

Chapter 2

How to Approach Reduction: Explication and Meta-science

Reduction, in the sense of ‘reduction’ we are interested in, reconciles diversity and directionality with strong unity. ‘Reduction’, however, can be given different interpretations; it is a technical term whose meaning largely depends on the theoretical goals of those who use the term. This chapter offers a disambiguation of the term, and it addresses methodological issues regarding the philosophical analysis of technical expressions such as ‘reduction’ in general. In order to establish a common ground, a taxonomy of notions in the vicinity of the concept of reduction is introduced, and the target of this book is located within this taxonomy. It is then argued that an appropriate characterization of the concept of reduction builds upon an explication, rather than an empirical investigation of alleged cases of reduction, or upon mere stipulation. Finally, the relevant terminology is introduced.

2.1 Reduction, Elimination, and Monism – A Taxonomy

How should we approach an appropriate characterization of the notion of reduction? This partly depends upon what we mean by ‘reduction’. Among philosophers, it is common courtesy to acknowledge that ‘reduction’ may come with many different intended meanings. The concept of reduction, as conceived of here, is supposed to reconcile diversity and directionality with strong unity, without relying on elimination. This characterization may help to some extent. However, it employs notions such as *unification* and *elimination*, so that one may fear that it tries to illuminate the obscure with the opaque. To ensure that the introduction of the present book’s target is not flawed by ambiguities in the use of ‘reduction’, ‘elimination’ and ‘unification’, or the distracting associations these terms may provoke, we should try to distinguish different “kinds” of reduction, or different conceptions thereof, thereby pointing to a taxonomy within which the target concept can be located. In passing, this will enable us to clarify the conservatism/eliminativism distinction and shed light on the connection between reductionist commitments and scientific realism.

2.1.1 *Reductionism and Monism*

Reductionism, as conceived of here, is a specific form of monism. Reductive physicalists are monists, and some radical idealists are monists as well. Monism, as conceived of here, is the doctrine that there is one kind of entities and properties *only*. How does the monist deal with the fact that there are different sciences, each of which purports to be concerned with reality, and each of which comes with its own models, its own experimental techniques, and its own conceptual schemes? Adopting *reductionism* is one of the options the monist has in response to this question: A monist may be a reductionist in the sense that she believes that the objects postulated by at least some of the different sciences exist, and that these objects are all of the same kind, e.g. all physical, or mental in nature. In this sense, a reductionist, as envisaged here, is a *realist* about the objects or properties postulated by reduced sciences. What other options does the monist have? Obviously, the monist may be an anti-realist about the objects of reduced sciences. To illustrate this point, let us pick an example and focus on the monist's stance towards *biology* only, assuming that her monism is not of a form of "biologism". How does the monist deal with the fact that there is a science of biology, on her view a non-fundamental science? The monist will take two aspects of the existence of a science of biology to be particularly pressing: first, its relation to reality, and second, its pragmatic or epistemic value. We have already distinguished two options the monist has regarding the metaphysics of biology. The monist may buy into the metaphysical commitments that go together with accepting biology, and, hence, adopt a metaphysically realist stance, or she may deny that biological properties or kinds exist, and, hence, adopt an anti-realist stance towards biology. Reduction, as conceived of here, reconciles diversity and directionality with unity without relying on elimination, that is: it reconciles diversity, directionality and *a realist interpretation of the reduced level* with unity (among objects at the reduced and the reducing level).¹ The eliminativist, in the sense of 'elimination' employed in the slogan, opposes the realist interpretation of the reduced level. This use of 'eliminativism' has to be distinguished from another use of the term.

2.1.2 *Two Forms of Eliminativism*

Both, a realist as well as an anti-realist interpretation of biology are compatible with two rival *pragmatic* assumptions about biology: the assumption that we should stick to biology, and the assumption that biology is a science we should dispense with. These views can be described in terms of the *eliminativism/conservatism* distinction as well. *Metaphysical* eliminativism can partially be cashed out in terms

¹Here, I merely refer to metaphysical anti-realism. Semantic realism might be combinable with metaphysical anti-realism. This distinction will be mainly ignored here.

of ontological commitment; *pragmatic* or *epistemic* eliminativism can partially be characterized in terms of whether or not a science (or, more broadly, a scientific discourse) is pragmatically appropriate:

- (*Metaphysical Eliminativism*) If metaphysical eliminativism about *F*-s is correct, then *F*-s do not exist.
- (*Pragmatic Eliminativism*) If pragmatic eliminativism about *a science of F*-s is correct then this is a science we can dispense with.

Conservatism, the denial of eliminativism, comes in two forms, too, corresponding to metaphysical and pragmatic eliminativism. For our present concern, these partial characterizations will be sufficient. Employing these different conceptions of eliminativism and conservatism as well as the notion of monism, we can give a sufficiently precise characterization of the notion of reduction we are interested in here and distinguish it from other notions of reduction. To do so, let me from now on restrict the use of ‘eliminativism’ (and, similarly, ‘conservatism’) to its pragmatic or epistemic form. Instead of employing the *metaphysical eliminativism/conservatism* terminology, I will talk about *anti-realism/realism* (for a specific domain).

2.1.3 Four Types of Monism – The Taxonomy

Based on these minimal characterizations of the distinctions between anti-realism/realism and eliminativism/conservatism, we can introduce a taxonomy of four monistic positions, all of which have been associated with some stance towards *reductionism*. I will again refer to the case of biology:

- (*Conservative Anti-Realism*) Monism is true, and there are no biological properties or kinds, but nevertheless, biology should be conserved.
- (*Eliminative Anti-Realism*) Monism is true, and there are no biological properties or kinds, and therefore, biology should be replaced by a lower-level science.
- (*Conservative Realism*) Monism is true, and there are biological properties or kinds and therefore, biology should be conserved.
- (*Eliminative Realism*) Monism is true, and there are biological properties or kinds, but nevertheless, biology should be replaced by a lower level science.

Let me briefly comment on these versions of monism, drawing distinctions captured in Table 2.1.

Eliminative Realism seems to underlie unificationist projects that are not committed to the assumption that high-level sciences are metaphysically flawed. If a unified science seems desirable on independent grounds, this may provide reasons for

Table 2.1 Monisms and reductionisms

Monisms and reductionisms	Anti-realism	Realism
Eliminativism	Monism-1: Eliminativism/replacementism	Monism-2: Unificationism
Conservatism	Monism-3: Epistemic/pragmatic non-reductivism	Monism-4: Conservative reductionism

abandoning metaphysically or semantically appropriate high-level special sciences.² This amounts to a reductionist form of what can be called *unificationism*. It will be argued later that in a sense *Eliminative Realism* can be combined with *Conservative Realism* – if we assume (as we should) that the normative components in these claims are to be qualified: In some respects, it is worth “conserving” (using) a reduced theory, in some others, reduced theories could be “eliminated” (it would be desirable if we could ignore them). There neither is conservatism nor eliminativism *simpliciter* (see Sect. 9.2).

Versions of *Conservative Anti-Realism*, according to which high-level sciences are not true, but, nevertheless, play an important role and have to be conserved, is usually justified by pragmatic considerations (Van Gulick 1992, 2001) or by the somewhat similar assumption that idealization plays a crucial role in the construction of high-level sciences (Sklar 2003). Even though theories of high-level sciences are not literally true or metaphysically adequate, they nevertheless yield a useful picture of reality, a picture that is maybe even more useful for some domains of inquiry than any picture lower-level sciences could possibly yield. A theory of good governance describes reality using a conceptual scheme we can easily handle, which is, by and large, appropriate, even though the propositions expressed by sentences in a theory of good governance are not, and cannot be, literally true of the matter at hand. Being confronted with a true description of the relevant situation, however, we would be utterly lost. Therefore, we should stick to the pragmatically appropriate high-level science. This stance can be regarded as one form of non-reductive monism, namely, a pragmatic or epistemic version of this doctrine.

²Maybe in the early years of positivist behaviorism, we can find a position according to which at least some high-level sciences are possibly true though eliminable. If we take positivist behaviorism to state that (i) psychological terms are to be *analyzed* in behavioral terms, then positivist behaviorism is committed to the claim that there are psychological kinds (because there are behavioral kinds represented by psychological concepts). So, psychology is a science that is at least possibly true. If we assume that these positivists claim that (ii) psychology should, in principle, be replaced by a science that is formulated in the language of physics, because we can hope for an analysis which bridges the gaps between scientific levels, then we have a position that instantiates *Eliminative Realism*. Truth or appropriateness of psychology is combined with in-principle-elimination. To mention just two authors who could be understood in a structurally similar way: Carnap, assuming that we should seek at *translation* of all scientific statements (Carnap 1934, 32), and Neurath, who argues that we can understand utterances concerning, for example, empathy as being translatable into physical language (Neurath 1959, 298).

In contrast, (Eliminative Anti-Realism) is justified by reference to the alleged fact that (some) higher level sciences are so utterly misleading that they have to be replaced by lower level sciences. Those who subscribe to this kind of eliminativism normally argue that *merely some* higher level sciences should be eliminated, that is: subscribing to a version of (Eliminative Anti-Realism), no actual philosopher commits herself to the strong claim that *all* high-level sciences are equally defective. Trying to replace psychology by neuroscience makes sense only if neuroscience is regarded as fundamentally different from psychology, even though it is a higher-level science (it is not physics). Therefore, those who subscribe to some sort of *partial eliminativism*, like the Churchlands (Patricia Churchland 1986; Paul Churchland 1981, 1985),³ should be regarded as subscribing to a weak or domain-restricted version of *Eliminative Anti-Realism*. This is classical eliminativism, a doctrine that usually cashes out the notion of reduction in terms of theory-replacement. To add another ‘-ism’ to the world of labels for philosophical doctrines (and a particularly ugly one), we could call this view ‘replacementism’.

I think that *Conservative Realism* is self-explaining: If the high-level science is metaphysically appropriate and, ideally, true, then why shouldn’t we accept it as appropriate? The more challenging question in the context of both versions of realism is this: What is the relation between the higher-level sciences and the lower-level sciences if both are true and the fundamental level’s entities and properties do all the work? Answers to this question range from ontological theories of supervenience and realization to theories of explanatory relations between theories. Note that, given the qualification of *Eliminative Anti-Realism* given above, you can be an eliminativist for some levels, and subscribe to *Conservative Realism* for others (and you can combine even more positions for different levels).⁴ Conservative monistic realism is conservative reductionism.

In these brief characterizations, the term ‘realism’ has been employed in order to get rid of an ambiguity in ‘eliminativism’. It should be noted that monism, and, hence, realism is metaphysically neutral; as already suggested, monism can be

³The Churchlands figure among the most prominent defenders of this sort of eliminative materialism. Eliminative materialism in this sense states that

[...] our common-sense conception of psychological phenomena constitutes a radically false theory, a theory so fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience. (P. Churchland 1981: 67)

It is an interesting feature of this formulation that eliminative materialism is contrasted with reduction. The pre-theoretical notion of reduction presupposes that the reduced theory is true or appropriate, rather than false and totally misguided.

⁴Note secondly that for some philosophers it might be difficult to judge whether or not they belong to the camp of *Conservative Realism* or *Conservative Anti-Realism*: If you take truth to be described in terms of pragmatism, then, under certain interpretations, the distinction breaks down.

embraced by idealists as well as by naturalists. Similarly, the notion of reduction is, in a sense, metaphysically neutral, though particular forms of reductionism are not.

2.1.4 *Reduction and Metaphysical Neutrality*

To see that reduction is, in a sense, metaphysically neutral, consider the following questions: Why not define reduction in terms of physicalism? After all, as Melnyk put it, ‘doctrines of physicalism [...] should be seen as competing responses to the many sciences problem’ (Melnik 2003, 2). Isn’t physicalism, or naturalism, the form of monism reductionists adopt? And can an appropriate explication of reduction ignore the question of what physicalism consists in? After all, it is physicalists who prominently employ this term! It is common to qualify certain things as being physical. There are *physical* events, *physical* properties, *physical* kinds, *physical* laws, *physical* objects, and the *physical* level of a hierarchy of sciences. Accordingly, I will talk about *physicalism*, about *physical objects* and about *ultimate physics*. We have an idea of how to apply them, but we may not be able to give a definition. That is all we need for the moment – that there is a coherent use of these expressions we can rely on.⁵ ‘Physicalism’, ‘naturalism’, and ‘materialism’ (but not its variants – there is no materialist level of a hierarchy of science) will be used interchangeably. I shall elaborate on related notions of being physical in Sect. 9.1.

Now, what is the relation between the concept of reduction, and doctrines of physicalism or naturalism? An appropriate definition of the notion of reduction could be given even if, in principle, we were unable to say anything about what physicalism actually states. Reductionism does not presuppose physicalism. This is due to the fact that the concept of *monism* is, in a sense, metaphysically neutral. Neither the notion of naturalism, nor the notion of idealism is built into the notion of monism. Since the concept of monism is metaphysically neutral, the concept of reduction is, too. That is: The reductive idealist and the reductive physicalist agree that there are relations of reduction in the actual world. They do not agree on the *direction* of reduction in the actual world. After all, an idealist who believes that everything reduces to the mental may employ the term ‘reduction’ in the way the physicalist employs the term! They do not disagree about the *notion* of reduction, but rather about which reduction relations actually hold, about how our world is shaped metaphysically, or what is more fundamental – the mental or the physical. Compare

⁵Definitions of being physical normally come in ontological terms, that is: in terms of what sort of objects and kinds are physical *simpliciter* and, sometimes, the physical is defined as the non-mental (Spurrett and Papineau 1999; Papineau 2002, ch. 1.10; Crook and Gillett 2001). Sometimes, philosophers defined the physical via reference to physics (Papineau 1993; Hellman and Thompson 1975, 553 ff.).

this dispute to the following dispute: Assume that two persons argue about whether or not the Eiffel-tower is taller than the Tower of London. They do not disagree about the meaning of ‘_is taller than_’, but rather about what actually is taller – the Tower of London or the Eiffel Tower.

The notion of reduction is independent of the notion of physicalism at least in this respect: It is conceptually possible that (a version of) reductionism is true, and that physicalism is false. Possession of any richer concept of reduction, that combines metaphysically neutral reduction with physicalism, requires possession of the topic-neutral concept. So, we should not define our general notion of reduction in terms of physicalism. The notion covers a specific dependence relation that is neutral with respect to what actually forms the fundamental stuff. Physicalism is a doctrine about what the fundamental stuff is. Thus, these two issues have to be kept separate. The concept of reduction is *not* metaphysically neutral in the sense that reductionism goes together with metaphysical commitment; *realism* suits the reductionist as conceived of here. So, what is the precise connection between scientific realism and reductionism?

2.1.5 *Reductionism and Realism*

It should be obvious that questions about reduction occur only, or become at least seriously more pressing if we assume that the aim of science is to provide us with true propositions about what actually is out there. The reductionist’s rationale for presupposing this picture is that if sciences were necessarily concerned with mind-dependent entities, or with constructions, which necessarily miss the way the world actually is, then questions regarding reduction would seem rather unimportant. That the best theory of the actual world might be dualistic, or that it might presuppose a mentalist monism, is threatening to the physicalist only if this theory’s furniture matches the furniture of the world we actually live in. Otherwise, the question of dualism would merely be a question about how we do or are bound to conceive of the world, rather than a question about how the world is. So, to give reductionism, and, hence, the concept of reduction an interpretation according to which issues of reductionism and anti-reductionism are obviously pressing, I will presuppose that reductionists are committed to scientific realism (for the relevant domains).⁶

⁶Note that the generality requirement underlying the debate between anti-realists and realists is not of much interest for us. Some sciences might be such that they aim at truths (to use van Fraassen’s idea, see below), whereas others do not. Thus, a weakened form of the realistic picture sketched above is to a certain extent compatible with modest forms of what is often labeled ‘anti-realism’. Van Fraassen’s idea of constructive-anti-realism, for example, is characterized by its taking theories to be understood in a literal sense (what I referred to as ‘semantic realism’). If we grant in addition that at least *some* theories are true or at least aim at truth without assuming that it is *the* aim of science in general “to give us, in its theories, a literally true story of what the world is like” (van Fraassen 1980, 8), then for these theories questions regarding reduction are equally pressing. The point is this: Questions regarding reduction, as these questions are understood here and will

Now, the different positions sketched above raise different conceptual issues. This book deals with *monistic realism regarding high-level science*. For the monist, the idea of reconciling diversity and directionality with strong unity without relying on (metaphysical) elimination, or anti-realism, is particularly pressing. This leads to an important distinction between issues related to *reductionism* on the one hand, and *eliminativism*, or *replacementism* on the other.

2.1.6 Two Philosophical Challenges: Reduction and Replacement

If we take reductionism to consist in the position that high-level sciences are (merely) to be replaced, then the question of how the *concept* of reduction should be characterized is not philosophically challenging at all. If one theory is replaced by another theory then one theory is abandoned, and the other is used instead. That's it. Interesting empirical questions arise, for example, when we try to find out under which conditions scientists actually replace theories, or when we are interested in the *epistemology* of theory-conservation and elimination. In the conservative case, questions may arise as to *pragmatic or representational issues*, which are concerned with a clarification of the idea that a false theory can serve a scientific purpose. So, we have to distinguish between two different issues: Even though the notion of theory-replacement is as clear as a notion can be, it is not so clear under which conditions scientists *believe to be (and are) justified in changing their theories*. Answering these epistemological and procedural questions, other questions concerning inter-theory relations turn out to be interesting, like the question of *how false scientific representation can serve a scientific purpose* or of *how a false theory can be similar to a correct theory (under certain conditions)*, or, more generally, *what it is for non-equivalent theories to resemble each other*. Investigating such questions is what is done under the heading 'reduction' in the sense of 'theory replacement'. However, this has got nothing to do with the *concept*

be developed below, are interesting only in the context of allegedly *true* theories. If people were able to give a true science of everything, no anti-realist would dismiss such a theory on the grounds that it is to be conceived of as being true. In this case, the question of how, intuitively speaking, higher-level theories relate to this fundamental science, is of much interest. If higher-level theories turned out to be accidentally true or appropriate (in the light of the fundamental theory) everybody, the realist as well as the anti-realist, would still have the problem of reduction. Nevertheless, even in this picture, a realistic understanding of science enters the game, even though it is a realistic picture of possible science, rather than of actual science.

Now, finally, even if you dismiss realism, you might still believe that the truth-predicate as well as 'exists' and similar expressions are interesting from an instrumentalist perspective. Then, again, you might be interested in a theory of reduction from an instrumentalist perspective, and just interpret 'truth' or 'exists' the way you prefer. Thereby, you will arrive at a structurally similar, anti-realist interpretation of reduction that still should play an explanatory role within the chosen framework.

of replacement – it has got to do with concepts of other relations connecting actual (or possible) pairs of theories, containing one element that has been or should be or should have been replaced by the other. The *concept* of reduction as replacement is not philosophically interesting, although *reductions* as replacements give rise to a number of philosophically challenging issues.

In contrast, consider positions of the type of monistic realism. For example, such positions give rise to the metaphysical question of what a scientific level is supposed to be, if sciences at different levels are true even though monism holds. This notion will be the target of the investigation to follow. It is a notion of reduction according to which the reduction of one science (or kind, or phenomenon, or term, or fragment of discourse) to another implies that for the relata of the reduction relation, there are no two distinct levels of reality, or different sets of kinds and substances. Schematically: *a* reduces to *b* only if *a* and *b* are (ontologically speaking) homogeneous (and not trivially so, due to being empty). That is: If ‘*a*’ and ‘*b*’ are substituted by expressions referring to theories, then these theories’ terms do refer, but do not refer to kinds at different levels, and if they are substituted by kind-terms, then the corresponding instance of ‘*a* = *b*’ is true. Thus, if we were able to model this notion, we would be in the position to give an answer to the question of *how it is possible that there are true, ontologically committal sciences (or, more liberally: appropriate descriptions and terms) that are seemingly unconnected, or are situated at different levels, although they are compatible with strict monism.*

The notion of identity-based reduction this book focuses on thus fits monistic realism: *The concept of reduction is supposed to reconcile diversity and directionality (mirrored by the idea of levels) with strong unity (mirrored by the idea of monism), without relying on (metaphysical) elimination (that is the idea of realism).* But isn’t this target, or this description of the target obviously ill conceived? Apparently, it does not do justice to the fact that in the philosophy of science, reduction is commonly regarded as a relation holding primarily between theories.

2.1.7 Theory Reduction and Ontological Reduction

Reduction in the philosophy of science is usually regarded as being a relation instantiated by pairs of theories, or scientific models, rather than by pairs of properties, event-types or states of affairs. It is thus time to distinguish between two versions of reduction – *theory* reduction and *ontological* reduction:

- | | |
|-------------------------|--|
| (Theory Reduction) | The predicate ‘_ reduces to _’ expresses a relation of theory reduction only if its arguments in true sentences refer to things such as theories, or scientific models, or fragments of theories. |
| (Ontological Reduction) | The predicate ‘_ reduces to _’ expresses a relation of ontological reduction only if its arguments in true sentences refer to things such as kinds, types, properties, events, substances, or individuals. |

Obviously, the notion of reduction that gives rise to the puzzle sketched in the introduction rests on a version of ontological reduction, whereas notions of theory reduction are more holistic in spirit and can account for diversity in a straightforward way, namely, by referring to differences between the reduced and the reducing theory. For the moment, we will ignore this issue. I will treat the notion of reduction discussed here as distinct from a notion of theory reduction. Whether or not it collapses into a notion of theory reduction, or whether or not it is totally alien to the notion of reduction pertinent in the philosophy of science will be discussed later. Both questions will be answered to the negative in Chaps. 6, 7, and 8; in fact, it will turn out that “ontological” reduction is more fundamental than theory reduction.

When it comes to characterizing notions of reduction, suggestions are made by philosophers taking methods of the philosophy of language tradition seriously (e.g. Kim 1992; Lewis 1972), whereas philosophers of science, trained in formalisms, frequently conceive of the problem as a problem to be solved in formal terms (Moulines 1984; Balzer 1984; Sneed 1971). Some adopt a pragmatist stance towards questions about some epistemic or other value of high-level sciences (Van Gulick 1992, 2001; Sklar 2003), whereas others feel pressed to back up notions of scientific levels and their interrelations by telling an ontological story (e.g. Fodor 1974; Esfeld and Sachse 2007; Bechtel and Richardson 1993; Bechtel 1994; Craver 2005). Such differences in the format of the characterizations pose difficulties, but they do not necessarily result in fundamental problems. In contrast, there is a more problematic issue, not concerning the way of designing the characterization (putting it into structuralist or pragmatist terms, or giving a formal or a non-formal definiens), but rather concerning the method of approaching the characterization. So, back to our initial question: How should we approach an appropriate characterization of the notion of reduction?

2.2 Ways of Defining Reduction

To characterize different version of definitions of reduction, we need an idea of what is meant here by ‘definition’. Assume that any candidate for a full-blown characterization comes in the following form:

x reduces to y iff xRy [...].

This is what I will call a *definition*.⁷ For what follows, every bi-conditional connecting open sentences will be regarded as a definition. Let us talk about *kinds*

⁷The way the notion of a definition is introduced here will serve our present purpose. It is not supposed to generalize to other contexts. I will also be tolerant about answers to the question of *what* can be defined. Let us take predicates as a starting point. We can then arrive at derivative uses of the term ‘definition’. A concept is defined iff some predicate expressing the concept is defined. A relation or property is defined iff the predicate picking out the relation or the property is defined (this might result in different true definitions of one and the same property, namely, when two

of *definitions* in order to distinguish between different interpretations of how a definition should be approached. There are basically three kinds of definitions philosophers have offered for the concept of reduction: *stipulating* a meaning for ‘reduction’, tying a characterization of the notion to an investigation of *alleged cases of reductions*, and, thus, taking an empirical dimension into account, and, thirdly and only implicitly, by way of *explication*. An explication can basically be conceived of as a procedure that tries to illuminate a term’s pre-theoretic use by (i) reflecting upon criteria of adequacy this use imposes on a definition, without relying on the assumption that we can come up with a full-blown conceptual analysis of that notion, and (ii) by *sharpening* that concept. This section argues that stipulations as well as empirically inspired investigations of actual theory-successions do not form interesting candidates for an appropriate definition of the concept of reduction. Rather, we should try to *explicate* the concept of reduction.

2.2.1 Definitions as Stipulations

In some contexts the meaning of the term ‘reduction’ is treated as being fixed by an explicit definition or stipulation. Philosophers of science sometimes talk that way when they take what has come to be known as the Nagel-model of reduction as being fixed (stipulated) by Nagel’s own definition of ‘reduction’ (utterly briefly): reduction is derivability plus bridge-laws, whose job is to connect the terms of the reducing science(’s laws) to the terms of the reduced science(’s laws), if the vocabulary of these sciences is heterogeneous (Nagel 1961, chapter 11). This has a puzzling effect: For example, Richardson in his (1979, 1982) seems to *presuppose* the Nagel model, rather than to take it as a *candidate* for a definition of the notion. Richardson argues that the Nagel-model is fulfilled by theory-derivation on the basis of bridge-laws that take the form of conditionals (something observed by Nagel himself (1961, 355, fn. 5)). Does this, without further ado, show that *reductions* can be based on bridge laws that take the form of mere conditionals? It does not; it shows that *Nagel’s characterization* can be satisfied in this specific way. Whether or not conditions on *reduction* can, and, thus, whether or not the definition captures reduction, is to be decided on independent grounds. If a given pair of, in this case, theories does, intuitively, not fall under the concept, but does fulfill the definiens offered for this concept then we have reason to deny that the definition actually covers the concept. Being an instance of some model or definition of a concept C is *not* to be treated as being the make or break for being an instance of C. On some occasions, Kim (1993, 150 & 248) and Fodor (1981, 150) seem to follow this way of talking about (Nagelian) reduction. Recently, Butterfield has argued in a similar spirit that reduction and emergence turn out to be compatible, on

predicates with different meanings pick out the same property). A general term is defined iff the predicate that is build by (the copula and) the general term is defined.

a Nagelian interpretation of reduction (Butterfield 2011a, b). Interestingly, some versions of *reductionism* have been criticized based on a similar strategy, i.e. (implicitly) taking candidate characterizations to be stipulations. This point has been brought out by Gene Witmer in his intriguing discussion of Steven Horst's book⁸ *Beyond Reduction: Philosophy of Mind and Post-Reductionist Philosophy of Science* (Horst 2007; Witmer 2008). Horst argues that reductionism is mistaken *because* reductions in the sense of the Nagel-model are rare (Horst 2007, 49). But what if the Nagel-model did not adequately capture the relevant relation? In this spirit, Witmer objects that failure of the Nagel-model does not translate into a failure of reductionism – he suggests that it 'should still be open to us to say that A is reductively explained by B even though a derivation is impossible' (Witmer 2008).

What has this got to do with stipulation? You cannot make a mistake *stipulating* that reduction is derivability plus bridge-laws. But you can make a mistake *proposing* that reduction is derivability plus bridge-laws. For stipulations, it is just pointless to ask questions about whether or not the definition is appropriate. If the meaning of 'reduction' were to be fixed by stipulation, then there would not be a substantive issue here. This sort of "approach" (which is barely worth this name in the case of reduction) will not loom large in the rest of this book. It just misses the target. As a desideratum, we get that a definition should be approached in a way that allows for mistakes – we want a serious discussion about whether or not a candidate definition is appropriate.

An alternative to stipulation is tied to the idea that part of the job of philosophers in the philosophy of science is to *rationaly reconstruct* science, or to do an exercise in *meta-science*: reflect upon actual cases of theory-successions to shape a model of reduction.

2.2.2 Empirically Inspired Definitions

Let us begin with a very radical point of view: that reduction is to be defined in accordance with a previously fixed set of examples. This idea appeared in one of the first formulations of how to approach the notion of reduction in the literature. Kemeny and Oppenheim (1956) write:

The label 'reduction' has been applied to a certain type of progress in science. As this process has been the subject of much philosophical controversy, it is the task of the philosopher of science to give a rational reconstruction of the essential features of reduction. We will discuss the basic features of this process informally, we will review two previous attempts to make the concept precise and we will offer certain improvements which we hope will *bring the philosophical characterization of reduction closer to what actually happens in science*. (Kemeny and Oppenheim 1956, 6, my emphasis)

Kemeny and Oppenheim clearly state that they believe the concept of reduction to be defined in a way that gives an insight into actual scientific processes – it is

⁸I am grateful to an anonymous reviewer who brought this book to my attention.

a technical term the definition of which has to be judged by how well it fits a set of previously fixed examples. In this context, the term ‘reduction’ is reserved for *whatever the relevant relation might be which holds between different theories or sciences, which stand in a succession-relation to each other (and share, maybe, some additional features)*. Balzer (1984, 331) explicitly reflects on this point without deriving a definite conclusion. According to him, we should conceive of approaches to the notion of reduction as consisting in reflections on actual examples, which he describes as *paradigmatic* cases from which a tentative definition is derived, which is then applied to alternative cases. If the definition does not cover a case we believed to be covered, we have to decide whether or not the case is to be counted as relevant. So, Balzer, Kemeny and Oppenheim seem to share the intuition that any definition of the notion of reduction is to be judged by its appropriateness with respect to a certain class of examples. This can be subsumed under the label of ‘*extension first*’ approaches to the definition of reduction. Even though the techniques employed are often formal (and the theory-relations so defined are formal), these models’ touchstone is the adequacy for a given set of actual theories. Finding out about whether or not they match these set’s elements is an empirical matter.

A cognate of this approach is a *mixtum compositum* of experimental philosophy and rational reconstruction. What John Bickle labels ‘meta-science’ (Bickle 2008) can roughly be characterized as follows: In order to understand what reduction is, we have to pay attention to how scientists use the term, i.e. what they claim to be a reduction. The philosopher’s job is to pay attention to how scientists use the term, and to reconstruct the cases so described by scientists, in a way that is similar to rational reconstruction as described above.

Note firstly that, as argued in Sect. 2.1.6, no *concept* of theory-change is philosophically challenging (although the epistemology or the logical relations between the *relata* of theory-changes are challenging). Similarly, Bickle’s project is not a conceptual one. The concept of reduction is, on this account, not challenging at all; the challenge lies in understanding the cases described by scientists as reductions. The main difference between classical rational reconstruction and meta-science seems to be this: In the classical case, it is the philosopher who decides which case of alleged reduction to include in the set of relevant cases. On Bickle’s interpretation, it is the scientists’ use of the term ‘reduction’ that defines this set.

Note secondly that results of this procedure will not be very satisfactory, if we aim at a definition of a notion that is already in use in specific explanatory contexts *in philosophy*. As long as there is no criterion by which we can judge whether or not the decision concerning the relevance of an example is appropriate, this sort of approach is at risk of involving an element of *ad hoc* decisions about what to count as belonging to the concept’s extension. Bickle suggests a criterion; but one may wonder what the scientist’s authority concerning the use of the term ‘reduction’ in philosophy is. The concept of reduction seems to play an important role in characterizing reductionist positions and in clarifying variants of non-reductive physicalism. The term ‘reduction’ thus has a use that is prior to any definition. It is hard to see how an investigation of actual cases of theory-successions (or similar

cases of explanations in science) can, without further ado, do justice to this pre-theoretic use. As a desideratum, we thus get that our method should do justice to the pre-theoretic use of the term.

However, in a sense, Kemeny's, Oppenheim's, Balzer's and Bickle's dialectical remarks should be taken seriously. As we shall see in Chap. 7, attempts to include replacement in a definition of reduction can be regarded as being carried out in the spirit of this way of talking about reduction. The very fact that Kemeny, Oppenheim and others use the term 'reduction' in this way is sufficient to claim that in fact, there is a technical term 'reduction' in the sense just outlined. But this is not the only one.

2.2.3 Definitions as Explications

Sarkar contrasts attempts to model a certain relation we have access to independent of empirical investigations, with attempts to come up with accurate descriptions of a set of actual scientific developments (Sarkar 1992, 169; see also Wimsatt 1976). We just dismissed the second approach for the context at hand. But what is the alternative? Doesn't the fact that intuitively, this approach should be independent of empirical considerations, at first sight, make it look like good old conceptual analysis?⁹

Block seems to deny that conceptual analysis forms a promising candidate for approaching a definition of reduction, stating that there is "much less interest in analyzing technical philosophical concepts [than there is in analyzing some ordinary language concepts]. We should use whatever technical concepts do the jobs we want done." (Block 1997, 112) This rough characterization already hints at the main idea underlying the concept of an explication: There is a job a technical concept is supposed to do. This job has to be identified. An appropriate characterization should reflect an appropriate job-description. Intuitively, we thereby combine an element of stipulation with reference to previously fixed criteria of adequacy – the latter characterizing the job we want our concept to do. This makes room for mistakes in definitions. Consider the following statement, where Dizadji-Bahmani, Frigg, and Hartmann argue that

... the syntactic view is unnecessary to get GNS [the *Generalized Nagel Schaffner Model*] off the ground [...] Where first order logic is too weak, we can replace it with any formal system that is strong enough to do what we need it to do. (Dizadji-Bahmani et al. 2010, 403)

⁹Nickles (1973) seems to at least partly rely on the idea that we should reflect upon how the term 'reduction' is used in different contexts in order to come up with an appropriate characterization. His main point is that there are different uses of the term in philosophical and scientific contexts. The notion focused on here is closer to what he labels 'reduction₁' rather than to 'reduction₂', that captures, intuitively, some aspects of actual scientific change.

The authors seem to adopt a view according to which it is possible that one might be *mistaken* about a definition's appropriateness. The authors rescue what they take to be the core idea of Nagel's (and Schaffner's) model of reduction by abstracting away from *irrelevant* or *misguided* positivistic aspects of the definition. This is possible only if the definition is not a stipulation. Secondly, they describe the definition achieved by the suggested amendments as *Nagelian* (being in accordance with the Nagel-model) – it is *Nagelian* in the sense that it captures what Nagel aimed at. Accordingly, there is a core idea which can be *distinguished from* the definition and which is *prior to* the definition.¹⁰

This idea is similar to the idea of (early) *Carnapian explication* (Carnap 1950, 1988 (1956), §2). The idea of explication is captured by the following passage, where the *explicatum* is a newly introduced expression, which is introduced to replace the *explicandum* in a certain context:

Generally speaking, it is not required that an explicatum have, as nearly as possible, the same meaning as the explicandum; it should, however, correspond to the explicandum in such a way that it can be used instead of the latter. (Carnap 1988 (1956), 8)

Carnap's conception of explication is not particularly precise. He illustrates the explication relation referring to the relation between the ordinary-language concept of heat and that of temperature (as a measurement-concept) and he introduces four criteria any appropriate explication should meet that can be summarized as follows: *similarity in use*, *simplicity of the definition of the explicatum*, *fruitfulness of the definition of the explicatum*, and *exactness of the explicatum* (1950, 7 ff.). Arriving at measurement concepts plays a key role in his conception of explication. His own suggestions for explications are, apart from his work on degrees of confirmation, often radically different from the heat-temperature example; he is concerned with the explication of ordinary language concepts, such as truth and necessity, by concepts defined for a formal language, that should be used in this formal language (the idea is *not* that they should replace the ordinary language notions in ordinary language contexts). We do not learn much about the typical deficiencies of concepts that should be explicated; Carnap just mentions that a concept that is worth explicating is usually vague (Carnap 1988 (1956), 7 – where 'vagueness' should probably not be understood in the linguist's sense of the term).

Quine (1960, § 53) takes up the Carnapian idea, arguing that explication should replace conceptual analysis as classically conceived – that is: as decomposition of meaning constituents (or something similar), which obeys criteria of synonymy, as he puts it (Quine 1960, 258) – in *any* context. Fortunately, his remarks on the relevant similarity relation between the explicandum and the explicatum are more precise than Carnap's. According to Quine, we should conceive of the relation between explicatum and explicandum in terms of *functional* similarity. Some functional features of the explicandum are to be preserved by the explicatum, whilst

¹⁰Similar ideas are pertinent in Endicott (1998), Waters (1990), and it is discussed in van Riel (2011).

others are to be neglected. Since these function-preservations are not to be achieved by complete meaning-conservations, an element of stipulation enters the game, as his discussion of the definition of the term ‘pair’ nicely illustrates (Quine 1960, §53). Quine suggests that the “ordinary” concept of a pair is just deficient – the metaphysical commitments that go together with the use of this concept are just untenable. Nevertheless, aspects of this concept are worth preserving. In successful explications of the concept of a pair, we get definitions that preserve the desired aspects and get rid of the misleading ones.¹¹ So, the upshot is this: If there is an expression which serves a certain important goal, but which poses serious philosophical problems, we should try to replace this term by another term (give this term another use), by introducing a definition which involves an element of stipulation, but which, nevertheless, is oriented towards the goal the primary use of the term determined. This orientation towards the same explanatory or descriptive goals is guaranteed by partial functional equivalence between explicandum and explicatum.¹²

In the present context, we are neither concerned with formal languages which are supposed to model natural languages (in certain respects), nor should we conceive of the task of defining the notion of reduction to be the task of replacing a not quite exact, or vague, or ontologically misleading expression by a term expressing an exact formal concept, which can be used instead of the previous expression in a certain context. To repeat: ‘Reduction’ is a technical term, and we just do not know how to define it, such that it can perform the job it is supposed to do. However, the procedure associated with explications seems promising when adopted for the case of ‘reduction’.

In order to ensure that our concept can perform the job it is supposed to do, we start with non-formal characterizations that define a set of criteria S (this corresponds, intuitively, to the concept’s *job-description*). We grasp these criteria reflecting on *meaning*, or reflecting on *use*. For example, the notion of reduction, as it is introduced in the literature, describes a directional, or a seemingly asymmetric relation. Whoever fully understands the pre-theoretic characterizations is in a position to know that if ‘ a reduces to b ’ expresses a truth, then ‘ b does not reduce to a ’ expresses a truth as well. Up to now, the project is similar to classical conceptual analysis. In a second step, however, we try to introduce a notion that satisfies a set of criteria $S^* \subseteq S$ (in the best case, $S^* = S$, if S does not comprise contradictory, irrelevant, or otherwise problematic criteria). The conceptual resources we use to match these criteria depend on our decisions and, occasionally, on reflections upon independent issues. The idea of reflection upon independent issues can be canvassed as follows: There might be a number of definitions which match the criteria equally well, but which differ in other important respects, especially with respect to how they conceive of *other* notions that enter the definition. Let me give an example. Assume that Nagelian or classical empiricist or syntactic reduction

¹¹ Say, by Kuratowski’s set-theoretical characterization according to which $\langle a, b \rangle = \{\{a\}, \{a, b\}\}$.

¹² For a discussion of the notion of explication, see also Strawson (1963) and Hanna (1968).

(R_N) and structuralist or (more broadly) semantic reduction (R_S) are equivalent (in a sense that will be specified in a moment). In this case, they differ primarily with respect to how they conceive of *theories* or *scientific representation*, and, correspondingly, of derivation, but are *equivalent in other respects*. In this case, we have to take independent considerations into account in order to decide between these definitions. Here is an idea of how we can conceive of a relevant kind of equivalence between two models of reduction in a very simplified manner. Assume that for any structuralist theory, t , there is some (possible) syntactic theory, t^* , which stands in some relation R to this structuralist theory, such that tRt^* . Moreover, there is no other structuralist theory $t' \neq t$, such that $t'Rt^*$. Then, there is an interesting similarity between the notions of reduction defined in terms of these different kinds of theory iff for any reduction of a structuralist theory, there is a corresponding reduction of a syntactic theory, and vice versa. Intuitively, we can think of such a relation between syntactic and structuralist theories in terms of *expression*. Let the structuralist theories be the structured semantic values of syntactic theories. Note that this is not quite appropriate, because classical empiricist theories do not simply express structuralist theories (or structures). If we assume that there are enough (possible) syntactic theories, however, *expression* is a perfect dummy. Here is a semi-formal version of this idea of an equivalence between the two reduction relations R_N (Nagelian or syntactic or positivist) and R_S (structuralist), where ' x ' and ' y ' range over theories in the structuralist sense and ' x^* ' and ' y^* ' range over theories as syntactic entities (ignoring modalities):

$$\forall x \exists x^* (x^* \text{ expresses } x \ \& \ \neg \exists y (x^* \text{ expresses } y \ \& \ y \neq x))$$

$$\forall x, y, x^*, y^* ((x^* \text{ expresses } x \ \& \ y^* \text{ expresses } y) \rightarrow (xR_S y \leftrightarrow x^*R_N y^*)).$$

This equivalence is interesting, because it will give rise to similar claims concerning hierarchies of sciences, concerning monism and so forth. If we assume that the notion of reduction does not presuppose a certain notion of a theory, then which relation we should choose to define reduction, R_N or R_S , depends on independent criteria, presumably concerning the metaphysics of theories. It might turn out that theories are to be conceived of as structures rather than syntactic entities, such that the structuralist definition is more appropriate in this respect. In this sense, an explication may involve aspects of stipulation.

Nevertheless, on this interpretation, explication has still got a lot to do with conceptual analysis. Intuitively, it aims at a partially *illuminating* definition of the target-concept, just like conceptual analysis, classically conceived, aims at an illuminating or meaning-revealing definition.

Definitions can be correct without being illuminating. A definition is correct iff the truth conditions fixed by the *definiens* capture the range of (possible) applications of the predicate so defined. However, there are correct though uninteresting definitions. Let us assume that the first natural kind I thought about this morning was the kind *horse*. Then, we can define the predicate ' is a horse' as follows: $x \text{ is a horse} \leftrightarrow_{\text{def.}} x \text{ belongs to the first kind I thought about today (in the actual world)}$. This is a correct though non-illuminating definition. Similarly, and, avoiding the use

of indexicals, we can define the concept of water as follows: $x \text{ is water} \leftrightarrow_{\text{def.}} x \text{ is } H_2O$. If water is identical to H_2O , then this definition fixes the truth-conditions of applications of the predicate ‘_is water’. If you assume that it suffices to fix truth-conditions for the actual world only, we could go a step further: Assume that being a renate (the property of having kidneys) is not being a cordate (the property of having a heart),¹³ despite the fact that all renates are cordates, and *vice versa*. Then: $x \text{ is a renate} \leftrightarrow_{\text{def.}} x \text{ is a cordate}$ is a perfect definition. Compare the examples given so far to the following case: $x \text{ is a vixen} \leftrightarrow_{\text{def.}} x \text{ is a female fox}$. Again, the truth conditions are fixed. But, in addition, we learn something about the meaning of the predicate ‘_is a vixen’ (as opposed to merely learning something about the predicate’s extension, intension, or the property signified by the predicate). Put differently: This is a definition based on intuitions concerning meaning rather than a definition which is true in virtue of coincidence or mere a posteriori metaphysical truth. We can describe this definition as *illuminating*. Note that there are two kinds of illuminating definitions, the first of which is instantiated by the vixen-definition. It can be described as being based on a decomposition of meaning: The meaning of the term ‘vixen’ is given in the *definiens*. However, there are definitions which are illuminating (in an interesting sense), but which do not build on decomposition of meaning, even though they are based on relations between the meanings of the terms employed in the *definiens* and the *definiendum*. Here is an example: $x \text{ is a natural number} \leftrightarrow_{\text{def.}} x \text{ is either odd or even \& } x \text{ is a positive number (or } x \text{ is zero)}$. This type of definition is based on reasoning about meaning, but we should not regard the *definiens* as consisting in the decomposition of the meaning of the predicate used in the *definiendum*. It is illuminating in the sense that it gives us some insights into the meaning-relations between different predicates, unlike the other examples, but it is not illuminating in the way the vixen-example is. Why is this distinction important? Even though meaning-decomposition might be a project not worth aiming at in the context of reduction, it might still be the case that a definition based on meaning considerations can illuminate some aspects of the notion as it is pre-theoretically, or informally characterized. Ideally, these characterizations yield the relevant *job-description*. By becoming clear about this description, we give a (partly) illuminating definition, which may involve stipulations. This approach does justice to the desiderata identified above: Candidate definitions can be mistaken, and in order to avoid mistakes, we should pay attention to the term’s pre-defined use, namely, by paying attention to the job-description. But which job-description can be extracted from the informal use of the term ‘reduction’?

¹³This example is taken from Quine (1986 (1970), 8 f.). Note that Quine is concerned with the question of *synonymy*. Here, we are concerned with *definitions*. Both questions are intimately connected, but they are distinct. For the present purpose, I slightly changed the example: Property designators (like ‘being a renate’) do not figure in the Quinean case. Talking about properties, it seems easier to construct the relevant example.

2.3 The Concept of Reduction as the Subject of Explication?

We are already familiar with a short description of the job ‘reduction’ is supposed to do: *The concept of reduction reconciles diversity and directionality with strong unity, without relying on elimination*. Let me briefly connect this slogan to informal characterizations of the term that can be found in the literature, and then turn to alleged paradigmatic cases of reduction. We thereby arrive at and motivate the relevant criteria of adequacy. Here are some of these informal or pre-theoretic characterizations, which will lead to criteria of adequacy an explication of the notion of reduction has to fulfill:

- (Kim) If Xs are reduced, or reducible, to Ys, there are no Xs over and above Ys. (cf. Kim 2006, 275 f., (given in a similar fashion by Smart 1959))
- (Wimsatt) Inter-level reductions are compositional. They localize, identify, and articulate mechanisms that explain upper level phenomena, relationships and entities. (Wimsatt 2006, 449)
- (Chalmers) [W]hen [in the context of a reductive explanation, *RvR*] we give an appropriate account of lower-level processes, an explanation of the higher-level phenomenon falls out. (Chalmers 1996, 42)¹⁴
- (Nagel-1) [In reductions, A] set of distinctive traits of some subject matter is assimilated to what is patently a set of quite dissimilar traits. (Nagel 1961, 339 f.)
- (Nagel-2) [The reduced] science deals with macroscopic phenomena, while the [reducing] science postulates a microscopic constitution for those macroscopic processes. (Nagel 1961, 340)
- (Sarkar) The reduced theory is explained by a reducing theory which is presumed to be more fundamental. (Sarkar 1992, 167)

These descriptions are often enriched by examples, which supposedly describe *paradigmatic cases* of reduction: (*human*) *pain* reduces to *C-fiber stimulation*, or *water* reduces to *H₂O* (the latter can be found in (Putnam 1975, 431) (Fodor 1981, 150) and (Kim 1992, 23)).¹⁵ Talk about levels, part-whole relations and explanation suggests *directionality towards*, or *priority of* the reducing item. Talk about assimilation and not-existing “over and above” suggests unity as well as

¹⁴According to Chalmers, reductive explanation is explanation of a phenomenon in terms of a set of properties on which the property of being an instance of this phenomenon globally logically supervenes (Chalmers 1996, 48).

¹⁵Note that these cases are paradigmatic in a sense different from the sense in which Balzer’s cases are paradigmatic. In Balzer’s case, the paradigmatic cases are actual cases, and if it turns out that an actual case, which was used as a paradigmatic case (in Balzer’s sense) does not match the definition, the definition is to be changed. In contrast, these paradigmatic cases can be used even if water is not H₂O.

non-elimination. This is also reflected in the *use* of the term ‘reduction’. Let me give just one example to illustrate what I am thinking of:

(Menzies) It is widely held among philosophers that macrolevel causal relations are reducible to microlevel causal relations. (Menzies 1988, 551)

This statement embeds a claim concerning the relation between two different sets of causal relations. The term ‘reduction’ is *used to articulate a thesis about a reductive relation*. It is not a characterization of the use of the term ‘reduction’. It should be clear that what underlies “what is widely held among philosophers” will, in this case, be some vague intuition concerning reduction, or a family of intuitions, rather than some clearly defined concept of reduction. Similar statements include ‘mental properties reduce to physical properties’, or ‘wholes reduce to their parts’.¹⁶

So, what role does this set of descriptions, paradigmatic examples and uses of the term ‘reduction’ play in the reduction debate? The answer is simple: Descriptions and paradigmatic cases serve the purpose of giving a first, tentative idea of what a well-defined notion of reduction should roughly consist in. These descriptions express intuitions and preliminary ideas of how the notion of reduction is to be conceived of. Paradigmatic cases can be regarded as the (or a more or less flexible version of a) touchstone for any definition of reduction. Using the term in philosophical discourse that is not concerned with an introduction of the term imposes constraints on the job-description.

We have thus sketched one reason to assume that the third variant of developing a definition is appropriate in the present context. There are pre-theoretic characterizations that should *guide* the explicit definitions. These characterizations describe the job our concept of reduction is supposed to do. It is not a set of examples, which does so, but rather a number of criteria we already possess when we try to define the notion of reduction. The role of examples, or paradigmatic cases is this: Independent of whether or not water actually is H₂O – we just assume that it is. Then, this case serves as a good example of a reduction. No empirical investigation of this example will give us a deeper understanding of the concept of reduction. This stands in stark contrast to what Kemeny and Oppenheim had in mind. As suggested above, the message of approaches that are in the spirit of Kemeny and Oppenheim can be captured by labeling them *extension first-approaches*. In contrast, we aim at a definition based on criteria that are gained from intuitive characterizations of the concept, or its explanatory role in philosophy – hence, the approach can be characterized as a *criteria first-approach*! The notion of reduction is, on this latter interpretation, at least partly determined by criteria that are gained independently of reflection on (an alleged) extension. Whereas *extension first-approaches* suit the goal of a rational reconstruction of scientific endeavors, such as *reductions*, any illuminating explication of the philosophical *concept* of reduction will have to rely on a *criteria first* approach.

¹⁶These statements do not perfectly fit other examples of reduction-talk; I will later refer to these statements as ‘generic’ statements – we need a special apparatus to deal with these constructions (see Sect. 5.6.2).

There is a use of the term ‘reduction’ we have not commented upon yet that clearly matches the goal of extension first approaches: When philosophers speak of *asynchronous* reduction, they have in mind a specific kind of theory-change. This concept captures a temporal relation that involves theories (or stages of one theory) at different times. In contrast, one may tentatively describe the notion of reduction we are interested in here as *synchronous*, or, in a more Platonist spirit, as *atemporal*: It is a relation that does not require its relata to instantiate a certain temporal relation, such as the relation of *occurring later than*.

Our slogan contains the core idea that defines the concept’s job (schematically):

(*Job-Description*): We need a definition that

- (i) accounts for the directionality, such that it gives a definition according to which if *a* reduces to *b*, then necessarily, *b* does not reduce to *a*, and
- (ii) accounts for the idea of unity (ideally in the sense of strong unity) without elimination.

In addition, it will be desirable that our definition enables us to

- (iii) illuminate the paradigmatic cases;
- (iv) explain the intuitive characterizations of the notion of reduction;
- (v) account for related topics, such as reduction and physicalism, reduction and scientific unification, and similar issues; i.e. the explication should yield a fruitful notion by fulfilling the explanatory task associated with ‘reduction’.

Let me add this proviso: In contexts of explication, one may feel free to drop one criterion in favor of another. One case in point is this: As already mentioned, in the philosophy of science, it is common to drop an assumption imposed by the paradigmatic cases, namely, that reduction should connect *kinds* or *phenomena* or *events* (in any primary sense of the term ‘reduction’). Approaches to reduction that take theories to be the primary *relata* of the reduction relation follow this idea. This is usually justified by arguing that the idea of “ontological” reduction is just confused. We will turn back to this assumption later on. Let me close this introduction with a few terminological remarks.

2.4 Terminology

The most important terminological distinction for the discussion to come can briefly be described as the distinction between *ontological* talk and *representational* talk (for the application of a similar distinction to the reduction-debate, see Van Gulick 2001). Here is a bunch of ontological items philosophers in the philosophy

Table 2.2 Ontological furniture

Ontological furniture	
<i>Abstract objects</i>	<i>Concrete objects</i>
<i>(not located in space or time)</i>	<i>(located in space and time)</i>
<i>Kinds</i>	<i>Particularized properties</i>
<i>Relations</i>	<i>Particularized events</i>
<i>Properties</i>	<i>Tokens</i>
<i>Types</i>	<i>Compositions</i>
<i>Sets</i>	<i>Sums (of concrete objects)</i>
<i>Events</i>	<i>Wholes</i>
	<i>Parts</i>
	<i>Individuals</i>

of science and in the philosophy of mind often refer to: properties, relations, kinds, types, events, sets, sums, compositions, wholes, parts, tokens, particularized properties, particularized events, and individuals. Let me propose a list that is structured with respect to the ontological location of its items. The list is Platonist in spirit, even though, with small changes, it allows for both, nominalist as well as constructivist readings. Table 2.2 gives an overview on our ontological furniture (I will comment on the representational furniture in a moment).

These categories are not mutually exclusive. Any individual horse, for example, is a token of the natural kind horse; we might conceive of compositions as a certain kind of wholes, for example as structured wholes, and of wholes and compositions as individuals; we could conceive of events as properties, and of particularized events as particularized properties.¹⁷ For what follows, ‘property’, ‘type’ and ‘kind’ will be used interchangeably. The term ‘relation’ refers to what can also be referred to by ‘relational property’, ‘relational type’ and ‘relational kind’, or, briefly, by ‘kind’, ‘property’ and ‘type’ (thus, there are n-ary kinds, types and properties with $n > 1$ as well as with $n = 1$). If you prefer a nominalist ontology, take properties to be sums of individuals (and replace the idea of properties not being located in space and time by some other idea that serves your purpose), if you are more liberal, take them to be sets of entities, if you are even more liberal, take them to be functions from worlds to extensions. For the discussion to come, it will be useful to think of properties in one of the intensional senses or, for most of the properties the bearers of which are individuals occupying a position in space and time, in functional terms.

The purpose of this list is primarily to indicate that if one of its items occurs, we have a case of ontological talk. Secondly, we can use this list to introduce relations instantiated by pairs of items on this list: Particularized properties and

¹⁷Lewis (1986) treats events as properties (Lewis interprets them as properties of space-time regions), Kim interprets them as property-exemplifications which involve a property, a substance and a time (Kim 1976). Davidson famously defended the idea that events are particulars (Davidson 1969, 1970). Here, I shall speak of events as follows: there are event-types and there are instances of event-types, namely, concrete particularized events. Thus, I shall assume that at least in principle, talk about kinds of events makes sense. Further differences will not matter for our present concern.

Table 2.3 Representational furniture

Representational furniture	
<i>Conceptual level</i>	<i>Vehicles</i>
<i>Contents</i>	<i>Words</i>
<i>Meanings</i>	<i>Sentences</i>
<i>Propositions</i>	<i>Theories</i>
<i>Concepts</i>	<i>Mental representations</i>
<i>Theories</i>	
<i>Events</i>	

particularized events *instantiate* properties or events. Parts *constitute* complex wholes and compositions, and different things *form* a sum of these things. In addition, we need a convention concerning ways of making these things available in discourse: We talk about such entities using terms, or, sometimes, in the case of properties, using predicates. I will say that terms and predicates *pick out* the entities we talk about using these linguistic devices. Whereas terms *designate* or *refer to* the entity they pick out, predicates *signify* the properties they pick out. Possible differences between *referring terms* and *descriptions* that *designate* objects will be neglected here.

In contrast, terms and predicates *express* their conceptual content, if any, which brings us to our second list containing the representational furniture, given in Table 2.3. It involves what is expressed as well as what expresses, that is, entities such as concepts on the one hand, and words and sentences on the other.

Note that these lists are not complete. They form a device used to indicate representational talk as being distinct from ontological talk. Obviously, we can talk about representational entities in an ontological way, for example, when we use a term to refer to an expression’s meaning, or to a word. That is: The representational entities can become part of the ontological furniture of a discourse, namely, a discourse that is concerned with representational issues. Meta-language is a paradigmatic example of this kind of discourse.

Relations of *picking out* can be regarded as being concerned with a language’s ontology, whilst the relation of *expressing* is here understood as being concerned with a language’s conceptual realm, or with a language’s meanings. I shall use the term ‘concept’ in constructions which are used to refer to meanings (like ‘the concept of a horse’). Intuitively, one can describe the difference between a language’s ontology and its conceptual realm in terms of the objects, facts, and properties a language is *about* on the one hand and the representational structure in virtue of which it is about these entities and properties on the other. However, whether or not some feature belongs to the language’s ontology or to its representational level is dependent on the discourse, not (primarily) on what the entity is. Note that this is not to say that any expression has a meaning, or that if an expression picks out an object or a property, it has a meaning beyond what it picks out. Directly referring terms lack meaning (in the sense of ‘conceptual content’).

The term ‘theory’ occurs twice in Table 2.3. This is neither to be regarded as a stipulation, nor as a substantial thesis about what theories are. Rather, it mimics two different ways of talking about theories. In the second sense (as linguistic entities or representational vehicles), theories can be regarded as sets of sentences, that is: sets of linguistic entities. Strictly speaking, they are therefore not subject to operations such as derivation, and they do not stand in relations like the implication relation to each other, relations which hold in the primary sense of the terms ‘derivation’ and ‘implication’ between linguistic entities only (no set is a linguistic entity). Nevertheless, I will, in accordance with the literature, say that sets of sentences *imply* sentences or other sets of linguistic entities.¹⁸ Moreover, I will ignore the aspect of theories as *developing* entities. Thus, I idealize, using a static notion of a theory.

Now, we are in a position to generate a number of operators, which will enable us to distinguish between ontological aspects and conceptual aspects of a certain linguistic item. Talking about properties, I will use the singular term forming operator ‘the property of being _’, which works on general terms (similarly, I will use ‘the _ relation’ for general terms which, if they are used to form a predicate in the standard way, can form an n-place predicate with $n > 1$).¹⁹ The result of an application of this operator to a general term is a property designator that refers to the property that is picked out by the general term (or is signified by the predicate that is composed of the copula (if any) and this general term). Sometimes, and when appropriate, I will use the suffix ‘-ness’ to form a property-designator. ‘Redness’ refers to the property of being red, the property that is picked out by ‘red’ and signified by ‘_ is red’. Note that sometimes, it will be suitable to treat terms like ‘water’ and ‘H₂O’ as referential expressions designating kinds; for example, when we say that *water* reduces to *H₂O*, this should in at least some contexts be interpreted as stating that the kind water reduces to the kind H₂O. (Similarly for ‘pain’ and ‘C-fiber stimulation’.) Thus, ‘water’ is, at least on some occasions, in the same boat as ‘the property of being water’; whether or not it is will become apparent from context.

Similarly, I will refer to conceptual contents or meanings using results of an application of the operators ‘the concept of a _’, in the case of sortals, and ‘the concept of _’ in the case of non-sortals. These operators also work on general terms. The referents of such terms are the conceptual contents expressed by the embedded terms. For example, the ‘concept of red’ refers to the conceptual content expressed by ‘red’.

¹⁸In order to bridge this gap one might think of theories as sums of sentences rather than sets. Even though this is metaphorical talk (if we take sums in the ordinary sense of the expression to be sums of individuals located in space and time), it captures the idea that the whole does not belong to a kind of entities possessing relevantly different properties from the properties possessed by the elements, parts or constituents.

¹⁹For a detailed discussion of how ‘the property of being_’, ‘_ness’ and the like function, see Schnieder (2006).

2.5 Conclusion

Reduction reconciles diversity and directionality with strong unity. We now have a better idea of what strong unity amounts to, and we have an idea of how the guiding question should be addressed:

- Q1: How can we reconcile *diversity* and *directionality* with strong *unity*?
 Th. 1: *Strong unity* is the unity of identity. *Method*: To answer Q1 is to come up with an *explication* of the concept of reduction.

Unity is the unity postulated by a monist who is a realist about objects at “higher levels”. Moreover, building on the model of explication proposed here, we are now in a position to start a discussion of the puzzle of reduction. A discussion of the puzzle suggests that diversity in reduction is conceptual in nature: Reduced and reducing item differ in how they are presented in a true reduction statement.

Bibliography

- Balzer, Wolfgang. 1984. On the comparison of classical and special relativistic space-time. In *Reduction in science: Structure, examples, philosophical problems*, ed. Wolfgang Balzer, David A. Pearce, and Heinz-Jürgen Schmidt, 331–357. Dordrecht: Reidel.
- Bechtel, William. 1994. Levels of descriptions and explanation in cognitive science. *Minds and Machines* 4: 1–25.
- Bechtel, William, and Robert Richardson. 1993. Emergent phenomena and complex systems. In *Emergence or reduction? Essays on the prospects of nonreductive physicalism*, ed. Ansgar Beckermann, Hans Flohr, and Jaegwon Kim, 257–288. Berlin: de Gruyter.
- Bickle, John. 2008. Real reduction in real neuroscience: Metascience, not philosophy of science (and certainly not metaphysics!). In *Being reduced*, ed. Jakob Hohwy and Jesper Kallestrup, 34–51. Oxford: Oxford University Press.
- Block, Ned. 1997. Anti-reductionism slaps back. *Philosophical Perspectives* 11: 107–132.
- Butterfield, Jeremy. 2011a. Emergence, reduction and supervenience: A varied landscape. *Foundations of Physics* 41: 920–959.
- Butterfield, Jeremy. 2011b. Less is different: Emergence and reduction reconciled. *Foundations of Physics* 41: 1065–1135.
- Carnap, Rudolf. 1934. *The unity of science*. London: Kegan Paul, Trench, Trubner, and Co.
- Carnap, Rudolf. 1950. *Logical foundations of probability*. Chicago: University of Chicago Press.
- Carnap, Rudolf. 1988 (1956). *Meaning and necessity. A study in semantics and modal logic*. Chicago: The University of Chicago Press.
- Chalmers, David. 1996. *The conscious mind*. Oxford: Oxford University Press.
- Churchland, Paul M. 1981. Eliminative materialism and the propositional attitudes. *The Journal of Philosophy* 78: 67–90.
- Churchland, Paul M. 1985. Reduction, qualia and the direct introspection of brain states. *Journal of Philosophy* 82: 8–28.
- Churchland, Patricia. 1986. *Neurophilosophy*. Cambridge, MA: MIT Press.
- Craver, Carl. 2005. Beyond reduction: Mechanisms, multifield integration and the unity of neuroscience. *Studies in the History and Philosophy of Biological and Biomedical Sciences* 36: 373–395.
- Crook, Seth, and Carl Gillet. 2001. Why physics alone cannot define the ‘physical’. *Canadian Journal of Philosophy* 31: 333–360.

- Davidson, Donald. 1969. The individuation of events. In *Essays in honor of Carl G. Hempel*, ed. Nicholas Rescher, 216–234. Dordrecht: Reidel.
- Davidson, Donald. 1970. Events as particulars. *Noûs* 4: 25–32.
- Dizadji-Bahmani, Foad, Roman Frigg, and Stephan Hartmann. 2010. Who is afraid of Nagelian reduction? *Erkenntnis* 73: 393–412.
- Endicott, Robert. 1998. Collapse of the new wave. *The Journal of Philosophy* 95: 53–72.
- Esfeld, Michael, and Christian Sachse. 2007. Theory reduction by means of functional sub-types. *International Studies in the Philosophy of Sciences* 21: 1–17.
- Fodor, Jerry A. 1974. Special sciences: Or the disunity of science as a working hypothesis. *Synthese* 28: 97–115.
- Fodor, Jerry A. 1981. *Representations: Philosophical essays on the foundations of cognitive science*. Cambridge, MA: MIT Press.
- Hanna, Joseph. 1968. An explication of ‘explication’. *Philosophy of Science* 35: 28–44.
- Hellman, Geoffrey P., and Frank W. Thompson. 1975. Physicalism: Ontology, determination, and reduction. *Journal of Philosophy* 72: 551–564.
- Horst, Steven. 2007. *Beyond reduction: Philosophy of mind and post-reductionist philosophy of science*. Oxford: Oxford University Press.
- Kemeny, John G., and Paul Oppenheim. 1956. On reduction. *Philosophical Studies* 7: 6–19.
- Kim, Jaegwon. 1976. Events as property exemplifications. In *Action theory*, ed. Douglas Walton and Myles Brand, 159–177. Dordrecht: Reidel.
- Kim, Jaegwon. 1992. Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52: 1–26.
- Kim, Jaegwon. 1993. *Supervenience and mind. Selected philosophical essays*. Cambridge: Cambridge University Press.
- Kim, Jaegwon. 2006. *Philosophy of mind*. Boulder: Westview Press.
- Lewis, David. 1972. Psychophysical and theoretical identifications. *Australasian Journal of Philosophy* 3: 249–258.
- Lewis, David. 1986. Events. In *Philosophical papers*, ed. David Lewis, 241–269. Oxford: Oxford University Press.
- Melnik, Andrew. 2003. *A physicalist manifesto: Thoroughly modern materialism*. Cambridge: Cambridge University Press.
- Menzies, Peter. 1988. Against causal reductionism. *Mind* 98: 551–574.
- Moulines, C. Ulises. 1984. Ontological reduction in the natural sciences. In *Reduction in science: Structure, examples, philosophical problems*, ed. Wolfgang Balzer, David A. Pearce, and Heinz-Jürgen Schmidt, 51–70. Dordrecht: Reidel.
- Nagel, Ernest. 1961. *The structure of science. Problems in the logic of explanation*. New York: Harcourt, Brace & World, Inc.
- Neurath, Otto. 1959. Sociology and physicalism. In *Logical positivism*, ed. Alfred J. Ayer, 282–317. New York: Free Press.
- Nickles, Thomas. 1973. Two concepts of intertheoretic reduction. *The Journal of Philosophy* 70: 181–201.
- Papineau, David. 1993. *Philosophical naturalism*. Oxford: Blackwell.
- Papineau, David. 2002. *Thinking about consciousness*. Oxford: Oxford University Press.
- Putnam, Hilary. 1975. On properties. In *Mathematics, matter and method. Philosophical papers*, ed. Hilary Putnam, 305–322. Cambridge: Cambridge University Press.
- Quine, Willard V.O. 1960. *Word and object*. Cambridge, MA: MIT Press.
- Quine, Willard V.O. 1986. *Philosophical logic*. Cambridge, MA: Harvard University Press.
- Richardson, Robert C. 1979. Functionalism and reductionism. *Philosophy of Science* 45: 533–558.
- Richardson, Robert C. 1982. How not to reduce a functional psychology. *Philosophy of Science* 48: 125–137.
- Sarkar, Sahotra. 1992. Models of reduction and categories of reductionism. *Synthese* 91: 167–194.
- Schnieder, Benjamin. 2006. Canonical property designators. *American Philosophical Quarterly* 43: 119–32.
- Sklar, Lawrence. 2003. Dappled theories in a uniform world. *Philosophy of Science* 70: 424–441.

- Smart, John J.C. 1959. Sensations and brain processes. *Philosophical Review* 68: 141–156.
- Sneed, Joseph. 1971. *The logical structure of mathematical physics*. Dordrecht: Reidel.
- Spurrett, David, and David Papineau. 1999. A note on the completeness of ‘physics’. *Analysis* 59: 25–28.
- Strawson, Peter F. 1963. Carnap’s views on constructed systems versus natural languages in analytic philosophy. In *The philosophy of Rudolf Carnap*, ed. A. Schilpp, 503–518. La Salle: Open Court.
- van Fraassen, Bas. 1980. *The scientific image*. Oxford: Clarendon.
- Van Gulick, Robert N. 1992. Nonreductive materialism and the nature of intertheoretical constraint. In *Emergence or reduction? Essays on the prospects of non-reductive physicalism*, ed. Ansgar Beckermann, Hans Flohr, and Jaegwon Kim, 157–179. Berlin: de Gruyter.
- Van Gulick, Robert N. 2001. Reduction, emergence and other recent options on the mind/body problem: A philosophic overview. *Journal of Consciousness Studies* 8: 1–34.
- van Riel, Raphael. 2011. Nagelian reduction beyond the Nagel-model. *Philosophy of Science* 78: 353–375.
- Waters, Kenneth. 1990. Why the anti-reductionist consensus won’t survive. The case of classical Mendelian genetics. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1: 125–139.
- Wimsatt, William C. 1976. Reductive explanation: A functional account. In *PSA: Proceedings of the biennial meeting of the Philosophy of Science Association*, 1974, ed. G. Pearce, A.C. Michalos, C.A. Hooker, and R.S. Cohen, 671–710. Dordrecht: Reidel.
- Wimsatt, William C. 2006. Aggregate, composed, and evolved systems: Reductionistic heuristics as means to more holistic theories. *Biology and Philosophy* 21: 667–702.
- Witmer, Gene. 2008. (Review) Beyond reduction: Philosophy of mind and post-reductionist philosophy of science. *Notre Dame Philosophical Reviews*. <http://ndpr.nd.edu/news/23449-beyond-reduction-philosophy-of-mind-and-post-reductionist-philosophy-of-science/>. Accessed 5 July 2013.

The Concept of Reduction

van Riel, R.

2014, XII, 226 p. 3 illus., Hardcover

ISBN: 978-3-319-04161-2