Armchair Philosophy Naturalized

Sebastian Lutz[*]

Preprint: 2017-09-24

Abstract

Carnap suggests that philosophy can be construed as being engaged solely in conceptual engineering. I argue that since many results of the sciences can be construed as stemming from conceptual engineering as well, Carnap's account of philosophy can be methodologically naturalistic. This is also how he conceived of his account. That the sciences can be construed as relying heavily on conceptual engineering is supported by empirical investigations into scientific methodology, but also by a number of conceptual considerations. I present a new conceptual consideration that generalizes Carnap's conditions of adequacy for analytic-synthetic distinctions and thus widens the realm in which conceptual engineering can be used to choose analytic sentences. I apply these generalized conditions of adequacy to a recent analysis of scientific theories and defend the relevance of the analytic-synthetic distinction against criticisms by Quine, Demopoulos, and Papineau.

Keywords: conceptual engineering; explication; methodological naturalism; Carnap sentence; Ramsey sentence; concept formation; theoretical terms; analytic-synthetic distinction

Dedicated to the memory of William Demopoulos

Contents

1   Two ways of naturalizing armchair philosophy

One motivation for naturalism in philosophy stems from the perceived gap between the number of results in the sciences and in philosophy. Science has achieved staggering successes in this respect, from enduring insights into the nature of the cosmos to conveniences like digitized music to cancer treatments. Philosophical insights, on the other hand, seem comparably fleeting (being disputed by the next or even contemporary generation of philosophers), seldom convenient, and even less often a matter of life and death.[1] So maybe philosophy can become better by being more like science, or one can show that philosophy already is better than its reputation suggests because it already is very similar to science. Either of these positions has been called 'philosophical naturalism'.

    Of course, not every similarity counts, as cargo-cult science shows (Feynman and Leighton 1989, 308–317), and so the philosophical naturalist must choose some features as particularly relevant. A proponent of *ontological naturalism* chooses the content of philosophical theories, demanding that philosophy be "concerned with the contents of reality, asserting that reality has no place for 'supernatural' or other 'spooky' kinds of entity", where the spooky kinds are typically those that are not condoned by science. A proponent of *methodological naturalism*, on the other hand, "is concerned with the ways of investigating reality, and claims some kind of general authority for the scientific method" (Papineau 2009a).

    In the following, I will focus on methodological naturalism. Typical arguments for methodological naturalism assume that scientific claims are *synthetic*,

---

1. What *counts* as death may of course be a philosophical question. But this would be a matter of 'life' and death', or a matter of LIFE and DEATH (the concepts).

that is, in some way about the world and thus factual. By establishing that philosophical claims are also synthetic and depend on the same kind of support as scientific claims (Papineau 2009b), one can conclude that philosophy and science must rely on the same methods. (Experimental philosophy arguably comes into play when one further assumes that the common philosophical methods are not as effective as the scientific ones.)

Instead of pursuing this line of thought, I will in the following argue that philosophical claims and a large number of scientific claims can be construed as *analytic*, and specifically *conventional*, that is, suggestions for (typically new) concepts rather than analyses of concepts we currently have. This means that suggestions for philosophical concepts can be justified in the same way that suggestions for scientific concepts can be justified, by what I will call 'conceptual engineering'. More specifically, I will argue that one can make a distinction between observational and non-observational claims and construe the non-observational claims as analytic. Methodological naturalism then either claims that philosophy chooses its analytic sentences the way the sciences do, or demands that philosophy choose its analytic sentences so. Carnap (1963, III.19), a paradigmatic armchair philosopher, stated exactly this demand.

In the following, I will first give an overview of the assumptions of conceptual engineering, with a focus on the role of empirical claims in the choice of analytic statements (section 2). The assumption that analytic statements are true by convention has important implications for the kind of argument that can be made for the claim that conceptual engineering is central to science (section 3). But before presenting this argument, I will discuss two attempts to establish the irrelevance of conceptual engineering: Papineau argues that analytic sentences contribute almost nothing to the content of philosophical theories. His argument, however, assumes Lewis's semantics for theoretical terms, an assumption that is in direct contradiction to those of conceptual engineering (section 4.1). Quine argues that there is no point to the specific kind of conceptual engineering known as 'explication', but his argument simply overlooks the goal of explication (section 4.2). That conceptual engineering plays a central role in the sciences has already been supported by some empirical investigations (section 5.1) and either directly or indirectly by some philosophical discussions of the sciences (section 5.2). I want to add to these latter discussions and show how little needs to be assumed to be able to introduce a distinction between analytic and synthetic components of scientific theories (section 5.3). An argument by Demopoulos that was meant to establish that under this distinction analytic sentences contribute more content to scientific theories than is plausible succeeds in showing the extent to which analytic sentences do so contribute, but fails to show that this is implausible (section 6). Thus scientific theories can be construed as engaging substantially in conceptual engineering, which allows for a methodologically naturalistic armchair philosophy in Carnap's sense.

3

2   Conceptual engineering

Carnapian conceptual engineering relies on the idea that it is possible to determine whether a sentence can be empirically tested (Bradley, forthcoming, §§2–3). It may be testable because the terms occurring in the sentence apply more or less immediately to perceptions (cf. Chang 2005), but more generally, the sentence's terms simply refer to concepts that are uncontroversially applicable and thus specifically not themselves under investigation (Reichenbach 1951, 49; cf. Lewis 1970, 428). These uncontroversial concepts I will call 'basic concepts', the terms I will call 'basic terms', and the sentence build up from the terms and logical constants I will call 'basic sentences'. Controversial terms, concepts, and sentences will be called 'auxiliary'. Whether sentences containing only basic terms are true, false, or indeterminate in their truth value can be determined by purely empirical investigations (for instance by experiments) or formal investigations (for instance by mathematical or logical proofs).[2] The construction of basic sentences out of basic terms with the help of logical constants assumes that the logical constants are uncontroversial. When they are not, one has to use basic sentences as primitives, but fortunately, the following discussion can proceed taking logical constants as uncontroversial.

Which concepts are basic is context dependent. In a dispute between two parties, for instance, the basic concepts are those that both parties accept. If a concept is challenged by either party, it becomes auxiliary. Such challenges must stop at observation concepts, which are therefore always basic. Concepts are observational if they can be learned by ostension, independently of one's background.[3] Quine (1975, 316) puts it as follows: "The really distinctive trait of observation terms and sentences is to be sought [. . .] in ways of learning. Observational expressions are expressions that can be learned ostensively. They are actually learned ostensively in some cases and discursively in others, but each of them could be learned by sufficiently persistent ostension." Observation concepts cannot be challenged because, as Przełęcki (1969, 36–38) points out, agreement on observation concepts is a purely psychological phenomenon: Either one has learned the concept in question or one has not; there cannot be a reasoned debate, only an empirical check.

The testing of a claim involves an investigation of the truth of the basic sentences it entails. If a sentence entails no basic sentences or only basic sentences whose truth is indeterminate, then it can be chosen to be true, which makes the sentence an analytic truth. A basic sentence with a determinate truth value, on the other hand, can be analytic, synthetic, or a combination thereof. It is analytic if its truth value follows from analytic basic sentences (which have been previously chosen to be true). It is synthetic if it is an observation sentence. It is a combination of analytic and synthetic if its truth value follows from observation

---

2. For a formal exposition of this claim, see Lutz (2012a, §3).

3. Schurz (2014, 2.9.2) provides a detailed definition of 'observation concept' and Eberle (1990) analyzes some of their logical aspects.

sentences and analytically true sentences.[4] Under this view of language, the only possibility of assigning meaning to terms is through observation and convention. I will follow (Przełęcki 1974, 402) and others in calling it 'semantic empiricism'.

Within semantic empiricism, pure observation statements are, strictly speaking, never analytic, but I will deviate slightly from this terminology in the following: I will always assume that logical constants are are uncontroversial and call sentences built up from observational terms and logical constants 'observational'. In this slightly deviant terminology there can thus be analytic observational sentences, namely logical truths.

Of course, one can even choose the observational terms and in general the observational language, but as Eklund (2009, §2) points out, this choice of language only leads to different ways of expressing the same propositions. The possibility of such redescriptions does not have any interesting philosophical implications, and I will ignore this aspect of language choice in the following.[5]

Two sets of sentences may entail the exact same basic sentences, in which case they are often called 'empirically equivalent'. I will reserve the term 'observational equivalent' for theories that entail the same observation sentences. Clearly, empirically equivalent theories are also observationally equivalent. Within semantic empiricism, the decision between two empirically equivalent theories is a matter of convention, as the correctness of the choice cannot be empirically tested and thus cannot be observationally tested. Rather, the choice only determines a convention about the use of the auxiliary terms.

The reason that the choice of convention is important is that, *pace* Lewis, it is not redundant to speak of 'arbitrary conventions', for a convention is often chosen with a specific goal in mind. That goal is often the accommodation of specific facts (Kyburg 1990, 161–162). Consider a simple example: We might have observed that some organism, when being interfered with in one of m ways, $\{D_i\}_{1 \leq i \leq m}$, reacts in one of n ways, $\{A_j\}_{1 \leq j \leq n}$:

$$\forall x \left(\bigvee_{1 \leq i \leq m} D_i x \rightarrow \bigvee_{1 \leq j \leq n} A_j x \right). \tag{1}$$

The question is now which new concepts should be introduced. One could, for example, group all of the ways of interfering as ways of inflicting damage, D, and

---

4. Przełęcki (1969, ch. 10) provides a formal semantics for basic sentences build up from observation sentences in this way.

5. Unfortunately, Eklund goes on to claim that the language choices (or choices of "linguistic frameworks") that Carnap (1950) describes consist entirely of this trivial kind, even for non-observational sentences. The alternative, Eklund (2009, 138) suggests, is that "a string of symbols can come out true in some languages but false in others, *while meaning what it actually means.*" But the Carnapian alternative is rather that the meaning of a string of symbols (the proposition it expresses) is completely dependent on the choice of language, and that besides the propositions expressed by observational sentences, there are no propositions to be expressed until the non-observational sentences are chosen. In short: The only language-independent propositions are those pertaining to observations. I thank André Carus for discussions of this point.

group all of the ways of reacting as ways of showing avoidance behavior, A. This would lead to

$$\forall x \Big(\bigvee_{1 \leq i \leq m} D_i x \leftrightarrow Dx \Big) \wedge \forall x \Big(\bigvee_{1 \leq j \leq n} A_j x \leftrightarrow Ax \Big) \wedge \forall x (Dx \rightarrow Ax) \,. \qquad (2)$$

But one might also consider the explicit definitions of 'damage' and 'avoidance behavior' too restrictive, since one might later want to classify other kinds of interference as damage (cf. Carnap 1936, 449) or one might later classify the organism's reactions not only as avoidance behavior. To allow for this, one could introduce a sufficient condition for damage and a necessary condition for avoidance behavior:[6]

$$\forall x \Big(\bigvee_{1 \leq i \leq m} D_i x \rightarrow Dx \Big) \wedge \forall x \Big(Ax \rightarrow \bigvee_{1 \leq j \leq n} A_j x \Big) \wedge \forall x (Dx \rightarrow Ax) \,. \qquad (3)$$

One could also introduce 'pain', P, as an intermediary between the different kinds of damage and avoidance behavior:

$$\bigwedge_{1 \leq i \leq m} \forall x (D_i x \rightarrow Px) \wedge \bigvee_{1 \leq j \leq n} \forall x (Px \rightarrow A_j x) \,. \qquad (4)$$

Finally, one could combine the previous two approaches, which seems to be closest to the path that was chosen in ordinary language:

$$\forall x \Big(\bigvee_{1 \leq i \leq m} D_i x \rightarrow Dx \Big) \wedge \forall x \Big(Ax \rightarrow \bigvee_{1 \leq j \leq n} A_j x \Big)$$
$$\wedge \forall x (Dx \rightarrow Px) \wedge \forall x (Px \rightarrow Ax) \,. \qquad (5)$$

The reason one can choose between these different sentences is that they all entail the same basic sentence (1), that is, all are empirically equivalent.

Finding empirically equivalent theories that are (or were) both actually used in the sciences is more difficult then the pain example may make it seem since finding appropriate conventions for the empirical results at hand is a difficult task. So difficult, in fact, that the conventions have often been taken to be discoveries. Nonetheless, a somewhat decent example may be the phlogiston theory and the oxygen theory, which were empirically equivalent at least within a restricted domain of application and until later auxiliary concepts were chosen that connected to the oxygen concept but not the phlogiston concept. It seems that most currently employed scientific theories that are empirically equivalent are also equivalent in some stronger sense: For instance, they are also definitionally

---

6. The definitions of 'damage' and 'avoidance behavior' are weakened differently because of the relation between the interferences and the reactions in the original observation: The mediators between the interferences and the reactions must be necessary for the former and sufficient for the latter.

equivalent (Glymour 1970, 280) or there exists an equivalence between their categories of models that preserves empirical content (Weatherall 2016, §5). In general it will also be rather difficult to identify empirically equivalent scientific theories because it is difficult to distinguish between the empirical and the conventional components of theories.

The reason that in the pain example above it is so easy to distinguish between empirical and conventional statements is that the empirical statements are identified by their vocabulary: By assumption, only the different kinds of damage $\{D_i\}$ and avoidance behavior $\{A_j\}$ have a fixed, uncontroversial meaning, while the newly introduced terms receive their meaning only through the sentences that introduced them. This assumption is not essential: As already noted, one could introduce basic sentences as a primitive as well, and I will discuss another way of avoiding the assumption of a basic vocabulary in section 5.2. For now, however, the restriction of the discussion to this simple case will suffice.

In the simple case, one can also find a simple answer to a question that is in some sense the inverse to the question which concepts should be introduced: If one is given a set of sentences Θ, what is the synthetic, observational component of the set and what is the analytic, conventional component? Since this is by assumption a simple case, the terms are bipartitioned into basic terms $\mathscr{B}$ and auxiliary terms $\mathscr{A}$. Specifically, the $\mathscr{B}$ terms are assumed to be observational, that is, every $\mathscr{B}$-sentence is synthetic. If Θ is a finite set, it has only finitely many $\mathscr{B}$ and $\mathscr{A}$ terms and can be written as $\Theta(B_1,\ldots,B_m,A_1,\ldots,A_n)$. Following a suggestion by Carnap (1963, 24.C–D), one can express the theory's synthetic component by its Ramsey sentence

$$\mathsf{R}_{\mathscr{B}}(\Theta) := \exists X_1,\ldots,X_n \bigwedge \Theta(B_1,\ldots,B_m,X_1,\ldots,X_n) \tag{6}$$

and its analytic component by its Carnap sentence

$$\mathsf{C}_{\mathscr{B}}(\Theta) := \mathsf{R}_{\mathscr{B}}(\Theta) \to \bigwedge \Theta \tag{7}$$

As I will discuss below, the Ramsey sentence entails all the $\mathscr{B}$-sentences entailed by Θ, while the Carnap sentence entails only tautological $\mathscr{B}$-sentences. Since furthermore the conjunction of Ramsey and Carnap sentences is equivalent to Θ, they are plausible synthetic and, respectively, analytic components of Θ. (For general $\mathscr{B}$ terms, the Carnap sentence is analytic, while the truth value of the Ramsey sentence is uncontroversial.) As required and easy to prove, the Ramsey sentences of all the different sentences motivated by the empirical relation between the interference with an organism and its reaction (1) are equivalent to that empirical relation itself.

3 The argument at issue

The bulk of the following discussion is devoted to showing that the above description of the role of conceptual engineering in the sciences is correct. The coarse

structure of the argument is as follows:[7]

1. A methodology that contributes essentially to the success of the sciences is methodologically naturalistic.

2. The introduction of non-observational sentences into the sciences contributes essentially to their success.

3. Non-observational sentences can be construed as being introduced into the sciences by conceptual engineering.

4. Therefore, conceptual engineering can be construed as contributing substantially to the success of the sciences.

5. Therefore, conceptual engineering can be construed as being methodologically naturalistic.

Premise 1 is a modification of the definition of 'methodological naturalism' given by Papineau (2009a). I added the demand that a methodology contribute *essentially* because, as noted above, not every similarity to the sciences counts: Cargo-cult sciences rely on methodologies that contribute to science's success (e. g., the use of journals, conferences, and citations), but these are not essential (Feynman and Leighton 1989, 308–317). For an example with immediate relevance for conceptual engineering, consider Papineau's position that science uses analytic sentences only for naming objects and properties that are already known to exist (Papineau 2009b, §III). A methodology relying on the choice of analytic sentences in Papineau's sense would thus be a pure labeling exercise (see §4.1).

The empirical premise 2 is very weak because few if any sentences in the sciences are purely observational. I take it that this premise is generally accepted (see, e. g., Carnap 1939, §24; Tuomela 1973, §1.1). The two conclusions, I think, follow analytically from the premises within the vagueness of natural language, especially those of the modal operator 'can be construed as'. This leaves premise 3 with that modal operator.

The modal operator does not appear in premise 3 as a cheap evasion of counterarguments, but rather because the stronger claim that non-observational statements *are* introduced by conceptual engineering is incompatible with the motivation of conceptual engineering given above. For the motivation relies on semantic empiricism, and thus on the claim that every statement that is not observationally testable is conventional, and therefore subject to conceptual engineering. This invites the argument that this claim itself is not testable, and therefore conventional, and therefore subject to conceptual engineering. (The argument is analogous to the argument that the verification criterion of meaningfulness identifies itself as

---

7. This section is a response to and in significant parts inspired by the comments of an anonymous reviewer.

meaningless, except that in this case, the criterion of conventionality identifies itself as conventional.)[8]

One could reply that a criterion does not have to be applied to itself, especially if the criterion applies to some object language and is phrased in the respective metalanguage. This could get the conventionality criterion, which can be phrased as a metalanguage claim, of the hook, but only on a technicality: The motivation for the criterion, after all, is that we do not want to claim absolute, convention-independent truth for statements that we cannot observationally test, and the conventionality criterion cannot be observationally tested, metalanguage or not. In other words, for each language, the motivation leads to a conventionality criterion in the language's respective metalanguage. Thus the modal operator 'can be construed as' cannot be avoided on pain of flouting the motivation of conceptual engineering: The operator expresses that premise 3 itself must be construed as a product of conceptual engineering, and thus as dependent on the choice of a language convention.

As used in the above argument, 'conceptual engineering' just means 'language choice as pursued in the sciences'. The rest of this paper is dedicated to showing premise 3, that the sciences indeed can be construed as choosing their theoretical concepts, rather than discovering them. Some of these arguments will be empirical, some of them conceptual. The conceptual arguments will support the stronger claim that it can be construed as being *necessary* that the sciences choose their non-observational statements.

The conceptual arguments also apply to non-sciences and to philosophy. However, they only establish that philosophy, insofar as it makes non-observational claims, can be construed as engaging in language choice. The demand that philosophy be methodological naturalistic amounts to the further demand that philosophy choose its languages in the way the sciences do.

4   Papineau and Quine against conceptual engineering

There have been a number of criticisms of conceptual engineering as a methodology for philosophy, of which I will discuss a recent one by Papineau and a particularly influential one by Quine.[9]

---

8. Whether the argument against the verification criterion succeeds is another matter, because it might rely on a careless statement of the verification criterion. For the demand that every *non-analytic* meaningful statement must be verifiable can also lead to the conclusion that the verifiability criterion is analytic and, depending on one's interpretation of 'analytic', conventional (Lutz 2017, §§5–6).

9. The best known criticism of conceptual engineering comes from Strawson (1963) and has been answered, to my mind successfully, by for instance Carnap (1963, §19) and Maxwell and Feigl (1961).

4.1   Papineau against conceptual engineering

Papineau (2009a, §2.1) argues that philosophy, like science, is about the world, and thus best pursued naturalistically (cf. Papineau 2009b, 3). There are no differences between the "aims and methods" of philosophy and science, but only in the specific topics. Many philosophical theories are very general (e. g. theories of "spatiotemporal continuants, universals and identity"), and

> unlikely ever to be decided between by some simple experiment, which is no doubt one reason that philosophers do not normally seek out new empirical data. Even so, the naturalist will insist, such theories are still synthetic theories about the natural world, answerable in the last instance to the tribunal of empirical data.
>
> What seems to identify [...] philosophical issues is that our thinking is in some kind of theoretical tangle, supporting different lines of thought that lead to conflicting conclusions. [...] Here too empirical data are clearly not going to be crucial in deciding theoretical questions—often we have all the data we could want, but can't find a good way of accommodating them. Still, methodological naturalists will urge [that an] empirical theory unravelled from a tangle is still an empirical theory, even if no new data went into its construction.

One could construe philosophy like this: Somehow, the existing empirical data force our hand and leave only one way out of the theoretical tangle. Of course, a proponent of this view has the unenviable task of showing a way of empirically testing the suggested solutions to the theoretical tangle, even though such tests might not be simple. By contrast, the conceptual engineer construes philosophy in a way that does not require this task: For the conceptual engineer, how to speak about continuants, universals, and identity is not an empirical question, because any answer is not testable and is thus a matter of choice. And if the unraveling of a theoretical tangle in an empirical theory indeed cannot be decided by any further empirical information, then it, too, is a matter of choice.[10] That of course does not mean that the theory in which the tangle occurred does not have an empirical component. It just means that the problem could be solved without changing the Ramsey sentence of the theory.

But Papineau (2009b, 9) has a counter to this reply: He argues that analytic sentences are not interesting. Thus, since philosophy is interesting, conceptual engineering cannot be philosophy. Using as an example a theory T with one auxiliary term F, he argues that only the Ramsey sentence, which cannot be changed by language choice, is philosophically interesting:

> From the perspective of this approach to concepts, the original theory $T(F)$ can be decomposed into the analytic Carnap sentence and

---

10. As noted, the choice can be better or worse, and the conceptual engineer claims that the best choice is likely to be based on empirical results and made in the manner of the sciences.

the synthetic "Ramsey sentence" of the theory—$\exists \Phi\, T(\Phi)$. The Ramsey sentence expresses the substantial commitments of the theory— there *is* an entity which ...—while the Carnap sentence expresses the definitional commitment to dubbing that entity 'F'.

According to Papineau, if the Ramsey sentence of a theory is not changed, the purported solution of some theoretical tangle can at best consist in the renaming of concepts. For the Ramsey sentence of a theory states which auxiliary concepts are actually realized, and the Carnap sentence only assigns them labels. But no tangle has ever been solved by renaming alone. Thus the Carnap sentence is very uninteresting (Papineau 2009b, 10):

> [T]he natural assumption is surely that it is the synthetic Ramsey sentences that matter to philosophy, not the analytic Carnap sentences. What makes philosophers interested in investigating further is the pretheoretical supposition that there *are* entities fitting such-and-such specifications, not just the hypothetical specification that *if* there were such entities, *then* they would count as free actions, or intentional states, or whatever.

As an argument for the irrelevance of the Carnap sentence, Papineau's argument fails even on technical grounds. For the Carnap sentence in his example has the form

$$\exists \Phi T(\Phi) \rightarrow T(F)\,, \tag{8}$$

that is, the variable Φ in the antecedent is bound by the existential quantifiers, and thus the F in $T(F)$ cannot refer back to Φ. This becomes obvious when looking at the open formula $T(\Phi) \rightarrow T(F)$, for it is not generally the case that Φ and F have the same reference: F is interpreted by some structure $\mathfrak{A}$, but Φ is interpreted by a variable assignment ν. For example, if $T(F)$ is the sentence $F b$, with $b \in \mathscr{B}$ and $F \in \mathscr{A}$, then $\exists \Phi \Phi b$, $F b$, and thus $\exists \Phi \Phi b \rightarrow F b$ are true in $\mathfrak{A}$ with $\mathrm{dom}(\mathfrak{A}) = \{1,2\}$, $b^{\mathfrak{A}} = 1$, and $F^{\mathfrak{A}} = \{1,2\}$, because $\mathfrak{A}, \nu \vDash \Phi b$ for $\nu(\Phi) = \{1\}$. But $F^{\mathfrak{A}} \neq \nu(\Phi)$.

Papineau probably does not have the Ramsey and Carnap sentences in mind after all, for he introduces the above discussion with the claim that "it is open to us to regard the concept F as having its reference fixed via the description 'the Φ such that $T(\Phi)$'. That is, F can be understood as referring to the unique Φ that satisfies the assumptions in T, if there is such a thing, and to fail of reference otherwise" (8). As Papineau (2009a, §2.3) states in a similar discussion, "the Ramsey sentence corresponding to $T(F)$ is '$\exists!\Phi(T(\Phi))$'". But this is incompatible with Papineau's claim that the "original theory framed using the concept F is [...] equivalent to the conjunction of the Ramsey and Carnap sentences" (Papineau 2009b, 9), which is generally false if the Ramsey sentence is defined in Papineau's way.

Take Papineau's own example of a "simple theory of pain", which is "constituted by the two claims that (a) bodily damage typically causes pains, and (b)

pains typically cause attempts to avoid further damage" (Papineau 2009b, 4). Simplifying even more by ignoring the 'typically' and replacing causation by a material implication, one gets

$$T(P) := \forall x(Dx \to Px) \land \forall x(Px \to Ax)\,, \tag{9}$$

with D, P, and A standing for 'is damaged', 'feels pain', and 'shows avoidance behavior', respectively, and $\mathscr{A} = \{P\}$. Then

$$\forall x(Dx \to Ax) \dashv\vdash \forall x(Dx \to Dx) \land \forall x(Dx \to Ax) \tag{10a}$$

$$\dashv\vdash \exists Y[\forall x(Dx \to Yx) \land \forall x(Yx \to Ax)] \tag{10b}$$

$$\dashv\vdash \mathsf{R}_{\mathscr{B}}(T) \tag{10c}$$

and accordingly, from the definition of the Carnap sentence (7),

$$\mathsf{C}_{\mathscr{B}}(T) \dashv\vdash \forall x(Dx \to Ax) \to \forall x(Dx \to Px) \land \forall x(Px \to Ax)\,. \tag{11}$$

Note that this derivation can proceed purely syntactically, without any assumptions about the specific semantics involved. And it is easy to see that, indeed, $\mathsf{R}_{\mathscr{B}}(T) \land \mathsf{C}_{\mathscr{B}}(T) \dashv\vdash T$. On the other hand,

$$\exists! Y\, T(Y) \vdash \exists Y \forall Z[T(Z) \leftrightarrow Z = Y] \tag{12a}$$

$$\vdash \exists Y \forall Z[\forall x[(Dx \to Zx) \land (Zx \to Ax)] \leftrightarrow Z = Y] \tag{12b}$$

$$\vdash \exists Y\big(\forall x[(Dx \to Yx) \land (Yx \to Ax)] \leftrightarrow Y = Y]$$

$$\land \forall x[(Dx \to Dx) \land (Dx \to Ax)] \leftrightarrow D = Y] \tag{12c}$$

$$\land \forall x[(Dx \to Ax) \land (Ax \to Ax)] \leftrightarrow A = Y]\big)$$

$$\vdash \exists Y(D = Y \land A = Y) \tag{12d}$$

$$\vdash \forall x(Dx \leftrightarrow Ax)\,, \tag{12e}$$

which is not entailed by T.[11] Hence $\exists! Y\, T(Y) \land \mathsf{C}_{\mathscr{B}}(T)$ is strictly stronger than T.

In fact, Papineau has silently switched from the Ramsey sentence to something akin to the Ramsey-Lewis sentence, which led to the incompatibility. Lewis (1970) introduces this sentence to allow for the explicit definition of all auxiliary terms. To achieve this, Lewis (1970, 429) first assumes that all auxiliary terms are constant names. By this assumption no "generality is lost, since names can purport to name entities of any kind: individuals, species, states, properties, substances, magnitudes, classes, relations, or what not." Thus all auxiliary terms can be replaced by constant names since, as Lewis (1970, 429) states, $\mathscr{B}$

provides the needed copulas:

---

11. Again, no assumptions about the semantics are required, and the proof does not even require the extensionality of predicates.

> ___ has the property ___
> ___ is in the state ___ at time ___
> ___ has ___ to degree ___

and the like.

Lewis (1970, 430) further assumes a logic in which constant names and definite descriptions without denotations in the domain refer to the same distinguished object. This object is not in the domain and thus lies outside of the scope of the normal quantifiers. Therefore identities between denotationless constant names or definite descriptions are true. On this basis, Lewis identifies the auxiliary terms of a theory $T(a_1, \ldots, a_n)$ with definite descriptions. That is, for each auxiliary term a_i,

$$a_i = \iota y_i \, \exists y_1 \ldots y_{i-1} y_{i+1} \ldots y_n \forall x_1 \ldots x_n$$
$$[T(b_1, \ldots, b_m, x_1, \ldots, x_n) \leftrightarrow (y_1 = x_1 \wedge \cdots \wedge y_n = x_n)] \,. \quad (13)$$

Because of Lewis's choice of logic, these equations can still be true even if the definite descriptions occurring within them do not have unique referents in the domain. For then the definite descriptions refer to the distinguished object (which is not in the domain) to which the theoretical terms have to refer as well. Under this assumption, and as is proper for definitions, the equations "do not imply any $[\mathscr{B}]$-sentences except logical truths" (438).[12]

However, Lewis's definitions of auxiliary terms typically do not follow from the theory T, for T entails equation (13) only if T also entails

$$\exists x_1 \ldots x_n \forall y_1 \ldots y_n [T(y_1 \ldots y_n) \leftrightarrow y_1 = x_1 \wedge \cdots \wedge y_n = x_n]$$
$$\dashv\vdash \exists! x_1 \ldots x_n \, T(b_1, \ldots, b_m, x_1, \ldots, x_n) \quad (14)$$

(Lewis 1970, 439). This is almost Papineau's notion of the Ramsey sentence, except that it does not involve higher order quantifiers and thus must rely on a rewritten theory. Lewis (1970, 432–434) assumes that scientific theories are *meant* to entail the unique realization of their auxiliary terms. But this completely trivializes the question of definability of theoretical terms: It is a condition of adequacy for the definition of constant names that the formula defining the symbol be uniquely realized (Hodges 1993, 59), and if the formula is uniquely realized, it can be used as definiens of the constant symbol. Lewis's assumption also completely trivializes the question of the possibility of conceptual engineering: If a formula is uniquely realized, there is exactly one object to which the corresponding theoretical term can refer; phrased in terms of properties, this means that there is exactly one property to which the term can refer, and thus there is no

---

12. As Schurz (2014, 306) observes, the use of the distinguished object "is just a 'formal trick' which cannot avoid the result that Lewis definitions […] do not determine the real reference of theoretical terms in worlds in which $[\mathsf{R}_{\mathscr{B}}(T)]$ is false." For one, any two definite descriptions or terms without unique referents refer to the same object.

choice of concepts whatsoever: Either the theory's auxiliary terms refer to these properties, or the theory is false.

As Papineau (1996, n. 4) expresses it, "we need to restrict the candidate referents for theoretical terms to *natural* properties or kinds, and exclude gerrymandered or gruesome entities." Lewis's rewriting of the theory so that auxiliary terms are constant names masks this assumption.[13] If there can be auxiliary predicate symbols, the uniqueness condition could be enforced by explicitly restricting quantification to natural predicates, which requires again a rewriting of the theory:[14] One could achieve the restriction by introducing the predicate 'is a natural kind' for predicates, for instance. This, however, in effect requires giving up on the Ramsey sentence because that predicate would also have to be treated as a basic term, even though it is clearly controversial in a debate with conceptual engineers (cf. Ainsworth 2009, §6).

Without rewriting the theory, the uniqueness assumption has consequences that are clearly not meant to be entailed by scientific theories: In the case of Papineau's simple theory of pain, the uniqueness condition entails that everyone in pain shows avoidance behavior *and* everyone who shows avoidance behavior is in pain. For real scientific theories the uniqueness assumptions has even more absurd consequences—at least if, as is assumed, the auxiliary terms are introduced by the theories themselves, so that the only postulates in which the auxiliary terms occur are those of the theory. For instance, Simon (1970, §2) shows in his analysis of Ohm's law that 'voltage' and 'internal resistance' can only be explicitly defined by 'resistance' and 'current intensity' if there are at least two electric circuits. Under a uniqueness assumption, Ohm's law therefore entails that there are at least two electric circuits. Simon (1947, 901) also lays out that in Newtonian particle mechanics, the component forces cannot be explicitly defined when a system contains more than four particles. Under a uniqueness assumption, Newtonian mechanics thus entails that there are at most four particles. The list could go on.

Papineau's argument for the irrelevance of the Carnap sentence thus either begs the question or is not sound: If one assumes that the theories are rewritten such that they fulfill the uniqueness condition, Papineau's conclusion follows; but it is based on assumptions that are explicitly denied by the conceptual engineer. If, on the other hand, the theories are not rewritten, the assumption that it fulfills the uniqueness condition is false. These are no idle technical considerations; rather, they are the technical guise of the philosophical point that empirical results do not determine new concepts, as seen in the simple theory about interference with an organism and the organism's reaction: There, very different auxiliary concepts have been introduced, all with logically equivalent Ramsey sentences. These different options are all possible because the Ramsey sentence, unlike the Carnap

---

13. I thank Adam Caulton who went beyond the call of duty to make me understand this point.

14. This point holds independent of any semantic assumptions that go beyond the standard formalism of higher order deduction, because quantifiers of type t can be instantiated by any formula of type t, even gerrymandered and gruesome ones.

sentence, does *not* introduce new concepts. And for this reason the Carnap sentence is conceptually interesting, even though it does not have empirical content.

4.2 Quine against conceptual engineering

Quine (1969) has given a highly influential argument for naturalism in epistemology, which he develops as a reaction to Carnap's preferred method of conceptual engineering: explication, or, as it is also called, 'rational reconstruction' (Brun 2016). Quine distinguishes between the "doctrinal side" of traditional epistemology (the attempt to justify all knowledge from sense experience) and its "conceptual side" (the attempt to explain all our concepts in sensory terms) (Quine 1969, 71–74) and presents Carnap's work *Der logische Aufbau der Welt* (Carnap 1928) as the most successful but still failed attempt at completing the conceptual side of epistemology by defining all concepts in sensory terms. Quine (1969, 75) identifies the *Aufbau* as an explicatory project, and wonders about the relevance of explication:

> But why all this creative reconstruction, all this make-believe? The stimulation of his sensory receptors is all the evidence anybody has had to go on, ultimately, in arriving at his picture of the world. Why not just see how this construction really proceeds? Why not settle for psychology?

Since this is a rhetorical question, Quine then proceeds to outline how psychology should replace epistemology. Of course, the question could be asked of any explication of a concept, including philosophical ones. Therefore this line of reasoning leads to a naturalization of all of philosophy.

The answer to Quine's rhetorical question is clear, however: Explication cannot be replaced by psychology because the goal of explication is not to find out about the actual concepts that humans have, but rather to find new concepts that are *better* than the actual concepts. A naturalized philosophy as described by Quine, and indeed any naturalized philosophy that only determines which concepts humans have, would only provide the first step of an explication (Lutz 2012a, §2.2): Getting clear about the concepts one wants to improve upon.

5 Conceptual engineering in the sciences

> Concepts that have proven useful in ordering things easily achieve such an authority over us that we forget their earthly origins and accept them as unalterable givens. Thus they come to be stamped as "necessities of thought," "a priori givens," etc. The path of scientific advance is often made impassable for a long time through such errors. For that reason, it is by no means an idle game if we become practiced in analyzing the long commonplace concepts and exhibiting those circumstances upon which their justification and usefulness depend, how they have grown up, individually, out of the givens of experience. By this means, their all-too-great authority will be broken. They will be removed if they cannot be properly legitimated, corrected

> if their correlation with given things be far too superfluous, replaced by others if a new system can be established that we prefer for whatever reason.
>
> <div align="right">Einstein (1916, 102), quoted by Howard (2015, §1)</div>

> [In Carnap's *Foundations of Logic and Mathematics* (Carnap 1939), designata] are admitted not only for concrete terms but also, in some cases at least, for abstract symbols and expressions. […] The reviewer would prefer a still more liberal admission of abstract designata, not on any realistic ground, but on the basis that this is the most intelligible and useful way of arranging the matter—it would apparently be meaningless to ask whether abstract terms really have designata, but it is rather a matter of taste or convenience whether abstract designata shall be postulated.
>
> <div align="right">(Church 1939, 822)</div>

The original motivation behind explication in philosophy was already methodologically naturalistic. Carnap's discussions very often rely on examples from the sciences, and he explicitly states that "[p]hilosophers, scientists, and mathematicians make explications very frequently" (Carnap 1962, §3). Hempel (1952, 664) similarly states:

> Explication is not restricted to logical and mathematical concepts […]. Thus, e. g., the notions of purposiveness and of adaptive behavior, whose vagueness has fostered much obscure or inconclusive argumentation about the specific characteristics of biological phenomena, have become the objects of systematic explicatory efforts. […] Similarly, the controversy over whether a satisfactory definition of personality is attainable in purely psychological terms or requires reference to a cultural setting centers around the question whether a sound explicatory or predictive theory of personality is possible without the use of sociocultural parameters; thus, the problem is one of explication.

Accordingly, science is teeming with explicata, such as 'temperature' explicating 'warm' (Carnap 1950, §4; Hempel 1952, §10), and completely new terms like 'phlogiston', 'oxygen', 'gene', and 'hydrochloric acid', which were introduced to account for empirical results described in basic terms like 'breathing', 'fire', 'child', and 'dissolving'.[15] I will argue that this view of scientific methodology is correct using examples from empirical research and conceptual arguments.

5.1 Empirical arguments

Empirical investigations of scientific methodology have been done with increasing rigor, and recent studies support the claim that science relies on conceptual

---

15. Schlick (1936, §1) analogously argues that Einstein arrived at his definition of 'simultaneity' by a method of concept formation that philosophers should follow. Unlike the later Carnap (1936, §15) and Hempel (1952, §6), however, Schlick assumes that the method must lead to explicit definitions of all terms in observation terms. I thank an anonymous reviewer for pointing out the relevance of Schlick's text.

engineering. Chang (2004), for example, has given his investigation of the formation of the temperature concept the title *Inventing Temperature*, and concludes on the basis of his investigation (206–208):

> It is very tempting to think that the ultimate basis on which to judge the validity of an operationalization should be whether measurements made on its basis yield values that correspond to the real values. But [an] unoperationalized abstract concept does not correspond to anything definite in the realm of physical operations, which is where values of physical quantities belong. […] Once an operationalization is made, the abstract concept possesses values in concrete situations. But we need to keep in mind that those values are products of the operationalization in question, not independent standards against which we can judge the correctness of the operationalization itself.

Thus there is no thing or property called 'temperature' in the world that is being found out. Rather, scientists develop and in effect *choose* this concept.

In an overview of biological concept formation, Stotz (2009, §3, footnote and reference removed) states that she and her colleagues

> have come to appreciate that conceptual change in science is rationally motivated by what scientists are trying to achieve, by their accumulated experience of how to achieve it, and by changes in what they are trying to achieve. […] The gene concept is a case in point: despite its ever-changing definition, the gene remains on the laboratory bench after a whole century because it has proved a flexible tool. This only makes sense if we think of concepts as tools of research, as ways of classifying the experience shaped by experimentalists to meet their specific needs. Necessarily these tools get reshaped as the scientists' needs change.

In other words, scientists choose their concepts according to the concepts' expedience. And specifically in biology, Stotz (2009, §4) comes to the following conclusion:

> It is simply that the molecular gene concept is not a good tool for some kinds of research. The instrumental, Mendelian gene remains the best tool in fields like medical genetics and population genetics. So while a particular scientific concept reflects the scientific knowledge at a point in time, this alone cannot explain the parallel use of several different concepts. For a full understanding of that phenomenon we need to see scientific concepts as tools for research, as much as glassware, microscopes or scales.

This is exactly the conclusion that Carnap (1963, 938–939) draws: "A natural language is like a crude, primitive pocketknife, very useful for a hundred different

purposes. But for certain specific purposes, special tools are more efficient, e.g., chisels, cutting-machines, and finally the microtome."

Thus there are good reasons to believe that some sciences, at least, rely on conceptual engineering. Justus (2012, abstract) explicitly makes this connection. On the basis of a case study of the concept of ecological stability, he argues that Carnap's theory of explication describes "how concepts should be characterized".

5.2   Conceptual arguments

There are some explicit arguments for semantic empiricism, the view that terms are assigned meaning through observation and convention. Rozeboom (1962), for instance, gives an elaborate defense. Carnap (1966, 187–188) also argues that any new concept of the scientific language (in my terminology: every auxiliary concept) must be chosen. Thus the sciences are forced to engineer concepts in the same way as philosophy:

> A working physicist is constantly coming upon methodological questions. What sort of concepts should he use? What rules govern these concepts? By what logical method can he define his concepts? How can he put his concepts together into statements and the statements into a logically connected system or theory? All these questions he must answer as a philosopher of science; clearly, they cannot be answered by empirical procedures.

Unfortunately, Carnap leaves the categorical statement of the last sentence without proof. But even so, it can be seen as a shifting of the burden of proof: To defend the position that science discovers rather than invents concepts like TEMPERATURE and GENE, one has to provide (and justify) an empirical procedure for deciding whether a concept is correct.[16]

Other assumptions about the sciences establish only indirectly that the sciences engage in conceptual engineering. In his argument for naturalized epistemology, Quine (1969, 81–82) concludes that "one has no choice but to be an empiricist so far as one's theory of linguistic meaning is concerned". He argues for the conventionality of language choice in the context of his argument for the indeterminacy of translation:

> [T]he linguist will end up with unequivocal translations of everything; but only by making many arbitrary choices [...]. By this I mean that different choices would still have made everything come out right that is susceptible in principle to any kind of check.

---

16. In effect, Carnap here relies on the same argument that he uses to criticize metaphysics (Bradley, forthcoming, §§2–3): If a sentence (like a sentence about the one correct meaning of a term) cannot be empirically tested, it cannot be a statement of fact. It can, however, be a pragmatically evaluable language suggestion (Lutz 2017, §6).

Very little is in principle susceptible to any kind of check, since the "all inculcation of meanings of words must rest ultimately on sensory evidence" (Quine 1969, 75). With this claim of the observational equivalence of different language choices, Quine provides pragmatic support for semantic empiricism, for even if words could non-conventionally apply to unobservable things or properties, whether they in fact do apply in a specific case must rely on observation. Without semantic empiricism, this view renders the truth of non-observation statements unknowable in principle due to lack of justification. Thus at a minimum, Quine's view entails that for all practical purposes, all non-observational concepts must be chosen. Arguably, however, Quine's position entails straightforward semantic empiricism, since Quine (1951, 41) explicitly embraces conventionalism:

> As an empiricist I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the situation as convenient intermediaries—not by definition in terms of experience, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. […] Both sorts of entities enter our conception only as cultural posits. The myth of physical objects is epistemologically superior to most in that it has proved more efficacious than other myths as a device for working a manageable structure into the flux of experience.

Since Quine here argues for choosing concepts the way the sciences do, he argues for conceptual engineering.

Quine's argument from the observational equivalence of alternative translations can be applied to other empiricist positions: In his constructive empiricism, van Fraassen (1980, §1.3) argues that non-observable objects are inaccessible to science, and thus science has to make do with empirical adequacy. The result is that there is no scientific means of determining the applicability or non-applicability of terms to unobservable objects (see Lutz 2014a, §3). As far as these objects are concerned, science has to rely on convention. Thus the statements of both philosophy and the sciences, whenever they go beyond observation, must be conventions about language use.

Similarly, Sober (1990, 404) states that according to his position of contrastive empiricism, "science is not in the business of discriminating between empirically equivalent hypotheses", where Sober considers theories empirically equivalent if they assign the same probabilities to all observation statements. Therefore, in particular the decision between two empirically equivalent theories with different concepts is a matter of choice.

5.3 A new conceptual argument

In the following, I want to add a new conceptual argument, or rather: Weaken the assumptions that are required for establishing a distinction between the analytic

and the synthetic component of a theory Θ in the simple case that allows the introduction of the Ramsey and the Carnap sentences. Carnap (1963, 23.D) argues for the adequacy of the Ramsey sentence $\mathsf{R}_{\mathscr{B}}(\Theta)$ as a theory's synthetic (in my terminology: uncontroversial) component σ and the Carnap sentence $\mathsf{C}_{\mathscr{B}}(\Theta)$ as a theory's analytic (conventional) component α based on the following conditions of adequacy for σ and α:

1. $\alpha \wedge \sigma \dashv\vDash \Theta$.

2. α entails only tautological $\mathscr{B}$-sentences.

3. σ does not contain any $\mathscr{A}$-terms.

4. σ is empirically equivalent to Θ.

I will show that there is a way of generalizing these conditions of adequacy such that they do not have to be restricted to the simple case in which basic sentences are identified via their vocabulary. I will assume a full (Tarski) semantics of higher order logic, as Carnap probably did as well (cf. Carnap 1956b, 43, 51, 62).

As first suggested by Caulton (2012), I will rephrase Carnap's conditions on sentences of the theory as conditions on the models of the theory: $\mathbf{T} := \{\mathfrak{A} : \mathfrak{A} \vDash \Theta\}$, $\mathbf{A} := \{\mathfrak{A} : \mathfrak{A} \vDash \alpha\}$, and $\mathbf{S} := \{\mathfrak{A} : \mathfrak{A} \vDash \sigma\}$. Then the following model-theoretic conditions of adequacy are (up to logical equivalent reformulations of the synthetic component) equivalent to the ones suggested by Carnap (Lutz 2014b, §3.2):

1. $\mathbf{A} \cap \mathbf{S} = \mathbf{T}$

2. $\forall \mathfrak{A} \exists \mathfrak{B} \in \mathbf{A} : \mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$

3. $\forall \mathfrak{A}, \mathfrak{B} : \mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}} \Rightarrow (\mathfrak{A} \in \mathbf{S} \Leftrightarrow \mathfrak{B} \in \mathbf{S})$

4. $\forall \mathfrak{A} : \exists \mathfrak{B}(\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}} \wedge \mathfrak{B} \in \mathbf{T}) \Leftrightarrow \exists \mathfrak{B}(\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}|_{\mathscr{B}} \wedge \mathfrak{B} \in \mathbf{S})$

Here '$\mathfrak{A}|_{\mathscr{B}}$' stands for the reduct of $\mathfrak{A}$ to the $\mathscr{B}$-vocabulary (Hodges 1993, 9). Note that in condition 2 it is implicitly assumed that the universal quantification ranges over structures of the right kind.[17] That the first and second conditions correspond to Carnap's conditions follows directly from basic Tarski semantics. The fourth condition demands that every $\mathscr{B}$-structure that is a reduct of a member of $\mathbf{T}$ is also a reduct of a member of $\mathbf{S}$ and vice versa. Roughly: $\mathbf{S}$ and $\mathbf{T}$ have the same reducts and are thus empirically equivalent. The third condition is the most complicated one: It expresses that whether a structure is in the synthetic component of a theory depends only on its reduct to the $\mathscr{B}$-vocabulary, which in turn means that the synthetic component does not restrict the interpretation of the $\mathscr{A}$-terms. Thus a sentence that is true in every structure of $\mathbf{S}$ must be equivalent to a sentence that does not contain any $\mathscr{A}$-terms.

---

17. To be somewhat more precise, all the structures in the conditions are assumed to have the same signature (Hodges 1993, 4–5).

It is straightforward to prove that the models of the Ramsey and Carnap sentences

$$\mathbf{S} := \{\mathfrak{A} : \mathfrak{A} \vDash \mathsf{R}_{\mathscr{B}}(\Theta)\} = \{\mathfrak{A} : \mathfrak{A}|_{\mathscr{B}} \in \mathbf{T}|_{\mathscr{B}}\}$$
$$\mathbf{A} := \{\mathfrak{A} : \mathfrak{A} \vDash \mathsf{C}_{\mathscr{B}}(\Theta)\} = \{\mathfrak{A} : \mathfrak{A}|_{\mathscr{B}} \notin \mathbf{T}|_{\mathscr{B}} \vee \mathfrak{A} \in \mathbf{T})\} \, ,$$

where $\mathbf{T}|_{\mathscr{B}} = \{\mathfrak{A}|_{\mathscr{B}} : \mathfrak{A} \in \mathbf{T}\}$, are adequate synthetic and, respectively, analytic components of $\mathbf{T}$ according to the model-theoretic conditions of adequacy.

The model-theoretic formulation is easily generalized when one considers the point of the reducts to the $\mathscr{B}$-vocabulary: Two structures have the same reduct if and only if they are empirically equivalent. What is important is thus the empirical equivalence of structures $\mathfrak{A}$ and $\mathfrak{B}$, which I will write as '$\mathfrak{A} \sim \mathfrak{B}$'. Then the following abstract conditions of adequacy for $\mathbf{A}$ and $\mathbf{S}$ suggest themselves:[18]

1. $\mathbf{A} \cap \mathbf{S} = \mathbf{T}$

2. $\forall \mathfrak{A} \exists \mathfrak{B} \in \mathbf{A} : \mathfrak{A} \sim \mathfrak{B}$

3. $\forall \mathfrak{A}, \mathfrak{B} : \mathfrak{A} \sim \mathfrak{B} \Rightarrow (\mathfrak{A} \in \mathbf{S} \Leftrightarrow \mathfrak{B} \in \mathbf{S})$

4. $\forall \mathfrak{A} : \exists \mathfrak{B}(\mathfrak{B} \sim \mathfrak{A} \wedge \mathfrak{B} \in \mathbf{T}) \Leftrightarrow \exists \mathfrak{B}(\mathfrak{B} \sim \mathfrak{A} \wedge \mathfrak{B} \in \mathbf{S})$

The overall class of structures could also be restricted to some class $\mathbf{C}$, so that $\mathbf{T}, \mathbf{A}, \mathbf{S} \subseteq \mathbf{C}$. Assuming that the equivalence relation ranges over $\mathbf{C}$, that is, $\mathfrak{A} \sim \mathfrak{B}$ only if $\mathfrak{A}, \mathfrak{B} \in \mathbf{C}$, the abstract conditions of adequacy do not have to be changed at all if one again silently assumes that the universal quantification in condition 2 ranges over structures of the right kind: the members of $\mathbf{C}$.

The previous model-theoretic conditions are a special case of these abstract conditions of adequacy (where $\mathfrak{A} \sim \mathfrak{B}$ if and only if $\mathfrak{A}|_{\mathscr{B}} = \mathfrak{B}|_{\mathscr{B}}$). The abstract conditions do not assume any specific language or any specific relation between a vocabulary and empirical statements. All they assume is that it is possible to determine when two structures are empirically equivalent, and this a very widely shared assumption

The abstract conditions of adequacy can be applied in a particularly nice way to a recent analysis of scientific theories: Andreas (2010, 538) develops a semantics for scientific theories that is holistic, and claims that

> it is rather misleading to construe relative holism as relying on the analytic-synthetic distinction. This becomes evident in light of the present account of semantic holism. In this account, only sentences qualifying as postulates are assumed to determine the meaning of theoretical terms. And the distinction between postulates and other theoretical sentences must clearly not be equated with the analytic-synthetic distinction. Analyticity is therefore no requirement for a sentence to determine the meaning of nonlogical symbols.

---

18. The conditions are based on joint work with Adam Caulton.

Since postulates can have both analytic and synthetic components, analyticity is clearly no requirement for determining the meaning of terms. The interesting question is, however, whether it is possible to distinguish between the analytic and synthetic component of postulates. I will now show that Andreas's account allows for the introduction of an analytic-synthetic distinction and thus for the conceptual engineering of the analytic component of theories.

Andreas (2010, 529–532) relies on a bipartition of the vocabulary into $\mathscr{B}$-terms that are observational and $\mathscr{A}$-terms that are theoretical, i.e. non-observational. Andreas (2010, 532) assumes a single intended $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ and a set U of unobservable objects, which are not in $O = \mathrm{dom}(\mathfrak{A}_{\mathscr{B}})$. $\mathrm{MOD}(\Theta)$ is the set of models of the postulates Θ, and $\mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U)$ the set of expansions to $\mathscr{V} = \mathscr{B} \cup \mathscr{A}$ of the extensions of $\mathfrak{A}_{\mathscr{B}}$ to $O \cup U$. In other words, $\mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U)$ is the set of all structures $\mathfrak{B}$ with domain $O \cup U$ that have $\mathfrak{A}_{\mathscr{B}}$ as a *relativized reduct*. The relativized reduct of $\mathfrak{B}$ to O and $\mathscr{B}$ is the restriction of the reduct $\mathfrak{B}|_{\mathscr{B}}$ to the domain O (cf. Hodges 1993, §5.1); I will write this as $\mathfrak{B}|O_{\mathscr{B}}$.[19] These concepts allow the following interpretation of the terms in $\mathscr{V}$ according to Θ and $\mathfrak{A}_{\mathscr{B}}$ (Andreas 2010, 533, my formulations)

Definition 1.

$$\mathbf{A}^*(\mathfrak{A}_{\mathscr{B}}) := \begin{cases} \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U) & \text{if } \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U) \neq \varnothing \\ \\ \mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U) & \text{if } \mathrm{MOD}(\Theta) \cap \mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U) = \varnothing \end{cases} \tag{15}$$

$\mathfrak{A}_{\mathscr{B}}$ combines empirical and conceptual information, because it assumes a fixed empirical state of the world and determines what, for this specific state, the interpretation of the auxiliary terms is. To express the interpretation of the auxiliary terms for *any* state of the world, define

$$\mathbf{A} := \bigcup \left\{ \mathbf{A}^*(\mathfrak{A}_{\mathscr{B}}) : \mathrm{dom}(\mathfrak{A}_{\mathscr{B}}) = O \right\} . \tag{16}$$

$\mathbf{A}$ contains the interpretations for each state of the world, since for any $\mathfrak{A}_{\mathscr{B}}$, $\mathrm{EXT}(\mathfrak{A}_{\mathscr{B}}, \mathscr{V}, U) \cap \mathbf{A} = \mathbf{A}^*(\mathfrak{A}_{\mathscr{B}})$. Thus, informally, restricting $\mathbf{A}$ to any $\mathfrak{A}_{\mathscr{B}}$ results in Andreas's semantics for $\mathfrak{A}_{\mathscr{B}}$. The previous expression (16) for $\mathbf{A}$ is equivalent to (Lutz 2014b, §4.3)

$$\mathbf{A} = \left\{ \mathfrak{A} : \mathrm{dom}(\mathfrak{A}) = O \cup U \wedge (\mathfrak{A}|O_{\mathscr{B}} \notin \mathbf{T}|O_{\mathscr{B}} \vee \mathfrak{A} \in \mathbf{T}) \right\} . \tag{17}$$

This already looks like an analogue to the models of the Carnap sentence and suggests a corresponding analogue to the Ramsey sentence:

$$\mathbf{S} := \left\{ \mathfrak{A} : \mathrm{dom}(\mathfrak{A}) = O \cup U \wedge \mathfrak{A}|O_{\mathscr{B}} \in \mathbf{T}|O_{\mathscr{B}} \right\} . \tag{18}$$

To check whether these two sets are adequate analytic and synthetic components of $\mathbf{T}$, we need to determine the conditions of adequacy for Andreas's semantics.

---

19. Hodges (1993, 202–203) defines relativized reducts as those substructures of a reduct that have the extension of some one-place predicate as their domain. I use a slight generalization.

Two structures $\mathfrak{A}$, $\mathfrak{B}$ are empirically equivalent in Andreas's semantics if and only if they have the same relativized reducts to $\mathscr{B}$ and O, $\mathfrak{A}|O_{\mathscr{B}} = \mathfrak{B}|O_{\mathscr{B}}$. Since Andreas restricts all structures to those with the domain $O \cup U$, the set $\mathbf{C}$ contains all and only structures with this domain. The abstract conditions of adequacy then give:

1. $\mathbf{A} \cap \mathbf{S} = \mathbf{T}$

2. $\forall \mathfrak{A} \exists \mathfrak{B} \in \mathbf{A} : \mathfrak{A}|O_{\mathscr{B}} = \mathfrak{B}|O_{\mathscr{B}}$

3. $\forall \mathfrak{A}, \mathfrak{B} : \mathfrak{A}|O_{\mathscr{B}} = \mathfrak{B}|O_{\mathscr{B}} \Rightarrow (\mathfrak{A} \in \mathbf{S} \Leftrightarrow \mathfrak{B} \in \mathbf{S})$

4. $\forall \mathfrak{A} : \exists \mathfrak{B}(\mathfrak{B}|O_{\mathscr{B}} = \mathfrak{A}|O_{\mathscr{B}} \wedge \mathfrak{B} \in \mathbf{T}) \Leftrightarrow \exists \mathfrak{B}(\mathfrak{B}|O_{\mathscr{B}} = \mathfrak{A}|O_{\mathscr{B}} \wedge \mathfrak{B} \in \mathbf{S})$

All conditions are indeed fulfilled: The proofs for the first and fourth conditions are straightforward. For the second condition, assume $\mathfrak{A} \in \mathbf{C}$. Then $\mathfrak{A}|O_{\mathscr{B}} \in \mathbf{T}|O_{\mathscr{B}}$ or $\mathfrak{A}|O_{\mathscr{B}} \notin \mathbf{T}|O_{\mathscr{B}}$. In the former case, there is some $\mathfrak{B} \in \mathbf{T}$ with $\mathfrak{B}|O_{\mathscr{B}} = \mathfrak{A}|O_{\mathscr{B}}$. Since $\mathfrak{B} \in \mathbf{T}$, it also holds that $\mathfrak{B} \in \mathbf{A}$. The latter case is trivial. For the third condition, assume $\mathfrak{A}|O_{\mathscr{B}} = \mathfrak{B}|O_{\mathscr{B}}$. $\mathfrak{A} \in \mathbf{S}$ if and only if $\mathfrak{A}|O_{\mathscr{B}} \in \mathbf{T}|O_{\mathscr{B}}$, which holds if and only if $\mathfrak{B}|O_{\mathscr{B}} \in \mathbf{T}|O_{\mathscr{B}}$ and thus if and only if $\mathfrak{B} \in \mathbf{S}$.

Thus the two sets of structures $\mathbf{A}$ and $\mathbf{S}$ fit with Andreas's semantics and are adequate analytic and synthetic components of the theory Θ. This gives an indication of the generality of the abstract conditions of adequacy, since they allow a distinction between the analytic and the synthetic components of theories for a semantics that was explicitly developed under the assumption that such a distinction may be impossible.[20]

In conclusion, Quine argues that the choice of language cannot be restricted with the help of an observational test, van Fraassen argues that there is no epistemic access to the applicability of terms to unobservables, Sober argues that science does not discriminate between observationally equivalent theories, and Andreas's semantics allows the distinction between an analytic and a synthetic component of theories. In all these cases, then, scientific theories can be partitioned into an observational component and a conventional component. How extensive this conventional component is can be inferred from an argument due to William Demopoulos.

6   Demopoulos against conceptual engineering

If not taken to simply beg the question against the conceptual engineer, Papineau argues, in effect, that the Ramsey sentence of a set of postulates Θ is so strong that there is nothing left for the Carnap sentence to do besides the labeling of concepts.

---

20. There are in fact two more ways of introducing a distinction between the analytic and the synthetic components of theories into Andreas's semantics (Lutz 2012b, 58–60; Lutz 2014b, §4.3), but the one presented here is the most general and seems to me the most natural.

Demopoulos (2003, 2007, 2008) on the other hand argues that the Ramsey sentence, with $\mathscr{B}$ as the set of observational terms and $\mathscr{A}$ as the set of theoretical, non-observational terms, is too weak even to capture the synthetic content of a theory.

In his first argument, Demopoulos (2003, 387) constructs an interpretation of Θ's $\mathscr{A}$-terms given a single intended $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$. Given that Θ is consistent, Θ has a model $\mathfrak{M}$, and he assumes that $M = \mathrm{dom}(\mathfrak{M})$ has the same cardinality as $A = \mathrm{dom}(\mathfrak{A}_{\mathscr{B}})$ without "any significant loss of generality or philosophical interest". Demopoulos (2003, 387, my notation) continues:

> It is therefore possible to extend the partial interpretation $[\mathfrak{A}_{\mathscr{B}}]$ to the theoretical vocabulary of Θ by letting each predicate of its theoretical vocabulary denote the image in A of its interpretation in $\mathfrak{M}$ under any one–one correspondence between M and A.[21] For example, suppose T is a binary theoretical relation of Θ. Then the interpretation $T^{\mathfrak{A}}$ of T in $\mathfrak{A}$ is *defined* as the image under φ, φ one–one from M on to A, of its interpretation $T^{\mathfrak{M}}$ in $\mathfrak{M}$. Since by construction $\langle a, b \rangle$ is in $T^{\mathfrak{M}}$ if and only if $\langle \varphi a, \varphi b \rangle$ is in $T^{\mathfrak{A}}$, φ is an isomorphism; and therefore, if $\mathfrak{M}$ is a model of Θ, so is $\mathfrak{A}$.

This is false:[22] What Demopoulos describes is a method with which, for any structure $\mathfrak{M}$ and any bijection φ from M to a set A, φ can be used to define a structure $\mathfrak{B}$ with domain A that is isomorphic to $\mathfrak{M}$. But there is no reason why the reduct $\mathfrak{B}|_{\mathscr{B}}$ of the resulting structure $\mathfrak{B}$ should be identical to a specific partial interpretation given by a $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ with domain A. As a simple example, choose the $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$ with $A = \{1, 2\}$ and $O^{\mathfrak{A}_{\mathscr{B}}} = \{\langle 1, 1 \rangle\}$, and theory $\Theta = \{\forall xy (Oxy \leftrightarrow Txy)\}$.[23] Demopoulos now claims that for any model $\mathfrak{M}$ of Θ and any bijection φ from M to A, the mapping of the extension $T^{\mathfrak{M}}$ to A by φ provides an extension that can be used in a expansion of $\mathfrak{A}_{\mathscr{B}}$ to a model of Θ. So choose $M = \{a, b\}$ and $O^{\mathfrak{M}} = T^{\mathfrak{M}} = \{\langle a, b \rangle\}$. Then $\mathfrak{M} \vDash \Theta$. But $\mathfrak{B}$ with $\mathfrak{B}|_{\mathscr{B}} = \mathfrak{A}_{\mathscr{B}}$ and $T^{\mathfrak{B}} = \{\langle \varphi a, \varphi b \rangle\}$ is not a model of Θ under any bijection $\varphi : M \longrightarrow A$ (let alone *all* bijections), because $\varphi a \neq \varphi b$, while Θ and $\mathfrak{A}_{\mathscr{B}}$ require that $T^{\mathfrak{B}} = \{\langle 1, 1 \rangle\}$. φ is an isomorphism from $\mathfrak{M}|_{\mathscr{A}}$ to $\mathfrak{B}|_{\mathscr{A}}$, but not from $\mathfrak{M}$ to $\mathfrak{B}$.

Demopoulos (2003, 387–388, my notation) goes on to criticize the Carnap sentence, since it entails that Θ is true whenever Θ's Ramsey sentence is true:

> Call the interpretation of Θ's $\mathscr{A}$-vocabulary in A that we have just described '$\mathfrak{A}_{\mathscr{A}}$' $[:= \mathfrak{A}|_{\mathscr{A}}]$. Any theory of knowledge and reference

---

21. Here and in the following, the original text does not distinguish between a structure and its domain (referring, for example, to the "the image in A of its interpretation in M under any one–one correspondence between M and A"). The distinction in my notation thus has to be taken as an interpretation of the quote.

22. Wagner (2009) was the first to criticize Demopoulos's proof, although the following argument differs from his.

23. Note that $\mathfrak{A}_{\mathscr{B}}$ has an expansion to a model of Θ.

that is incapable of distinguishing truth from truth under $\mathfrak{A}_{\mathscr{A}}$ is committed to the implication that Θ is true if Θ is true under $\mathfrak{A}_{\mathscr{A}}$. But modulo our assumption about cardinality, that Θ is true under $\mathfrak{A}_{\mathscr{A}}$ is a matter of model theory. $\mathfrak{A}_{\mathscr{A}}$ is arbitrary; the construction which employs it is clearly unacceptable, since it trivializes the question whether Θ is true. [...] By equating truth with truth under $\mathfrak{A}_{\mathscr{A}}$ we rob our knowledge of the truth of our theoretical claims of its a posteriori character: modulo a single assumption about cardinality, the theoretical statements of an empirically adequate theory come out true as a matter of metalogic.

Given the problem with Demopoulos's construction of $\mathfrak{A}_{\mathscr{A}}$, the criticism is too strong. The last quoted sentence states that *if* a theory is empirically adequate (that is, it entails only true $\mathscr{B}$-sentences), then it is true. The claim that $\mathfrak{A}_{\mathscr{A}}$ is arbitrary, however, would only be true if $\mathfrak{A}_{\mathscr{B}}$ were arbitrary as well, and that is not the case. In the counterexample given, $T^{\mathfrak{A}_{\mathscr{A}}} = T^{\mathfrak{A}}$ is fixed completely by $\mathfrak{A}_{\mathscr{B}}$ and the explicit definition of T by O that is given in Θ.

In his other arguments against the Carnap sentence, Demopoulos (2008, 376–377; 2011, 186–187) starts from a technical observation in first order logic (van Benthem 1978, lemma 3.2): For any $\mathscr{B}$-structure $\mathfrak{A}_{\mathscr{B}}$, if all $\mathscr{B}$-sentences entailed by Θ are true in $\mathfrak{A}_{\mathscr{B}}$, then there exists an expansion of an elementary extension of $\mathfrak{A}_{\mathscr{B}}$ that is a model of Θ. This trivially entails that there exists an expansion of an extension of $\mathfrak{A}_{\mathscr{B}}$ that is a model of Θ, and in Tarski semantics this entails that $\mathsf{R}_{\mathscr{B}}(\Theta)$ is true in an extension of $\mathfrak{A}_{\mathscr{B}}$. (If the basic sentences are assumed to be of higher order, this result is trivial, for then $\mathsf{R}_{\mathscr{B}}(\Theta)$ itself is a basic sentence entailed by Θ.)[24]

Demopoulos's conceptual point is the following (Demopoulos 2008, 377): If the Carnap sentence of Θ is taken to be an analytic truth, then the truth of Θ follows analytically from the truth of the basic sentences entailed by Θ—*modulo* a cardinality assumption (and in higher order logic without this assumption). This is correct. However, Demopoulos (2008, 381) further argues that the Carnap sentence

is incapable of accurately representing the truth of theoretical claims because it takes their truth to collapse into satisfiability in a sufficiently large domain. This is hardly what we take the truth of theoretical claims to consist in, since we characteristically—and rightly—distinguish them from those of pure mathematics. A reconstruction which fails to acknowledge this, is not merely odd, it misses what is arguably one of the chief desiderata of an adequate philosophy of the exact sciences.

---

24. In this argument, Demopoulos (2011, 187) also mentions an arbitrary mapping function (like φ in the previous argument), but restricts it to the objects that are newly introduced by the elementary extension of $\mathfrak{A}_{\mathscr{B}}$. Thus the arbitrariness of the interpretation of the theoretical terms is restricted to objects that are newly introduced by the elementary extension.

This conclusion is again too strong, for mathematical claims typically differ significantly from theoretical claims according to the Carnap sentence. For one, the truth of mathematical claims does not depend on the truth of any empirical claims, while the truth of theoretical claims does. Extending the above example, define $\Theta' := \Theta \cup \{\exists xy\, Txy\}$. Then $\mathsf{C}_{\mathscr{B}}(\Theta') \dashv\vdash \exists xy\, Oxy \rightarrow \Theta'$, so that the truth of any theoretical claim of Θ' depends on the $\mathscr{B}$-sentence $\exists xy\, Oxy$ being true. This is in stark contrast to mathematical claims. Furthermore, as the original example already shows, $\mathscr{A}$-terms can differ significantly from mathematical terms in that they may be explicitly definable in $\mathscr{B}$-terms. Finally, note that Demopoulos starts from the bare intuition that there must be something more substantial to scientific theories than only their empirical implications and their conventions. But this stance only contradicts the basic idea of semantic empiricism, without providing an argument.

Since Demopoulos's argument fails to show anything wrong with taking $\mathsf{R}_{\mathscr{B}}(\Theta)$ as the empirical content of Θ, his results can be turned around: They show how much in a theory is a matter of convention. Possibly *modulo* a cardinality assumption, the conceptual apparatus of a theory can be chosen at will, as long as the $\mathscr{B}$-sentences come out true.

7   Some implications

Both artificial language philosophy and science can be construed as relying on language choice, at least as far as choices between empirically equivalent sentences are concerned. Thus philosophical language choice can be practiced like language choice in the sciences and thus be naturalistic. In this way, the application of scientific methods to solve philosophical problems escapes Kim's criticism that a naturalized epistemology loses the normative component of traditional epistemology (Kim 1988). According to Kim, epistemology determines what, for instance, JUSTIFICATION should be, while, say, psychology only determines how people think about justification. But when relying on conceptual engineering, scientists, like philosophers, can develop a new concept called 'justification' that is fruitful in their research. And since this new concept is *suggested*, its development has a normative component.

Since in conceptual engineering analytic claims are typically suggested with the goal of accommodating specific facts, one can draw a number of tentative further conclusions. For one, the distinction between a priori and a posteriori claims is not particularly helpful: It is true that concepts could be suggested without any empirical input, but the resulting concepts are likely to be useless. The reason that we have the concept MOMENTUM, for instance, is that the product of mass and velocity is conserved. If it were not, there would be no reason to use the concept MOMENTUM, any more than there would be a reason to use a concept for the product of mass and age divided by the number of US presidents. This, I think, also defuses those of Quine's criticisms of analyticity that rely on the

idea that analytic sentences are true "come what may" (cf. Richardson 2003, 8): Analytic sentences are not the most certain of all sentences, they are rather the most malleable. And they are not held true come what may, but rather given up in response to empirical results.

Since the sciences rely on language conventions whenever they go beyond observational claims, most entities and properties referred to by scientific theories (for instance momentum) are *introduced* rather than discovered. According to methodological naturalism, philosophy then can also introduce entities and properties, and indeed *should* introduce them if their introduction is well-motivated. But this goes against ontological naturalism, which demands that philosophy not introduce properties or entities not condoned by the sciences. Thus, somewhat surprisingly, ontological and methodological naturalism are in tension.

Finally, if the sciences and philosophy are both involved in conceptual engineering, this allows for the cooperation between science and metaphysics, between science and ethics, and between science and aesthetics, for instance (Lutz 2012a, §5). So one of the paradigmatic armchair philosophies can fulfill naturalized philosophy's promise of a strong cooperation between the sciences and philosophy without rendering philosophy superfluous or subservient to scientific results. For conceptual suggestions in the sciences are, *prima facie*, on a par with conceptual suggestions in philosophy. Their adoption is justified, but not forced, by their usefulness.

References

Ainsworth, Peter M. 2009. "Newman's Objection." *The British Journal for the Philosophy of Science* 60:135–171.

Andreas, Holger. 2010. "Semantic Holism in Scientific Language." *Philosophy of Science* 77 (4): 524–543. doi:10.1086/656007.

Bradley, Darren. Forthcoming. "Carnap's Epistemological Critique of Metaphysics." *Synthese.* doi:10.1007/s11229-017-1335-x.

Brun, Georg. 2016. "Explication as a Method of Conceptual Re-engineering." *Erkenntnis* 81:1211–1241. doi:10.1007/s10670-015-9791-5.

Carnap, Rudolf. 1928. *Der logische Aufbau der Welt.* Berlin-Schlachtensee: Weltkreis-Verlag. References are to the translation (Carnap 1967).

———. 1936. "Testability and Meaning." *Philosophy of Science* 3 (4): 420–468.

———. 1939. *Foundations of Logic and Mathematics.* I:139–213. Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science 3. Chicago and London: University of Chicago Press. References are to the two-volume edition.

Carnap, Rudolf. 1950. "Empiricism, Semantics, and Ontology." *Revue Internationale de Philosophie* 4:20–40. References are to the slightly modified reprint (Carnap 1956a, appendix A).

———. 1956a. *Meaning and Necessity.* 2nd ed. Chicago: University of Chicago Press.

———. 1956b. "The Methodological Character of Theoretical Concepts." In *The Foundations of Science and the Concepts of Psychology and Psychoanalysis,* edited by Herbert Feigl and Michael Scriven, 1:38–75. Minnesota Studies in the Philosophy of Science. Minneapolis, MN: University of Minnesota Press.

———. 1962. *Logical Foundations of Probability.* 2nd ed. Chicago: University of Chicago Press.

———. 1963. "Replies and Systematic Expositions." In Schilpp 1963, 859–1016.

———. 1966. *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science.* Edited by Martin Gardner. New York and London: Basic Books, Inc.

———. 1967. *The Logical Structure of the World. Pseudoproblems of Philosophy.* Translated by Rolf A. George. Berkeley and Los Angeles: University of California Press.

Caulton, Adam. 2012. "Theoretical Analyticity, Revisited." Talk presented at the 2012 *Annual Conference* of the *British Society for the Philosophy of Science*, University of Stirling, 5 July 2012.

Chang, Hasok. 2004. *Inventing Temperature: Measurement and Scientific Progress.* Oxford Studies in the Philosophy of Science. New York: Oxford University Press.

———. 2005. "A Case for Old-fashioned Observability, and a Reconstructed Constructive Empiricism." *Philosophy of Science* 72 (5): 876–887. doi:10.1086/508116.

Church, Alonzo. 1939. "Review: Rudolf Carnap, *Foundations of Logic and Mathematics.*" *Bulletin of the American Mathematical Society* 45 (11): 821–822.

Demopoulos, William. 2003. "On the Rational Reconstruction of our Theoretical Knowledge." *The British Journal for the Philosophy of Science* 54:371–403.

———. 2007. "Carnap on the Rational Reconstruction of Scientific Theories." In *The Cambridge Compation to Carnap,* edited by Michael Friedman and Richard Creath, 248–272. Cambridge: Cambridge University Press.

———. 2008. "Some Remarks on the Bearing of Model Theory on the Theory of Theories." *Synthese* 164 (3): 359–383. doi:10.1007/s11229-008-9355-1.

Demopoulos, William. 2011. "Three Views of Theoretical Knowledge." *The British Journal for the Philosophy of Science:* 177–205. doi:10.1093/bjps/axq021.

Eberle, Rolf. 1990. "Classification by Comparison with Paradigms." *American Philosophical Quarterly* 27 (4): 295–304.

Einstein, Albert. 1916. "Ernst Mach." *Physikalische Zeitschrift* 17:101–104.

Eklund, Matti. 2009. "Carnap and Ontological Pluralism." In *Metametaphysics: New Essays on the Foundations of Ontology,* edited by David Chalmers, David Manley, and Ryan Wasserman, 130–156. Oxford: Oxford University Press.

Feynman, Richard P., and Ralph Leighton. 1989. *"Surely You're Joking, Mr. Feynman!" Adventures of a Curious Character.* Edited by Edward Hutchings. New York: Bantam Books.

Glymour, Clark. 1970. "Theoretical Realism and Theoretical Equivalence." Edited by Roger C. Buck and Robert S. Cohen, (Dordrecht), Boston Studies in the Philosophy of Science, no. VIII: 275–288.

Hempel, Carl G. 1952. *Fundamentals of Concept Formation in Empirical Sciences.* Vol. II,7. Foundations of the Unity of Science. Toward an International Encyclopedia of Unified Science. Chicago and London: University of Chicago Press. References are to the two-volume edition.

Hodges, Wilfrid. 1993. *Model Theory.* Vol. 42. Encyclopedia of Mathematics and its Applications. Cambridge: Cambridge University Press. Digitally printed in 2008.

Howard, Don A. 2015. "Einstein's Philosophy of Science." In *The Stanford Encyclopedia of Philosophy,* Winter 2015, edited by Edward N. Zalta. Stanford: The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University.

Justus, James. 2012. "Carnap on Concept Determination: Methodology for Philosophy of Science." *European Journal for Philosophy of Science* 2 (2): 161–179. doi:10.1007/s13194-011-0027-5.

Kim, Jaegwon. 1988. "What Is "Naturalized Epistemology"?" *Philosophical Perspectives* 2:381–405.

Kyburg, Henry E. 1990. "Theories as Mere Conventions." In Savage 1990, 158–174.

Lewis, David. 1970. "How to Define Theoretical Terms." *The Journal of Philosophy* 67 (13): 427–446. doi:10.2307/2023861.

Lutz, Sebastian. 2012a. "Artificial Language Philosophy of Science." *European Journal for Philosophy of Science* 2 (2): 181–203. doi:10.1007/s13194-011-0042-6.

Lutz, Sebastian. 2012b. "Criteria of Empirical Significance: Foundations, Relations, Applications." PhD diss., Utrecht University. http://philsci-archive.pitt.edu/id/eprint/9117.

———. 2014a. "Generalizing Empirical Adequacy I: Multiplicity and Approximation." *Synthese* 191:3195–3225. doi:10.1007/s11229-014-0440-3.

———. 2014b. "The Semantics of Scientific Theories." In *Księga pamiątkowa Marianowi Przełęckiemu w darze na 90-lecie urodzin,* edited by Anna Brożek and Jacek Jadacki, 33–67. Lublin: Nobertinum. Authoritative preprint: http://philsci-archive.pitt.edu/id/eprint/9630.

———. 2017. "Carnap on Empirical Significance." *Synthese* 194:217–252. doi:10.1007/s11229-014-0561-8.

Maxwell, Grover, and Herbert Feigl. 1961. "Why Ordinary Language Needs Reforming." *The Journal of Philosophy* 58 (18): 488–498.

Papineau, David. 1996. "Theory-Dependent Terms." *Philosophy of Science* 63 (1): 1–20.

———. 2009a. "Naturalism." In *The Stanford Encyclopedia of Philosophy,* Spring 2009, edited by Edward N. Zalta. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University.

———. 2009b. "The Poverty of Analysis." *Aristotelian Society Supplementary Volume* 83 (1): 1–30. doi:10.1111/j.1467-8349.2009.00170.x.

Przełęcki, Marian. 1969. *The Logic of Empirical Theories.* Monographs in Modern Logic Series. London/New York: Routledge & Kegan Paul/Humanities Press.

———. 1974. "On Model Theoretic Approach to Empirical Interpretation of Scientific Theories." *Synthese* 26:401–406.

Quine, Willard Van Orman. 1951. "Two Dogmas of Empiricism." *The Philosophical Review* 60 (1): 20–43. doi:10.2307/2181906.

———. 1969. "Epistemology Naturalized." In *Ontological Relativity and Other Essays,* 69–90. The John Dewey Essays in Philosophy 1. New York and London: Columbia University Press.

———. 1975. "On Empirically Equivalent Systems of the World." *Erkenntnis* 9:313–328.

Reichenbach, Hans. 1951. "The Verifiability Theory of Meaning." *Proceedings of the American Academy of Arts and Sciences* 80 (1): 46–60.

Richardson, Alan W. 2003. "Logical Empiricism, American Pragmatism, and the Fate of Scientific Philosophy in North America." In *Logical Empiricism in North America,* edited by Gary L. Hardcastle and Alan W. Richardson, 1–24. Minnesota Studies in the Philosophy of Science 18. Minneapolis, MN: University of Minnesota Press.

Rozeboom, William W. 1962. "The Factual Content of Theoretical Concepts." In *Scientific Explanation, Space, and Time,* edited by Herbert Feigl, Michael Scriven, and Grover Maxwell, 3:273–357. Minnesota Studies in the Philosophy of Science. Minneapolis, MN: University of Minnesota Press.

Savage, C. Wade, ed. 1990. *Scientific Theories.* Vol. 14. Minnesota Studies in the Philosophy of Science. Minneapolis, MN: University of Minnesota Press.

Schilpp, Paul Arthur, ed. 1963. *The Philosophy of Rudolf Carnap.* The Library of Living Philosophers 11. Chicago and LaSalle, IL: Open Court Publishing Company.

Schlick, Moritz. 1936. "Meaning and Verification." *The Philosophical Review* 45 (4): 339–369.

Schurz, Gerhard. 2014. *Philosophy of Science. A Unified Approach.* New York, NY, and London: Routledge.

Simon, Herbert A. 1947. "The Axioms of Newtonian Mechanics." *Philosophical Magazine* 38 (7).

———. 1970. "The Axiomatization of Physical Theories." *Philosophy of Science* 37:16–26.

Sober, Elliott. 1990. "Contrastive Empiricism." In Savage 1990, 392–410.

Stotz, Karola. 2009. "Experimental Philosophy of Biology: Notes From the Field." *Studies In History and Philosophy of Science Part A* 40 (2): 233–237.

Strawson, P. F. 1963. "Carnap's Views on Constructed Systems versus Natural Languages in Analytic Philosophy." In Schilpp 1963, 503–518.

Tuomela, Raimo. 1973. *Theoretical Concepts.* Library of Exact Philosophy 10. Wien: Springer.

Van Benthem, Johann. 1978. "Ramsey Eliminability." *Studia Logica* 37:321–336.

Van Fraassen, Bas C. 1980. *The Scientific Image.* The Clarendon Library of Logic and Philosophy. Oxford: Clarendon Press.

Wagner, Pierre. 2009. "Carnap's Rational Reconstruction of Theories and Newman's Argument." Talk presented at the *EPSA 09,* Amsterdam, October 2009.

Weatherall, James Owen. 2016. "Are Newtonian Gravitation and Geometrized Newtonian Gravitation Theoretically Equivalent?" *Erkenntnis* 81 (5): 1073–1091.