Axiomathes (2022) 32:101–121
https://doi.org/10.1007/s10516-020-09511-w

ORIGINAL PAPER

Behaviourism in Disguise: The Triviality of Ramsey Sentence Functionalism

T. S. Lowther[1]

Received: 17 April 2020 / Accepted: 27 August 2020 / Published online: 11 September 2020

Abstract
Functionalism has become one of the predominant theories in the philosophy of mind, with its many merits supposedly including its capacity for precise formulation. The most common method to express this precise formulation is by means of the modified Ramsey sentence. In this article, I will apply work from the field of the philosophy of science to functionalism for the first time, examining how Newman's objection undermines the Ramsey sentence as a means of formalising functionalism. I will also present a formal variation on Newman's objection through mathematical induction. Together, these proofs suggest that functionalism formalised by the Ramsey sentence trivially reduces to a kind of behaviourism plus a cardinality constraint on the number of relations holding between mental-relevant behaviours. As most functionalists see functionalism as a distinct theory of mind from behaviourism, this suggests that the modified Ramsey sentence cannot form a satisfactory formalism for functionalism.

Keywords Mind · Functionalism · Behaviourism · Triviality · Induction

1 Introduction

Since its inceptions in the 1960s, functionalism has risen to be one of the predominant theories in the philosophy of mind (Churchland 2005, p. 33). The core thesis of functionalism is that mental states can be fully characterised as a set of functions defined by their inputs, which are both input mental states and material stimuli, and their outputs, which are output mental states and material responses (Block 1978, p. 262; Churchland 2005, p. 33), such that each function captures the causal relation holding between any one mental state and its inputs, outputs, and relations to other mental states (Lewis 1972, p. 256). The appeal of functionalism is

✉ T. S. Lowther
toby.lowther@ling-phil.ox.ac.uk

[1] Department of Experimental Psychology, University of Oxford, Oxford, UK

grounded in its conceptual clarity, its amenability to formal representation, and its promise to reconcile the multiple realizability of mental states with token–token physicalist intuitions.

Our present discussion focuses on the second of these appeals: the claimed amenability to formal representation of functionalist theories. With its express focus on functional definability and relational structure, it is clear why functionalism should be amenable to formalism: functions and relations are exactly the kind of property best captured by formal systems. Conversely, this focus on functions and relations also gives reason to doubt a functionalist theory which cannot produce a satisfactory formalism: relations and functions are the kinds of things which are typically captured by formal systems, and hence, if no such formal system can be presented, this raises serious questions over what, exactly, the claims of the functionalist thesis are taken to be. Thus, there is a twofold motivation for the functionalist to pursue formal presentation: on the positive side, to provide clarity to the theory; and on the negative side, to demonstrate the substance of the theory.

Since at least the work of Lewis (1972), the mainstream answer to the problem of formalising functionalism has been found in the Ramsey sentence. The Ramsey sentence is a method of formalising theories developed through the work of Ramsey (1929) to formalise the claims of structuralist theories in the philosophy of science. The Ramsey sentence was widely adopted for the development of structural realism, the theory that scientific theories truly describe the structure of the world, even insofar as the theoretical substance of a scientific theory may be false (cf. Worrall 1989). The Ramsey sentence was thus intended to capture the structure encoded in a theory, and thus, a way of capturing the part of a scientific theory claimed to be amenable to truth evaluation by the structuralist thesis. The parallels between structuralist theories of science and the functionalist theory of mind are evident, and thus, Lewis and others adopted the Ramsey sentence as a method for capturing the ‘structure’ of the psychological theories with which functionalism is concerned.

Although Ramsey sentence formalisms for functionalism remain popular in the philosophy of mind, the popularity of Ramsey sentence structuralist theories has rapidly declined in the philosophy of science, in no small part due to the work of M.H.A. Newman. Newman’s 1928 article provided an informal argument against Russell’s structuralist theory, arguing that all such structuralist theories trivially reduce to the ‘observable’ parts of the theory plus a cardinality constraint on the number of theoretical terms. In 2004, Jeffrey Ketland provided a formal proof demonstrating that the same arguments held for the Ramsey sentence method, which has widely been considered a potential death-knell for at minimum Ramsey sentence structuralist theories.

Despite these damning arguments against the Ramsey sentence method of capturing the structure of the theory, the Ramsey sentence remains the predominant way to formalise functionalist theories. It is against this state of affairs that the present argument takes aim.

The goals of the present study are threefold. Firstly, I will demonstrate that Newman’s Objection is not confined to the philosophy of science but applies equally against functionalist theories which make use of the Ramsey sentence method. Secondly, I will present a formal variation on the Newman-Ketland argument, demonstrating that the issue at stake is deeper than Newman's or Ketland's presentations of the argument and represents a true and fundamental issue with the Ramsey sentence as a formalism. Thirdly and finally, I will begin to explore the implications for functionalism if the Ramsey sentence formalism fails: both the prospects of other approaches to formalising functionalism, and the prospects of rejecting a formalism altogether.

I will conclude that the proofs found in Newman's and Ketland's work, together with the further proof presented here, conclusively show that a functionalism based on the Ramsey sentence formalism trivially reduces to a certain variety of behaviourism. This leaves the functionalist with three options: pursue a different formalism; reject the possibility of formalising functionalism; or claim that these results (and those of Godfrey-Smith's triviality criticisms of computational functionalism) point to a deeper triviality issue which undermines the functionalist paradigm altogether. I will provide some brief initial thoughts on the prospects of each approach, but a full survey of these implications will require further research.

The major limitations of the study primarily result from limitations of scope, especially with respect to the wider implications of the results here developed in application to functionalist theory as a whole. Due to the limited space of a single article, I am unable to here properly examine the implications of these results beyond their immediate impact on functionalism as formalised by the Ramsey sentence to wider functionalist theories, which will need to be the purview of future research.

2 Defining Terms

Before we can turn to the arguments under consideration, our discussion will be much aided by a clarification and definition of key terminology.

Functionalism shall be taken to be the theory that mental states can be fully characterised by their inputs, outputs, and relations to other mental states.

A formalism is defined as a method whereby an informal theory is translated into some formal expression, typically in logic.

A Ramsey sentence formalism is thus the method whereby an informal theory is converted into a Ramsey sentence. The method proceeds in the following manner. First, we begin with some theory, defined broadly as a set of propositions which say something which may be either true or false of the world and which a person may or may not believe, which consists of a set of propositions relating certain instances or events. In the case of functionalism, this will be a psychological theory, formed either from explication of our ordinary concepts or from empirical science, which gives a set of propositions relating mental states to their inputs and outputs.

When they reformulate this theory in second-order predicate logic into a conjunction of predicates, of the form Θ:

$$\Theta[S_1 \ldots S_n, I_1 \ldots I_n, O_1 \ldots O_n]$$

where S_i is a mental state predicate, I_i is an input proposition, and O_i is an output proposition. Next, we replace each occurrence of a given term for a mental state with an appropriate variable and introduce an existential quantifier for that variable over our theory, producing the Ramsey sentence of Θ, $\Re(\Theta)$:

$$\Re(\Theta) = \exists X_1 \ldots \exists X_n \Theta[X_1 \ldots X_n, I_1 \ldots I_n, O_1 \ldots O_n]$$

where X_i is an appropriate variable. A Ramsey sentence is therefore defined as the product of a Ramsey sentence formalism. The Ramsey sentence of a theory is defined as the product of applying a Ramsey sentence formalism to that theory. We shall then define Ramsey sentence functionalism thus:

> Ramsey sentence functionalism: The functionalist theory that mental states are fully characterised by the Ramsey sentence produced by applying the Ramsey sentence formalism to an appropriate psychological theory.
> Ramsification is defined as the process of a Ramsey sentence formalism: a Ramsey sentence is reached by the Ramsification of a theory.

The notion of the modified Ramsey sentence was further introduced by Lewis (1972). A modified Ramsey sentence is defined as a Ramsey sentence in which the existential quantifiers are replaced with a modified existential quantifier, introduced by Lewis in parallel with Russell's iota operator, such that that $\exists_1 \times \varphi[x]$ abbreviates the expression $\exists y \forall x (\varphi[x] \leftrightarrow y = x)$ (Lewis 1972, p. 253). Thus, the modified Ramsey sentence of Θ above is $\Re'(\Theta)$:

$$\Re'(\Theta) = \exists_1 X_1 \ldots \exists_1 X_n \Theta[X_1 \ldots X_n, I_1 \ldots I_n, O_1 \ldots O_n]$$

where $\exists_1 X_1$ is a modified existential quantifier expression. We can then define the modified Ramsey sentence formalism as the formalism which has an informal theory as its input and a modified Ramsey sentence as its output, and we can define modified Ramsey sentence functionalism as follows:

> **Modified Ramsey sentence functionalism**: The functionalist theory that mental states are fully characterised by the modified Ramsey sentence produced by applying the modified Ramsey sentence formalism to an appropriate psychological theory.

Strictly, it is not Ramsey sentence functionalism which is widespread, but modified Ramsey sentence functionalism, as seen in the work of Lewis (1972), Block (1978), and Shoemaker (1981), among others. Most of the arguments that follow will apply equally to Ramsey sentence functionalism and modified Ramsey sentence functionalism: however, as the work of Ketland specifically addresses the Ramsey sentence and not the modified Ramsey sentence, it is worth keeping the distinction between the two in mind.

3 The Case for Modified Ramsey Sentence Functionalism

Before examining arguments against modified Ramsey sentence functionalism, and Ramsey sentence functionalism more generally, we should examine the reasons why this particular formalism has proven so popular amongst functionalist theories, and why it has taken so long for the issues with the Ramsey sentence observed by

Newman's objection to be acknowledged in the field of functionalism. I will present three such apparent virtues of the Ramsey sentence formalism as applied to functionalist theories.

Firstly, the Ramsey sentence is well established as a method of formalisation and is therefore transparent. As the method of Ramsification is to some degree standardised, it prevents ad hoc formal variations, which aids in the clarity of discussion around functionalist theories.

Secondly, the Ramsey sentence formalism is simple. As such, the Ramsey sentence (in theory) should not introduce any unexpected implications for 'higher' issues of the philosophy mind, such as chauvinism/liberalism or mental causation, which did not exist in the base theory. This is important for functionalists because it ensures that the debates being had are questioning functionalism itself, rather than some particular formalism, which is often seen as secondary to the core theory.

Thirdly, the Ramsey sentence formalism makes minimal assumptions on the organisation of the base theory for its expression. All that is required of the theory to be Ramsified is that it can be expressed as a set of relational predicates, which if it is to be an appropriate theory for functionalist theses is necessary in any case. This is particularly important for formalism, where the proper theoretical base for formalism forms one of the major debates in the theory, couched in the debate between Functionalism and Psychofunctionalism, as Block (1978, p. 269) terms them. Contrast this with a computational formalism, where the differences in organising principles between folk psychology and scientific theory may require different methods of abstraction to reach the necessary abstract structures and mapping principles.

Functionalism is served by having a transparent, simple, and minimally demanding formalism, and this appears to be exactly what the Ramsey sentence formalism offers. This explains some of the popularity of Ramsey sentence functionalism.

4 Newman's Objection Applied to Functionalism

4.1 Newman's Informal Objection

In 1928, M.H.A. Newman presented a criticism of Russell's 'Causal Theory of Perception'. Although directed towards Russell's particular theory, the fundamentals of this criticism can be expanded to any structuralist theory of this kind, including those which are based on the Ramsey sentence. Let us summarise the objection with an example.

Suppose we have an event space consisting of any four events, which we denote by {A, α, β, γ}. For a mentalistic example, suppose we are concerned with a pain state and its outputs: {A} is the state of 'in pain', {α} is pulling a body part away from the source of pain, {β} is shouting in pain, and {γ} is a disposition to describe oneself as being in pain. Now it should be clear that three causal relations are captured in our rudimentary theory of pain: R(A,α), R(A,β), and R(A,γ).

Now let us call this rudimentary pain theory T, and let us take the modified Ramsey sentence $\Re'$ (T):

$$\Re'(T) = \exists_1 X \exists_1 (x) T[X(x,\alpha), X(x,\beta), X(x,\gamma)]$$

This is the modified Ramsey sentence of our rudimentary pain theory. However, it is *also* the modified Ramsey sentence of various other theories. For example, suppose we define a relation P over the same set of events, where P is the relation 'an event denoted by letters from different alphabets'. It should be immediately clear that P has exactly the same extension as R above: out theory of events gives us the relations P(A,α), P(A,β), and P(A,γ). As such, if assume that A still constitutes a 'mental' event, the theory T_1, which is based on our P relation, will also have the same modified Ramsey sentence as our rudimentary pain theory T:

$$\Re'(T_1) = \exists_1 X \exists_1 (x) T[X(x,\alpha), X(x,\beta), X(x,\gamma)] \ = \ \Re'(T)$$

Newman demonstrates that this is not a problem only with these toy examples. Rather, from the basic principles of set theory, it can be shown that for any given aggregate A, a system of relations between its members can be found with *any* assigned structure compatible with a cardinality constraint, where systems A and B are defined as having the same structure iff a 1-to-1 correlation can be set up between the members of A and B such that, if two (or more) members of A instantiate a given relation R, their correlates in B instantiate a given relation S, and vice versa (Newman 1928, p. 139).

If this is so, it is clear that structures such as the Ramsey sentence do not capture as much information as we should suppose. Consider our examples above: if $\Re'$ (T_1) and $\Re'$ (T) are identical, then the modified Ramsey sentence formalism must only contain information common to both of our toy theories: namely, that some system of unique relations relates four entities.

Thus, Newman's argument holds that the only information captured in the statement that there exists a systems of relations on A with a certain structure is the cardinality of A (Newman 1928, p. 140); *any* given collection of things can be organised to have a certain given structure, provided there are the right number of them (Newman 1928, p. 144).

4.2 Ketland's Formal Proof

Newman's argument was informal in nature and targeted towards Russell's particular structuralist theory. However, Demopoulos and Friedman (1985) argue that these same criticism apply directly to the Ramsey sentence method, and this was demonstrated formally in 2004 by Jeffrey Ketland.

Ketland's original formulation was targeted towards the Ramsey sentence as applied in epistemic structural realism (ESR), a particular theory in the philosophy of science. In what follows, I will summarise Ketlands' formal proof appropriate adapted to application to a functionalist context.

We begin by differentiating three kinds of predicate in our theory (Ketland 2004, p. 289), which in a functionalist context will be: mental terms, which refer to the names, properties, and relations of mental states; observational terms, which refer to the names, properties, and relations of physical and behavioural states; and mixed terms, which refer to the input–output relations between mental states and physical (behavioural) states.

In order to pre-emptively respond to one criticism of Ketland's proof, Ketland's formulation requires that theoretical (here, mental) terms only apply to theoretical entities, observational terms only apply to physical entities, and only mixed terms apply to both. Cruse (2005) objects to this distinction when applied to structural realism, arguing that in the sciences, it is entirely possible for an observable property to be applied to a theoretical entity, or vice versa: a 'red rose' and a 'red bloodcell' have the same property of *redness*.

This may be a valid criticism of Ketland's approach when applied to the sciences, but it is not at all clear that the same criticism applies in the functionalist case. If thoughts can be 'red', it is not at all clear that they can be 'red' in the same sense that a rose is 'red'. Mental predicates apply to mental entities; physical predicates do not apply to mental entities. Thus, it seems to be in accordance with functionalist theories that the only predicates which should apply to both mental and physical entities are those causal predicates relating a given mental states to its physical inputs and outputs.

Let us now proceed to the proof. We begin by developing a formal framework in which we have an interpreted, two-sorted, second-order language, with individual variables ranging over two domains (observable and mental entities) and predicates, referring to properties and relations, of three types (observable, mental, and mixed).

We then let (D_O, O) be the structure associated with the intended interpretation of the observational part of the language, such that D_O is the set of observable objects and O represents the set {O_1, O_2, …} of the (sets of the) observable properties and relations of the observable predicates of the language (Ketland 2004,p. 296). We then take an arbitrary full structure of the language ((D_1, D_2), R_O, R_M, R_T), where (D_1, D_2) are the two-sorted domains of the language and R_O is the set {$R_{1.1}$, $R_{1.2}$, …} of observational predicates over D_1, R_T is the set {$R_{3.1}$, $R_{3.2}$, …} of mental predicates over D_2, and R_M is the set of mixed predicates over D_1 x D_2. From this, we define what it is for a structure to be *empirically correct*:

Definition 1 A structure ((D_1, D_2), R_O, R_M, R_T) is *empirically correct* iff its reduct (D_1, R_O) is isomorphic to (D_O, O). (cf. Ketland's 'Definition E', 2004, p. 296; Ainsworth's 'Definition 1', 2009, p. 145.)

That is, a structure is empirically correct iff the reduct of the observational part of the structure is isomorphic to the structure of the observable world, relative to the relevant predicates (Ainsworth 2009, p. 145). Let us then assume that the Ramsey sentence of a theory in this language is obtained by Ramseyfying both theoretical and mixed predicates (that is, replacing each with the appropriate variables and existentially quantifying as appropriate). It follows, then, that:

Theorem 1 *The Ramsey sentence of a theory A is true iff there is some sequence of relations (R_M, R_T) such that ((D_O, D_t), O, R_M, R_T) $\models$ A (cf. Ketland's 'Theorem 4',* 2004, p. 293; Ainsworth's 'Theorem *1*', 2009, *p. 146.)*

where D_T is the set of mental objects in the world. We then define a cardinality condition, *T-cardinality correctness*, which will function as the cardinality condition of Newman's Objection:

Definition 2 $((D_1, D_2), R_O, R_M, R_T)$ is *T-cardinality correct* iff $|D_2| = |D_T|$. (cf. Ketland's 'Definition G', 2004, p. 298; Ainsworth's 'Definition 2', 2009, p. 146.)

In other words, a full structure in the language is *T-cardinality correct* iff the mental domain of the structure has the same cardinality as the actual mental domain, i.e., the domain of mental objects in the world. We can then prove the following theorem:

Theorem 2 *The Ramsey sentence of a theory A is true iff A has a model that is empirically correct and T-cardinality correct. (cf. Ketland's 'Theorem 6', 2004, p. 298; Ainsworth's 'Theorem 2', 2009, p. 146.)*

Ketland provides proofs of both Theorems 1 and 2 above (Ketland 2004, pp. 292–293, 298–299), while Ainsworth provides a proof of Theorem 2, taking Theorem 1 to be intuitively clear (Ainsworth 2009, pp. 146–147). A proof is also given by Votsis (2003, pp. 881–882). Ketland's proof of the crucial Theorem 2 is as follows:

Proof For the left-to-right direction, suppose $\Re(\Theta)$ is true. So, there exists a (full) expansion $((D_O, D_T), O, R_M, R_T)$ of the reduct $((D_O, D_T), O)$ of the intended structure, such that $((D_O, D_T), O, R_M, R_T)$ satisfies Θ. This full model of Θ is empirically correct and T-cardinality correct.

For the right-to-left direction, suppose that Θ has a T-cardinality correct and empirically correct full model $\mathbf{M}$, such that $M = ((D_1, D_2), R_O, R_M, R_T)$. Because this model is empirically correct, the observational reduct (D_1, R_O) is isomorphic to the actual observational content: $(D_1, R_O) \approx (D_O, O)$, given by some bijection φ: $D_1 \rightarrow D_O$. Because this model is T-cardinality correct, there is another bijection ψ: $D_2 \rightarrow D_T$. We use these bijections φ and ψ to define a (full) structure $\mathbf{M} = ((D_O, D_T), O, R_M{*}, R_T{*})$ and show that this satisfies Θ. We use the isomorphisms φ: $D_1 \rightarrow D_O$ and ψ: $D_2 \rightarrow D_T$ to define new mixed relations $(R_M)_i^*$, and new mental relations $(R_t)_i^*$ as follows:

$$(R_M)_i^* =_{\mathrm{df}} \left\{ \left(\left(\varphi(\underline{x}), \psi\left(\underline{y}\right) \right) : \left(\underline{x}, \underline{y}\right) \right) \in (R_M)_i \right\}$$

$$(R_T)_i^* =_{\mathrm{df}} \left\{ \psi\left(\underline{y}\right) : \left(\underline{y}\right) \in (R_T)_i \right\}$$

where x and y are sequences of appropriate lengths. We have now combined φ and ψ into a 2-sorted isomorphism (φ, ψ):

$$(\varphi, \psi) : ((D_1, D_2), R_O, R_M, R_T) \rightarrow ((D_O, D_T), O, R_M{*}, R_T{*})$$

Hence, $((D_O, D_T), O, R_M{*}, R_T{*}) \models \Theta$. Hence, $((D_O, D_T), O) \models \Re(\Theta)$. So, $\Re(\Theta)$ is true. (Ketland 2004, pp. 298–299.) □

The implication of the theorem is that the truth of the Ramsey sentence of a theory is guaranteed by its empirical correctness and a simple cardinality constraint;

the truth of the Ramsey sentence is trivial under these conditions, and whatever *structural* content is carried by a Ramsey sentence beyond the observational content of the theory just *is* the cardinality constraint (Ketland 2004, p. 299).

4.3 Conclusions of the Objections with Respect to Functionalism

The conclusion of these objections—Newman's objection and Ketland's formal proof thereof—is thus that any theory which can be Ramsified without loss of empirical content is trivial. Whatever 'structural' information is captured by a Ramsey sentence amounts to no more than one or another kind of cardinality constraint.

5 A Formal Variation of the Objection

5.1 The Necessity of Formal Variation and a Further Argument

The informal argument from Newman and Ketland's formal proof should provide sufficient evidence for the challenges I wish to present against Ramsey sentence functionalism. However, two possible objections may be raised against an argument based on these alone.

The first possible concern is one of formal variation. Based on these arguments alone, it remains possible that the issue at stake may be with how Ketland's formal system was developed, rather than with the Ramsey sentence itself. Indeed, many of the criticisms that have been presented against Ketland's proof in the field of the philosophy of science (cf. Ainsworth 2009, p. 163; Cruse 2005; Melia and Saatsi 2006; Psillos 1999) amount to accusations that how Ketland defines the parts of his proof misrepresents what is intended by the use of the Ramsey sentence in the ESR context.

It is therefore necessary to present a formally distinct argument to the same conclusion, which would effectively demonstrate that the issue at stake is a substantive issue with the Ramsey sentence formalism itself, rather than an issue with the particular formal structure which Ketland developed to debate the triviality of the Ramsey sentence.

The second concern is that neither Newman's nor Ketland's arguments were targeted at the kind of modified Ramsey sentence formalism that Lewis advocates and which has been widely adopted in functionalist thought, but rather at the simple Ramsey formalism. As the introduction of a uniqueness criterion is a substantive modification of the formal structure, it is worth examining in finer detail whether the same results can be obtained regarding the modified Ramsey formalism.

Both of these concerns may be assuaged in what follows: a formally distinct argument, based on mathematical induction, which takes aim at the modified Ramsey sentence formalism directly.

5.2 Complexity and the Strong Induction Principle

Proofs by mathematical induction in the metatheory of logic rely on properties of the set of natural numbers. Under the Peano arithmetic, the following 'strong' principle of induction applies to all subsets of the natural numbers:

$$\text{If, for each}\, n \in \mathbb{N}, \mathrm{F}(m)\text{for all }\, m < n \text{ implies}\, \mathrm{F}(n), \text{then}\, \mathrm{F}(n)\text{for all}\, n \in \mathbb{N}$$

In other words, if a certain property applies to the lowest value in any subset of natural numbers, and it can be shown that, for any arbitrary number above that lowest value in the subset, the property belonging to *all* lower values entails it applying to that arbitrary value, it follows that the property belongs to *all* values in the subset.

In order to use this principle in our proof, we must therefore have a means by which we relate our various possible modified Ramsey sentences of mental theories to a subset of the natural numbers. Traditionally in logical metatheory, this is done by assignment of complexity.

Each modified Ramsey sentence φ can be assigned a numerical complexity, C[φ], by counting the number of connectives and quantifiers. In calculating the complexity of a given modified Ramsey sentence, it is important to remember that the modified existential quantifier $\exists_1 x\varphi[x]$ abbreviates the expression $\exists y\forall x(\varphi[x] \leftrightarrow y = x)$, and thus counts as four connectives/quantifiers for the sake of complexity.

Thus, let us take the case of the modified Ramsey sentence reached from a theory containing exactly one mental state, one input, and one output, which we shall call $\mathfrak{R}_{\mathrm{IMO}}$:

$$\mathfrak{R}'_{\mathrm{IMO}} = \exists_1 X\boldsymbol{\Theta}[X, I, O]$$

According to the standard procedure for forming the Ramsey sentence, the predicates within the theory are simply conjoined. Thus, $\mathfrak{R}'_{\mathrm{IMO}}$ includes one modified existential quantifier and two conjunctions, and hence, C[$\mathfrak{R}'_{\mathrm{IMO}}$] = 6.

Complexity forms a subset of the natural numbers, and hence, the strong principle of induction applies across complexity. Therefore, there are two things we must show to have our proof by natural induction. Firstly, we must show that in some 'base case', the modified Ramsey sentence with minimum complexity, that this modified Ramsey sentence has the property of encoding nothing more than the 'observable' parts of the underlying theory plus a cardinality constraint on the number of mental states. Secondly, we must demonstrate that for some Ramsey sentence of arbitrary complexity, that if all modified Ramsey sentences of lower complexity has this property of encoding only 'observable' predicates plus a cardinality constraint, then the modified Ramsey sentence of arbitrary complexity has this property. Once this has been demonstrated, the strong principle of induction tells us that *all* modified Ramsey sentences have this property; namely, that all modified Ramsey sentences encode only 'observable' predicates plus a cardinality constraint on the mental.

5.3 Proof in the Base Case

Let us start by defining our base case. For the inductive proof to hold, this case should be that modified Ramsey sentence, based on a psychological theory, which has minimal complexity. Now there is a slight complication here, because there are six different cases which could all be argued to form a base case, giving us four different complexity values for that base.

The mathematically simplest case would be the modified Ramsey sentence formed from a psychological theory including only exactly one input, or exactly one output. In both cases, the complexity comes out as zero:

$$\begin{aligned}&\mathfrak{R}'_{\mathrm{I}} = \boldsymbol{\Theta}[\mathrm{I}],\ \mathfrak{R}'_{\mathrm{O}} = \boldsymbol{\Theta}[\mathrm{O}]\\&\mathrm{C}\left[\mathfrak{R}'_{\mathrm{I}}\right] = \mathrm{C}\left[\mathfrak{R}'_{\mathrm{O}}\right] = 0\end{aligned}$$

However, it could be argued that a given theory only counts as a *psychological* theory if it has at least one mental predicate. Thus, it could be argued that the true base case is the modified Ramsey sentence formed from a psychological theory including exactly one mental state. This gives a complexity value of four:

$$\begin{aligned}&\mathfrak{R}'_{\mathrm{M}} = \exists_1 X\boldsymbol{\Theta}[X]\\&\mathrm{C}\left[\mathfrak{R}'_{\mathrm{M}}\right] = 4\end{aligned}$$

Further, one could argue that these cases do not present a truly *functionalist* modified Ramsey sentence theory, because the functionalist requires that mental states be defined in terms of inputs and outputs. Thus, an argument could be made that the base case for a functionalist modified Ramsey sentence is either the case of a modified Ramsey sentence based on a single mental state plus a single input or output; or, if we believe that every mental state is only functionally defined if given inputs *and* outputs, the base case would be that modified Ramsey sentence based on a psychological theory with a single mental state, and single input, and a single output.

$$\begin{aligned}&\mathfrak{R}'_{\mathrm{IM}} = \exists_1 X\boldsymbol{\Theta}[I, X], \mathfrak{R}'_{\mathrm{MO}} = \exists_1 X\Theta[X, O], \mathfrak{R}'_{\mathrm{IMO}} = \exists_1 X\boldsymbol{\Theta}[I, X, O]\\&\mathrm{C}[\mathfrak{R}'_{\mathrm{IM}}] = \mathrm{C}[\mathfrak{R}'_{\mathrm{MO}}] = 5\\&\mathrm{C}[\mathfrak{R}'_{\mathrm{IMO}}] = 6\end{aligned}$$

So that no one may accuse me of using an inappropriate base case in this proof, I shall demonstrate how the property under question—the property of encoding only ‘observables’ plus a mental cardinality constraint—applies to each of these cases in turn.

The demonstration is simple enough in the first three cases. In the case of $\mathfrak{R}'_{\mathrm{I}}$ and $\mathfrak{R}'_{\mathrm{O}}$, it is evident that the modified Ramsey sentence only encodes observational content, as the only content encoded is inputs and outputs, which at least under traditional functionalism would be described physically as either behaviour (under analytic functionalism) or perhaps some kind of neural impulse (under varieties of empirical functionalism).

In the case of $\Re'_M$, there is no observable content. However, only a little reflection on the meaning of the modified existential quantifier should reveal that the only content encoded here is the number of states. All that $\Re'_M$ states is that there exists some unique state *X*. However, this could clearly be fulfilled by *any* unique state whatsoever. We could reach this modified Ramsey sentence from formalising *any* theory which tells us that there exists some unique state, and therefore it encodes nothing more than that.

We turn now to the more difficult cases. As observable contents remain specified in the modified Ramsey sentence, let us assign values to I and O across $\Re'_{IM}$, $\Re'_{MO}$, and $\Re'_{IMO}$. Which specific values we assign to I and O are irrelevant to the argument at hand. To take a common example, let us take the input, *I*, to be 'pricking one's finger with a needle' and the output, *O*, to be 'exclaiming 'ouch''.

Now it is clear that with these inputs and outputs, $\Re'_{IM}$, $\Re'_{MO}$, and $\Re'_{IMO}$ could be theories of pain: namely, (i) that pain is caused by pricking one's finger, (ii) that being in pain causes one to exclaim 'ouch', and (iii) that pain is caused by pricking one's finger and causes one to exclaim 'ouch'.

However, given our early discussions of the content actually encoded in $\Re'_I/\Re'_O$ and $\Re'_M$, it should be clear that any number of relations could hold between any given input, output, and quantified predicate variable. The modified Ramsey sentence $\Re'_{IM}$ requires to be true only that there must be some unique predicate X that stands in some relation to the input I. However, it should be obvious that any number of theories could contain predicates that produce the same modified Ramsey sentence, so long as I was given as the input.

Thus, for our assignment of I and O, $\Re'_{IM}$ could be a formalism of a folk theory about the cause of a certain pain state. But it could also be a formalism of a folk theory of discomfort, or a physiological theory of drawing blood, or a neurological theory of skin pain receptors, and so on. Similarly, $\Re'_{IMO}$ could be a formalism of a folk theory of a certain kind of pain, but it could also be a formalism of a folk theory of discomfort, or a physiological theory of reflexes, or a neurological theory, and so on. The same applies to $\Re'_{MO}$.

Any of these theories could give us the structures of $\Re'_{IM}$, $\Re'_{MO}$, and $\Re'_{IMO}$ respectively. Therefore, each modified Ramsey sentence can't encode more information than whatever is common to all those theories which could produce that modified Ramsey sentence. If this were not the case, the modified Ramsey sentence would not be a true formalism, as for one or more of those theories it would either add or change content in the theory. What is common among those theories, of course, is simply the observational elements (I and O, as appropriate), and the statement that there exists some one unique predicate which relates to them—in other words, a cardinality constraint on the number of theorised predicates relating between them, which in functionalist terms amounts to a cardinality constraint on the number of mental predicates.

Thus, regardless of which base case we choose, we find that the property applies in the base case: the modified Ramsey sentence encodes nothing more than the observational elements (inputs/outputs) and a cardinality constraint on the number of mental states.

5.4 Proof to all Cases

We now turn to the inductive proof to all cases. First, let φ be some modified Ramsey sentence of arbitrary complexity.

For the purposes of our proof, let us now make an assumption, which I shall denote as our Induction Hypothesis (IH). The assumption is as follows:

> **IH**: Every modified Ramsey sentence ψ with complexity C[ψ] < C[φ] encodes only the 'observable' part of the theory (inputs and outputs) plus the number of uniquely quantified ('mental') predicates.

In other words, we assume that every modified Ramsey sentence ψ which has less complexity (fewer connectives and quantifiers) than φ has the same properties we observed in $\Re'_{\text{IMO}}$ and our other base cases: namely, that any number of theories could be found which produced the same modified Ramsey structure by formalisation, and hence, that the modified Ramsey sentence encoded nothing more than that which was common to all of them—the inputs, the outputs, and the number of uniquely quantified predicates, which under a functionalist interpretation are mental predicates.

Now for set of two modified Ramsey sentences {φ, ψ}, there are exactly three possible cases whereby C[ψ] < C[φ], namely:

C1. φ includes an additional conjunct on ψ, namely an additional input or output
C2. φ includes additional quantifiers on ψ, namely an additional quantified predicate
C3. Some combination of C1 and C2 in various degrees, i.e., φ includes some combination of additional inputs, outputs, and quantified predicates to ψ.

Let us examine each case in turn. Our purpose is to determine whether, in every case, φ would preserve from ψ the property of encoding only the 'observable' aspects of the theory and the number of quantified predicates.

Considering C1, it should be clear that if ψ has this property, φ does too. Adding an additional 'observable' predicate, be that an input or an output, will only increase the 'observable' content of the theory: it will have no impact on the mental implications of the theory. To see that this is the case, imagine we take the underlying theory which gave us $\Re'_{\text{IMO}}$ above, namely, < pricked finger, pain, 'ouch!' > , and we add another output, say, pulling the finger away, so that our final modified Ramsey sentence will have a structure like ℜ*:

$$\Re^{*} = \exists_{1}X\Theta[\mathit{pricked\,finger},\ X,\ \mathit{saying\ "ouch!"},\mathit{pulling\,finger\ away}]$$

Just as we could produce any number of alternative base theories which gave us the structure of $\Re'_{\text{IMO}}$, so too we can produce any number of alternative base theories which give us the structure of ℜ*: neurological theories, theories of proximate behaviour, even theories of the alphabetic ordering of sentences. This is always going to be true of the addition of observables, because although adding information may narrow down the number of relations that can hold between all the observables specified, we can always find more relations that could hold between those inputs and outputs, and as the modified Ramsey sentence only states that there is some unique relation (within a given domain) that *does* hold, it can’t make such alternative relations inadmissible.

Thus, we find that in case 1, at least, it seems that φ will only encode ‘observable’ predicates plus a cardinality constraint on the uniquely quantified predicates, if ψ does.

We may now turn to C2. In this case, we introduce a new quantified predicate to ψ, which is equivalent to having taken a base theory for φ with an additional mental state to those expressed in the base theory which gave us ψ. While it may not be as immediately clear in this case, I would argue that in this case also if ψ has the property under question, then so will φ. To demonstrate this point, let us take an example.

Suppose ψ is taken to be the modified Ramsey sentence produced by theory Θ_1, which includes two inputs, two outputs, and two functionally defined mental states which hold between these inputs and outputs. For the sake of argument, let Θ_1 have the inputs {‘pricked finger’, ‘eating cake’}, the outputs {‘saying ‘ouch!’’, ‘smiling’}, and the mental predicates ‘pain’ and ‘satisfaction’, which are defined such that ‘pain’ has the structure < ’pricking one’s finger’, ‘saying ‘ouch!’’ > and ‘satisfaction’ has the structure < ‘eating cake’, ‘smiling’ > . Ψ would therefore have the structure:

$$\Psi = \exists_1 X \exists_1 Y \boldsymbol{\Theta}[\mathit{prickedfinger},\ \mathit{eating\ cake}, X, Y,\ \mathit{saying\ "ouch!"}, \mathit{smiling}]$$

Now suppose we have another theory, Θ_2, which we formalise to give φ. Θ_2 is exactly like Θ_1, except that in addition to the predicates mentioned above, it has a ‘sadism’ state with the structure < ’pricking one’s finger’, ‘smiling’ > . φ would therefore have the structure:

$$\varphi = \exists_1 X \exists_1 Y \exists_1 Z \Theta[\mathit{prickedfinger},\ \mathit{eating\ cake}, X, Y, Z, \mathit{saying\ "ouch!"}, \mathit{smiling}]$$

However, it should be immediately clear that any number of predicate relations could satisfy Z, exactly as any number could satisfy X or Y. A neurological theory of our (presumably sadistic) participant could be formed to produce the same structure, as could any number of orthographic theories (for example, X relates events whose written expressions in English end in ‘r’ to those that end in punctuation, Y relates events whose written expressions in English end in a vowel to those ending in ‘g’, while Z relates events whose written expression in English ends in ‘r’ to those that end in ‘g’).

Now this example is of course a toy example, and the psychological theories with which functionalism concerns itself would include many more terms and a much broader domain. However, I hope that the mode of argumentation presented makes clear that the points here are general. In any given case, if ψ has the property under question—if ψ only encodes ‘observable’ predicates and a cardinality constraint on the mental—adding one more uniquely and existentially quantified predicate to ψ is not going to change this property, because it does not add any new information about how the observables and quantified predicates relate to each other: it only tells us that there is another one of those relations, whatever they may be. Thus, just as in

C1, we find that while the addition of a quantified predicate may narrow down the list of potential theories we can easily come up with that would form the basis for Ramsification, there will still be innumerably many such theories, and hence, if ψ has the relevant property, so too will φ.

C3 presents no issue, as it can always be broken down into repeated applications of C1 and C2, and as I have already argued that C1 and C2 preserve the property under question, it follows under reasonable assumptions that repeated application of these moves will also preserve that property.

Hence, given our induction hypothesis IH, that for all ψ such that $C[\psi] < C[\varphi]$, ψ encodes only the 'observable' content of its base theory plus a cardinality constraint on the number of quantified predicates, it follows that φ, too, encodes only the 'observable' content of its base theory plus a cardinality constraint. By the logical principle of conditional introduction, this entails that:

> if ψ encodes only 'observable' content plus a cardinality constraint for all ψ: $C[\psi] < C[\varphi]$, then φ encodes only 'observable' content plus a cardinality constraint.

You will recall that given the strong principle of induction, this conditional—plus the proof to the base case demonstrated in Sect. 4.2 above—is all that is required to conclude that for all modified Ramsey sentences φ, φ encodes only 'observable' content plus a cardinality constraint (on the number of quantified predicates).

Now I must stress that despite taking the form of a proof by mathematical induction, this is not a strictly formally complete proof. I have argued that we can always produce alternative theories which give the same modified Ramsey sentence formalism, but I have not strictly shown this in a mathematical or formal sense. This being said, I believe that the arguments I have given do demonstrate informally a fact that is a more or less direct product of the basic principles of the theory of sets and structures: that if we define a structure by means of existentially quantified relations, that structure will be instantiated by every possible relation holding between the relata which the existentially quantified relation holds between. In terms of modified Ramsey sentences, this means that our modified Ramsey sentence cannot contain more information than the fact that some relation holds between the relevant input and output, which in turn means that the modified Ramsey sentence is instantiated by every relation which holds between the input and the output. If this is true, it seems pretty clear that the modified Ramsey sentence does not include any information above and beyond the fact that the input and the output stand in some relation to each other, and from this, the proof follows.

6 Implications for Functionalism

6.1 Implications for Ramsey Sentence Functionalism

We thus have two formal variations which point to the same substantive conclusion: at least as far as functionalist theories go, a modified Ramsey sentence trivially reduces to the observable, or behavioural, part of the theory, plus a certain cardinality constraint on the number of mental states which the theory asserts.

The implications for Ramsey sentence functionalism seem clear and dire: if Ramsey sentence functionalism trivially reduces to a mental theory composed of its behavioural elements plus a cardinality constraint, then it is equivalent to a behaviourist theory—worse, an overcommitted behaviourist theory which requires a strict specification of the number of possible mental states a creature may instantiate.

For the behaviourist philosopher, this may be a positive outcome: it shows that a kind of Ramsey sentence functionalism can serve as a means of precisifying the behaviourist theory. However, much of the momentum of functionalist theories have been based on criticisms of the behaviourist model, and as such, if the functionalist wishes to remain a *functionalist*, they must reject Ramsey sentence functionalism.

This leaves the functionalist with two options: she can either pursue an alternative formalism, or she can reject formalism entirely. In the following sections I will briefly offer initial exploration of each of these approaches, although a thorough discussion and exploration of the prospects of either alternative formalisms or rejecting formalism falls well beyond the scope of this article.

6.2 Prospects of Alternative Formalisms

Although the Ramsey sentence formalism is the most common means of formalising functionalist theories, it is not the only formalism found in the literature, and the functionalist may seek to preserve a formal functionalist theory by adopting an alternative formalism. Besides variations of the Ramsey sentence, the other major class of formalisms for functionalism is that of computational formalisms (Godfrey-Smith 2009, p. 275).

The computational formalism begins by specifying a functional profile as a set of relations between abstract entities, understanding realisation in terms of a mapping between abstract and physical structures (Godfrey-Smith 2009, p. 275). The relevant notion of realisation is made explicit with reference to the concept of a combinatorial state automaton (CSA), an abstract structure in which the total inner state of a system is represented as a vector, or list, of substates (Godfrey-Smith 2009, p. 281; also cf. Chalmers 1996). We then define realisation of a functional system as follows:

> A physical system realises a given CSA during a time interval iff there is a mapping from states P of the physical system onto substates C of the CSA and from inputs and outputs of the physical systems onto inputs and outputs of the CSA, such that: for every state-transition $(\langle C_1, C_2, \ldots C_n \rangle, I) \rightarrow (\langle C'_1, C'_2, \ldots C'_n \rangle, O)$ of the CSA, if the physical system were to be in a combination of states $\langle P_1, P_n \ldots P_n \rangle$ that map to $\langle C_1, C_2, \ldots C_n \rangle$ during this time period, and received input I* that maps to CSA input I, then it would transition to a combination of substates $\langle P'_1, P'_n \ldots P'_n \rangle$ that map respectively to

> $\langle C'_1, C'_2, \ldots C'_n \rangle$, and would emit an output O* that maps to CSA output O. (Godfrey-Smith 2009, p. 282.)

where C_n is an internal state of the CSA and P_n is a physical state of the realiser.

This kind of formalism is not subject to the arguments raised above—Newman's, Ketland's or my own. However, it is not exempt from triviality concerns, as demonstrated by Godfrey-Smith (2009).

Godfrey-Smith notes that any functionally characterised physical systems can be broken down into a *transducer layer*, which connects the system to its environments (such as sensors and muscle fibres), and a *control system*, which is everything that is functionally important other than the transducer layer (2009, p. 284). Godfrey-Smith then demonstrates that any sufficiently complex physical system can be made into a behavioural duplicate of an intelligent agent solely through changes to the transducer layer of that system. Given the simple mapping criterion for realisation above, any system that can be given the behavioural profile of an intelligent agent is thereby made to realise the functional profile of that agent.

A change to the means of input and output should not alter *whether* a system has mental properties, even if it alters *which* mental properties a system has, and thus, every complex physical system already has the functional features that give intelligent agents mental properties (Godfrey-Smith 2009, p. 284). Thus, functionalism combined with a simple mapping account of realisation collapses into triviality of the same kind: if a system is behaviourally identical, it is functionally identical, and the result is a kind of behaviourism.

The arguments presented in this article demonstrate that functionalism formalised by the Ramsey sentence method reduces trivially to a variety of behaviourism. Godfrey-Smith's arguments demonstrate that the same is true for computational formalisms.

There remains a possibility of some other means of formalising functionalism. However, these arguments demonstrate that the two most popular kinds of formalism for functionalism are trivial, it is difficult to conceive of an alternative method which would not fall into the failings of either of the above methods—any method based on predicate logic and functions is likely to be subject to the same structuralist criticisms of Newman, Ketland, and the present argument, while a method based on mapping functions, such as a network approach, would likely fall into the same problems which Godfrey-Smith raises against a computational method. Thus, while it may be possible to find some alternative formalism, the prospects seem somewhat dire.

6.3 Prospects of Rejecting Formalism

One further possibility presents itselfperhaps the call for formalism itself is misplaced.[1] Under this position, the functionalist's response to the failure of Ramsey sentence functionalism and of the computational formalism should not be

[1] I owe this observation to an anonymous reviewer.

to pursue some alternative formalism, but should instead be to forgo formalism entirely.

Although I cannot here examine in detail the arguments surrounding this direction, it seems to me that there are initially two major issues that present themselves here, which any philosopher following this track would need to account for. These are, of course, only beginning exploratory considerations.

Firstly, it is unclear how functionalism can be made clear and precise without formalism. Philosophers such as Hinckfuss, Searle (1990), Putnam (1988), Chalmers (1996), and Copeland (1996) have presented arguments that functionalism trivially reduces to behaviourism regardless of the formalism used. With a formal description in hand, the functionalist can use this formalism to attempt to demonstrate the ways in which functional structure differs from a behavioural description; without such a formalism, it becomes more difficult to argue against these criticisms. The purpose of a formalism is to provide clarity. Without it, it may still be possible to make the theory clear and precise, but the degree of clarity will always be limited by the incoherence of the natural language chosen for its expression.

Secondly, however functionalism may be made precise, the failure of formalism would seem to indicate that it does not say what it purports to say.[2] Functionalism, as the name suggests, is intended to be concerned with functions: input–output functions relating mental states to one another, and to physical inputs and outputs. Functions, at least defined extensionally, are exactly the sort of thing which logico-mathematical formalism is intended to capture, and successfully captures in its applications in the sciences and elsewhere in philosophy. If mental states are *fully* characterised by their relations to inputs, outputs, and other mental states, these relations should be extensional functions. If these relations are intensionally defined, then a mental state is only fully characterised by these relations *plus* whatever intensional aspects ensure these relations are of the right kind, in which case functionalism, at least as it is normally portrayed, is false. Thus, we have an impasse: if functionalism, in the sense of the thesis that mental states are fully characterised by their relations to inputs, outputs, and other mental states, is true, then it should be amenable to formalisation; if functionalism is not amenable to formalisation, either it is false, or it means something other than what it claims to mean.

Thus, while it is eminently possible that functionalism is all of true, nontrivial, and not amenable to formalism, if this is the case, then it is apparent that functionalism is not the theory it claims to be, and it is eminently unclear how the theory can be made precise without formalism to determine what, in fact, it is.

6.4 Further Prospects for Functionalism

The arguments presented herein are not intended as a death-knell for the functionalist project as a whole, and indeed several options remain open to the

[2] I owe this point significantly to the contributions of F. D. C. Willard.

functionalist. However, these options are severely limited by the conclusions of this article.

One option is to claim that we simply have not yet found the correct formalism for functionalism, and that both the Ramsey sentence and the computational formalism were simply red herrings and dead ends. If this is so, then further research should be taken to determine what formalism may be developed which would not reduce trivially to a variety of behaviourism.

The second option is to reject formalism entirely. However, this raises the issue of how to make the theory precise without formalism, and how to account for the failure of the functionalist ‘functions’ to be captured by the logico-mathematical notion of a function. Further research in this direction should examine how to make an informal functionalist theory precise in such a way as to satisfy the precision of a formal theory, as well as providing some theoretical account as to what about functionalism makes it ill-suited to formalism and why it seemed so pre-eminently suited to formal treatment before.

The third and final option is to claim that the triviality of formal treatments of functionalism reveal a deeper issue within the theory itself, one which only became apparent when functionalism was made clear and precise through formal treatment—that while functions may provide a means for organising and precisifying information, they do not provide us with any information above and beyond that encoded in the inputs and outputs of those functions, effectively reducing functionalism to a particularly clear and precise variety of behaviourism.

7 Conclusions

7.1 Conclusions of the Present Study

In conclusion, then, I have carried over the insights from Newman’s objection to the philosophy of mind, demonstrating that the same objection Newman raised against structuralist theories in the philosophy of science applies to Ramsey sentence functionalism. By adapting Newman’s and Ketland’s arguments to a functionalist context and providing a formal variation of my own, I have demonstrated that functionalism formalised by the Ramsey sentence trivially reduces to behaviourism plus a cardinality constraint on the number of mental states.

With respect to methodology, these conclusions should ensure that the Ramsey sentence as a method of formalism for functionalism is roundly defeated, and where a formal treatment of functionalism is desired, an alternative treatment is preferred.

With respect to its implications for wider functionalist thought, these conclusions consign us to one of three paths. We can try to find an alternative formalism for functionalism, although the present arguments and those given by Godfrey-Smith against the computational formalism make the prospects for this direction suspect at best. We can try and reject formal treatments of functionalism entirely, although this opens up the question of how to make the functionalist thesis precise and requires a philosophical account for why the functions of functionalism are not amenable to logical treatment and why it appeared so suited to formalism before. Lastly, we can claim that the systematic failings of formal treatments of functionalism reveal a deeper issue with the theory, concluding that the triviality of Ramsey sentence formalism is more than ‘skin deep’, and that functionalism as a whole will ultimately reduce to behaviourism, however we make it precise.

I make no assertion as to which of these routes will prove most profitable, and I leave it to others to examine how to save functionalism without forgoing the clarity and precision which Ramsey sentence functionalism promised.

When I first discovered the functionalist thesis, it appeared to me a shining theory, a perfect union of multiple realizability and physicalist sentiment. I do not raise the criticisms of this article because I want to do away with functionalism, but because I believe if functionalism is to be all it once promised, it must stand the test and show that its formal basis is as sound as its theoretical. Whether or not it will, only time will tell. My hope is that from this little article, both functionalist and anti-functionalist alike will further appreciate the importance of our choices of formalism in the philosophy of mind.

7.2 Directions for Further Research

These findings suggest two major directions for further research. Firstly, there is clearly value in examining whether we can develop a formalism for functionalism which is not subject to the criticisms raised here against the Ramsey sentence or the criticisms given by Godfrey-Smith against the computational formalism. Secondly, it is clear that the field would benefit from further research into the implications of abandoning formalism for functionalism—the degree to which, if at all, this poses a problem for the theory, and how the theory could be made precise if this approach is to be taken.

Acknowledgements I would like to thank Dr. Tushar Menon for bringing Newman’s Objection to my attention. I would also like to thank Dr. Will Davies for encouraging me to pursue this work, and Lara Scheibli for her encouragement and critique. This work would not have been possible without the patient support of Jason Lowther (University of Birmingham). None of those acknowledged should be blamed for the work that follows.

Compliance with Ethical Standards

Conflict of interest There are no conflicts of interest to declare.

References

Ainsworth PM (2009) Newman's objection. Br J Philos Sci 60:135–171

Block N (1978) Troubles with functionalism. Minesota Stud Philos Sci 9:261–325

Chalmers D (1996) Does a rock implement every finite-state automaton? Synthese 108:309–333

Churchland P (2005) Functionalism at forty: a critical perspective. J Philos 102(1):33–50

Copeland J (1996) What is computation? Synthese 108:335–359

Cruse P (2005) Ramsey sentences, structural realism and trivial realization. Stud Hist Philos Sci 36:557–576

Demopoulos W, Friedman M (1985) Bertrand Russell's the analysis of matter: its historical context and contemporary interest. Philos Sci 52:621–639

Godfrey-Smith P (2009) Triviality arguments against functionalism. Philos Stud 145:273–295

Ketland J (2004) Empirical adequacy and ramsification. Br J Philos Sci 55:287–300

Lewis D (1972) Psychophysical and theoretical identifications. Australas J Philos 50(3):249–258

Melia J, Saatsi J (2006) Ramseyfication and theoretical content. Br J Philos Sci 57:561–585

Newman MHA (1928) Mr. Russell's "causal theory of perception". Mind 37(146):137–148

Psillos S (1999) Scientific realism: how science tracks truth. Routledge, London

Putnam H (1988) Reality and representation. MIT Press, Cambridge, MA

Ramsey FP (1929) Theories. In: Braithwaite RB (ed) Foundations of mathematics. Routledge, New York, pp 212–236

Searle J (1990) Is the brain a digital computer? Proc Addresses Am Philos Assoc 64:21–37

Shoemaker S (1981) Some varieties of functionalism. Philos Top 12:93–119

Votsis I (2003) Is structure not enough? Philos Sci 70(5):879–890

Worrall J (1989) Structural realism: the best of both worlds? Dialectica 43(1–2):99–124

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.