30TH OXFORD GRADUATE PHILOSOPHY CONFERENCE

In Defence of Carnapian Constitutive Functionalism

Ether Lin

14 November 2026

Functionalism

- Roughly, functionalism is the view that the meaning of a term is the role it plays in the theory.
  - e.g., pain is whatever state is caused by damage and causes avoidance (Lewis 1972), and a gene is whatever is passed down from parents and makes a difference to a heritable trait.
- It is traditionally associated with philosophy of mind, but has recently attracted a lot of attention in philosophy of physics.
  - spacetime functionalism (Knox 2014; Lam and Wüthrich 2018)
  - inter-theoretic reduction (Butterfield and Gomes 2023)
  - Knox and Wallace's *Functionalism Fit for Physics* (forthcoming)
- But there is more than one way of understanding "whatever plays the role".

Two kinds of functionalism

Knox and Wallace (forthcoming) distinguish between

- **causal-role functionalism**, where a term's definition depends on identifying a unique realiser of the role, e.g., pain is C-fibre firing if that is what plays the pain role. The identification is a defeasible hypothesis about the lower level;
- **constitutive functionalism**, where a system's functioning in the way the theory describes is analytically sufficient for the concept to apply, however that functioning is realised.
- Knox and Wallace argue that the second is the right fit for physics.
- I agree, but I also disagree with their account substantially. In this talk I want to defend a Carnapian version of it, for reasons I will come to.

The Ramsey sentence and Lewis

Causal-role functionalism has a standard formal version due to Lewis.

- Take a theory T with O-terms, understood independently of T, and T-terms, introduced by it (e.g., ν). Its Ramsey sentence says that something plays the roles without saying what.

$$R(T) := \exists X_1 \ldots \exists X_n \, T(X_1, \ldots, X_n)$$

- Lewis (1970, 1972) modifies this to require a unique realiser, and defines each T-term by a description.

$$\exists! x \, T(x), \qquad \tau := \iota x \, T(x)$$

- A T-term then names the unique occupant of its role, or nothing. So multiply realised concepts (pain) are on a par with unrealised ones (phlogiston), as both are denotationless.

Butterfield and Gomes: mandatory bridge laws

- Butterfield and Gomes (2023) use Lewis's definition for inter-theoretic reduction. To reduce fluid dynamics to molecular dynamics T^*, we need a bridge law $\nu = \theta$ with θ in T^*.
- Such bridge laws are normally just posited, but substituting Lewis's definition gives

$$\theta = \iota x\, T(x),$$

- which contains only O-terms and T^*-terms. If T^* entails it, $\nu = \theta$ follows by transitivity of identity, so the bridge law is **mandatory**, i.e., derived rather than added as a separate hypothesis.
- This of course only works if the role has a unique occupant in T^*.

Knox and Wallace on viscosity

Knox and Wallace, however, object to this picture. Their main example is viscosity in the Navier–Stokes equation.

$$\frac{\partial u}{\partial t} + (u \cdot \nabla)u - \nu\nabla^2 u = -\frac{1}{\rho}\nabla p + g$$

- A causal-role functionalist would need some molecular quantity θ that uniquely plays the ν-role. But there is no obvious candidate, as viscosity is read off as a coefficient from a heavily mathematised derivation (p. 6).
- If unique realisation were required, ν would name nothing, and the liquids we observe would be zombie liquids (pp. 7–8).

The constitutive alternative

Knox and Wallace's constitutive functionalism does not require any realiser.

Knox and Wallace (p. 7)

"[N]othing more is required for it to have viscosity [ν] than for that to be the value that the viscosity must be set to for the N-S equation to reproduce the actual behavior of the system."

- So no molecular realiser is needed, and *zombie liquids*, i.e., things that behave like liquids but are not liquids, are ruled out, since fitting the equation just is being a liquid.
- Calling ν viscosity is a further step of what they call "explication", tied to antecedent notions such as stickiness (p. 11).

Knox and Wallace and the semantic view

- Knox and Wallace also pair each view with a conception of theories (p. 2). Causal-role functionalism gives "a route to Nagelian reduction of theories conceived syntactically", whereas constitutive functionalism "maps onto a semantic conception".
- The semantic view in question is Wallace's (2022) maths-first realism, where theories are primarily represented by classes of mathematical models, and only then described in language by "predicate precisification" (Wallace 2024).
- On their view, then, nothing more than the fit of the model is required to constitute the concept.

If so, the Ramsey sentence seems to be on the wrong side.

Where I agree, and where I don't

- I agree with Knox and Wallace that unique realisation should go, and that constitutive functionalism is the right fit for physics.
- But in my view their account is far from complete.
  - The alliance with the semantic view is unnecessary, as the two functionalisms are distinguished by analyticity, not by how theories are presented. It also confines the view to physics.
  - Their account says when a concept applies, but not (1) what makes that analytic, or (2) what fixes the concept beyond its application rule.
- I think a Carnapian account can answer both, applies beyond physics, and does not depend on the syntactic view either.

Syntactic or semantic?

- The distinction is drawn by analyticity, so nothing about the presentation of theories follows from it. The mapping is motivated by what account of inter-theoretic reduction works in physics, which is orthogonal to the question about concepts.
- On top of that, maths-first realism (which Wallace aligns with semantic) is particularly ill-suited, considering functionalism is a more general thesis not restricted to physics, and it is hard to imagine “pain” being captured by mathematical models.
- Butterfield and Gomes cite Lutz (2017) as evidence of there being no material difference between the syntactic and semantic conceptions. It should however be noted that Lutz’s argument is that the two are equally expressive in *higher-order logic*, whereas Lewis and Butterfield and Gomes rely on first-order terms.

(1) What makes it analytic?

Knox and Wallace's criterion has the form of $\forall x\,(\varphi x \rightarrow Vx)$, where φx says that N-S fits x, and Vx that x has a viscosity.

- Knox and Wallace say the identification is analytic (p. 1), and that nothing more than fitting N-S is required (p. 7). They do not say where the analyticity comes from, although on a maths-first picture the natural answer is the model.
- But setting up the model is no free stipulation. We must first find variables that track real degrees of freedom and fix how they relate. Viscosity makes no sense unless some variable actually contributes to the equation.
- As a more general critique, the way maths-first realism is presented masks the conventionality in mathematical representation, alongside its need for empirical adequacy.

(2) Application or meaning?

If the account is only about when a concept applies, it is hardly saying anything interesting. Knox and Wallace clearly also aim at meaning, not just application. To have a property is "nothing more or less than" to fit the role (p. 4), and what it is to be a liquid is to satisfy certain equations (p. 7).

- Lewis's uniqueness was what fixed a full meaning, i.e., an extension in every possible situation, at the price of empty extensions when nothing plays the role.
- Without uniqueness, a role only partially defines a term. Knox and Wallace's account only says about model fit but not what happens when it fails, which leaves the term's meaning underdetermined (e.g., phlogiston).
- So the slogan that a term's meaning is fixed by its role becomes ambiguous, and their rule of application is not yet a meaning.

So what fixes the concept beyond this application rule?

The Carnap sentence

The proposal is reminiscent of Carnap's own apparatus. Carnap's (1963) use of the Ramsey sentence requires no unique realiser. He pairs it with the Carnap sentence

$$C(T) := R(T) \rightarrow T(\tau_1, \ldots, \tau_n), \qquad C(T) \wedge R(T) \dashv\vdash T$$

- $R(T)$ carries the empirical commitments, and $C(T)$ assigns the terms if $R(T)$ is met.
- A very primitive proposal, then, is that a concept applies analytically if $R(T)$ holds. This is already constitutive functionalism in sentential form.
- Nothing here asks for a unique realiser or a reducing theory, which is exactly what Knox and Wallace drop from the "Ramsey-Carnap-Lewis method" (pp. 5–6).

But does $C(T)$ give Knox and Wallace what they want?

Against the Carnap sentence

$C(T)$ as it stands only says something once the whole theory fits the world.

- Everything $C(T)$ makes analytic has the form $R(T) \to \alpha$ (Lutz 2014, Claim 6). So even if T entails the criterion, $C(T)$ only gives

$$R(T) \to \forall x\,(\varphi x \to Vx).$$

- So a system that N-S fits has a viscosity only if the whole fluid theory is right about the world.
- And if the theory fails anywhere, $C(T)$ leaves the terms completely unrestricted (Lutz 2014, p. 5), which is the same critique raised against Knox and Wallace.

So the part that defines "viscosity" must say more than $C(T)$. Is that allowed?

Carnap's adequacy conditions

Carnap's (1963) conditions on an analytic component A and a synthetic component S (cf. Lutz 2014) are

Adequacy conditions

1. $A \wedge S \dashv\vdash T$.
2. A has no O-content, i.e., if $A \models O$ for an O-sentence O, then $\models O$.
3. S is empirically equivalent to T.
4. S contains no T-terms.

- $R(T)$ is the unique S, and $C(T)$ the weakest A. Any A with $T \models A \models C(T)$ and $\models R(A)$ is adequate (Lutz 2014, Cor. 10), so stronger components are allowed.
- With $S = R(T)$, condition 1 amounts to $T \models A \models C(T)$, and condition 2 to A adding nothing observable.

The same conditions over models*

The conditions can also be given in a semantic form (Przełęcki and Wójcicki 1969; Lutz 2019, 2020). Bipartition the vocabulary into O and T, and let $\mathfrak{T} := \mathrm{mod}(T)$, $\mathfrak{A} := \mathrm{mod}(A)$ and $\mathfrak{S} := \mathrm{mod}(S)$.

1. $\mathfrak{A} \cap \mathfrak{S} = \mathfrak{T}$.
2. $\forall \mathcal{O}\, \exists \mathcal{A} \in \mathfrak{A}\, (\mathcal{O} = \mathcal{A}|_O)$, i.e., every O-structure expands to a model in $\mathfrak{A}$.
3. $\mathfrak{S}|_O = \mathfrak{T}|_O$.
4. $\forall \mathcal{M}, \mathcal{N}\, (\mathcal{M}|_O = \mathcal{N}|_O \Rightarrow (\mathcal{M} \in \mathfrak{S} \Leftrightarrow \mathcal{N} \in \mathfrak{S}))$.

- The models of $R(T)$ and $C(T)$ also fulfil these, and so does any $\mathfrak{A}$ with $\mathfrak{T} \subseteq \mathfrak{A} \subseteq \mathrm{mod}(C(T))$ meeting (2).
- So the apparatus is not restricted to the syntactic view. For the sake of simplicity I will continue to work on the syntactic notation.

Knox and Wallace's criterion as a component

So add their criterion without the condition.

$$A_{KW} := C(T) \wedge \forall x\,(\varphi x \rightarrow Vx)$$

- Condition 1 holds, as the theory already says this, so $T \models A_{KW} \models C(T)$.
- Condition 2 holds, as it doesn't contradict with any observation.
- Now fit alone settles viscosity. If water fits N-S, it has a viscosity, even if N-S fails for oobleck, and a zombie liquid that fits N-S without a viscosity is ruled out by what "viscosity" means.
- The same also applies for the biconditional $A_{\leftrightarrow} := C(T) \wedge \forall x\,(\varphi x \leftrightarrow Vx)$.

This is my answer to (1). Their criterion is analytic because it adds nothing to what the theory says about what we observe.

Carnapian constitutive functionalism

- $R(T)$ carries the empirical commitments, and an analytic component A carries the conventional ones. A constitutive identification belongs to A, and is analytic because A is O-conservative.
- The meaning of a term is then what A says about it, i.e., which interpretations of the term it rules out. With $C(T)$ alone, a failed theory rules out nothing, so "phlogiston" means nothing in particular. With "whatever burning releases is phlogiston" in A, we know what would count, even though nothing does.
- This answers (2). What fixes the concept beyond the application rule is which of the theory's claims we put in A, and the theory itself does not decide that.

The view

A concept is the constraint a chosen analytic component places on how the term may be interpreted relative to the theory it belongs to.

“But you still need realisers”

Objection. The Carnapian view still demands an existence claim in $R(T)$. It allows multiple realisers but not none, which is exactly Knox and Wallace’s case, as nothing molecular plays the viscosity role.

Reply.

- The very idea of “unique realiser” only matters in a first-order language, where terms are realised by objects in the domain, as we see in Butterfield and Gomes’s framework. In a higher-order language properties can be claimed “for free”.
- The kind of existential quantifier in $R(T)$ is closer to empirical fit than ontological commitment, so to say $R(T)$ is met is the same as saying the empirical content is fit, which is also what Knox and Wallace commit to.

"Why the O/T distinction?"

Objection. Knox and Wallace's account does not need the O/T distinction, and the whole distinction itself is contestable.

Reply.

- It need not be understood as observational/theoretical as Carnap originally intended. Lewis (1970) understands it as old/theory-introduced, and Button and Walsh (2018) as okay/troublesome.
- Even though Knox and Wallace didn't make the distinction, there is still an implicit commitment in their account. The very fact that we target viscosity rather than pressure when empirical fit fails in N-S suggests that it plays a different role in the theory.
- It is even more problematic if we do without the distinction. If all concepts are treated as naïvely functional, i.e., theory-introduced, nothing grounds the theory essentially.

"It is *ad hoc*"

Objection. The modification of *A* is *ad hoc*, since you have to stipulate the concept yourself, which goes against the spirit of functionalism.

Reply.

- Nothing is added to the theory. *A* must be something *T* already says (condition 1), so the meaning still comes from the theory alone.
- Compare with Lewis's own functionalist account: he also made specific modifications on the T-terms for them to be uniquely realised. The same goes for Butterfield and Gomes's account. Carnapian constitutive functionalism is similarly placing further requirements on the T-terms.

Conclusion

- Knox and Wallace are right to give up unique realisation, but their account only tells us under what conditions a concept applies, without explaining what makes this analytic.
- The Carnap sentence has the right shape, but it is too weak, as every verdict depends on the whole theory.
- Their claim is better understood as a stronger analytic component, which places further constraints on the theory-introduced concepts.
- This needs no molecular realiser, applies beyond physics, and is not necessarily tied to the semantic view.

Thank you!

Ether Lin
14 November 2026