Definability and Reducibility
—Two Case Studies—

Sebastian Lutz*

2008–05–23

Abstract

I review some general features of definitions and how the traditional discussion about reducibility can be phrased in terms of the concept of definability. Using two case studies, I argue that contemporary discussions about reducibility still hinge on certain features of definitions and apply results from definition theory. First, I show how the discussion between Batterman and Belot revolves around the implicit definability of a reduced theory's terms and the translatability of the reduced theory's formulas, and I argue that as a result, Batterman's position hinges on his concept of explanation. Second, I argue that the discussion about New Wave Reductionism in the philosophy of mind is implicitly also one about definability, and I suggest that this discussion suffers from ambiguities between the intensional and psychological concepts of understanding and between actual and logically possible research.

1 Introduction

There are at least two steps in the reduction of a theory T_2 to a theory T_1. The first step is the recovery of some of T_2's concepts by the concepts of T_1. The second is the recovery of some part of T_2 with the help of T_1. While the second step has been explicated in much depth with the help of concepts in logic, mathematics, mereology, and explanation, the first step has not even received a thorough treatment in logic. It is the goal of this paper to do this in two case studies. The first case study concerns the discussion between Batterman and Belot about the relevance of reduced theories for the explanation of phenomena (§ 4), and the second concerns the mind-body problem and the discussion about New Wave Reductionism (§ 5). To show the role that specific properties of definitions play in both cases, I give a short overview of results from definition theory first (§ 2). An overview of the role of definability in reduction follows (§ 3).

*Department of Philosophy, University of Western Ontario. sebastian.lutz@gmx.net. This paper was written during a visiting fellowship at the Tilburg Center for Logic and Philosophy of Science. I would like to thank TiLPS for the support.

2 Some aspects of definability

I will consider the concepts of a theory to be the (partly primitively interpreted of) terms that theory, so that T_1 is formulated in a set $\mathscr{V}_1$ of terms (non-logical constants) and T_2 in a set $\mathscr{V}_2$ of terms. Since $\mathscr{V}_2$ is to be recovered from $\mathscr{V}_1$, define the set of basic terms $\mathscr{B} := \mathscr{V}_1$ and the set of auxiliary terms $\mathscr{A} := \mathscr{V}_2 \backslash \mathscr{V}_1$. Define also $\mathscr{V} := \mathscr{V}_1 \cup \mathscr{V}_2$. Sentences that contain only $\mathscr{V}_1$-terms are $\mathscr{V}_1$-sentences, analogously for formulas and for $\mathscr{B}$, $\mathscr{A}$, $\mathscr{V}_2$ and $\mathscr{V}$.

In early discussions about reduction between theories (Nagel 1951, Woodger 1952), the demand was arguably that each term of the reduced theory be explicitly definable in terms of the reducing theory (Kemeny and Oppenheim 1952). Focussing for simplicity just on predicates, this means that there is a set Δ of definitions such that for each n-place predicate $P \in \mathscr{A}$ there is a $\mathscr{B}$-formula ψ with n free variables such that

$$\Delta \vDash \forall x_1, \ldots, x_n [P x_1, \ldots x_n \leftrightarrow \psi(x_1, \ldots, x_n)] \quad ,$$

that is, Δ entails an explicit definition of P in $\mathscr{B}$. Here, the left-hand side of the biconditional is the definiendum and the right hand-side is the definiens.

Since the definitions in Δ are not supposed to contain empirical information about the realm of $\mathscr{B}$, Δ also has to be conservative (or non-creative) in $\mathscr{B}$, which is a common necessary condition for any kind of meaning postulate: A set of $\mathscr{V}$-sentences Δ is $\mathscr{B}$-conservative with respect to a set Λ iff for any set Γ of $\mathscr{B}$-sentences and for any $\mathscr{B}$-sentence β, $\Delta \cup \Lambda \cup \Gamma \vDash \beta$ only if $\Lambda \cup \Gamma \vDash \beta$. Δ is conservative in $\mathscr{B}$ iff Δ is conservative in $\mathscr{B}$ with respect to the empty set. For first order logic, the use of the set Γ in the definition is superfluous.

The explicit definability of all $\mathscr{A}$-terms is desirable because it ensures the translatability of each $\mathscr{V}$-formula into $\mathscr{B}$ (cf. Rantala 1977, ch. 14). More precisely, for each $\mathscr{V}$-formula φ, there is a $\mathscr{B}$-formula ψ such that

$$\Delta \vDash \varphi \leftrightarrow \psi \quad .$$

Note that this is just the demand for the eliminability of all terms of $\mathscr{A}$ by Δ.

According to a theorem by Padoa and Beth, a term P is explicitly definable in Δ iff it is implicitly definable, that is, iff any interpretation of all terms of Δ except P in which all sentences of Δ are true uniquely determines the interpretation of P (cf. Slomson and Bell 2006, § 8.4). The theorem only holds for finite Δ in higher order logic (Shapiro 2000, § 6.6.3).

Explicit definability of a term is probably the strongest plausible condition for the reducibility of a term. The strengthening of this condition by restricting the definiens to molecular formulas is a remnant of the idea that confirmability, meaningfulness, and reducibility to observation terms are the same (Gemes 1998). However, even if this restriction is plausible for criteria of confirmation, it is completely implausible for criteria of reduction, where the sentences in Δ are already assumed to be inductively confirmed.

Of more interest are the weakenings of explicit definability to conditional definability (cf. Essler and Trapp 1977), which ensures eliminability if some condition is fulfilled, and piecewise definability (explicit definability up to disjunction), which ensures eliminability in all consistent complete extensions of Δ (cf. Rantala 1977, ch. 8,14.2).

This bit of definition theory already goes a long way towards analyzing some problems of reduction. The most important aspects of definitions will be the dependence of definability on the set Δ and the intensionality of definitions.

The latter point may seem puzzling at first, since the semantics of predicate logic is extensional (two predicates are equivalent in a model iff they have the same extension). But as, for example, Belnap (1993) has pointed out, if definitions were extensional, then any (possibly contingent) true sentence could be translated into any other (possibly contingent) true sentence. What is needed for translatability of formulas is their intensional equivalence, that is, their extensional equivalence in all models of Δ.

While definitions give the intension of the defined term, they do not give the extension of the term if the definiens' extension is not given. More importantly, a definition does not give the intention of the defined term, that is, it is not hyper-intensional (cf. Belnap 1993). This limitation has to be taken into account, for example, whenever a theory in $\mathscr{B}$ is definitionally extended to explain a phenomenon described in $\mathscr{B}$: While intensionally, the theory is as informative as its definitional extension with respect to $\mathscr{B}$, it could very well be that, for some person, the definitional extension allows for a derivation of the phenomenon's description that is understandable, while the original theory does not.

A similar point can be made for the meaning of words. It is clear that two formulas cannot have the same meaning if they are not intensionally equivalent. But if meaning is understood as a psychological or any other hyper-intensional concept, then the meaning of two intensionally equivalent formulas may indeed differ. Hence the explication of the meaning of a formula as its interpretation in all models is not a psychological one.

3 Some aspects of reducibility

The dependence of definability on a set Δ is unavoidable, since for $\Delta = \emptyset$, only tautologies are derivable so that any $\mathscr{A}$-term appears only trivially. If a theory's vocabulary is extended by stipulative definition, Δ only contains the definitions of the new terms, but in reductions, matters are somewhat more complicated. In the easiest case, a theory T_1 that reduces a theory T_2 does not correct but rather extends T_2. A straightforward explication of this intuition is that T_2 is conservative with respect to T_1 in the theories' common terms $\mathscr{C} := \mathscr{V}_1 \cap \mathscr{V}_2 \subseteq \mathscr{B}$. In first order logic, this means that all $\mathscr{C}$-sentences entailed by T_2 are also entailed by T_1.

Initially, $\mathscr{C}$ was assumed to contain only observation terms (Kemeny and Oppenheim 1952), but Sklar (1967) suggests the above conception of $\mathscr{C}$, and Darrigol (2007) shows how, historically, the same non-observable concepts are used in different physical theories. It is in light of these terms, which are interpreted in the common background of the two theories, that one theory can be "empirically" more or as successful. The remaining terms of T_2, $\mathscr{A}$, are now partially interpreted by T_2 itself if they have any bearing on $\mathscr{C}$-statements (cf. Wójcicki 1966). Note that T_2 cannot be taken to be a set of definitions only, because it is conservative in $\mathscr{C}$ with respect to T_1, not with respect to $\emptyset$. This is unproblematic because $T_2 \cup T_1$ can still be seen as a definitional extension of T_1. Interestingly enough, connections between $\mathscr{A}$ and $\mathscr{C}$ are wholly in $\mathscr{V}_2$ and hence plainly in the domain of T_2. In this sense, an extension of the reducing theory improves the chances of its reducibility, that is, the definability of its terms in $\mathscr{B}$, and hence in terms of T_1.

One could now argue that, since T_2 already contains empirically established statements, it is not a qualitatively new step to allow further empirical research to establish additional connections between $\mathscr{A}$-terms and $\mathscr{B}$-terms. This seems true to me, but may brush over an important point: It may be that the empirical research necessary to connect the two theories lies within the realms of application of T_1 and T_2, not in the realm of some overarching theory containing both T_1 and T_2. This is because each theory is considered sufficiently interpreted by itself, without the need for the respective other theory. These interpretations determine the intended models for each theory, and these intended models can be compared *without further empirical research*. If the intended models are not determined accurately enough, this is a problem within each theory, not a problem specific to the reduction between the two theories.

This point is obvious if the interpretation of $\mathscr{A}$-terms is given by meaning postulates connecting them to $\mathscr{C}$-terms. Then, since $\mathscr{C}$ terms are interpreted in both theories, there is no need for further research. Some $\mathscr{A}$-terms may have an interpretation or a range of possible interpretations in all models of T_2 without the interpretations being connected to $\mathscr{C}$. But if the terms have interpretations, then these interpretations can be compared to those of the terms defined in $\mathscr{B}$ without further research.

4 Limiting cases: Batterman and Belot

From a variety of examples in physics, Batterman (2002) concludes that even when a theory T_3 has been superseded by some other theory T_1—in the sense that T_1 explains every individual phenomenon at least as accurately as T_3—there are still some explanations that require the use of concepts from T_3 and that therefore cannot be given by T_1. Specifically, while T_1 may explain each individual phenomenon that T_3 explains, the explanation of patterns of phenomena of a given type may require a new theory T_2, which is found by taking some parameter in T_1 to a limit. The resulting limiting theory T_2 is such that "elements from both theories seems to be intimately intertwined" (Batterman 2002, p. 94; see also Batterman 2007, § 4). Since T_1 does not explain everything that T_2 explains, and T_2 relies on concept from T_3, T_3 cannot be reduced to T_1.

Belot (2005) discusses two of Batterman's major case studies, the explanation of the rainbow by wave and geometrical optics and the explanation of the energy spectra of chaotic systems by quantum and classical mechanics. Arguing that wave theory (T_1) is sufficient to explain the rainbow pattern, Belot shows that the limit procedures investigated by Batterman allow for the explicit definition of those concepts of geometrical optics (T_3) that occur in T_2. The concepts of T_3 are hence eliminable in favor of those of T_1. He argues similarly for the explanation of the energy spectra of chaotic systems. Batterman (2005) responds that applying the equations of T_1 to the phenomena is impossible without concepts from T_3.

Both Batterman (2002, pp. 3, 7) and Belot (2005, p. 129) note that for Batterman, understanding about physics is gained by attending to scientific details, while understanding within physics is gained by ignoring details. While I think that there is no way around detailed case studies to learn about science, I will ignore a lot of these details in my discussion of Batterman's and Belot's exchange and show how this leads to an understanding of the role of definability in the exchange.

The first step of my discussion is to adequately explicate the concept of explanation as it is used in the exchange. Despite Belot's repeated reference to understanding (e. g. pp. 129, 133, 152), I think it is obvious that the concept of explanation considered by Belot is not a hyper-intensional one. In order to prove his point with a hyper-intensional meaning of 'explanation', Belot would not be able to rely on definitions because a hyper-intensional context does not in general allow for equivalent reformulation of equations. Specifically, if an explanation were to provide psychological understanding, Belot would have to show that the definitions he introduces do not aid this understanding. And this seems false: The definitional extension of wave theory makes it easier to understand the derivation of the rainbow's description, and it is plausible that at least some people will be able to understand the resulting explanation, while being unable to understand a direct explanation without the use of intermediate definitions.

Batterman (2005) does not explicitly question Belot's presumption that the relevant concept of explanation is intensional, but he does make a distinction between explanation and understanding:

> Belot's strategy in arguing against [the explanatory inadequateness of T_1] is to demonstrate that the mathematics of the fundamental theory [T_1] contains, in an appropriate sense, the mathematics of the emeritus theory [T_2]. [...] If this is right, and if such explanation provides understanding, then fundamental theories (at least those considered in the book and by Belot) are perfectly explanatorily adequate. (p. 154)

I will discuss below what the "appropriate sense" of containment may be. For now, it is important to see that for Batterman, explanations can come with and without understanding, and this understanding is relevant (maybe necessary) for T_1 to be explanatorily adequate. Unfortunately, Batterman does not give his understanding of 'understanding'. If what he refers to is psychological understanding, then Belot's argument is incomplete, as argued above. In that case, however, Batterman faces two difficulties: First, his arguments for his position seem to miss the point, because the properties of limit procedures and singularities that he adduces are only relevant insofar as he can also show that they make a difference in psychological understanding. Furthermore, they are quite probably not needed to prove his point, since he just has to show that explanations based on T_2 are psychologically easier to understand than those based on T_1, which is probable for wave optics and its limiting theory, since as Belot (2005, p. 140) himself points out, no analytic solutions of the equation of wave optics are known for the general phenomenon of rainbows:

> We cannot, at this time, say very much about the behavior of light—even in its qualitative aspects—in problems closely related to the physical problem for perfectly spherical raindrops. We do not have any overarching theorems. [...] But we can do better in the high frequency limit.

Second, Batterman's argument is as incomplete as Belot's, for Belot relies on the intensional equivalences given by mathematical derivations to show that T_1 explains the phenomena, and Batterman relies on the intensional equivalences to show that T_2 explains the phenomena.

Summing up, T_1 is in many cases, and very probably in the cases discussed by Batterman and Belot, explanatorily inadequate if explanation is understood to be a psychological notion. In this case, the arguments by Batterman and Belot also miss the point. So a more caritable—and clearly more likely—view is that Batterman and Belot do not use this concept of explanation. And although it is not completely clear from the exchange what concept of explanation Batterman and Belot use, their extensive use of mathematical machinery suggests that for the cases under discussion, a theory T can be taken to explain a phenomenon described by φ if T entails φ. Batterman's claim that "asymptotic explanation", the explanation of phenomena by means of limit procedures, is, "at the very least, quite different from standard philosophical accounts" (2005, p. 157), is quite compatible with this result: His remark is mainly directed at the deductive-nomological account by Hempel and Oppenheim and Hempel's position that "the notion of explaining particular instances of a given kind of occurrence is the primary one" (1965, p. 423). Nothing in the above suggests, however, that φ has to describe a particular instance. It may well be the description of a general kind.

Taking explanation to mean entailment, Batterman's position would be that T_1 entails a description of one specific phenomenon, but not a generic description of phenomena of the same type, while T_2 entails the generic description. Take φ_i to be a description of an individual phenomenon, and a representative of a type of phenomena $\{\varphi_i\}_{i\in I}$ for some index set I. Batterman's claim then seems to be that

$$T_1 \vDash \varphi_i \quad \text{for all } i \in I \quad ,$$

but

$$T_1 \nvDash \forall i \in I : \varphi_i \quad . \tag{1}$$

If explanation were a question of provability, not entailment, there would be cases like this if the language of T_1 contains arithmetic, which it does if T_1 is the wave theory of light. But this would mean that a sentence stating the existence of rainbows in their generic description is true but unprovable in T_1. This would be an impressive result, but does not seem to be warranted without an actual proof. There is also no obvious reason to restrict the inferential rules of physics to provability, especially since physicists do not seem to restrict themselves to purely syntactical derivations either.

Batterman's argument sets in earlier, however: He argues that the statement '$\forall i \in I : \varphi_i$' cannot be given in the terms of T_1, as is clear from his response to Belot's discussion of the rainbow:

> If one is going to say how the perturbation of the shape of a macroscopic object, characterized by variations in [the drop's radius] r, figures in the stability of the fringe spacings and intensities [of the rainbow], then it seems one needs to refer to the relationship between that macroscopic object and the lines of concentrated intensity exiting the object as its shape is altered. In other words, the very investigation of the stability under perturbation of the equations involves considering perturbations to the boundaries understood in terms of structures foreign to the fundamental theory. (Batterman 2005, p. 160)

And,

> the theory required to characterize [the initial and boundary conditions] as appropriate for the rainbow problem in the first place is the theory of geometrical optics. (p. 159)

Note that Batterman explicitly refers to the perturbation of the drop's shape, and hence to a class of boundary conditions. If one were to take geometrical optics to be necessary for individual boundary conditions, it would be impossible to derive the specific description of a single rainbow from T_1 without T_2, as Belot (2005, fn. 44) points out. Note also that the shape of the drop cannot be derived from either ray or wave theory, but at best, the conditions under which a rainbow occurs can be so derived.

To spell out Batterman's position in more detail, one can split up each phenomenon φ_i into the statement that there is a rainbow, ρ, and a description of the initial and boundary conditions, $\Gamma_i, i \in I$, leading to a rainbow. The earlier paraphrase of T_1's explanatory inadequateness (1) thus becomes

$$T_1 \not\models \forall i \in I : \Gamma_i \rightarrow \rho \quad ,$$

which is equivalent to

$$T_1 \not\models (\exists i \in I : \Gamma_i) \rightarrow \rho \quad . \tag{2}$$

For this paraphrase to be well formed, the entailment relation has to be defined over $\mathscr{V}_1 \cup \mathscr{V}_2 = \mathscr{B} \cup \mathscr{A}$. This is a harmless assumption, since it does not change T_1. Batterman's argument is then that '$\exists i \in I : \Gamma_i$' is a term of T_2 that cannot be recovered in T_1.

Belot's argument attempts to show that the claim of explanatory inadequateness (2) is false. Its starting point is Batterman's assumption that T_2 explains the generic phenomenon, that is

$$T_2 \vDash (\exists i \in I : \Gamma_i) \rightarrow \rho \quad .$$

Belot (2005) gives a set Δ of explicit definitions for the terms of T_2 in terms of T_1, that is, he defines $\mathscr{V}_2$ in $\mathscr{V}_1$. He shows that $T_1 \cup \Delta \vDash T_2$ and can therefore conclude that with the help of his definitions Δ, T_1 also explains the phenomenon,

$$T_1 \cup \Delta \vDash (\exists i \in I : \Gamma_i) \rightarrow \rho \quad .$$

There might be a problem with the definitions Δ, however, and this point seems to form the core of Batterman's response to Belot. Outlining Belot's argument, Batterman (2005) states that if

> the mathematics of the fundamental theory contains, in an appropriate sense, the mathematics of the emeritus theory [, then the] appeal to the physically interpreted mathematics of the emeritus theory in explanation is eliminable in favor of the physical interpretation of the mathematics of the pure fundamental theory.(p. 154)

But the antecedent of this conditional is false, because

> [i]n order to see what boundary conditions to impose on the partial differential equation [of the wave theory] in the first place, we must conceptualize the problem as one in which (to a first approximation) we are considering specular reflection off the back of the raindrop. It involves, that is, thinking about light behaving as rays on the physical boundaries. (p. 159)

In short, T_1 does not contain T_2 in an appropriate sense, because the interpretation of '$\exists \Gamma_i \in I$' is given by T_2, not T_1, in spite of Δ. Whether this claim is correct depends on Batterman's conception of 'conceptualize' and 'thinking about'. For reasons analogous to those given above in connection with Batterman's conception of 'explanation', I think it is reasonable to take the conceptions to be intensional. Hence, the theory of definition is applicable.

Under this assumption, I can see two ways in which the existence of the definitions Δ may not be enough for a derivation of the phenomenon from T_1. For one, Batterman's argument can be construed in the following way: While '$\exists i \in I : \Gamma_i$' can be defined in $\mathscr{V}_1$, this does not show that the new terms so defined actually have the same meaning as the ones of $\mathscr{V}_2$ without using $\mathscr{V}_2$. But only the terms of $\mathscr{V}_2$ actually allow to give an explanation and thus, $\mathscr{V}_2$ must be used. This, however, is provably incorrect. Since Δ is a set of explicit definitions, it determines the interpretation of '$\exists i \in I : \Gamma_i$' in all models, and hence its intension, because of the theorem of Padoa and Beth.

The other possible way in which to interpret Batterman's response is that Δ itself is dependent on T_2, and hence $T_1 \cup \Delta$ is dependent on T_2 as well. The analyticity of bridge laws would not weaken Batterman's position, because the models of T_2 are still needed to determine Δ. The eliminability of $\mathscr{A}$-terms by explicit definitions, however, entails that every sentence of $\mathscr{V}_1 \cup \mathscr{V}_2$ can be translated into $\mathscr{V}_1$, so that there is a $\mathscr{V}_1$-sentence Γ such that

$$T_1 \cup \Delta \vDash (\exists i \in I : \Gamma_i) \leftrightarrow \Gamma$$

and

$$T_1 \vDash \Gamma \rightarrow \rho \quad . \tag{3}$$

Since Γ is a $\mathscr{V}_1$ sentence, its intension is given solely within T_1, and the derivation (3) does not in any way depend on T_2. Hence an explanation of the occurrence of rainbows is completely given in T_1.

To summarize: If explanation is considered to be of a psychological nature, and if the ray theory explains the rainbow, Batterman's conclusion is right, but not essentially because of features of limit procedures, but rather because of features of human psychology. However, the extensive use of mathematical transformations by both Batterman and Belot suggests that explanation is considered to be intensional. Under this assumption, Belot's position is correct, because in the cases discussed, T_1 entails descriptions of individual phenomena and types of phenomena, and these descriptions' intensions are given exclusively within T_1.

5 On a critique of New Wave Reductionism

New Wave Reductionism concerns the relation of psychology and neuroscience, and specifically the two theories' explanatory powers. More specifically, the claim is that the explanations and predictions of psychological theories (T_3) can be given more accurately by neuroscientific theories (T_1). Typically, the predictions of T_3 will be corrected, so that with respect to the theories' common terms, there is a theory T_2 that is conservative in $\mathscr{C}$ with respect to T_1 and that approximates T_3. The explication of the approximation relation is difficult, but not obviously impossible. The approximation relation is also not necessarily dependent on the entailment relation, and in the subsequent discussion, I will assume its independence and proceed with T_2 instead of T_3.

Now, as discussed above, many of the terms of T_2 will be at least partially defined in $\mathscr{C}$-terms by T_2 itself, and much of recent research is about further connecting the remaining psychological terms $\mathscr{A}$ and the neurological and physical terms $\mathscr{B}$. In a sense, the most difficult step, the first connection of psychological terms with physical terms, has already been taken via psychology's and neuroscience's common terms $\mathscr{C}$.

Arguments against the feasibility of a reduction like this by van Eck et al. (2006, p. 174) focus on three claims by its proponents: (1) T_2 serves merely a heuristic function for developing T_1 with no persisting explanatory influences on T_1, (2) T_2 and its explanations will eventually be displaced by T_1, and (3) the explanations of the relevant phenomena by T_2 can be replaced by those of T_1 without explanatory loss.

The first claim is ambiguous, as Nagel (1951, p. 330) points out in connection with the reduction of biology to physics and chemistry:

> [O]rganismic biologists are on firm ground if what they maintain is that all biological phenomena are not explicable thus far physico-chemically, and that no physico-chemical theory can possibly explain such phenomena until the descriptive and theoretical terms of biology meet the condition of definability. On the other hand, nothing in the facts surveyed up to this point warrants the conclusion that biology is in principle irreducible to physico-chemistry.

Nagel distinguishes between the current state of research and the possible results of future research, and he points out that current research has not progressed far enough for a reduction. But there is a stronger claim to be made: If the first claim is that in actual research, at one point T_2 will cease to play a role, the reductionist position seems implausible as well, considering that even in a comparably simple field like the wave theory of light, geometrical optics is still a helpful guide in the search for solutions of equations. In the case of neuroscience, life on earth may very well vanish before psychology becomes superfluous in guiding neuroscientific research. Accordingly, van Eck et al. (2006) give a thorough analysis of the role that psychological results play in neuroscientific research, and come to the conclusion that there is no reason to believe that this influence will cease.

A different question is whether a complete consistent extension of neuroscience in $\mathscr{B}$ is powerful enough to predict everything that psychology can predict. In this case, and contrary to Nagel's claim, piecewise definability is enough to ensure eliminability, and whether it is possible cannot be answered solely by extrapolating from current research.

For a hyper-intensional conception of explanation, all three reductionist claims are implausible for the same reasons given in connection with the exchange between Batterman and Belot, and van Eck et al. (2006) give many examples to the effect that psychological $\mathscr{A}$-terms, defined in neurological $\mathscr{B}$-terms, are being used systematically in neuroscience for understanding. However, if 'explanation' is taken to mean 'derivation', the discussion rests on intensional concepts, and the very use of $\mathscr{A}$-terms as definienda of definitional extensions of neuroscience makes the definability of $\mathscr{A}$-terms in $\mathscr{B}$-terms more plausible.

6 Concluding remarks

It has become clear that in the two case studies given, the most glaring disagreements are foundational. These disagreements are not found within the respective sciences, but in the domain of general philosophy of science. In the debate between Batterman and Belot, the explication of 'explanation' and 'conceptualization' as neither intensional nor psychological notions may save Batterman's position. In the criticism of New Wave Reductionism by van Eck et al., each position can be right, either because explanation is taken to be intensional and the reducibility of psychology to neurobiology a matter of derivability between complete theories, or because explanation is taken to be a hyper-intensional, possibly psychological notion, and the reducibility between the theories is taken to be a matter of the actual development of the sciences.

The case studies also illustrate that these foundational questions cannot be answered without the analysis of the relevant specific sciences. The existence of Belot's definitions Δ is not obvious, and the concepts neurobiologists use to explain or develop their theories are not determined by definition theory either.

In that sense, definition theory again mirrors the theory of inference: It is helpful for foundational questions that inevitably arise when tackling problems in the philosophy of special sciences, but cannot replace the analysis of the special sciences itself.

References

Robert W. Batterman. *The Devil in the Details. Asymptotic Reasoning in Explanation, Reduction, and Emergence*. Oxford Studies in Philosophy of Science. Oxford University Press, Oxford, New York, 2002.

Robert W. Batterman. Response to Belot's "Whose devil? Which details?". *Philosophy of Science*, 72(1):154–163, jan 2005.

Robert W. Batterman. Intertheory relations in physics. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. The Metaphysics Research Lab, Center for the Study of Language and Information, Stanford University, Stanford, sep 2007. URL `http://plato.stanford.edu/archives/fall2007/entries/physics-interrelate/`.

Nuel Belnap. On rigorous definitions. *Philosophical Studies*, 72:115–146, 1993.

Gordon Belot. Whose devil? Which details? *Philosophy of Science*, 72(1):128–153, January 2005.

Olivier Darrigol. The modular structure of physical theories. *Synthese*, 2007. doi: 10.1007/s11229-007-9181-x.

Wilhelm K. Essler and Rainer Trapp. Some ways of operationally introducing dispositional predicates with regard to scientific and ordinary practice. *Synthese*, 34(4):371–396, apr 1977.

Ken Gemes. Logical content and empirical significance. In Paul Weingartner, Gerhard Schurz, and Georg Dorn, editors, *Proceedings of the 20th International Wittgenstein Symposium*, Vienna, 1998. Hölder-Pichler-Tempski.

Carl G[ustav] Hempel. Aspects of scientific explanation. In *Aspects of Scientific Explanation And Other Essays in the Philosophy of Science*, pages 331–496. The Free Press, New York, 1965.

John G. Kemeny and Paul Oppenheim. On reduction. *Philosophical Studies*, 7(1–2):6–19, jan 1952.

Ernest Nagel. Mechanistic explanation and organismic biology. *Philosophy and Phenomenological Research*, 11(3):327–338, mar 1951.

Veikko Rantala. Aspects of definability. *Acta Philosophica Fennica*, 29(2–3), 1977.

Stewart Shapiro. *Foundations without Foundationalism: A Case for Second-order Logic*. Oxford University Press, Oxford, 2000.

Lawrence Sklar. Types of inter-theoretic reduction. *The British Journal for the Philosophy of Science*, 18:109–124, 1967.

Alan B. Slomson and John L. Bell. *Models and Ultraproducts: An Introduction*. Dover Publications, Mineola, New York, 2006.

Dingmar van Eck, Huub Looren De Jong, and Maurice K. D. Schouten. Evaluating new wave reductionism: The case of vision. *British Journal for the Philosophy of Science*, 57:167–196, 2006.

Ryszard Wójcicki. Semantical criteria of empirical significance. *Studia Logica*, 19:75–102, 1966.

John Henry Woodger. *Biology and Language: An Introduction to the Methodology of the Biological Sciences, Including Medicine*. Cambridge University Press, Cambridge, 1952.

Set with pdfLaTeX 2_ε in 11pt Fourier.