first-order parameters is excluded, then not all Henkin structures are full structures. The basic observation here—which is a standard piece of model theory—is that if a set is definable (without first-order parameters), then it is invariant under automorphisms. That is, let h be an automorphism of the frame $\mathcal{F}$, i.e., a bijection $h : |\mathcal{F}| \to |\mathcal{F}|$ such that for every $E \in \mathcal{E}^{\mathcal{F}}$,

$$\langle a_1, \ldots, a_n \rangle \in E \Leftrightarrow \langle h(a_1), \ldots, h(a_n) \rangle \in E \tag{16}$$

It will take only a straightforward proof by induction to show that, for any formula $\psi(x_1, \ldots, x_n, Y_1, Y_2, \ldots)$,

$$\mathcal{F}[Y_i \mapsto P_i, x_j \mapsto a_j] \models \psi \Leftrightarrow \mathcal{F}[Y_i \mapsto P_i, x_j \mapsto h(a_j)] \models \psi \tag{17}$$

Thus, if the extension D is defined by ψ (with respect to the second-order parameters P_i), then for any $a_1, \ldots, a_n \in |\mathcal{F}|$,

$$\langle a_1, \ldots, a_n \rangle \in D \Leftrightarrow \langle h(a_1), \ldots, h(a_n) \rangle \in D \tag{18}$$

But if $\mathcal{F}$ admits some non-trivial automorphism (i.e., an automorphism which is not the identity map), then not all sets will be invariant under all automorphisms; from which it follows that not all sets are definable. So we are not facing the same level of trivialisation as we had before.

Nevertheless, that does not mean that there is no threat of trivialisation. Melia and Saatsi, after making the above argument, go on to claim that it only provides a short

[19] [Melia and Saatsi, 2006, p. 577]
[20] [Melia and Saatsi, 2006, p .578]

respite for the Ramsey-philiac:[21]

> Unfortunately, though attractive, restricting the quantifiers to qualitative properties is too weak to stave off Newman-style arguments for very long. True, restricting the second order quantifier in this manner implies that it is not necessarily true that whenever you have a set of objects there is one and only one property that those objects instantiate. The possibility of symmetric worlds [i.e., models with non-trivial automorphisms] demonstrated that. But worlds showing such symmetry are extremely rare. Where a world lacks this level of symmetry, it will be the case that, again, for every set of objects there is a qualitative property that the members of this set, and only the members of this set, instantiate. If the world is such that every object has a unique qualitative property then, by forming the relevant disjunction, every set of objects will correspond to a unique qualitative property too.[22]

On our assimilation of qualitative properties to the notion of definability without first-order parameters, this argument requires the converse of the principle discussed above: i.e., it requires the claim that if an extension E over a frame $\mathcal{F}$ is invariant under every automorphism of $\mathcal{F}$, then E is definable (without first-order parameters). Again, so long as we are willing to grant the defining language as much expressive power as necessary, then this claim seems plausible. It will indeed then follow that in rigid frames (those with no non-trivial automorphisms), every extension is definable, so that we face triviality once again.

In addition to this, I want to adduce one more problem for the move to Henkin semantics, even if we do fix on some limited conception of definability (and hence avoid the trivialisation results). The concern is that even with such limits, the notion of equivalence associated with Ramsey sentences interpreted by Henkin semantics is too weak: even with a fairly limited defining language, there are theories which seem intuitively inequivalent, which nevertheless generate Ramsey sentences that are equivalent (with respect to the Henkin semantics). Note that this problem will not be resolved by appeal to a richer notion of definability. As we enrich the defining language, we make it harder for something to be a Henkin structure (since it is harder for it to be closed under definability); we therefore make it easier for a pair of Ramsey

[21] Newman has his own reply to this response (at [Newman, 1928, pp. 145–146]); I confess, however, that I don't fully understand his reply.

[22] [Melia and Saatsi, 2006, p. 578]

sentences to have all their Henkin models in common, i.e., move to a weaker notion of equivalence. (To put it another way, the problem with full semantics was that the associated notion of equivalence was too weak, and the problem with frame semantics was that the associated notion of equivalence was too strong. Enriching the defining language moves us towards full semantics, i.e., in the direction associated with a weaker notion of equivalence.)

More specifically, consider again the Henkin semantics generated by first-order definability with second-order parameters. We saw above that bi-interpretability is a sufficient condition for Ramsey-equivalence with respect to this semantics. The following example shows, however, that it is not a necessary condition: there are Ramsey-equivalent theories (with respect to this semantics) which are not bi-interpretable.

Example 3. Let $\Sigma_1 = \{P_0, P_1, \dots, \}$, and let $\Sigma_2 = \{Q_0, Q_1, \dots\}$. Let $T_1 = \varnothing$, and $T_2 = \{\forall x(Q_0 x \to Q_i x)\}_{i\in\mathbb{N}}$. T_1 and T_2 are not bi-interpretable: there is no way of defining Q_0 in terms of Σ_1. However, their Ramsey-sentences,

$$T_1^R = \top \tag{19}$$

$$T_2^R = \exists X_0 \exists X_1 \dots [\forall x(X_0 x \to X_1 x) \land \forall x(X_0 x \to X_2 x) \land \cdots] \tag{20}$$

are Henkin-equivalent, since T_2^R is a theorem of Henkin semantics. For, given any Henkin structure $\mathcal{H}$, the set H is definable by the formula $x = x$. So if G is any variable-assignment such that $X_i \mapsto H$ for all $i \in \mathbb{N}$, then $\mathcal{H}[G] \models T_2^*$, and hence $\mathcal{H} \models T_2^R$.

This is a problem if one thought that bi-interpretability was a plausible necessary condition on theoretical equivalence.

6 Modalising the Ramsey sentence

So, one way or the other, it does not look like Henkin semantics is the magic bullet we might have hoped for. To think about a solution, start with the last problem we raised for Henkin semantics. Why might one think that bi-interpretability is an appropriate necessary condition for equivalence? [Halvorson, 2012], in discussing an example very similar to example 3 above, observes that the two theories are intuitively inequivalent because "the first theory tells us nothing about the relations between the predicates, but the second theory stipulates a nontrivial relation between one of the predicates and the rest of them."[23] That is, T_2 (unlike T_1) says that there is some

[23][Halvorson, 2012, p. 193]

predicate such that in each model, anything satisfying that predicate satisfies all other predicates. That content gets lost in T_2^R, since all T_2^R says is that in each model, there are some (definable) predicates such that anything satisfying the first satisfies all the others. (That is, we move from a claim of the form "there is ... such that for all ..." to a claim of the form "for all ... there is ...".)

In more metaphysical terms, the problem is that the original theory T_2 cared about the transworld identity of properties in a way that T_2^R does not. Putting things this way suggests an association to the account that [Melia and Saatsi, 2006] give of the problems with the Ramsey sentence:

> The properties postulated in scientific theories are typically taken to stand in certain intensional relations to various other properties. Some properties *counterfactually depend* on others, some are *correlated in a law-like manner* with others, some are *independent* of others, and some are *explanatory* of others. In the model theoretic arguments considered so far, the logical frame-work in which Ramseyfication takes place is not capable of saying that such relations between properties hold.[24]

So, how might we seek to capture such relations between properties (or predicates) in the form of a Ramsey-sentence-style construction? Melia and Saatsi make the following suggestions:

> There are a number of ways to formalise intensional relations between properties. The simplest way would be to introduce new higher order relational predicates into one's theory. Alternatively, since many of the relations between properties that are of interest to us are certain kinds of *modal* associations, one could augment the relevant formal system with modal operators and use them to express these modal relations. So, for instance, let L_P express 'it is physically necessary that...'. Then $\exists X L_P \forall x(Xx \leftrightarrow Gx)$ says that there is a property which is *lawfully* coextensive with G.[25]

Here, I consider the latter suggestion: that we formulate our Ramsey-style sentences using modal operators. Specifically, suppose that we supplement our language $L^2(\Omega)$ with the sentential necessity operator $\Box$, thereby obtaining a modal second-order language $ML^2(\Omega)$. The intended interpretation of this operator is that of expressing physical necessity (just like Melia and Saatsi's L_P operator). Taking inspiration from Melia

[24] [Melia and Saatsi, 2006, pp. 579–580]
[25] [Melia and Saatsi, 2006, pp. 580–581]

and Saatsi's example sentence, let us say that the *Melia-Saatsi* sentence of a theory T is the $ML^2(\Omega)$ sentence

$$T^{MS} := \exists X_0 \exists X_1 \ldots \Box T^* \tag{21}$$

where T^* is as before.

As before, this raises the question of what semantics we are using to interpret the Melia-Saatsi sentence. Given that Melia and Saatsi argue vigorously against the cogency of restricting the second-order quantifiers, the natural suggestion is that they envision the Melia-Saatsi sentence being interpreted by something like *full Kripke semantics*. A full Kripke structure $\mathcal{K}$, for a signature Σ, consists of the following data:

- A set W (of worlds)
- For each $w \in W$, a set D_w (of individuals at that world)
- For each n-ary $\Pi \in \Sigma$, an *intension over* $\mathcal{K}$: i.e., a map $\Pi^\bullet : w \in W \mapsto \Pi^w \subseteq D^n$

Thus, for our language $ML^2(\Omega)$, a full Kripke structure may be written in the form $\langle W, D_\bullet, \{S_i^\bullet\}\rangle$. Such a full Kripke structure $\mathcal{K} = \langle W, D_\bullet, \{S_i^\bullet\}\rangle$ evaluates formulae of $ML^2(\Omega)$ relative to a world $w \in W$, a first-order variable-assignment $g : \mathrm{Var} \to D_w$, and a *modal* second-order variable-assignment Γ. A modal second-order variable-assignment assigns each second-order variable to some intension I over $\mathcal{K}$, i.e., to some map of the form $w \mapsto I_w \subseteq D_w$.

Unfortunately, however, the Melia-Saatsi sentence (interpreted by full Kripke semantics) suffers from a Newman-style problem. More specifically, we have the following result.[26]

Proposition 4. Let T be a satisfiable theory. Suppose that $\mathcal{K} = \langle W, D_\bullet, \{S_i^\bullet\}\rangle$ is a full Kripke structure such that for each $w \in W$, there is a model $\mathcal{M} \in \mathrm{Mod}(T)$ such that $\mathcal{M}$ is Ω-isomorphic to to w. Then $\mathcal{K} \models T^{MS}$.

Proof. Let $f : W \to \mathrm{Mod}(T)$ be a map taking each world to an Ω-isomorphic model; and for each $w \in W$, let $h_w : D_w \to D_{f(w)}$ be an Ω-isomorphism. Let Γ be a modal second-order variable-assignment such that

$$\Gamma(R_i) : w \mapsto h^*(R_i^{f(w)}) \tag{22}$$

where h^* denotes pullback under h. It then follows that for any $w \in W$, $\mathcal{K}[\Gamma, w] \models T^*$; hence, that $\mathcal{K}[\Gamma] \models \Box T^*$; and hence, that $\mathcal{K} \models T^{MS}$. □

[26] This result generalises the observations of [Yudell, 2010, §7].

Hence, in order for a full Kripke structure to be a model of the Melia-Saatsi sentence, it suffices that each of its worlds be Ω-isomorphic to some model of the theory. Clearly, this is still a rather weak condition; so the Melia-Saatsi sentence, interpreted by full Kripke semantics, is not a plausible candidate for capturing the structural content of a theory. But there is a natural next move: combine the tricks we have considered so far. That is, perhaps if we both modalise *and* restrict the range of the quantifiers, then we will get somewhere. I now turn to evaluating this proposal.

7 Modal Henkin semantics

Let a *modal frame* $\mathcal{F}$ for the language $ML^2(\Omega)$ consist of

- A set W
- For each $w \in W$, a set D_w
- For each $n \in \mathbb{N}$, a set $\mathcal{I}_n^{\mathcal{F}}$ of n-ary intensions over $\mathcal{F}$; let $\mathcal{I}^{\mathcal{F}} = \cup_{n \in \mathbb{N}} \mathcal{I}_n^{\mathcal{F}}$
- For each n-ary $S_i \in \Omega$, an intension $S_i^{\mathcal{F}} \in \mathcal{I}_n^{\mathcal{F}}$

Given a world W, a modal second-order variable-assignment Γ and a first-order variable-assignment g, a modal frame evaluates formulae involving the second-order quantifier via the clause

- $\mathcal{F}[w, \Gamma, g] \models \exists X^{(n)} \phi$ iff for some $I \in \mathcal{I}_n^{\mathcal{F}}$, $\mathcal{F}[w, \Gamma_I^X, g] \models \phi$

A modal frame can be written in the form $\langle W, D_\bullet, \{S_i^\bullet\}, \mathcal{I} \rangle$. Given a modal frame $\mathcal{F}$, an n-ary intension I over $\mathcal{F}$ is definable just in case there is some n-place formula ϕ such that the extension of ϕ at each $w \in W$ coincides with the value of the intension for w. It is definable with second-order parameters if there is some formula ϕ with n free first-order variables and k free second-order variables and some intensions $J_1, \ldots, J_k \in \mathcal{I}^{\mathcal{F}}$, such that when the second-order variables are assigned to $J_1, \ldots, J_k$, the extension of ϕ at each $w \in W$ coincides with the value of the intension for w. (Definability with first-order parameters gets a little complicated in the modal case; given that we have already rehearsed reasons to avoid such definability, I will not bother formulating it.) A modal frame semantics does not seem like it would be helpful (for the same reasons as discussed in section 4 above); but we can consider modalised Henkin semantics, by restricting to frames whose intension-universes are closed under definability (in some language, and possibly with second-order parameters).

An immediate advantage of this approach is that we do not have the trivialisation problems that arose for non-modal Henkin semantics. Recall that the problem was that even if we rule out definability with first-order parameters, any sufficiently non-symmetric Henkin model will (assuming an arbitrarily powerful defining language) be a full model. This objection is not compelling in this case. Even if we require that every world of a modal Henkin model be non-symmetric (i.e., admit no non-trivial automorphisms), it will not follow that every intension is definable, as the following example illustrates.

Example 4. Let $\Omega = \{P_1^{(1)}, P_2^{(1)}\}$, and consider the following modal Ω-frame $\mathcal{F} = \langle W, D_\bullet, \{P_i^\bullet\}_{i=1,2}, \mathcal{I}\rangle$:

$$
\begin{aligned}
W &= \{w_1, w_2\} \\
D_{w_1} &= \{1, 2\} \\
D_{w_2} &= \{3, 4, 5\} \\
P_1^{w_1} &= 1 \\
P_2^{w_1} &= 2 \\
P_1^{w_2} &= 3 \\
P_2^{w_2} &= 4 \\
\mathcal{I} &= \{P_1^\bullet, P_2^\bullet\}
\end{aligned}
$$

Note that no world in W admits any non-trivial automorphisms. Now consider the intension I defined by

$$
\begin{aligned}
I_{w_1} &= \{1\} \\
I_{w_2} &= \{4\}
\end{aligned}
$$

This intension is not definable in $\mathcal{F}$. For, consider the map $f : w_1 \to w_2$ such that

$$
\begin{aligned}
f(1) &= 3 \\
f(2) &= 4
\end{aligned}
$$

This map is a homomorphism: it is a standard result from model theory, therefore, that for any 1-place formula $\phi(x) \in L^1(\Omega)$ and any $a \in D_{w_1}$, if $\mathcal{F}[w_1, x \mapsto a] \models \phi$ then $\mathcal{F}[w_2, x \mapsto f(a)] \models \phi$. Since $1 \in I_{w_1}$ but $3 \notin I_{w_2}$, it immediately follows that no such formula ϕ can define I. Moreover, given that the intension-universe of $\mathcal{F}$ is just $\{P_1^\bullet, P_2^\bullet\}$, the same reasoning applies to any first-order formula ϕ with free second-order variables. Thus, the intension I is not definable over $\mathcal{F}$ (with or without second-order parameters).

This approach also overcomes the second problem canvassed above for the Henkin-semantics approach to Ramsey sentences (i.e., that bi-interpretability was not necessary for Ramsey-equivalence). This time, we are able to show that, with an appropriate choice of defining language, a pair of theories are Ramsey-equivalent if and only if they are bi-interpretable. The "appropriate choice" is the modal analogue of the defining language used in Proposition 3. More specifically, given a modal frame $\mathcal{F}$, over signature Ω, we say that an intension I over $\mathcal{F}$ is *first-order definable with second-order parameters* if there is some first-order formula $\psi \in L^2(\Omega)$ with free first-order variables $x_1, \dots, x_n$ and free second-order variables $Y_1, \dots, Y_m$, and there are some intensions $I_1, \dots, I_m \in \mathcal{I}^{\mathcal{F}}$, such that, for any $w \in W_{\mathcal{F}}$, for any $a_1, \dots, a_n \in D_w$,

$$\langle a_1, \dots, a_n \rangle \in I^w \Leftrightarrow \mathcal{F}[Y_i \mapsto I_i, w, x_j \mapsto a_j] \models \psi \tag{23}$$

We then have the following result.

Proposition 5. Suppose that T_1 and T_2 are two theories, in signatures $\Sigma_1 = \Omega \cup \Theta_1$ and $\Sigma_2 = \Omega \cup \Theta_2$ respectively (where $\Omega = \{S_i\}_i$, $\Theta_1 = \{R_j^1\}_j$, and $\Theta_2 = \{R_k^2\}_k$). Then T_1 and T_2 are bi-interpretable if and only if T_1^{MS} (wherein each R_j^1 has been replaced by a variable X_j^1) is logically equivalent to T_2^{MS} (wherein each R_k^2 has been replaced by a variable X_k^2), on the Henkin semantics generated by first-order definability with second-order parameters.

Proof. The left-to-right direction strongly resembles the proof of Proposition 3. So, suppose that T_1 and T_2 are bi-interpretable, but that their Melia-Saatsi sentences are not logically equivalent (under the given semantics). Without loss of generality, we can suppose that there is some Henkin structure $\mathcal{H}$ such that $\mathcal{H} \models T_1^{MS}$ but $\mathcal{H} \not\models T_2^{MS}$, where $\mathcal{H}$ is closed under first-order definability with second-order parameters.

Hence, $\mathcal{H}[\Gamma] \models \Box T_1^*$, for some modal second-order variable-assignment Γ; and hence, for each $w \in W$, $\mathcal{H}[\Gamma, w] \models T_1^*$. So now, for each w, consider the Σ_1-structure

$\mathcal{M}_w$ defined by

$$|\mathcal{M}_w| = D_w \tag{24}$$

$$(S_i)^{\mathcal{M}_w} = (S_i)^w \tag{25}$$

$$(R_j^1)^{\mathcal{M}_w} = \Gamma(X_j^1)^w \tag{26}$$

Clearly, $\mathcal{M}_w \models T_1$.

Since T_1 and T_2 are bi-interpretable, we can use the translation manual Δ for T_2 in terms of T_1 to construct a model $\Delta(\mathcal{M}_w)$ of T_2. But now let Γ' be any modal second-order variable-assignment such that, for any second-order variable X_k^2 occurring in T_2^{MS}, for each $w \in W$,

$$\Gamma'(X_k^2)^w = (R_k^2)^{\Delta(\mathcal{M}_w)} \tag{27}$$

For any other n-ary second-order variable X, let $\Gamma'(X)$ be some arbitrary element of $\mathcal{I}_n^{\mathcal{H}}$.

I now show that Γ' is a variable-assignment for $\mathcal{H}$. Given any $R_k^2 \in \Theta_2$, we know that the formula $\tau_{R_k^2}$ defines the extension $(R_k^2)^{\Delta(\mathcal{M}_w)}$ in $\mathcal{M}_w$. Now observe that for each j, $\Gamma(X_j^1) \in \mathcal{I}^{\mathcal{H}}$ (since Γ was an assignment for $\mathcal{H}$); that means, by equation (26), that the intension $(w \mapsto (R_j^1)^{\mathcal{M}_w}) \in \mathcal{I}^{\mathcal{H}}$. So now consider the formula $\tau^*_{R_k^2} := \tau_{R_k^2}[Y_j/R_j^1]$, i.e., the formula obtained by uniformly substituting variables Y_j for the predicates R_j^1 in Θ_1. Note that this formula is first-order (albeit with second-order variables), and of signature Ω. If the second-order variables Y_j are assigned to the parameters $(w \mapsto (R_j^1)^{\mathcal{M}_w})$, then $\tau^*_{R_k^2}$ defines the intension $(w \mapsto (R_k^2)^{\Delta(\mathcal{M}_w)})$. But since all of those parameters are in $\mathcal{I}^{\mathcal{H}}$, that means that $\tau^*_{R_k^2}$ parametrically defines $(w \mapsto (R_k^2)^{\Delta(\mathcal{M}_w)})$ in $\mathcal{H}$. Given that $\mathcal{H}$ is closed under first-order definability with second-order parameters, it follows that $(w \mapsto (R_k^2)^{\Delta(\mathcal{M}_w)}) \in \mathcal{I}^{\mathcal{H}}$. By equation (27), this means that $\Gamma'(X_k^2) \in \mathcal{I}^{\mathcal{H}}$. Therefore, for any second-order variable X, if X occurs in T_2^R, then $\Gamma'(X) \in \mathcal{I}_n^{\mathcal{H}}$; and by stipulation, if X does not occur in T_2^R, then $\Gamma'(X) \in \mathcal{I}_n^{\mathcal{H}}$. So Γ' is a variable-assignment for $\mathcal{H}$.

But now, given that $\Delta(\mathcal{M}_w) \models T_2$ for every w, it is clear that $\mathcal{H}[\Gamma', w] \models T_2^*$ for each w; hence, that $\mathcal{H}[\Gamma'] \models \Box T_2^*$; and hence, that $\mathcal{H} \models T_2^{MS}$. This contradicts our assumption, so the left-to-right half of the proposition follows.

Before proving the converse direction, we need the following lemma.

Lemma 1. *Let $\mathcal{F}$ be a modal frame over signature Σ, such that $\mathcal{I}^{\mathcal{F}}$ consists of all and only those intensions definable (without parameters) by some formula of $L^1(\Sigma)$. Let $\mathcal{H}$ be the reduct* *of $\mathcal{F}$ to Ω, i.e., the frame such that*

$$W_{\mathcal{H}} = W_{\mathcal{F}} \tag{28}$$

$$D^{\mathcal{H}}_w = D^{\mathcal{F}}_w \tag{29}$$

$$\mathcal{I}^{\mathcal{H}} = \mathcal{I}^{\mathcal{F}} \tag{30}$$

$$(P^{\mathcal{H}})^w = (P^{\mathcal{F}})^w, \textit{for each } P \in \Sigma \tag{31}$$

Then $\mathcal{H}$ is a Henkin structure in the sense considered here: that is, $\mathcal{I}^{\mathcal{H}}$ is closed under first-order definability with second-order parameters.

Proof. Suppose that the lemma were false; that is, that there were some n-ary intension I which is not in $\mathcal{I}^{\mathcal{H}}$, but is definable over $\mathcal{H}$ by some first-order formula with second-order parameters. Then there is some $\psi \in L^2(\Omega)$, with free variables $x_1, \ldots, x_n$ and $Y_1, \ldots, y_m$, such that ψ define I with respect to some parameters $I_1, \ldots, I_m \in \mathcal{I}^{\mathcal{H}}$. It follows that for each I_i, there is some formula $\psi_i \in L^1(\Sigma)$ which defines I_i. So now consider the formula $\psi[\psi_i/Y_i]$. For any world $w \in W_{\mathcal{H}}$,

$$\mathcal{H}[Y_i \mapsto I_i, w, x_j \mapsto a_j] \models \psi \text{ iff } \mathcal{H}[w, x_j \mapsto a_j] \models \psi[\psi_i/Y_i] \tag{32}$$

and hence, $\psi[\psi_i/Y_i]$ is a formula of $L^1(\Omega)$, which defines the intension I. It follows that $I \in \mathcal{I}^{\mathcal{H}}$ after all; so by contradiction, the lemma follows. □

We can now prove the converse half of our result. First, we define a frame $\mathcal{F}$, of signature Σ_1, as follows:

$$W_{\mathcal{F}} = \mathrm{Mod}(T_1) \tag{33}$$

$$D^{\mathcal{F}}_{\mathcal{M}} = M, \text{ for any } \mathcal{M} \in \mathrm{Mod}(T_1) \tag{34}$$

$$(S^{\mathcal{F}}_i)^{\mathcal{M}} = (S_i)^{\mathcal{M}} \tag{35}$$

$$((R^1_i)^{\mathcal{F}})^{\mathcal{M}} = (R^1_i)^{\mathcal{M}} \tag{36}$$

and $\mathcal{I}^{\mathcal{F}}$ consists of all and only those intensions over $\mathcal{F}$ which are definable by some formula of $L^1(\Sigma_1)$. Then, let $\mathcal{H}$ be the reduct of $\mathcal{F}$ to Ω; by the lemma, $\mathcal{H}$ is a modal Henkin structure.

Let Γ be any modal second-order variable-assignment such that, for any $\mathcal{M} \in \mathrm{Mod}(T_1)$,

$$\Gamma(X^1_i)^{\mathcal{M}} = (R_i)^{\mathcal{M}} \tag{37}$$

Then for any $\mathcal{M} \in W_{\mathcal{H}}$, $\mathcal{H}[\Gamma, \mathcal{M}] \models T_1^*$, so $\mathcal{H}[\Gamma] \models \Box T_1^*$, so $\mathcal{H} \models T_1^{MS}$.

Therefore, by the equivalence of T_1^{MS} and T_2^{MS}, $\mathcal{H} \models T_2^{MS}$. So for some modal second-order variable-assignment Γ' for $\mathcal{H}$, and for any $\mathcal{M} \in \mathrm{Mod}(T_1)$,

$$\mathcal{H}[\Gamma', \mathcal{M}] \models T_2^* \tag{38}$$

Since Γ' is a variable-assignment for $\mathcal{H}$, we know that for any k, $\Gamma'(X_k^2) \in \mathcal{I}^{\mathcal{H}}$; hence, for some $\phi_k \in L^1(\Sigma_1)$, ϕ_k defines $\Gamma'(X_k^2)$ over $\mathcal{H}$. Hence, the formula

$$\forall x_1 \ldots \forall x_n (R_k^2 x_1 \ldots x_n \leftrightarrow \phi_k) \tag{39}$$

is an explicit definition of R_k^2 in terms of Σ_1. By repeating this process for all k, we obtain a dictionary Δ for Θ_2 in terms of Σ_1.

Finally, note that given any $\mathcal{M} \in \mathrm{Mod}(T_1)$,

$$(R_k^2)^{\Delta(\mathcal{M})} = (\phi_k)^{\mathcal{M}} = \Gamma'(X_k^2)^{\mathcal{M}} \tag{40}$$

and hence, by equation (38), $\Delta(\mathcal{M}) \models T_2$. So Δ is not just a dictionary, but a translation manual for T_2 in terms of T_1. By repeating the above in the other direction, we can construct a translation manual for T_1 in terms of T_2. Therefore, T_1 and T_2 are bi-interpretable.

□

This result is a reason to think that of the options canvassed so far, this proposal (or something like it) is the best hope for something like a Ramsey-sentence approach. What it shows is that the Melia-Saatsi sentence of a theory, if interpreted using modal Henkin semantics, captures precisely the content of a theory which is invariant under bi-interpretability: if you prefer, precisely the content which is shared by a pair of bi-interpretable theories. At first glance, this might seem like a fairly reasonable condition, especially given the following fact: a dictionary Δ for Θ_2 in terms of Σ_1 is a translation manual for T_2 in terms of T_1 if and only if for every consequence ϕ of T_1, $\Delta\phi$ is a consequence of T_2—where $\Delta\phi$ is the Σ_2-formula obtained by uniformly substituting the formula τ_R for every occurrence of R in ϕ, for all $R \in \Theta_2$. So, given a pair of theories T_1 and T_2, a translation manual for T_2 in terms of T_1 gives a means of "embedding" all the content of T_1 into that of T_2: everything that T_1 says is converted, by the translation manual, into something that T_2 says. So if everything that T_1 says can be converted into something that T_2 says and vice versa (i.e., if T_1 and T_2

are bi-interpretable), does it not follow that T_1 and T_2 say the same thing?

Unfortunately, the answer turns out to be no—or at least, to follow only with an implausible conception of "saying the same thing". The following example shows that bi-interpretability is an implausible criterion of theoretical equivalence, since there are bi-interpretable theories which one has good reason to consider inequivalent.[27]

Example 5. Let $\Omega = \varnothing$, $\Theta_1 = \{P\}$, and $\Theta_2 = \{Q, R\}$ (where P, Q and R are all unary predicates). Consider the following pair of theories:

$$T_1 = \{\forall x(Px \vee \neg Px)\}$$
$$T_2 = \{\forall x(Qx \to Rx)\}$$

Intuitively, T_1 and T_2 are inequivalent: T_1 is a triviality, true in every Σ_1-model, whilst T_2 is not. Yet they are bi-interpretable. Consider the following dictionary Δ for Θ_2 in terms of Σ_1:

$$\forall x(Qx \leftrightarrow Px)$$
$$\forall x(Rx \leftrightarrow Px)$$

This is a translation manual for T_2 in terms of T_1. For, given any model $\mathcal{M}$ of T_1 (i.e., any Σ_1-structure), $\Delta(\mathcal{M})$ is a Σ_2-structure in which whatever set constituted the extension of P in $\mathcal{M}$ is now the extension of both Q and R. Thus, $\Delta(\mathcal{M}) \models T_2$. On the other hand, consider the following dictionary Δ' for Θ_1 in terms of Σ_2:

$$\forall x(Px \leftrightarrow Qx) \tag{41}$$

This is a translation manual for T_1 in terms of T_2. For, given any model $\mathcal{N}$ of T_2, $\Delta'(\mathcal{N})$ will be a Σ_1-structure—and hence, a model of T_1. So T_1 and T_2 are bi-interpretable.[28]

[27]Further examples besides this one may be adduced: for instance, the pair of theories in Example 4 of [Barrett and Halvorson, MS] are bi-interpretable. This shows that completeness of theories is not preserved under bi-interpretability.

[28]Thus, by Proposition 5, they have equivalence Melia-Saatsi sentences. This may be seen directly by observing that the Melia-Saatsi sentence of T_2 is

$$T_2^{MS} = \exists X_1 \exists X_2 \Box \forall x(X_1 x \to X_2 x) \tag{42}$$

which is a logical validity under modal Henkin semantics. For, in any modal Henkin structure $\mathcal{H}$, if Γ is an assignment such that $\Gamma(X_1) = \Gamma(X_2)$, then $\mathcal{H}[\Gamma] \models \Box T_2^*$.

8 Conclusion

One response, of course, is to think that some further ingenious tweak to the notion of a Ramsey sentence will see off the problem. In this final section, I wish to explain why I am pessimistic about the prospects of doing so.

First, note that Example 5 will persist as a problem if we move to any modal second-order semantics which is stronger than that employed here. If we use any form of semantics in which frames are closed under first-order definability with second-order parameters, then the left-to-right direction of Proposition 5 will apply, and bi-interpretability will be sufficient for logical equivalence of the Melia-Saatsi sentences. To put the same point another way: strengthening the semantics means allowing fewer frames to count as structures; that makes it easier for two theories to have logically equivalent Melia-Saatsi sentences (i.e., Melia-Saatsi sentences satisfied by exactly the same structures); and hence, it leads to a more liberal criterion of equivalence. But Example 5 is exactly a concern that the criterion of equivalence we have arrived at is too liberal—so further liberalising it will hardly help!

The alternative, then, is to weaken the semantics. This will lead to a stricter criterion of theoretical equivalence, which may then enable us to rule out Example 5. The problem is that we are then apt to run into cases like Example 2: i.e., cases showing that the associated criterion of theoretical equivalence is stricter than definitional equivalence. The following proposition gives a sufficient condition for this to occur.

Proposition 6. Suppose that $\mathcal{F}$ is a modal frame over signature $\Omega = \{S_i\}_i$, and that there is a formula ψ with free variables $x_1, \ldots, x_n$ and $Y_1, \ldots, Y_m$ such that for any intensions $I_1, \ldots, I_m \in \mathcal{I}^{\mathcal{F}}$ (with I_i of the same arity as Y_i), the intension I_ψ defined by ψ with respect to the parameters $I_1, \ldots, I_m$ is *not* in $\mathcal{I}^{\mathcal{F}}$. Then, for any semantics under which $\mathcal{F}$ is admissible as a structure, one can have definitionally equivalent theories whose Melia-Saatsi sentences are not logically equivalent (with respect to that semantics).

Proof. First, define $\Theta_1 = \{R^1_1, \ldots, R^1_m\}$, and let $\Sigma_1 = \Omega \cup \Theta_1$. Then take some particular sequence of intensions $I_1, \ldots, I_m \in \mathcal{I}^{\mathcal{F}}$, with I_i of the same arity as Y_i, and for every

$w \in W_{\mathcal{F}}$, define a first-order Σ_1-structure $\mathcal{M}_w$ as follows:

$$|\mathcal{M}_w| = D_w \tag{43}$$
$$S_i^{\mathcal{M}_w} = (S_i^{\mathcal{F}})^w \tag{44}$$
$$R_i^{\mathcal{M}_w} = I_i^w \tag{45}$$

Now, let

$$T_1 = \{\phi : \mathcal{M}_w \models \phi \text{ for every } w \in W_{\mathcal{F}}\} \tag{46}$$

Second, let $\Theta_2 = \{R_1^2, \ldots, R_m^2, R\}$, and let

$$T_2 = \{\phi[R_i^2/R_i^1] : \phi \in T_1\} \cup \{\forall x_1 \ldots \forall x_n (Rx_1 \ldots x_n \leftrightarrow \psi[R_i^2/Y_i](x_1, \ldots, x_n)) \tag{47}$$

Clearly, T_1 and T_2 are definitionally equivalent.

I now show that $\mathcal{F} \models T_1^{MS}$. First, let G be a modal second-order variable-assignment such that $G(x_i^1) = I_i$. For any world $w \in W_{\mathcal{F}}$, a straightforward proof by induction will demonstrate that for any $\phi \in L^1(\Sigma_1)$ and any first-order variable-assignment g for $\mathcal{M}_w$, if $\mathcal{M}_w[g] \models \phi$ then $\mathcal{F}[G, w, g] \models \phi[X_i^1/R_i^1]$. Thus, since $\mathcal{M}_w \models \phi$ for all sentences $\phi \in T_1$, and given the definition of T_1^*, we have that $\mathcal{F}[G, w] \models T_1^*$. Since this holds for all $w \in W_{\mathcal{F}}$, we obtain $\mathcal{F}[G] \models \Box T_1^*$; and hence, $\mathcal{F} \models T_1^{MS}$.

It remains only to show that $\mathcal{F} \not\models T_2^{MS}$. So suppose that this were not the case. Then for some modal second-order variable-assignment G' for $\mathcal{F}$, $\mathcal{F}[G'] \models \Box T_2^*$. Thus, in particular,

$$\mathcal{F}[G', w] \models \forall x_1 \ldots \forall x_n (Xx_1 \ldots x_n \leftrightarrow \psi[X_i^2/Y_i](x_1, \ldots, x_n)) \tag{48}$$

for every $w \in W_{\mathcal{F}}$. But now let $I_i' := G'(X_i^2)$, and let $I' := G'(X)$. Then it follows that for any $w \in W_{\mathcal{F}}$ and any $a_1, \ldots, a_n \in D_w$,

$$\langle a_1, \ldots, a_n \rangle \in (I')^w \iff \mathcal{F}[Y_i \mapsto I_i', w, x_j \mapsto a_j] \models \psi \tag{49}$$

Hence, I' is defined by ψ with respect to the parameters $I_1', \ldots, I_m'$. But then by hypothesis, $I' \notin \mathcal{I}^{\mathcal{F}}$, and so G' is not a variable-assignment for $\mathcal{F}$ after all. So by contradiction, the proposition follows. □

The hypothesis of Proposition 8 is stronger than merely supposing that $\mathcal{F}$ is not closed under first-order definability with respect to second-order parameters: the latter condition would require only that for *some* sequence of parameters, the intension defined by ψ with respect to those parameters is not in $\mathcal{I}^{\mathcal{F}}$. Still, it is hard to see how structures of the form of $\mathcal{F}$ could be decisively ruled out without imposing that closure condition; and as soon as we have done so, we are subject to counterexamples of the form of Example 5. This suggests that fiddling with exactly which semantics to employ is unlikely to help.

Another way of seeing this is to observe that the difference between definitional equivalence and bi-interpretability is not based on a difference over how to unpack the notion of definition: by contrast, the same notion of definition is used in both. Definitional equivalence strengthens bi-interpretability, not by changing the conditions on what is apt to count as a translation, but by requiring that the pair of translations relate to one another in a certain kind of way: namely, that they are inverse to one another.[29] This is a distinction that the Ramsey-sentence approaches considered in this paper are simply blind to.

In other words, the problem here is not that we are failing to consider the right kind of Ramsey-sentence construction, or the right kind of semantics for interpreting such a construction. Rather, the issue is something more fundamental: it does not appear that one can isolate some specific construction that it is appropriate to identify as *the* structural content of a theory. This suggests an important methodological lesson. Structural realists are often challenged to explicate their view, and to explain exactly what they mean by the "structural content" of a theory. The way they have typically sought to do this is by writing down some new theory, which (they claim) captures all and only the structural content of the old, without any of the descriptive fluff. If the above is correct, then this kind of approach is misguided.

So much the worse for structural realism? Not necessarily, for the analysis above suggests an alternative. I observed above that if the Ramsey sentence (or something like it) captures the "real" content of a theory, then that naturally induces an associated criterion of theoretical equivalence;[30] much of my critique was based on showing that the criteria so obtained were implausible criteria for theoretical equivalence. So perhaps the problem is that we are putting the cart before the horse. After all, the notion of a fully fluff-free presentation of a theory is a fantasy; any presentation of a theory will incorporate *some* inessential representational features (the colour of the ink, the typeface, etc.). So rather than questing after such chimeras, perhaps we should

[29] See [Barrett and Halvorson, MS].

[30] cf. [Coffey, 2014]

specify the content of a theory by saying what it would take for two theories to agree in their content: that is, by endorsing a criterion of theoretical equivalence. For structural realists, this means specifying a criterion of structural equivalence. Translational equivalence (perhaps with a fixed specification of how to translate between the observational vocabularies of the two theories) seems like a plausible candidate, at least for the kinds of theories discussed in this paper. In a slogan: the philosophers have only *extracted* the content of a theory, in various ways; the point, however, is to *abstract* it.

References

[Ainsworth, 2009] Ainsworth, P. M. (2009). Newman's Objection. *The British Journal for the Philosophy of Science*, 60(1):135–171.

[Barrett and Halvorson, MS] Barrett, T. W. and Halvorson, H. (MS). Glymour and Quine on Theoretical Equivalence. Unpublished draft.

[Coffey, 2014] Coffey, K. (2014). Theoretical Equivalence as Interpretative Equivalence. *The British Journal for the Philosophy of Science*, 65(4):821–844.

[Glymour, 1970] Glymour, C. (1970). Theoretical Realism and Theoretical Equivalence. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 275–288.

[Glymour, 1977] Glymour, C. (1977). The Epistemology of Geometry. *Noûs*, 11(3):227–251.

[Halvorson, 2012] Halvorson, H. (2012). What Scientific Theories Could Not Be. *Philosophy of Science*, 79(2):183–206.

[Ketland, 2004] Ketland, J. (2004). Empirical Adequacy and Ramsification. *The British Journal for the Philosophy of Science*, 55(2):287–300.

[Lewis, 1983] Lewis, D. (1983). New work for a theory of universals. *Australasian Journal of Philosophy*, 61(4):343–377.

[Manzano, 1996] Manzano, M. (1996). *Extensions of First Order Logic*. Cambridge University Press.

[Maxwell, 1971] Maxwell, G. (1971). Structural Realism and the Meaning of Theoretical Terms. *Minnesota Studies in the Philosophy of Science*, 4:181–192.

[Melia and Saatsi, 2006] Melia, J. and Saatsi, J. (2006). Ramseyfication and Theoretical Content. *The British Journal for the Philosophy of Science*, 57(3):561–585.

[Newman, 1928] Newman, M. H. A. (1928). Mr. Russell's "Causal Theory of Perception". *Mind*, 37(146):137–148.

[Psillos, 1999] Psillos, S. (1999). *Scientific Realism: How Science Tracks Truth*. Routledge.

[Redhead et al., 2001] Redhead, M., Lipton, P., Douven, I., and Bueno, O. (2001). Quests of a Realist. *Metascience*, 10(3):341–366.

[van Fraassen, 1980] van Fraassen, B. C. (1980). *The Scientific Image*. Oxford University Press, Oxford, UK.

[Votsis, 2003] Votsis, I. (2003). Is Structure Not Enough? *Philosophy of Science*, 70(5):879–890.

[Yudell, 2010] Yudell, Z. (2010). Melia and Saatsi on structural realism. *Synthese*, 175(2):241–253.

[Zahar, 2004] Zahar, E. G. (2004). Ramseyfication and structural realism. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, 19(1):5–30.