Ramsey equivalence

April 17, 2017

In the literature over the Ramsey-sentence approach to structural realism, there is often debate over whether structural realists can legitimately restrict the range of the second-order quantifiers, in order to avoid the Newman problem. In this paper, I argue that even if they are allowed to, it won't help: even if the Ramsey sentence is interpreted using such restricted quantifiers, it is still an implausible candidate to capture a theory's structural content. To do so, I use the following observation: if a Ramsey sentence did encode a theory's structural content, then two theories would be structurally equivalent just in case they have logically equivalent Ramsey sentences. I then argue that this criterion for structural equivalence is implausible, even where frame or Henkin semantics are used.

1 Introduction

In the literature over the Ramsey-sentence approach to structural realism, there is often debate over whether structural realists can legitimately restrict the range of the second-order quantifiers, in order to avoid the Newman problem. In this paper, I argue that even if they are allowed to, it won't help: even if the Ramsey sentence is interpreted using such restricted quantifiers, it is still an implausible candidate to capture a theory's structural content.

The structure of the argument is as follows. In section 2, I introduce the Ramsey sentence and the standard semantics used to interpret it. In section 3, I explain how the Newman problem poses difficulties for the realist credentials of a view that wants to use the Ramsey sentence as the content of a theory. Among other things, we will see that the Ramsey-sentence approach, combined with full semantics, leads to an implausibly weak criterion of theoretical equivalence. In section 4, I consider a first pass at a response based on restricting the range of the second-order quantifiers: one which uses so-called *frame* semantics. I show that this leads to an implausibly strong criterion of theoretical equivalence (*viz.*, one stronger than definitional equivalence). So in section 5, I consider the use of Henkin semantics as a putative "Goldilocks" option, and show that it leads to a criterion of theoretical equivalence that is (at least) not too strong a criterion, since it is strictly weaker than definitional equivalence. However, section 6 shows that Henkin semantics is at risk of collapsing back into full semantics (with the associated resurrection of the Newman problem); and even if this could be resisted, section 7 argues that the associated criterion of theoretical equivalence is too weak anyway (and that it is highly unlikely that any way of amending it is to be found).

2 The Ramsey sentence

In what follows, I will suppose that the theory T with which we begin comprises a set of sentences of first-order logic.[1] Obviously, this isn't a realistic assumption, but it's appropriate 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 §5 below). Finally, we suppose that the vocabulary of T is bifurcated: it consists of a "benign" vocabulary β and a (wholly disjoint) "problematic" vocabulary π.[2]

The Ramsey sentence of T is then 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 $R_i \in \pi$ occurring in $\bigwedge T$ by an n-ary second-order variable X_i, thereby obtaining an open second-order sentence; we

[1] I won't require that T be deductively closed. Nothing major hinges on this, since including the deductive consequences does not essentially alter the Ramsey sentence so obtained: if $T \vDash \phi$, then $T^R \equiv (T \cup \phi)^R$, even if the standard of equivalence is that given by frame semantics (see below).

[2] I borrow this terminology from (Frigg and Votsis, 2011, p. 247)

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 T contains κ-many sentences, and if λ-many predicates from π occur in T, then T^R must be in a second-order language that permits κ-size conjunctions, and permits the introduction of λ-many second-order quantifiers. Since we will be looking at models, it is important to specify the vocabulary of the Ramsey sentence's ambient language (since syntactically identical sentences of different languages will have distinct classes of models). I will think of the Ramsey sentence itself, T^R, as a sentence of $L^2(\beta)$—that is, of second-order logic (with whatever infinitary resources are necessary) over vocabulary β. But at points, it will also be helpful to consider $\widetilde{T}^R$, the syntactically identical sentence of $L^2(\beta \cup \pi)$.

In order to assess the content of the Ramsey sentence, we need to specify how second-order sentences are to be interpreted: that is, we have to specify a second-order semantics. The standard semantics (also known as the *full* semantics) for second-order logic goes as follows. A *full structure* $\mathcal{M}$ for a language with vocabulary ξ consists of the same data as a first-order structure for vocabulary ξ: that is,

- A non-empty set $|\mathcal{M}|$
- For each $n \in \mathbb{N}$, for each n-ary $R \in \xi$, a set $R^{\mathcal{M}} \subseteq |\mathcal{M}|^n$

We call the set $R^{\mathcal{M}}$ the extension of R in $\mathcal{M}$; in general, for any n, I'll refer to any set of n-tuples of elements of $|\mathcal{M}|$ as an (n-ary) *extension* over $\mathcal{M}$ (whether or not it is the extension of some particular predicate). Recall that if two such structures $\mathcal{M}^+$ and $\mathcal{M}$, for vocabularies ξ^+ and $\xi \subset \xi^+$ respectively, are such that $|\mathcal{M}| = |\mathcal{M}^+|$ and for every $R \in \xi$, $R^{\mathcal{M}} = R^{\mathcal{M}^+}$, then $\mathcal{M}^+$ is said to be an *expansion* of $\mathcal{M}$ to ξ^+; and that $\mathcal{M}$ is said to be the *reduct* of $\mathcal{M}^+$ to ξ, and denoted $\mathcal{M}\restriction_{\xi}$.

A full structure evaluates formulae of the second-order language relative to a first-order variable-assignment g (a map from the first-order variables to elements of $\mathcal{M}$), and a second-order variable-assignment G (an arity-respecting map from the second-order variables to extensions over $\mathcal{M}$). For atomic formulae using second-order variables, the relevant clause is

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

whilst for formulae formed using the second-order quantifiers, the clause (for X an n-ary variable) is

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

where G^X_E is a variable-assignment just like G, save that it assigns E to X. In other words, according to the full semantics, the second-order quantifiers range over all extensions over the structure.

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, it follows from the above that a full structure $\mathcal{M}$ for the vocabulary β is a model of the Ramsey sentence T^R iff there is an expansion $\mathcal{M}^+$ of $\mathcal{M}$ which is a model of T:[3] that is, if there is some way to define extensions for the π-predicates on $\mathcal{M}$ in such a way as to satisfy T, then $\mathcal{M}$ satisfies T^R. Correspondingly, a full structure $\mathcal{N}$ for the vocabulary $\beta \cup \pi$ is a model of $\widetilde{T}^R$ iff there is an expansion $\mathcal{N}^*$ of the reduct $\mathcal{N}\restriction_\beta$ which is a model of T: that is, if there is some way to redistribute the extensions of the π-predicates on $\mathcal{N}$ in such a way as to satisfy T, then $\mathcal{N}$ satisfies $\widetilde{T}^R$.

The standard way of explaining why this result is problematic is as follows. Let $\mathcal{W}$ be "the world"—or, better, let it be a structure for the vocabulary $\omega \cup \theta$ which faithfully represents the world (i.e., which has the extensions of the predicates distributed in just the fashion that the actual relations are actually distributed in the world). Say that a theory is *observationally adequate* if it has a model $\mathcal{M}$ which is observationally isomorphic to $\mathcal{W}$: i.e., which is such that $\mathcal{M}\restriction_\omega \cong \mathcal{W}\restriction_\omega$.[4] The above result then shows that $\widetilde{T}^R$ is true (of the world, i.e. is satisfied by $\mathcal{W}$) if and only if T is observationally adequate.[5] Or, equally: T^R is satisfied by $\mathcal{W}\restriction_\omega$ iff T is observationally adequate. This is then a Bad Thing, if the Ramsey sentence was supposed to be part of a realist strategy,

[3](Ketland, 2004, Theorem 2)

[4]This notion of observational adequacy bears some resemblance to van Fraassen's notion that a theory is empirically adequate if it has an empirical substructure that can be embedded in the world (van Fraassen, 1980, chap. 3)—although (as a referee pointed out) van Fraassen rejects the bifurcation of a vocabulary into theoretical and observational predicates in the manner applied here. For further discussion, see (Ainsworth, 2009, pp. 145–146) or (Lutz, 2012, §4.2).

[5](Ketland, 2004, Theorem 6), (Ainsworth, 2009, Theorem 2)

since realists (by definition) are those committed to more than just the observational adequacy of scientific theories.[6]

There's an alternative way of getting at what's problematic about Ramseyfication: one framed not in terms of the relationship between theories and the world, but rather in terms of theoretical equivalence. I prefer this way of putting things, since it enables us to finesse difficult questions about representation and the theory-world relation,[7] and it's this way of setting up the problem that I'll refer back to in later sections. 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.

However, 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 $\omega \cup \theta_1$ and $\omega \cup \theta_2$ respectively.[8] 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-isomorphic model of T_2, and vice versa. It is straightforward to show from the above that 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. That is, two theories have equivalent Ramsey sentences if and only if their classes of models agree on how many things there might be, and on how the ω-structure could be distributed over those things.

Either of these ways of putting the problem points to the same conclusion: that if the structural realist uses this Ramseyfication procedure to elaborate their view, then structural realism collapses into ω-realism (realism only about the observational structure of the world). In particular, if we Ramseyfy *all* the predicates of the theory (i.e. if $\beta = \varnothing$) then the only information retained by the theory is, at best, information about cardinality. However, it will be worth going into a little more detail about exactly how Ramseyfication is in tension with realism, since that will enable us to introduce some apparatus to be used later.

First, it doesn't follow from the above that in all cases, T^R must be regarded as expressing less than T. For example, consider the following theory of colours, T_c. It

[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. I don't have space to discuss this here: see Ainsworth (2009) for (what I take to be) convincing replies.

[7]For a sense of quite how much there is to be said about these questions, see Frigg and Nguyen (2017).

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

has two observable predicates, R (for "is red") and B (for "is blue"), and two theoretical predicates S ("is made of scarlatineal particles") and A ("is made of azurean particles"). The theory has the following axioms:

$$\begin{aligned} &\forall x \neg (Rx \wedge Bx) \\ &\forall x (Rx \leftrightarrow Sx) \\ &\forall x (Bx \leftrightarrow Ax) \end{aligned} \tag{1}$$

In this case, there is a one-to-one correspondence between the models of the theory and those of its Ramsey sentence T_c^R: given any model $\mathcal{M}$ of the Ramsey sentence, there is a *unique* expansion $\mathcal{M}^+$ which is a model of T_c (and as per usual, given any model $\mathcal{N}$ of T_c, its unique reduct $\mathcal{N}\restriction_\omega$ is a model of T_c^R). So there is a sense in which in this case, Ramseyfication does not wash out any theoretical content: the models of T_c^R are just as fine-grained as those of T_c.

However, T_c is pretty special (and, indeed, pretty dubious): it implicitly defines its theoretical vocabulary in terms of its observational vocabulary.[9] That is, recall that given a vocabulary ξ and some predicate $R \notin \xi$, a theory T in vocabulary $\xi \cup \{R\}$ *implicitly defines* R *in terms of* ξ iff for any two models $\mathcal{M}$ and $\mathcal{N}$ of T, if $\mathcal{M}\restriction_\xi = \mathcal{N}\restriction_\xi$ then $\mathcal{M} = \mathcal{N}$.[10] As a result, once we know the observational structure of a model of T_c, we already know exactly which model of T_c we are dealing with.

It is clear both that T_c implicitly defines S and A in terms of R and B, and that it is exactly for this reason that the models of T_c^R (which, recall, are just the ω-reducts of the models of T_c) have unique expansions to models of T_c. If a theory T does not implicitly define its theoretical vocabulary in terms of its observational vocabulary, then the passage from T to T^R obliterates some "merely theoretical" distinctions among the models of T. For, suppose that T has a pair of non-isomorphic models $\mathcal{M}$ and $\mathcal{N}$, such that $\mathcal{M}\restriction_\omega \cong \mathcal{N}\restriction_\omega$. It is natural to describe this by saying that $\mathcal{M}$ and $\mathcal{N}$ describe a pair of possibilities which are observationally equivalent, but theoretically distinct. In moving from T to T^R, however, $\mathcal{M}$ and $\mathcal{N}$ are reduced to just the one model (up to isomorphism)—so T^R, unlike T, does not countenance distinctions between possibilities which differ only with regards to how the theoretical relations

[9] For further details of definability theory, see (Hodges, 1997, chap. 2) or (Suppes, 1957, chap. 8).

[10] If one is worried about the fact that this depends on identity between models, rather than just isomorphism, then one can use the following definition instead: T implicitly defines R in terms of ξ iff for any two models $\mathcal{M}$ and $\mathcal{N}$ of T, for any ξ-isomorphism $f : \mathcal{M} \to \mathcal{N}$, f is an isomorphism from $\mathcal{M}$ to $\mathcal{N}$. (Note the resemblance to the definition of strong global supervenience (McLaughlin and Bennett, 2014).) These two definitions are easily shown to be equivalent.

are distributed.

There is a second way in which T_c is peculiar: it explicitly defines all its theoretical vocabulary in terms of its observational vocabulary. That is, recall that given a vocabulary ξ and some n-ary predicate $R \notin \xi$, an *explicit definition of R in terms of ξ* is a formula δ_R of the form

$$\forall x_1 \dots \forall x_n (Rx_1 \dots x_n \leftrightarrow \tau_R(x_1, \dots, x_n)) \tag{2}$$

where τ_R is an n-place formula in the language over ξ. A theory T with vocabulary $\xi \cup \{R\}$ explicitly defines R in terms of ξ iff $T \vDash \delta_R$, for some explicit definition δ_R of R in terms of ξ. Finally, given disjoint vocabularies ξ and η, a theory T with vocabulary $\xi \cup \eta$ explicitly defines η in terms of ξ iff for every $R \in \eta$, T explicitly defines R.

Clearly, T_c explicitly defines S and A in terms of R and B. In fact, given that T_c is a first-order theory, this follows from the fact that it implicitly defines them: Beth's Theorem tells us that if a first-order theory T implicitly defines η in terms of ξ, then it explicitly defines η in terms of ξ. (The converse direction—that if T explicitly defines η in terms of ξ, it implicitly defines η in terms of ξ—is immediate, and holds in many more logical contexts than Beth's Theorem does.)

These resources let us give a fresh gloss to the problem facing the Ramseyfying realist. At least traditionally, scientific realism is associated with three commitments.[11] Metaphysically, the realist is committed to the existence of an external, mind-independent world, which possesses more than observational structure. Epistemically, they are committed to the possibility of knowledge of that world, and of (at least some aspects of) its non-observational structure. And semantically, they are committed to a literal interpretation of scientific claims: they deny that scientific discourse is reducible to or definable in terms of a mere observation-language.

So the Ramseyfying realist is faced with a dilemma. Either they think that (in general) the theoretical predicates of a scientific theory are explicitly definable in terms of the observational predicates, or they do not. If they do, then they fail to meet the semantic criterion; and, moreover, they are saddled with a view about scientific theories which is implausible on empirical grounds. If they do not, then by Beth's theorem they also think that the theoretical predicates are not implicitly definable, from which it follows that the Ramsey sentence fails to preserve theoretical structure: it fails to admit possibilities which agree on their observable structure, but disagree on

[11] I take these from (Chakravartty, 2016, §1.2)

theoretical structure. So if the Ramsey sentence is the sum total of what we can know *vis-a-vis* such a theory, then our Ramseyfying realist either fails to meet the epistemic criterion (if they believe that possibilities can indeed differ in such purely theoretical ways) or the metaphysical criterion (if they do not).

Now, one could challenge the sharpness of this dilemma, by pointing out that it relies upon the original theory being first-order: in second-order logic (for example), implicit definability does not entail explicit definability. So, it would be possible to have a second-order theory which does not explicitly define its theoretical predicates in terms of its observational predicates, and yet which is such that there is a one-to-one correspondence between its models and those of its Ramsey sentence. Two comments are worth making in response. First, the Beth property is not unique to first-order logic; so this escape route will require committing to the claim that theories ought (in general) to be formulated in logics lacking the Beth property. In the absence of an independent justification for that claim, it looks like a rather *ad hoc* move. Second, even the claim that any scientific theory implicitly defines all its theoretical terms is rather implausible. It amounts to the claim that in a respectable scientific theory, models of the theory are uniquely individuated by the totality of their observational structure. But this is certainly false, at least of our best current theories: it is entirely possible, for instance, to have a pair of distinct solutions to a particle-scattering event that would result in the same pattern of detector excitations, given that such detectors are not perfect (nor omnipresent).

One virtue of setting things up this way is that the analogous dilemma can be posed for other ways of cashing out the Ramsey-sentence approach. For example, Melia and Saatsi (2006) argue that we need not Ramseyfy away *every* non-observable predicate—merely that we need to Ramseyfy away all theoretical predicates. That is, suppose that one took the vocabulary to be *tri*furcated, into "observational" predicates (those which apply only to observable things, e.g. "is red"), "theoretical" predicates (those which apply only to unobservable things, e.g. "has colour charge red"), and "generic" predicates (those which apply to both, e.g. "has mass").[12] Melia and Saatsi (2006) argue that the structural realist is not obliged to Ramseyfy the mixed predicates, but only the theoretical ones:

> There is nothing in the spirit of structural realism that implies that *all*

[12] This provides some reason to avoid a two-sorted framework as a means of formalising the distinction between observables and unobservables (contra Ketland (2004)): if they are represented as sorts, then one cannot have generic predicates in this sense.

> predicates which can apply to unobservables should be Ramseyfied away. The structural realist thinks we cannot know certain aspects of the nature of the unobservable world, but that structural aspects of the unobservable world can be known. This is quite compatible with the structural realist retaining some interpreted predicates for unobservables.[13]

In the terms of this essay, they argue that β should be identified with the observational *and mixed* vocabulary, and π with the observational vocabulary.

But even if we accept this, and allow the Ramseyfying realist to refrain from Ramseyfying away the mixed predicates, this solution only helps insofar as we think that one's realist credentials are retained by this more moderate form of Ramseyfication (as Melia and Saatsi note). Clearly, the formal results outlined above transfer immediately, so we have the same pair of options. On the one hand, it could be that, in general, theories implicitly (or explicitly) define their theoretical predicates in terms of observational and mixed predicates; on the other, it could be that they do not. So the Ramseyfying realist needs to tell us which horn of the dilemma they choose to grasp, and to defend the claim that grasping that horn is consistent with all three realist commitments.[14] Although there is much to be said about the permissibility of doing so, it's not my main concern in this paper, so I forebear from further discussion.

4 Frame semantics

The strategy I wish to focus on instead is that of restricting the range of the second-order quantifiers, so that they only range over "real" rather than "fictitious", or "important" rather than "trivial", relations. Such a strategy was canvassed by Newman himself, who argued that although this would solve the Newman problem, such a distinction was inadmissible.[15] It seems fair to say that the consensus within the philosophy-of-science literature has followed Newman on these two points: that this strategy would indeed rescue the Ramsey-sentence approach, but that (at least insofar as the Ramsey sentence is supposed to be expressive of a *structuralist* position) the strategy is impermissible, since it relies on a non-structural distinction between different classes of objects. For instance, Psillos argues as follows:

[13] (Melia and Saatsi, 2006, p. 569)

[14] Melia and Saatsi's "ray theory" example may be understood as an argument that the first horn is untenable: that is, as providing an example of a recognisably scientific theory which does not implicitly define its theoretical predicates in terms of its mixed and observable predicates.

[15] See (Newman, 1928, 145ff.).

> In order to pick out as important one among the many relations which generate the same structure on a domain, we have to go beyond structure and talk about *what* these relations are, and *why* some of them are more important than others.
>
> One thing should be clear. If triviality is to be avoided, then *some* restrictions should be imposed on the relations defined on a given domain: not all subsets of the power set of the domain of discourse should be taken into account. Some of them must be excluded. Instead of a 'full structure', the domain should already possess a more restricted, but more definite, structure. In other words, the domain should be structured by a *definite* relation. The natural suggestion here is that among all those relations-in-extension which generate the same structure, only those which express *real* relations should be considered. But, as I have already noted, specifying which relations are real requires knowing something *beyond* structure, viz. which extensions are 'natural', i.e. which subsets of the power set of the domain of discourse correspond to natural properties and relations. Having specified these *natural* relations, one may abstract away their content and study their structure. But if one begins with the structure, then one is in no position to tell *which* of the relations one studies and *whether* or not they are natural.[16]

It is not so obvious to me that this strategy is impermissible: insofar as I have an intuitive grasp on what counts as "structural" or not, I find it hard to see why the presence of objective unity among certain classes of objects or tuples (but not others) counts as a non-structural feature of the world. (After all, it's natural to think that an abstract algebraic group is a "purely structural" notion; yet that is consistent with—indeed, consists in—the fact that certain sets of tuples over the domain of the group are privileged.)

However, whether Psillos' argument succeeds is by the by for my purposes. I want to argue that *even if* the structural realist is allowed to distinguish between the natural classes and the non-natural classes, the Ramsey sentence still fails to deliver an account of theoretical content which is acceptable to structural realists. I will show that even if the Ramsey sentence is interpreted using more restricted (rather than full) structures, the account of theoretical equivalence which is delivered is either implausibly strong

[16](Psillos, 1999, p. 66); see also the discussion in (Ainsworth, 2009, §6).

or implausibly weak. I take such a result to be useful to those (like me) who find it hard to get a clear sense of whether the use of non-full structures is allowed for the structural realist: whatever the outcome of that debate might be, the use of such structures will not (by itself) be enough to save the Ramsey-sentence approach to structural realism.

To look at the details of this, we need to get more precise about what the proposal looks like. The intuitive idea is that we add a little more data to the structures used to interpret second-order sentences: we include information about which sets of tuples are privileged (i.e., over which such sets the second-order quantifiers will range). More formally, a structure equipped with this extra data is known as a *frame* (or a *pre-structure*) for second-order logic.[17] A frame $\mathcal{F}$ (for vocabulary ξ) consists of

- A non-empty set $|\mathcal{F}|$
- For each $n \in \mathbb{N}$, a set $\mathcal{E}_n^{\mathcal{F}}$ of subsets of $|\mathcal{F}|^n$; let $\mathcal{E}^{\mathcal{F}} := \cup_{n\in\mathbb{N}}\mathcal{E}_n^{\mathcal{F}}$
- For each $n \in \mathbb{N}$ and n-ary $R \in \xi$, a set $R^{\mathcal{F}} \in \mathcal{E}_n^{\mathcal{F}}$

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 $L^2(\xi)$: each set $\mathcal{E}_n^{\mathcal{F}}$ gives the range of the second-order n-ary quantifiers. 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 (for X an n-ary variable) via the clause

- $\mathcal{F}[G, g] \models \exists X\phi$ iff for some $E \in \mathcal{E}_n^{\mathcal{F}}$, $\mathcal{F}[G_X^A, g] \models \phi$

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 much stronger if interpreted over frame semantics rather than full semantics. Certainly, it is strong enough to block the Newman problem as discussed above, as the following example illustrates.

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

$$T = \{\exists! x(x = x), \forall x Rx\} \tag{3}$$

Clearly,

$$T^R = \exists X\, (\exists! x(x = x) \wedge \forall x Xx) \tag{4}$$

[17]The below follows (Manzano, 1996, chap. 4).

Since $\beta = \varnothing$, β-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:[18] 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.

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 in the way of definability theory.[19] Suppose we have a single benign vocabulary β but disjoint problematic vocabularies π_1 and π_2. A *dictionary* for π_2 in terms of $\beta \cup \pi_1$ is a set Δ of explicit definitions, one for each $R \in \pi_2$, in terms of $\beta \cup \pi_1$. Given a dictionary Δ for π_2 in terms of $\beta \cup \pi_1$, any first-order $(\beta \cup \pi_1)$-structure $\mathcal{S}$ can be converted into a $(\beta \cup \pi_2)$-structure $\Delta(\mathcal{S})$, by first taking the unique expansion of $\mathcal{S}$ to $\beta \cup \pi_1 \cup \pi_2$ that satisfies Δ, and then taking the reduct to $\beta \cup \pi_2$.

Now suppose that T_1 and T_2 are a pair of first-order theories, in signatures $\beta \cup \pi_1$ and $\beta \cup \pi_2$ respectively. A dictionary Δ for π_2 in terms of $\beta \cup \pi_1$ 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

[18]In (something like) the sense of Lewis (1983).
[19]Again, see (Hodges, 1997, chap. 2) or (Suppes, 1957, chap. 8) for more details.

terms of T_2, then we will say that T_1 and T_2 are *mutually interpretable*. 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 β includes, at least, all the observational vocabulary (even if we follow Melia and Saatsi in also including the mixed 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.[20] 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 equivalence. If they are to see off the pessimistic meta-induction,[21] 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.[22] 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 $T_1 = \{\exists x Fx\}$, $T_2 = \{\exists x \neg Gx\}$ (where $\beta = \varnothing$, $\pi_1 = \{F\}$, and $\pi_2 = \{G\}$). T_1 and T_2 are definitionally equivalent, by the dictionaries

$$\begin{aligned} \Delta &= \{\forall x(Gx \leftrightarrow \neg Fx)\} \\ \Delta' &= \{\forall x(Fx \leftrightarrow \neg Gx)\} \end{aligned} \tag{5}$$

However, their Ramsey-sentences

$$\begin{aligned} T_1^R &= \exists X \exists x Xx \\ T_2^R &= \exists X \exists x \neg Xx \end{aligned} \tag{6}$$

[20]For arguments to this effect, see Glymour (1970), Glymour (1977).

[21]Recall that an originating motivation for structural realism (in Worrall (1989)) was to explain how the realist could resist the pessimistic meta-induction: the concern that, given that many of our past theories have been radically wrong about the ontological content of the world, we should conclude that our current best theories are similarly strewn with radical ontological mistakes (Laudan, 1981).

[22]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)

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 (n-ary) 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*. However, as we change what language ϕ may be written in, we will get different conceptions of definability, and hence different notions of Henkin structure.[23] Moreover, we may want consider not just (mere) definability, but the broader notion of *definability with parameters*. An (n-ary) extension E over a frame $\mathcal{F}$ is definable with parameters if there is (a) some formula ψ with $k \geq n$ free first-order variables and $m \geq 0$ second-order variables, (b) $k - n$ individuals from $|\mathcal{F}|$, and (c) m 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 $|\mathcal{F}|$ is definable over $\mathcal{F}$ by a formula such as $x = x$, and so it at least would have to be included in $\mathcal{E}^{\mathcal{F}}$. So, we might hope that by a judicious choice of definability, we can ensure that definitionally equivalent theories will generate equivalent Ramsey sentences. And indeed, this is the case: 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, it will turn out that not even definitional

[23] 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.

equivalence, but merely mutual interpretability, is sufficient. In the remainder of this section, I demonstrate this.

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 ξ, we say that an n-ary extension E over $\mathcal{F}$ is *first-order definable with second-order parameters* if there is some first-order formula $\psi \in L^2(\beta)$ 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 $a_1, \ldots, a_n \in |\mathcal{F}|$,

$$\langle a_1, \ldots, a_n \rangle \in E \Leftrightarrow \mathcal{F}[Y_i \mapsto E_i, x_j \mapsto a_j] \models \psi \tag{7}$$

where $Y_i \mapsto E_i$ denotes any second-order variable-assignment G such that $G(Y_i) = E_i$ (for all $1 \leq i \leq m$), and $x_j \mapsto a_j$ denotes any first-order variable-assignment g such that $g(x_j) = a_j$ (for all $1 \leq j \leq n$). We then have the following result.

Proposition 1. Suppose that T_1 and T_2 are two mutually interpretable theories, in signatures $\beta \cup \pi_1$ and $\beta \cup \pi_2$ respectively (where $\beta = \{S_i\}_i$, $\pi_1 = \{R^1_p\}_p$, and $\pi_2 = \{R^2_q\}_q$). Then T_1^R (wherein each R^1_p has been replaced by a variable X^1_p) is logically equivalent to T_2^R (wherein each R^2_q has been replaced by a variable X^2_q), on the Henkin 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 β) 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. In the below, we will derive a contradiction, by showing that $\mathcal{H} \models T_2^R$.

To begin, we have that $\mathcal{H}[G] \models T_1^*$, for some second-order variable-assignment G. So consider the first-order structure $\mathcal{M}$ for vocabulary $\beta \cup \pi_1$ defined by

$$|\mathcal{M}| = |\mathcal{H}| \tag{8}$$

$$S_i^{\mathcal{M}} = S_i^{\mathcal{H}} \tag{9}$$

$$(R^1_p)^{\mathcal{M}} = G(X^1_p) \tag{10}$$

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

Since T_1 and T_2 are mutually 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. We use this model to generate a second-order variable assignment G' such that $\mathcal{H}[G] \models T_2^*$ (which will then show that

$\mathcal{H} \models T_2^R$).

So, let G' be the second-order variable-assignment such that:

- for any second-order variable X_q^2 occurring in T_2^R,

$$G'(X_q^2) = (R_q^2)^{\Delta(\mathcal{M})} \tag{11}$$

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

I show that G' is a variable-assignment for $\mathcal{H}$, i.e., that for any second-order variable X, $G'(X) \in \mathcal{E}^{\mathcal{H}}$. Clearly, this is true by stipulation for any X not occurring in T_2^R. So, suppose that X does occur in T_2^R, i.e. that $X = X_q^2$ for some q.

For the associated predicate $R_q^2 \in \pi_2$, we know that the formula $\tau_{R_q^2}$ (occurring in the definition $\delta_{R_q^2} \in \Delta$ of R_q^2 in terms of $\beta \cup \pi_1$, as per equation (2)) defines the extension $(R_q^2)^{\Delta(\mathcal{M})}$ in $\mathcal{M}$. It does not define the extension in $\mathcal{H}$, however, since $\mathcal{H}$ is of signature β and $\tau_{R_q^2}$ is of signature $\beta \cup \pi_1$.

But now observe that for each p, $G(X_p^1) \in \mathcal{E}^{\mathcal{H}}$ (since G was an assignment for $\mathcal{H}$); that means, by equation (10), that $(R_p^1)^{\mathcal{M}} \in \mathcal{E}^{\mathcal{H}}$. So now consider the formula $\tau^*_{R_q^2} := \tau_{R_q^2}[Y_p/R_p^1]$, i.e., the formula obtained by uniformly substituting variables Y_j for the predicates R_p^1 in $\tau_{R_q^2}$. Note that $\tau^*_{R_q^2}$ is a first-order formula of $L^2(\beta)$: i.e., it has second-order variables but no second-order quantifiers, and is of signature β.

If the second-order variables Y_p are assigned to the parameters $(R_p^1)^{\mathcal{M}}$, then $\tau^*_{R_q^2}$ defines the extension $(R_q^2)^{\Delta(\mathcal{M})}$. But since all of those parameters are in $\mathcal{E}^{\mathcal{H}}$, that means that $\tau^*_{R_q^2}$ parametrically defines $(R_q^2)^{\Delta(\mathcal{M})}$ in $\mathcal{H}$, relative to assigning Y_p to $(R_p^1)^{\mathcal{M}}$.

Given that $\mathcal{H}$ is closed under first-order definability with second-order parameters, it follows that $(R_q^2)^{\Delta(\mathcal{M})} \in \mathcal{E}^{\mathcal{H}}$. By equation (11), this means that $G'(X_q^2) \in \mathcal{E}^{\mathcal{H}}$, as we were required to show. 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. □

Since definitional equivalence entails mutual interpretability, this proposition shows that (this form of) Henkin semantics will rule out problematic examples such as Example 2. 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, in the final two sections, I turn to two problems with Henkin semantics: the first that it collapses back into full semantics, and the second that—even if such a collapse is resisted—it is associated with an implausibly weak criterion of equivalence.

6 Trivialisation

First, the risk of collapse. The basic form of the worry is that with a sufficiently liberal conception of definability, perhaps it will turn out that all extensions over a frame are definable, and hence 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
>
> $$x \text{ is } a\text{, and } y \text{ is } \alpha \text{ or } \beta \text{ or } \gamma \tag{12}$$

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 of signature ξ, and that for every $a \in |\mathcal{H}|$, there is some constant $\alpha \in \xi$ 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 definability with first-order parameters, 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 definable with first-order parameters.

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

$$\bigvee_{\lambda}(x_1 = y_1^\lambda \wedge x_2 = y_2^\lambda \wedge \cdots \wedge x_n = y_n^\lambda) \tag{13}$$

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,[24] 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 (12) above.[25]

Melia and Saatsi (2006) consider a similar proposal: rather than the "real/fictitious" distinction, they consider the distinction between *qualitative* properties (those "tied

[24] 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.

[25] (Newman, 1928, p. 145)

to what the objects are *like*, the kinds of *features* that they have, the *qualities* that they possess")[26] and non-qualitative properties (properties such as "*being identical to a, b or c*")[27].

The most natural way of making this precise is to restrict ourselves to definability *without* first-order parameters. After all, as we have seen, we do not need first-order parameters for Proposition 1. 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 $E \in \mathcal{E}^{\mathcal{F}}$,

$$\langle a_1, \ldots, a_n \rangle \in E \Leftrightarrow \langle h(a_1), \ldots, h(a_n) \rangle \in E \tag{14}$$

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

$$\mathcal{F}[Y_i \mapsto P_i, x_j \mapsto a_j] \models \psi \Leftrightarrow \mathcal{F}[Y_i \mapsto P_i, x_j \mapsto h(a_j)] \models \psi \tag{15}$$

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

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

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.

[26](Melia and Saatsi, 2006, p. 577)
[27](Melia and Saatsi, 2006, p .578)

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 respite for the Ramsey-philiac:[28]

> 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.[29]

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. Thus, the very coherence of a position that uses Henkin structures rather than full structures is in question: barring restrictions on the defining language, a rigid Henkin structure just is a full structure.

7 The weakness of mutual interpretability

Finally, 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

[28] 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.

[29] (Melia and Saatsi, 2006, p. 578)

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 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 mutual interpretability is a sufficient condition for Ramsey-equivalence with respect to this semantics. But the following example shows that mutual interpretability is an implausible criterion of theoretical equivalence, since there are mutually interpretable theories which one has good reason to consider inequivalent.[30]

Example 3. Let $\beta = \varnothing$, $\pi_1 = \{P\}$, and $\pi_2 = \{Q, R\}$ (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 structure of signature $\beta \cup \pi_1$, whilst T_2 is not. Yet they are mutually interpretable. Consider the following dictionary Δ for π_2 in terms of $\beta \cup \pi_1$:

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

This is a translation manual for T_2 in terms of T_1. For, given any model $\mathcal{M}$ of T_1 (i.e.,

[30] Further examples besides this one may be adduced: for instance, the pair of theories in Example 4 of Barrett and Halvorson (2015) are mutually interpretable. This shows that completeness of theories is not preserved under mutual interpretability.

any structure of signature $\beta \cup \pi_1$), $\Delta(\mathcal{M})$ is a structure of signature $\beta \cup \pi_2$ 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 π_1 in terms of $\beta \cup \pi_2$:

$$\forall x(Px \leftrightarrow Qx) \tag{17}$$

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 structure of signature $\beta \cup \pi_1$—and hence, a model of T_1. So T_1 and T_2 are mutually interpretable.[31]

One response, of course, is to think that some further ingenious tweak to the notion of a Ramsey sentence will see off the problem. I'll now take a moment to explain why I am pessimistic about the prospects of doing so.

First, note that Example 3 will persist as a problem if we move to any 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 Proposition 1 will apply, and mutual interpretability will be sufficient for logical equivalence of the Ramsey 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 Ramsey sentences (i.e., Ramsey sentences satisfied by exactly the same structures); and hence, it leads to a more liberal criterion of equivalence. But Example 3 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 3. 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 2. Suppose that $\mathcal{F}$ is a frame of signature $\beta = \{S_i\}_i$, 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 extensions

[31] Thus, by Proposition 1, they have Henkin-equivalent Ramsey sentences. This may be seen directly by observing that the Ramsey sentence of T_2 is

$$T_2^R = \exists X_1 \exists X_2 \forall x(X_1 x \rightarrow X_2 x) \tag{18}$$

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

$E_1, \ldots, E_m \in \mathcal{E}^{\mathcal{F}}$ (with E_i of the same arity as Y_i), the extension defined by ψ with respect to the parameters $E_1, \ldots, E_m$ 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 Ramsey sentences are not logically equivalent (with respect to that semantics).

Proof. First, define $\pi_1 = \{R_1^1, \ldots, R_m^1\}$. Then take some particular sequence of extensions $E_1, \ldots, E_m \in \mathcal{I}^{\mathcal{F}}$, with E_i of the same arity as Y_i, and define a first-order structure $\mathcal{M}$ of signature $\beta \cup \pi_1$ as follows:

$$|\mathcal{M}| = |\mathcal{F}| \tag{19}$$

$$S_i^{\mathcal{M}} = S_i^{\mathcal{F}} \tag{20}$$

$$R_i^{\mathcal{M}} = E_i \tag{21}$$

Now, let

$$T_1 = \{\phi : \mathcal{M} \models \phi\} \tag{22}$$

Second, let $\pi_2 = \{R_1^2, \ldots, R_m^2, R\}$, and let

$$T_2 = \{\phi[R_i^2/R_i^1] : \phi \in T_1\} \cup \{\forall x_1 \ldots \forall x_n(Rx_1 \ldots x_n \leftrightarrow \psi[R_i^2/Y_i](x_1, \ldots, x_n)) \tag{23}$$

In other words, we obtain T_2 by uniformly substituting the R_i^2 for the R_i^1 everywhere in T_1, and then adjoining an explicit definition of R in terms of $\beta \cup \pi_2$—where τ_R is given by ψ, with the R_i^2 uniformly substituted for Y_i. Clearly, T_1 and T_2 are definitionally equivalent (since T_2 is a definitional extension of a notational variant of T_1).

I now show that $\mathcal{F} \models T_1^R$. First, let G be a second-order variable-assignment such that $G(X_i^1) = E_i$ (which is permitted, since all $E_i \in \mathcal{E}^{\mathcal{F}}$). A straightforward proof by induction will demonstrate that for any $\phi \in L^1(\beta \cup \pi_1)$ and any first-order variable-assignment g for $\mathcal{M}$, if $\mathcal{M}[g] \models \phi$ then $\mathcal{F}[G, g] \models \phi[X_i^1/R_i^1]$. Thus, since $\mathcal{M} \models \phi$ for all sentences $\phi \in T_1$, and given the definition of T_1^*, we have that $\mathcal{F}[G] \models T_1^*$; and hence, $\mathcal{F} \models T_1^R$.

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

$$\mathcal{F}[G'] \models \forall x_1 \ldots \forall x_n(Xx_1 \ldots x_n \leftrightarrow \psi[X_i^2/Y_i](x_1, \ldots, x_n)) \tag{24}$$

But now let $E_i' := G'(X_i^2)$, and let $E' := G'(X)$. Then it follows that for any $a_1, \ldots, a_n \in$

$|\mathcal{F}|$,

$$\langle a_1, \ldots, a_n \rangle \in E' \iff \mathcal{F}[Y_i \mapsto E'_i, w, x_j \mapsto a_j] \models \psi \tag{25}$$

Hence, E' is defined by ψ with respect to the parameters $E'_1, \ldots, E'_m$. But then by hypothesis, $E' \notin \mathcal{E}^{\mathcal{F}}$, and so G' is not a variable-assignment for $\mathcal{F}$ after all. So by contradiction, the proposition follows. □

The hypothesis of Proposition 7 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 3. This strongly 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 mutual 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 mutual 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.[32] 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 captures the

[32]See Barrett and Halvorson (2015).

"real" content of a theory, then that (together with a choice of second-order semantics) naturally induces an associated criterion of theoretical equivalence;[33] 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 can only specify the content of a theory insofar as we can say 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: hitherto the philosophers have only *extracted* the content of a theory, in various ways; the point, however, is to *abstract* it.

References

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

Barrett, T. W. and Halvorson, H. (2015). Glymour and Quine on Theoretical Equivalence. *Journal of Philosophical Logic*, 45(5):467–483.

Chakravartty, A. (2016). Scientific Realism. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

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

Frigg, R. and Nguyen, J. (2017). Models and representation. In Magnani, L. and Bertolotti, T., editors, *Springer Handbook of Model-Based Science*. Springer, Berlin; New York.

Frigg, R. and Votsis, I. (2011). Everything you always wanted to know about structural realism but were afraid to ask. *European Journal for Philosophy of Science*, 1(2):227–276.

[33] cf. Coffey (2014)

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

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

Hodges, W. (1997). *A Shorter Model Theory*. Cambridge University Press, Cambridge, UK.

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

Laudan, L. (1981). A confutation of convergent realism. *Philosophy of science*, 48(1):19–49.

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

Lutz, S. (2012). *Criteria of Empirical Significance: Foundations, Relations, Applications*. PhD thesis, University of Utrecht.

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

McLaughlin, B. and Bennett, K. (2014). Supervenience. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2014 edition.

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

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

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

Suppes, P. (1957). *Introduction to Logic*. Van Nostrand Reinhold, New York.

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

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

Worrall, J. (1989). Structural Realism: The Best of Both Worlds? *Dialectica*, 43(1-2):99–124.

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