Philos Stud (2020) 177:3571–3593
https://doi.org/10.1007/s11098-019-01383-0

Structural problems for reductionism

Stephan Leuenberger[1]

Published online: 9 December 2019

Abstract Universal reductionism—the sort of project pursued by Carnap in the *Aufbau*, Lewis in his campaign on behalf of Humean supervenience, Jackson in *From Metaphysics to Ethics*, and Chalmers in *Constructing the World*—aims to reduce everything to some specified base, more or less austere as it may be. In this paper, I identify two constraints that a promising strategy to argue for universal reductionism needs to satisfy: the *exhaustion constraint* and the *chaining constraint*. As a case study, I then consider Chalmers' *Constructing the World*, in which a priori implication, or "scrutability", plays the role of reduction. Chalmers first divides up the total vocabulary of our language into different families, and then argues, for each family separately, that truths involving expressions in that family are scrutable from the putative base. He does not systematically address the question whether "cross-family sentences"—sentences involving expressions from more than one family—are scrutable. I shall argue that this lacuna cannot be filled, since scrutability does not allow for the exhaustion constraint and the chaining constraint to be jointly satisfied. I further suggest that Carnap's account, in which definability plays the role of reduction, has better prospects of meeting these constraints.

Keywords Scrutability · Supervenience · A priori · Reduction · Chalmers

1 Layer-cake reductionism

To reduce everything to an austere base is a familiar philosophical ambition. Such universal reductionism comes in a variety of versions. The desired austerity in the base may be ontological, in the sense of involving few entities or categories of

✉ Stephan Leuenberger
stephan.leuenberger@glasgow.ac.uk

[1] Philosophy, School of Humanities, University of Glasgow, Glasgow G12 8QQ, UK

entities; conceptual, in the sense of requiring few concepts or families of concepts; or epistemic, in the sense of assuming only modest knowledge claims. I consider the views advocated in Carnap (1928a), Lewis (1986), Jackson (1998), Chalmers (2012), and Chalmers and Jackson (2001) as paradigmatic representatives of universal reductionism.[1]

Despite their differences, reductionist projects tend to share a certain picture of the world as layered, with the putative base forming the bottom layer. Every non-basic layer is supposed to be reducible to the ones below it. Moreover, it is a guiding assumption of reductionist picture-thinking that reduction admits of a certain kind of "chaining": layers higher up do not just count as reducible to the ones below them, taken together, but also to the bottom layer all by itself.

To illustrate: if the biological reduces to the chemical and the physical together, and the chemical reduces to the physical, then it follows that the biological reduces to the physical alone. Or to vary and expand the example, consider the question whether the realm of colours—the "chromatic"—reduces to the physical. A natural strategy in arguing for a positive answer will appeal to the phenomenal in an auxiliary role. It is plausible that the realm of colours reduces to the physical and the phenomenal together. Here, the physical includes properties or facts involving light, surfaces, retinas, brains, etc, while the phenomenal includes colour qualia or facts about what it is like to have colour experiences. It is, of course, controversial whether the phenomenal reduces to the physical. But we would expect that in tandem, the two reductive claims will settle our question. That is, we would expect the argument with premises (1) and (2) and conclusion (3) to be valid:

(1) The chromatic reduces to the phenomenal and the physical together.
(2) The phenomenal reduces to the physical.
(3) The chromatic reduces to the physical.

What is more, we would expect that argument to be valid purely in virtue of the logic of reduction, and thus remain valid no matter what we substitute for "the chromatic", "the physical", and "the phenomenal."

In addition to such "vertical chaining" in the layer-cake, reduction should also allow for what we might call "horizontal accumulation", which is exemplified by the following argument:

(4) The chromatic reduces to the physical.
(5) The evaluative reduces to the physical.
(6) The chromatic and evaluative together reduce to the physical.

Again, we would expect that argument to be valid purely in virtue of the logic of reduction.

[1] Some philosophers treat a universal reductionist thesis such as physicalism as a background assumption, not as a hypothesis to be argued for. This may well be a sensible policy in some philosophical contexts. In this paper, however, my concern is with the structure of *arguments* for a universal reductionist conclusion.

Since sweeping arguments for universal reductionism are rarely available, arguments for it tend to proceed in a piecemeal manner. Without the availability of vertical chaining and horizontal accumulation, such a piecemeal approach is not going to succeed. It is thus a constraint on an explication of reduction that it guarantees the validity of arguments such as the above—that is allows chaining, broadly construed. (I shall be more precise about what this *chaining constraint* amounts to in Sect. 2.)

The significant role that chaining plays in arguments for reductionist theses can be further exemplified with the case of physicalism. Confronted with high-level phenomena, studied by special sciences such as sociology or economics, the physicalist has little prospect of finding a *direct* way of establishing their reducibility.[2] In practice, they take an indirect route, perhaps through psychology and neuroscience. The situation is no different when we consider philosophical rather than scientific reductionist projects. David Lewis' strategy when arguing for Humean supervenience—a close relative of physicalism—is a case in point. Humean supervenience claims that everything supervenes on the Humean mosaic, understood as a spatiotemporal arrangement of local intrinsic qualities. When sketching an argument for the Humean supervenience of counterfactuals, Lewis implicitly appeals to the kind of chaining I mentioned:

> To the extent that this similarity [which determines whether counterfactuals are true] consists of perfect match in matters of particular fact, it supervenes easily on the arrangement of qualities; and to the extent that it consists of ... conformity by one world to the laws of the other, it supervenes if the laws do. (Lewis 1986, xii)

Lewis' idea here is that counterfactuals can be shown to supervene on the Humean mosaic by verifying that they supervene on the laws and the mosaic together, and also establishing that the laws supervene on the mosaic. The argument has the same form as the one from (1) and (2) to (3) displayed above.

A martial metaphor may help to make the rather abstract chaining constraint more vivid.[3] Suppose a fictitious country, Physicalia, is trying to devise a strategy to conquer the whole world. Its leaders might draw up a list of territories, and decide upon a sequence in which they are to be brought under control. A prudent plan for them would be to first conquer the neighbouring countries, then build bases in them, and only then invade *their* neighbouring countries that are not its own neighbours; and so on. Otherwise the supply lines from Physicalia would be too long. The idea of chaining is that every country thus conquered will count as controlled by Physicalia alone, since any countries that helped had themselves been conquered by it.

[2] I shall take the physicalist to be a paradigmatic universal reductionist. The term "reduction" gets used in many different ways, of course. In this paper, I use it in a rather broad sense, ignoring among other things some of the finer distinctions at issue in the debate between "reductive" and "non-reductive" physicalism.

[3] It is revealing that Lewis talks of the "plan of battle" (p. xi) he is executing in his "campaign" (p. ix) on behalf of Humean supervenience.

The point of the metaphor is that a reductionist physicalist will wish to use the resources of neighbouring disciplines of physics when attempting to reduce the likes of sociology and economics. To satisfy the chaining constraint, the relevant notion of reduction must have the right structural features, such as being transitive—a theme that will be explored in more detail in the next section.

Of course, the most carefully chosen sequence of conquests will not enable Physicalia to achieve its goal of world domination unless the list of territories drawn up is complete, in the sense of adding up to the whole world. This *exhaustion constraint* may appear obvious as introduced here. But it has real bite when we leave the war metaphor behind, and ask the question what a physicalist needs to provide a reduction for. One natural answer is that she needs to provide a reduction for all truths. Plausibly, there are infinitely many truths, and verifying their reducibility one by one is not an option. The reductionist can only explicitly consider a finite number of truths. Her strategy satisfies the exhaustion constraint if she can successfully argue that if the cases she considers are reducible, then all cases are reducible.

In this paper, I shall argue that otherwise promising approaches to reduction fail to satisfy these constraints. In a nutshell, and with gross simplification, the point is this: if the relata of reduction are classes of predicates or properties, then the exhaustion constraint is violated; if they are classes of truths—be they sentences, propositions, or facts—then the chaining constraint is violated. (Note the ambiguity between properties and facts in "the realm of colours", "the physical" and "the phenomenal" above.)

My stalking-horse in this paper is the universal reductionist project recently articulated in Chalmers (2012). I choose it because of its virtues: it sets out the logic of reductionist reasoning more explicitly, and in more detail, than other representatives of the genre, for example Jackson (1998) or Lewis (1986).[4] My aim is not to refute the theses that Chalmers is putting forward, but rather to identify a limitation in the argumentative strategy he uses in supporting them.[5] The lessons to be drawn are not restricted to Chalmers' particular proposal, however, and extend to universal reductionist projects more generally: such projects need to be designed in such a way that both the chaining constraint and the exhaustion constraint are satisfied.

In the next section, I introduce Chalmers' project, and the key notion of scrutability. In Sects. 3–6, I argue that the project is caught between the Scylla and Charybdis of the exhaustion and chaining constraints. The final Sect. 7 discusses what lessons the case study holds for universal reductionist projects more generally.

[4] In both Lewis' and Jackson's case, the notion of supervenience in play is insufficiently specified; see Stalnaker (1996) and Williamson (2001) for discussion.

[5] Whether this is a problem for the project of Chalmers (2012) depends on whether that project is to positively establish (with the sort of cogency that philosophical theses aspire to) the key thesis of the book, or whether it is merely to argue that certain putative counterexamples can be resisted. I shall not discuss that interpretive question in this paper.

2 Definability versus scrutability

Chalmers presents his account as an updated and improved version of Carnap (1928a), which many have considered a heroic failure at best, and a quixotic folly at worst. While Carnap's and Chalmers' projects differ in a number of respects, one is of particular relevance for the present paper: Carnap argues that all terms are defined in terms of base terms, while Chalmers argues that all truths are implied by base truths.[6] Implication differs from definition in two salient ways: its relata are sentences rather than words, and it requires only a sufficient condition, rather than one that is both necessary and sufficient.

The shift from definition to implication is motivated by reflection on Gettier cases and other counterexamples to proposed definitions. We can recognise that knowledge cannot be defined as justified true belief, the thought goes, because Gettier's description of his case implies that Smith does not know that Jones owns a Ford or is in Barcelona. As Chalmers (2012, p. 13) puts it, "it is striking that in many cases, specifying a situation in terms of expressions that do not include 'knowledge' or its cognates (synonyms or near-synonyms) enables us to determine whether or not the case involves knowledge." He thinks that this is typical: terms do not have definitions in base vocabulary, but a full enough specification of a situation in base terms tells us whether they apply.

The central claim of Chalmers (2012) is that all truths about the world are a priori implied by truths statable in some highly restricted vocabulary. While this view is not neatly classified as either ontological, conceptual or epistemological reductionism, it has important connections to all these views, which the book maps out in detail.

Commenting on the broader philosophical orientation of their respective projects, Chalmers (2012, xviii) summarises the contrast with Carnap as follows: "To oversimplify, one might say that where Carnap leans toward empiricism, I lean toward rationalism." While he does not elaborate on the contrast, I take empiricism and rationalism to respond differently to the worry that truths in areas such as metaphysics and mathematics are unknowable. Empiricists give a deflationist interpretation of the sentences whose truth is in question—e.g. "God exists," "people have immortal souls," "our will is free," "there are inaccessible cardinals"—or they deny the intelligibility of these sentences altogether. Rationalists, in contrast, take the sentences at face value, and argue that we do, in principle, have the capacity to come to know whether they are true.

As I see it, the difference in outlook between empiricism and rationalism is reflected in the difference between definability claims and a priori implication claims. "Definition" often connotes "verbal", "conventional", or "non-substantive." The claim that terms belonging to a certain discourse are definable is naturally accompanied by a deflationist conception of the subject matter of that discourse—a conception I have associated with empiricism. Correspondingly, the claim that

[6] I am glossing over subtleties of Carnap exegesis here. What is defined in Carnap (1928a) are officially objects or concepts, not terms.

truths belonging to a certain discourse are a priori implied fits a view that takes the sentences expressing these truths at face value—a conception I have associated with rationalism.

In this section and the two subsequent ones, shall argue that the shift from terms to truths raises a structural problem. In Sect. 6, I shall return to the contrast between rationalism and empiricism.

To be able to state Chalmers' central thesis more precisely, we need to define the key notion of scrutability:

Definition 1 A sentence ϕ is *scrutable* from a class of sentences Γ iff for some subset Γ' of Γ, the material conditional $\bigwedge \Gamma' \rightarrow \phi$ is a priori. A class Δ is scrutable from Γ just in case every member is.

Here, $\bigwedge \Gamma'$ is the (possibly infinite) conjunction of all the members of Γ'.[7]

Given a scenario w—which we may think of as a centred possible world—we are often interested in classes Γ of sentences that are true in w, and such that every truth of w is scrutable from Γ. I shall call such a class a *base* for w.

Chalmers aims to identify candidate bases that are parsimonious, and to argue that they are indeed bases. The relevant kind of parsimony concerns the number of expressions from which the sentences in the base are composed. According to him, a base is *compact* if "it includes expressions from only a small number of families" (p. 21).[8] The term "small" in the characterisation of compactness is vague, and we may not be sure about how to individuate and count families. However, these issues will not matter for our discussion.

Chalmers goes on to consider various candidate bases, most extensively one called *PQTI* that is (in a slightly different form) already introduced in Chalmers and Jackson (2001). In this article, I shall not be concerned with any base in particular. Accordingly, I shall not try to present counterexamples to Chalmers' scrutability theses. Rather, I wish to highlight certain abstract and structural features of the proposal, and draw lessons for how reductionist projects ought to be pursued.

Scrutability as defined is a relation between a sentence and a class of sentences, and also between classes of sentences. In Sect. 1, I noted that any relation that is to play the role of reduction needs to satisfy the chaining constraint. It is now time to be more precise about what kind of chaining ought to be licensed.

In Sect. 1, I illustrated the chaining constraint with the argument from (1) and (2) to (3), concerning the reduction of the realm of colours. The structural feature of reduction exploited in that argument can be schematically formulated as follows:

$$[\text{Quasi-transitivity}] \qquad \frac{x \Rightarrow y \qquad x \cup y \Rightarrow z}{x \Rightarrow z}$$

[7] This notion of scrutability is called "a priori scrutability" by Chalmers; since other notions of scrutability he defines will not play a role in this article, I shall use the shorter term.

[8] In fact, Chalmers adds another condition for a base to be compact, that it "includes no trivializing mechanisms." This complication is not relevant to this paper, and I shall ignore it.

In words: for all x, y, and z, if y reduces to x, and z reduces to x and y together, then z reduces to x. The symbol $\Rightarrow$ is a place-holder for a predicate expressing a reductive relation (scrutability, for example), with the reducing items written on the left, and the items to be reduced on the right. Moreover, $x \cup y$ denotes the class-theoretic union of x and y, which stands for "x and y together". To obtain the argument from (2) and (1) to (3) as an instance, we can interpret z as "the chromatic", y as "the phenomenal", and x as "the physical."

Quasi-transitivity licenses a move that Chalmers calls "minimizing the base", and that pervades his chapter 7: if it has been established that $x \cup y$ is a base—take z to be the universal class—and that x is scrutable from y, we can conclude that y is a base.

This structural feature of the relation $\Rightarrow$ is sometimes simply called "transitivity"; I opt for the label "quasi-transitivity" since I wish to reserve the more familiar term for the following feature:

$$\text{[Transitivity]} \qquad \frac{x \Rightarrow y \quad y \Rightarrow z}{x \Rightarrow z}$$

Just like the chaining constraint requires a candidate relation of reduction to be quasi-transitive, it requires it to be transitive.

In general, a quasi-transitive relation need not be transitive, nor vice versa. However, transitivity follows from quasi-transitivity among relations that are *monotonic*, in the sense that $x \Rightarrow y$ entails $x' \Rightarrow y$ for every x' of which x is a subclass.[9] Conversely, quasi-transitivity is entailed by transitivity together with a further feature, which I shall call "self-adjunction"[10]:

$$\text{[Self-Adjunction]} \qquad \frac{x \Rightarrow y}{x \Rightarrow x \cup y}$$

I shall not attempt to state conditions for satisfying the chaining constraint that are individually necessary and jointly sufficient. I do claim, however, that being transitive and quasi-transitive are necessary, since these features license forms of reasoning that are practically indispensable in the context of universal reductionist arguments. Among relations that are reflexive and monotonic, I take transitivity and quasi-transivity also as jointly sufficient conditions, since they jointly entail certain further features that license natural forms of reasoning.[11] In particular, they entail accumulation, the feature that guarantees the validity of the argument from (4) and (5) to (6) in Sect. 1[12]:

[9] Suppose $x \Rightarrow y$ and $y \Rightarrow z$. By monotonicity, $x \cup y \Rightarrow z$, and by quasi-transitivity, $x \Rightarrow z$.

[10] Suppose $x \Rightarrow y$ and $x \cup y \Rightarrow z$. Since $\Rightarrow$ is self-adjunctive, $x \cup y \Rightarrow y$, and by transitivity, $x \Rightarrow z$.

[11] What about relations that are transitive and quasi-transitive, but either not reflexive or not monotonic? For the purposes of this paper, it does not matter which ones among them shall count as satisfying the chaining constraint.

[12] To derive this, suppose $x \Rightarrow y$ and $x \Rightarrow z$. By monotonicity, $x \cup z \Rightarrow y$; by reflexivity, $y \cup z \Rightarrow y \cup z$, and by monotonicity again, $y \cup (x \cup z) \Rightarrow y \cup z$. Quasi-transitivity applied to this and to $x \cup z \Rightarrow y$ yields $x \cup z \Rightarrow y \cup z$; and applied again with $x \Rightarrow z$, we get $x \Rightarrow y \cup z$.

$$\text{[Accumulation]} \qquad \frac{x \Rightarrow y \quad x \Rightarrow z}{x \Rightarrow y \cup z}$$

So much for what the chaining constraint requires of a given relation which is to play the role of reduction. As it turns out, the relation of scrutability passes that test. It is reflexive—for any member ϕ of any Γ, $\phi \rightarrow \phi$ is a priori—and monotonic—if $\bigwedge \Gamma' \rightarrow \phi$ is a priori and $\Gamma' \subseteq \Gamma$, then $\bigwedge \Gamma \rightarrow \phi$ is a priori. Moreover, it can be shown to be self-adjunctive and transitive, which guarantees its quasi-transitivity.[13]

3 Sentential reduction and the exhaustion constraint

A "brute force" method of establishing a universal reductionist conclusion would consist in going through all terms or sentences, and reduce them to the base, one by one. As I emphasised in Sect. 1, and as should anyway be obvious, the brute force method is not promising. In practice, the realm of what is to be reduced is divided up in different sub-realms.

Traditionally, the sub-realms are different disciplines, like chemistry, biology, psychology, and sociology, or different levels, perhaps corresponding to such disciplines (Oppenheim and Putnam 1958). Chalmers' sub-realms are the truths expressible in certain families of expressions. In chapter 6 of his book, entitled "Hard Cases", there are sections called "Deferential terms", "Names", and "Indexicals and Demonstratives". Many other sections have the form '*X*-truths', where '*X*' indicates the subject matter: mathematical, normative and evaluative, ontological, modal, intentional, social, and metalinguistic. These subject-matters correspond closely to families of expressions.

If truths involving only vocabulary in family *A* have been shown to be scrutable from truths involving *B*-vocabulary, and the latter from *C*-truths, then we can conclude that *A*-truths are scrutable from *C*-truths. No further consideration of the relation between *C*-truths and *A*-truths is needed, since the result follows from the transitivity of scrutability. More generally, there is no problem with chaining on this strategy.

However, chaining only gives us so much. Suppose we have shown that *A*-truths and *B*-truths are both scrutable from *C*-truths. Then by accumulation, it follows that the class consisting of both the *A*-truths and the *B*-truths is scrutable. But neither accumulation nor any of the other chaining principles considered will give us the result that a truth mixing vocabulary from *A* and *B* will also be scrutable. Any binary

[13] To verify that scrutability is self-adjunctive, suppose that Γ is scrutable from Δ, and suppose $\phi \in \Gamma \cup \Delta$. If $\phi \in \Gamma$, then it is scrutable by assumption. If $\phi \in \Delta$, then it is scrutable from Δ since $\phi \rightarrow \phi$ is a priori.

To establish transitivity, suppose that Γ is scrutable from Δ, and Δ from Λ. Pick any $\phi \in \Gamma$. Then there is a class $\Delta' \subseteq \Delta$, $\Delta' = \{\psi_i : i \in I\}$, such that $\bigwedge \Delta' \rightarrow \phi$ is a priori. For every $i \in I$, there is $\Lambda' \subseteq \Lambda$ such that $\bigwedge \Lambda' \rightarrow \psi_i$ is a priori. Fix such a class Λ_i for every i. Then since what is a priori is closed under conjunction and under tautological consequence, $\bigwedge \bigcup_{i \in I} \Lambda_i \rightarrow \phi$ is a priori. Hence Γ is scrutable from Λ.

sentential operator can be used to generate such *cross-family sentences*, and many of them will be true. This reveals a problem with the strategy of partitioning the vocabulary into different families, and then establishing, for each family, that the sentences formed using it are scrutable. Due to the existence of cross-family sentences, that strategy is not suitable to show that *all* truths are scrutable from a candidate base—it fails the exhaustion constraint.

Given that conjunction is the paradigmatic means of forming cross-family sentences, it may appear that no serious problem results. Suppose that sentences ϕ and ψ are formulated in different vocabularies, and that they are both scrutable from Γ. Consider now the conjunction $\phi \wedge \psi$ as an example of a cross-family sentence. Since there are subsets Γ' and Γ'' of Γ such that $\bigwedge \Gamma' \rightarrow \phi$ and $\bigwedge \Gamma'' \rightarrow \psi$ are a priori, $\bigwedge \Gamma' \wedge \bigwedge \Gamma'' \rightarrow \phi \wedge \psi$ is also a priori. So conjunctions of scrutable sentences are themselves scrutable. The same argument, *mutatis mutandis*, applies to disjunctions of scrutable sentences, and to true material conditionals involving them.

However, that kind of argument crucially relies on these connectives linking sentences in a truth-functional way. As I will show in the next section, it does not generalise to sentential connectives that are not truth-functional, nor to certain other logical operators, such as quantifiers. Therefore, the argument does not provide an easy fix for the problem. So even granted that each section of chapter 6 of Chalmers (2012) demonstrates the scrutability of truths involving the relevant families of expressions, and granted that those families jointly exhaust the vocabulary of our language, the scrutability of all truths has not been established: cross-family sentences may have escaped the net.

Still, we might think that the exhaustion constraint can be satisfied if we are more careful with our book-keeping. The problem arises, the thought goes, because the relevant relation holds between classes of sentences. The realm of sentences is infinite, and there does not seem to be any useful way of partitioning it according to the subject matter of the sentences, or the families of expressions occurring in them. But perhaps we can reformulate the reductive project in such a way that it is the class of all atomic expressions, or atomic concepts, that needs to stand in a suitable relation to a base. In contrast to the realm of sentences, that of atomic expressions is arguably finite, and there appear to be natural ways of partitioning it—something philosophers have tried their hand at ever since Aristotle wrote the *Categories*. Indeed, chapter 6 of Chalmers (2012) provides a good example of such a partition. We should not expect to fall afoul of the exhaustion constraint if our relata of reduction are classes of atomic expressions.

As it turns out, there is a relation among classes of expressions that is closely related to scrutability. The relation is structurally similar to the more familiar one of global supervenience, and it may be heuristically useful to think of it in these terms. But since there are competing formal explications of global supervenience that differ from each other in ways that are relevant here, and since the term 'supervenience' has connotations I wish to avoid in this context, I prefer to introduce a new term of art for this relation: 'scriability', derived from 'scry', which is Chalmers' "preferred verb form of 'scrutable'" (p. 30).

The definition of the relation requires some set-up. For a class of expressions C, let $\mathcal{L}(C)$ be the language whose vocabulary consists of C plus logical expressions.[14] I shall use $t(C, w)$ to denote the class of sentences of $\mathcal{L}(C)$ that are true in w ($t(C)$, with no scenario specified, is the class of $\mathcal{L}(C)$-sentences that are actually true, perhaps relative to a designated individual and time).

Definition 2 A class of expressions A is *scriable* from a class of expressions B in w iff $t(A, w)$ is scrutable from $t(B, w)$.

As this definition makes clear, everything that can be said using 'scriable' could be said using 'scrutable', so in the ensuing discussion we will still be concerned with scrutability. I am introducing 'scriable' only to have a convenient way of highlighting certain problems relating to chaining.

It is also true that an important range of scrutability claims can be equivalently formulated in terms of scriability. Suppose that a putative scrutability base Γ consists of all the truths only involving a certain vocabulary B ($\Gamma = t(B)$).[15] Then the claim that Γ is a scrutability base is equivalent to the claim that B is a scriability base.

So we have transformed a claim about the reducibility of all sentences into an equivalent one about the reducibility of all atomic expressions. While the existence of cross-family sentences means that we cannot exhaust the realm of sentences, there is no particular reason to think that we cannot exhaust the realm of atomic expressions. However, cross-family sentences continue to cause trouble, this time with chaining, as I shall now argue.

4 Term reduction and the chaining constraint

Suppose that we frame the reductive project in terms of scriability. The desired conclusion is that a given class B is a scriability base (such that $t(B)$ is a scrutability base). Let us grant that the families of expressions that Chalmers considers in chapter 6 jointly exhaust the total vocabulary; and that each family is scriable from the putative base B. Assuming that mathematical and colour vocabulary are among the families, this means that every sentence formed exclusively from mathematical vocabulary is scrutable from $t(B)$, and likewise every sentence formed exclusively from colour vocabulary.

It is clear that if we wish to infer the desired conclusion from the premises just granted, we shall need to rely on certain chaining principles. These would allow us, for example, to infer that a truth that mixes mathematical and colour vocabulary—a kind of cross-family sentence—is also scriable from B. Just like cross-family sentences raised the exhaustion problem for the project as formulated in terms of scrutability, they seem to prima facie raise the chaining problem for its variant using scriability. The same underlying problem manifests itself in different ways for the

[14] I shall specify the background logic when I introduce cases.

[15] Chalmers' main candidate base, *PQTI*, does not satisfy that condition. See footnote 30.

two strategies. In this section, I shall argue in detail that scriability fails to satisfy the chaining constraint.

Specifically, I shall show that scriability fails to be quasi-transitive, and that it fails to accumulate. To do that, it suffices to establish that it is not self-adjunctive.[16] For scriability is reflexive, and among reflexive relations, quasi-transitivity entails self-adjunction, as does accumulation.[17]

One class of counterexamples to self-adjunction involves cross-family *quantifications*. I shall introduce the problem first in a schematic way, and subsequently suggest a couple of instances.

Consider a scenario w_{ab} with exactly two things a and b, with a being F and b being G; and w_{aa}, which is like $w\ ab$ except that a is both F and G, and b is neither. Thus the two scenarios both contain an F and a G, but they differ on whether the F is the same as the G.

More formally, the following descriptions are true in these scenarios:

$$Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb \wedge \forall z(z = a \vee z = b) \qquad \text{(Scenario } w_{ab})$$
$$Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb \wedge \forall z(z = a \vee z = b) \qquad \text{(Scenario } w_{aa})$$

These descriptions are couched in a language $\mathcal{L}(\{F, G, a, b\})$, with F, G, a, and b as the only non-logical terms, and with the logical resources of first-order predicate logic with identity. We shall leave aside the question whether names are scriable, and focus on how w_{ab} and w_{aa} can be described without them. This move is dialectically permissible, since the present aim is only to show that self-adjunction has false instances, so I am free to specify the languages in question. Moreover, reductionists typically do not consider names, or sentences containing names, to be in the reduction base.[18]

I shall argue that $\{F, G\}$ is not scriable from $\{F\}$ in w_{ab}. For note that the following sentence of $\mathcal{L}(\{F, G\})$ is true in w_{ab} but not in w_{aa}:

(7) $\exists x \exists y(Fx \wedge \neg Gx \wedge \neg Fy \wedge Gy \wedge \forall z(z = x \vee z = y))$

However, there is no sentence of $\mathcal{L}(\{F\})$ that distinguishes the scenarios. All we can say about them with such limited vocabulary is that there are two things, exactly

[16] These points about the structural features of scriability are in some sense analogous to ones that have been made in the context of assessing different explications of the notion of global supervenience. When A and B are classes of predicates, A is scriable from B in all scenarios just in case the class of relations expressed by the members of A *weakly* globally supervenes on the class of relations expressed by the members of B. In a paper defending the usefulness of the concept of weak global supervenience, Sider (1999, p. 917, fn.11) notes that weak global supervenience is not self-adjunctive and also fails a further condition closely related to quasi-transitivity. In Leuenberger (2009, pp. 119–120), I argue that since weak global supervenience does not have the right formal features, it is not a good explication of the intuitive notion of global supervenience.

[17] Suppose $x \Rightarrow y$. Since $\Rightarrow$ is reflexive, $x \cup y \Rightarrow x \cup y$. By quasi-transitivity (setting $z = x \cup y$), $x \Rightarrow x \cup y$.
To show that accumulation and reflexivity entail self-adjunction, suppose $x \Rightarrow y$. By reflexivity, $x \Rightarrow x$, and by accumulation, $x \Rightarrow x \cup y$.

[18] This is true of Chalmers, Jackson, and Lewis, for example.

one of which is F. Hence $t(\{F\}, w_{ab}) = t(\{F\}, w_{aa})$.[19] This allows us to infer the desired conclusion that $\{F, G\}$ is not scriable from $\{F\}$ in w_{ab}. For *reductio*, suppose otherwise. Then for some $\Gamma \subseteq t(\{F\}, w_{ab})$, the material conditional with $\bigwedge\Gamma$ as antecedent and (7) as consequent is a priori. But since $t(\{F\}, w_{ab}) = t(\{F\}, w_{aa})$, $\bigwedge\Gamma$ is true in w_{aa}. We then get the false result that (7) is true in w_{aa}.

Since $\{F, G\}$ is not scriable from $\{F\}$ in w_{ab}, we will have a failure of self-adjunction if $\{G\}$ is scriable from $\{F\}$ in w_{ab}. But that may very well be the case—for some instances, there will be an a priori connection between how things stand F-wise with how things stand G-wise. More formally, what is needed is that the following sentence is a priori:

(8) $\exists x \exists y(Fx \wedge \neg Fy \wedge \forall z(z = x \vee z = y)) \rightarrow \exists x \exists y(Gx \wedge \neg Gy)$

All we can say about w_{ab} in language $\mathcal{L}(\{G\})$ is that there are two things, exactly one of which is G. Assuming that (8) is a priori, this will be a priori implied by its antecedent $\exists x \exists y(Fx \wedge \neg Fy \wedge \forall z(z = x \vee z = y))$, which belongs to $t(\{F\}, w)$. It then follows that $t(\{G\}, w)$ is scrutable from $t(\{F\}, w)$, i.e. that $\{G\}$ is scriable from $\{F\}$ in w_{ab}. This shows that scriability is not self-adjunctive.

This concludes the schematic argument that cross-family quantifications may lead to a violation of the chaining constraint. I shall now consider instances of the schematic letters F and G.

Suppose that the predicates are interpreted as follows:

- F is true of x in w iff x has proton mass in w, and w includes at least one object that has unit negative charge, and one that does not have unit negative charge.
- G is true of x in w iff x has unit negative charge in w.

For these instances, (8) is indeed a priori: the antecedent implies that there is something that has unit negative charge, and something that does not—which is what the consequent says.

We obtain scenarios verifying the relevant sentences by supposing that a is a proton and b an electron in w_{ab}, and that a is an anti-proton—which has the mass of a proton and the (negative) charge of an electron—and b a positron—which has the mass of an electron and the (positive) charge of a proton—in w_{aa}.

Once we see the pattern, we can easily generate further instances. Suppose that as in Fig. 1, a is a black square in w_{ab}, and a black circle in w_{aa}; while b is a grey circle in w_{ab}, and a grey square in w_{aa}.

There are no things apart from a and b. Interpret the predicates thus:

- F is true of x in w iff x is black in w, and w includes at least one circle and one non-circle.
- G is true of x in w iff x is a circle in w.

[19] A rigorous argument for this claim would proceed by induction on the complexity of formulas of $\mathcal{L}(\{F\})$.

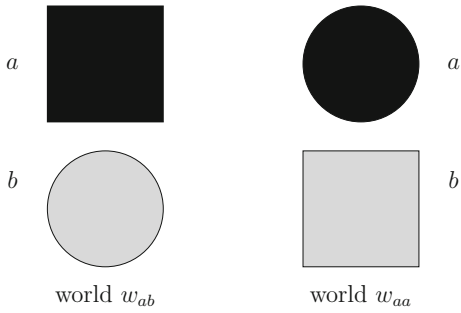

Fig. 1 Scenarios indistinguishable in colour-language and in shape-language, but distinguishable in colour-and-shape language

Again, it will then follow that $\{G\}$ is scriable from $\{F\}$ in w_{ab}, but $\{F, G\}$ is not.[20]

The significance of the counterexample might be disputed on the grounds that the interpretation of the predicate F is contrived, picking out a highly extrinsic conjunctive property. This feature is inessential, however. We could work with less contrived predicates, and compensate by increasing the *syntactic* complexity of the a priori conditionals. In the case of the first interpretation, we could add a binary predicate which applies to x and y in w iff they are oppositely charged in w, for example.

Quantification is not the only means to form cross-family sentences that leads to failures of self-adjunction. I shall consider one further example, a cross-family *counterfactual*, linking sentences p and q that are atomic expressions.

Consider two scenarios in which p and q are both false, but which differ on the counterfactual "if p had been the case, then q would have been the case". Formally:

$$\neg p \wedge \neg q \wedge p \,\Box\!\!\rightarrow q \qquad \text{(Scenario } v\text{)}$$
$$\neg p \wedge \neg q \wedge \neg(p \,\Box\!\!\rightarrow q) \qquad \text{(Scenario } v'\text{)}$$

These sentences belong to the language $\mathcal{L}(\{p, q, \Box\!\!\rightarrow\})$, with p and q as the only atomic sentences, and containing the connectives of propositional logic in addition to $\Box\!\!\rightarrow$. The scenarios cannot be distinguished in that language. For the atomic sentence p is false in both, and counterfactuals such as $p \,\Box\!\!\rightarrow p$ or $\neg p \,\Box\!\!\rightarrow p$ have the same truth-value in both scenarios. With an argument analogous to the one given in the case of cross-family quantification, we can conclude that $\{p, q, \Box\!\!\rightarrow\}$ is not scriable from $\{p, \Box\!\!\rightarrow\}$ in v (nor in v').

Consider now the following:

(9) $\neg p \rightarrow \neg q$

[20] This is structurally the same as the "many-property problem" that Jackson (1977, 64ff) influentially raises against adverbialism. If the arguments of the present paper are correct, that problem afflicts Jackson's own later account of reductive explanation, developed in joint work with Chalmers (Chalmers and Jackson 2001).

If this material conditional is a priori, then $\{q\}$ is scriable from $\{p\}$ in v, and *a fortiori* from $\{p, \Box\!\!\rightarrow\}$. For $\neg q$ a priori implies every member of $t(\{q\}, v)$. Hence $\{q\}$ is scriable from $\{p, \Box\!\!\rightarrow\}$ in v. Since $\{p, q, \Box\!\!\rightarrow\}$ is not, as we have seen above, self-adjunction fails.

It is not hard to find instances where $\neg p \rightarrow \neg q$ is a priori. Let p mean "there is precipitation", and q "it is snowing". Surely it is a priori that if there is no precipitation, it is not snowing. A dry and cold place can serve as our scenario v, and a dry and warm place as our v'.

The examples presented in this section show that scriability fails to satisfy the chaining constraint.[21]

It might be suggested that scriability does not chain in the right way since it is defined in terms of a priori implication, rather than a more robust relation such as ground (Bliss and Trogdon 2016). I shall briefly consider the proposal that we work with ground as our reductive relation.

Let *ground-scrutability* be defined as in Definition 1, except that "Γ' grounds ϕ" replaces "the material conditional $\bigwedge \Gamma' \rightarrow \phi$ is a priori"; and likewise for *ground-scriability* and Definition 2. Then my counterexamples to self-adjunction arguably do not go through: $\exists x Fx$ surely does not ground $\exists x Gx$, where G is "has negative charge", and F is true of x iff has mass and inhabits a world with at least one object that has unit negative charge, and one that does not have unit negative charge.

Still, this modification does not go to the heart of the problem. We are facing the same structural problem when we replace talk of scrutability with talk of ground. Since claims about ground tend to be a great deal more controversial than claims about scrutability, it is a bit harder to find clear violations of the chaining constraint. But here is a fictitious example that might work.

Suppose that there is a relation of marriage*, which is like marriage, except that it can only hold between a woman (the wife*) and a man (the husband*). In scenarios w and v, Andrea and Blake are married* (Mab), and this truth grounds the truth that there is a husband* and a wife* ($\exists x Hx \wedge \exists y Wy$). Specifically, $\{\mathrm{W}\}$ is ground-scriable from $\{a, b, M\}$ in w. Now Andrea is a wife* in w and a husband* in v, and Blake a husband* in w and a wife* in v—gender is a contingent characteristic. It is then plausible that Mab does not ground Wa in w.[22] Hence $\{W, a\}$, and *a fortiori*

[21] We could define a relation scriability* in such a way that self-adjunction is guaranteed: a class of expressions A is *scriable** from a class of expressions B in w iff $t(A \cup B, w)$ is scrutable from $t(B, w)$. It turns out that that relation is also quasi-transitive. However, it is not transitive, and thus still fails to satisfy the chaining constraint. Given the interpretation sketched, $\{q, \Box\!\!\rightarrow\}$ is scriable* from $\{q\}$, and $\{s\}$ from $\{p\}$; yet $\{q, \Box\!\!\rightarrow\}$ is not scriable* from $\{p\}$ in either scenario. The bump in the carpet has only been moved.

Just as scriability is closely related to the notion of weak global supervenience (see note 16), scriability* is closely related to the notion of *intermediate* global supervenience, which Bennett (2004) and Shagrir (2002) introduce as a natural explication of global supervenience. When A and B are classes of predicates, A is scriable* from B in all scenarios just in case the class of relations expressed by the members of A *intermediately* globally supervenes on the class of relations expressed by the members of B. In Leuenberger (2009, 119), I show that intermediate global supervenience is neither monotonic, nor transitive, nor accumulative.

[22] This is a consequence of necessitarianism about grounding, but even those who think that grounds are not *always* necessitating should accept the verdict in this particular case.

$\{W, a, b, M\}$, is not ground-scriable from $\{a,b,M\}$ in w—a violation of self-adjunction, given that {W} is ground-scriable from $\{a, b, M\}$ in w. The move to ground has thus failed to solve the problem, and I will not consider it further in this paper.

5 Analysing arguments that appeal to chaining

We have considered two relations that might be used to construct a reductionist layer-cake. Scrutability, which takes classes of sentences as relata, does not let us divide the realm to be reduced into finitely many blocks that are usefully treated separately. Scriability takes classes of expressions as relata, but lacks the structural features that would legitimise certain forms of inference that we consider natural in such a context.

These relations have so far been examined in an abstract and general manner. In this section, I am returning to the argument from (1) and (2) to (3), concerning the reducibility of the realm of colours. I pointed out in section 1 that we would expect arguments of that form to be valid, and that reductionist projects are in trouble if they are not. With the sharpened tools in hand, I would like to have a closer look at that argument.

We need to disambiguate the generic "reduces to." I shall argue that there is a problem whether we read it as "is scriable from" as well as if we read it as "is scrutable from".

Suppose that we re-write the argument replacing "reduces to" with "is scriable from":

(1′) The chromatic is scriable from the phenomenal and physical together.
(2′) The phenomenal is scriable from the physical.
(3′) The chromatic is scriable from the physical.

Accordingly, we take "the chromatic", "the physical" and "the phenomenal" to be classes of expressions. The validity of this arguments turns on the quasi-transitivity of scriability. As argued in the last section, however, scriability fails to be quasi-transitive, so the argument is invalid.

Suppose now that we replace "reduces" with "is scrutable from":

(1″) The chromatic is scrutable from the phenomenal and physical together.
(2″) The phenomenal is scrutable from the physical.
(3″) The chromatic is scrutable from the physical.

Accordingly, we take the relata to be classes of sentences. However, we are not yet done disambiguating, since it is not immediately clear what class of truths the "phenomenal and physical together" picks out. Premise (1″) has two natural readings: if B is the class of physical and B' the class of phenomenal expressions, "the phenomenal and physical together" can mean either $t(B \cup B')$ or $t(B) \cup t(B')$. On the former reading, (1″) is equivalent to (1′), and the argument as a whole is just a notational variant of the one from (1′) and (2′) to (3′). As such, it will share its fate of being invalid.

On the reading of "the phenomenal and physical together" as $t(B) \cup t(B')$, the argument is valid—scrutability is quasi-transitive, as we have seen. However, premise (1″) is not the plausible claim that (1) was advertised to be.[23] To know what colours things have in a scenario, we need to know not just their physical type and the distribution of phenomenal colour experiences, but also how the two match up. I will illustrate this with two scenarios w and v that are physically alike and yet phenomenally inverted with respect to each other.

In both w and v, there are exactly five things: two observers x and x', two observed things y and y', and a fifth thing z that is neither observed nor observing.[24] The observed y and the unobserved z are of the same physical type F; y' is of the incompatible physical type F'; and the observers x and x' are of the same physical type F''. Object x observes y but not y' ($Oxy \wedge \neg Oxy'$), and x' observes y' but not y ($Ox'y' \wedge \neg Ox'y$). The physical facts in w and v are represented in Fig. 2, with arrows drawn between observers and what they observe.

The two scenarios are indistinguishable in physical language: if $\{F, F', F'', O\}$ is the class B of physical expressions, then $t(B, w) = t(B, v)$.

Phenomenally, w and v are also indistinguishable: in both, there is one observer that has a green experience, and one that has a red experience. Then if C is the class of phenomenal expressions, we have $t(C, w) = t(C, v)$.[25]

However, the observers in the two scenarios are phenomenally inverted relative to each other. In both scenarios, y' is the only F' thing. In scenario w, x', the observer of y', has a red experience, and x, observing an F-thing, has a green experience. In scenario v, x' has a green experience, and x a red experience. Suppose further that in both scenarios, there are law-like, counterfactual-supporting connections between observed physical types and phenomenal properties instantiated in the observer. It is then plausible to say that there is at least one red thing and one green thing in both scenarios: in w, y' is red and y is green, while y is red and y' green in v. But what is the colour of the unobserved z? The natural answer, it seems to me, is that z has whatever colour y has, since they are of the same physical type F.[26] So z is green in w and red in v. If that is correct, "there are two green things" is a chromatic truth of w that is not scrutable from $t(B, w) \cup t(B', w)$.[27]

[23] A similar diagnosis, *mutatis mutandis*, applies to Chalmers' claim that 'All truths are a priori scrutable from F-truths and indexical truths' follows from 'All super-rigid truths are a priori scrutable from F-truths' and 'All truths are a priori scrutable from super-rigid truths and indexical truths', "given the transitivity of a priori scrutability" (Chalmers 2012, 406). (He uses the label 'transitivity' for a stronger condition than I have done in this paper.)

[24] I here take observation to be a physical relation, but this is not essential to the example; we could also take it to be a phenomenal relation, or both physical or phenomenal, or neither.

[25] Neither B nor C includes names of the things in these scenarios. In Sect. 4, I have argued that excluding them is legitimate.

[26] There may well be philosophical accounts of colour which do not vindicate these judgments. My point is merely that premise (1″) is open to objections which would not seem to threaten (1) on the more natural reading.

[27] Such inversion scenarios which spell trouble for (1″) are compatible with the truth of (2″): for all that has been said, $t(C, w)$ may well be scrutable from $t(B, w)$.

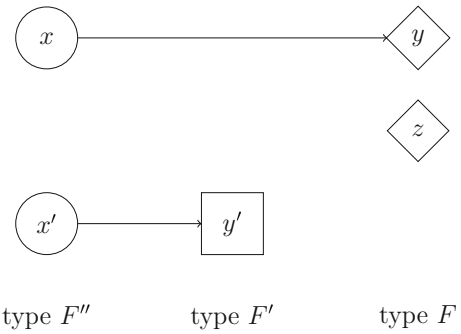

Fig. 2 Physical types of things in *w* and *v*

6 Beyond toy languages?

I have argued that Chalmers' strategy to verify a given class of sentences as a scrutability base is problematic. This is not to say, of course, that the sort of candidate classes he considers, consisting of physical, phenomenal, indexical and totality truths, are not in fact bases. When illustrating the failure of the chaining constraint, I worked with highly impoverished toy languages. This prompts the question whether the considerations of this paper "scale up" to the question whether all truths are scrutable from the truths in a plausible candidate base. It seems to me that they do.

Sentences formulated in both the language of arithmetic and the language of analysis are traditionally taken to be a priori knowable. But now consider the following cross-family sentence, using vocabulary from both languages:

(10) Some natural number is a real number.

While (10) seems pre-theoretically plausible, it is incompatible with the set-theoretic definitions of number systems that mathematics students are taught in their first semester. There is no agreement among philosophers of mathematics about whether (10) is true. This does not show establish that this sentence is unknowable. Nonetheless, I think the contrast of their status with that of sentences of either arithmetic or analysis is suggestive. It must be at least doubtful whether (10) is a priori knowable, or scrutable from a plausible base.

My second potential counterexample involves counterfactual conditionals.[28] Suppose we are given a certain description Γ of the world that includes all truths stateable in physical, neurophysiological, phenomenal, and behavioural terms. It says, among other things, that at a given time Ann's C-fibres are firing (*Ca*), that Ann is in pain (*Pa*), and that Ann's hand jerks back from the stove (*Ja*). I now wish to show that given certain further assumptions, the truth-values of counterfactuals that involve any two of these sentences are scrutable from Γ, but that some that involve all three are not.

[28] My exposition will assume a standard semantics for counterfactuals in the style of Lewis and Stalnaker, but I suspect that the problem does not depend on the correctness of that semantics.

Suppose that Γ also includes sentences a priori equivalent to the following six counterfactuals:

(11) $\neg Pa \,\Box\!\!\rightarrow \neg Ca$
(12) $\neg Pa \,\Box\!\!\rightarrow \neg Ja$
(13) $\neg Ca \,\Box\!\!\rightarrow \neg Pa$
(14) $\neg Ca \,\Box\!\!\rightarrow \neg Ja$
(15) $\neg Ja \,\Box\!\!\rightarrow Ca$
(16) $\neg Ja \,\Box\!\!\rightarrow Pa$

Sentences (11) and (12) tell us what would happen if Ann were not in pain: her C-fibres would not be firing, and her hand would not jerk back from the stove. Similarly, (13) and (14) tell us what would happen if Ann's C-fibres were not firing: she would not be in pain, and again, her hand would not jerk back from the stove. We may take (11) and (13) to reflect a *realisation* relationship between the firing of Ann's C-fibres and Ann's pain. In contrast, (12) and (14) are symptomatic of a *causal* relationship between them and Ann's hand jerking back from the stove. Counterfactuals (15) and (16) are true since the hand movement occurs later than the onset of pain and C-fibre firing, and since the counterfactuals are read in a non-backtracking way.

Further, Γ includes twelve sentences a priori equivalent to negated counterfactuals with a contradictory consequent, of the form $\neg(p \wedge q \,\Box\!\!\rightarrow \bot)$, where Ca or $\neg Ca$ may take the place of p, Ja or $\neg Ja$ the place of q, and Pa or $\neg Pa$ either the place of p or of q (but where G is not occurring both the p-slot or the q-slot). These sentences express the non-vacuity of counterfactual suppositions involving any two of Ca, Pa, and Ja. One such sentence is $\neg(Ca \wedge \neg Pa \,\Box\!\!\rightarrow \bot)$. Given the semantics, it entails that there are worlds where Ann's C-fibres are firing without her being in pain.

Any counterfactual that involves, along with logical vocabulary, only two from among the neurophysiological truth Ca, the phenomenal truth Pa, and the behavioural truth Ja, is thus scrutable from Γ. The same is not in general true, I shall argue, for cross-family counterfactuals that involve all three of these sentences. While one example of inscrutability from Γ would be enough in the present dialectic, it is instructive to consider a pair of them, following Karen Bennett (2008, 2003)[29]:

(17) $Pa \wedge \neg Ca \,\Box\!\!\rightarrow Ja$
(18) $Ca \wedge \neg Pa \,\Box\!\!\rightarrow Ja$

Both these counterfactuals have antecedents that are only true in worlds where the link between C-fibre firing and pain is broken: we have either pain without C-fibre firing, or C-fibre firing without pain. What is their truth-value in a world in which all members of Γ are true?

[29] Bennett discussed these counterfactuals in the context of the causal exclusion problem.

I shall introduce four metaphysical hypotheses about the world that are all a priori compatible with Γ, and yet predict different truth-values for (17) and (18). Three of these views are versions of dualism according to which C-fibre firing is nomically connected to pain in a way that supports counterfactuals such as (11) and (13) (Chalmers 1996).

Consider *overdeterminist dualism* first. If that view is true, [Ann's C-fibres are firing] and [Ann is in pain] are genuine causes of [Ann's hand jerks back from the stove], and each one would be causally sufficient on its own. (I write [*p*] for 'the fact that *p*'.) Hence both (17) and (18) are true. Indeed, Bennett suggests that the truth of these two counterfactuals is diagnostic of overdetermination.

Next, *interactionist dualism.* In a world where that view is true, [Ann is in pain] is the genuine cause of [Ann's hand jerks back from the stove], while *Ca* is only causally relevant to the latter, via its metaphysically contingent nomic relationship to [Ann is in pain]. So once we move to the closest worlds where that latter relationship does not hold, *Ja* co-varies with *Pa*, not with *Ca*. Hence (17) is true and (18) false.

Third, suppose that *epiphenomenalist dualism* holds. Then [Ann's C-fibres are firing] is the genuine cause of [Ann's hand jerks back from the stove], and accordingly, (17) is false and (18) true.

The fourth view is a kind of *physicalism* that takes the connection between C-fibre firing and pain to be contingent in two respects. First, it allows that firing C-fibres do not realise pain if they fail to be appropriately integrated into a nervous system—if their environment is a petri dish, for example (Shoemaker 1981). Second, it allows that there are merely possible worlds where physicalism is false, including worlds where Ann is in pain without having C-fibres that are firing—she might be a disembodied ghost. Such a view is compatible with Γ. Specifically, it is compatible with (12), since it allows that a realised property like pain can be a counterfactual difference-maker (List and Menzies 2009). We can argue that both our cross-family counterfactuals are false, on such a view. Of the two, (18) is easier to evaluate: in some of the the closest worlds where $Ca \wedge \neg Pa$ is true, physicalism still holds, but the C-fibres are wired up very differently from how they actually are in Ann's body. In this different environment, they fail to cause *Ja*, such that (18) is false. The truth-value of (17) depends on what the closest worlds where $Pa \wedge \neg Ca$ is true are like. We can suppose, consistently with Γ and physicalism, that C-fibres are the only possible physical realisation of pain. If so, physicalism is false in any world where $Pa \wedge \neg Ca$ is true. Arguably, some of the closest such worlds are ones where Ann has an epiphenomenal pain, such that (17) is false.

If I am right about these four views, then the counterfactuals (17) and (18) that involve *Ca*, *Pa*, and *Ja* are not scrutable from Γ, even though Γ settles the truth-values of counterfactuals that involve any two of them. We might have thought that if the truth-values of certain simple counterfactuals are settled, those of more complex ones are thereby also determined.[30] But this is not so, as we have seen. If

[30] When Chalmers describes the candidate base he calls *PQTI*, he includes a limited range of counterfactual conditionals in it: "We can allow *P* to include … statements of lawful regularities and counterfactual dependence among microphysical and macrophysical truths. … We can allow *Q* to include

the antecedent of a counterfactual conditional negates a nomic connection between sentences in different families, it will be true only at far-away worlds. So the truth-value of the conditional will depend on remoteness comparisons among those worlds which are irrelevant to evaluate counterfactuals with antecedents from one family only. Those comparisons may in turn depend on metaphysical questions about the actual world.

Any such example raises important points that could be discussed further. But I recommend that we move beyond a mere case-by-case evaluation of potential counterexamples to a scrutability thesis. We need to ask not only whether we have encountered a counterexample, but also how evidentially significant our inability to find a counterexample is. As Eddington famously pointed out, some generalisations seem to be confirmed not because they are true, but because falsifying instances evade our means of detection:

> Let us suppose that an ichthyologist is exploring the life of the ocean. He casts a net into the water and brings up a fishy assortment. Surveying his catch, he proceeds in the usual manner of a scientist to systematise what it reveals. He arrives at two generalisations: (1) No sea-creature is less than two inches long. (2) All sea-creatures have gills. These are both true of his catch, and he assumes tentatively that they will remain true however often he repeats it. (Eddington 1938, 16)

The point is, of course, that the holes of the net are too big to catch smaller fish. We need to ask whether there is any reason to think that on the assumption that there are cross-family sentences falsifying a certain scrutability thesis, we have not been able to recognise them as such. It seems to me that there are two such reasons. First, we are prone to ignore cross-family sentences, or only consider special cases of them. Second, such sentences tend to be complex, and they are often particularly hard to reach stable intuitive verdicts about. Moreover, what verdicts we favour may well turn on where we stand with respect to certain contentious metaphysical questions.

In response to this point, Chalmers might claim that the relevant cross-family sentences are indeterminate in truth-value. At any rate, there is a passage that suggests that this would be his line about (10):

> An exception [to the thesis that inscrutability of reference does not lead to indeterminacy in truth-value] may be quasi-philosophical statements such as 'the number two is a set of sets', and the like. But now the issue is restricted to a few isolated sentences in the metaphysical domain. (Chalmers 2012, 37)

Perhaps it is indeed indeterminate whether some natural number is a real number. However, pleading indeterminacy may be less attractive in connection with (17) and (18), let alone in general. On the face of it, metaphysics, and philosophy more

Footnote 30 continued
any … truths concerning lawful regularities and counterfactual dependence between the phenomenal truths … and microphysical or macrophysical truths" (Chalmers 2012, pp. 110–111). Crucially, *PQTI* does not include counterfactuals with cross-family antecedents such as (17) and (18).

generally, is shot through with cross-family sentences. Both "some minds are brain" and its negation mix physical and mental vocabulary; both "all sentient organisms have rights" and its negation mix biological and normative vocabulary. The relevant class is extensive, and hardly consists only of a few isolated sentences.

Chalmers has not given us reasons to think that such sentences are scrutable from the bases he considers. If they are indeterminate in truth-value, then a class of sentences may be a base even though neither they nor their negations are scrutable from it. If so, the scrutability thesis would not be threatened.[31] But stipulating widespread indeterminacy significantly detract from the interest of his thesis. As Chalmers emphasises, scrutability limits in principle ignorance. But it is no news that knowability claims can be defended by denying that apparently meaningful philosophical claims have a truth-value. Carnap (1928b) and Ayer (2001) have taught us that much. Chalmers would hardly be more rationalist than Carnap.

It is a natural thought that Chalmers could avoid joining Carnap in the anti-rationalist ranks by instead taking a leaf from him with respect to another issue that separates them—the role of definitions. Carnap's idea was that all terms are to be defined using only terms in the base. That strategy does not face either of the problems I mentioned. The number of terms, as opposed to the number of sentences, is finite, so the realm to be reduced can be usefully sub-divided. But unlike scriability, the relation that holds between classes of terms if every member of the former is definable in terms of the latter does satisfy the chaining constraint.

Given plausible assumptions, the claim that every term is definable using terms in B entails the claim that B is a scriability base.[32] So we should certainly not give more credence to the claim that there is a compact definability base than to the claim that there is a compact scrutability base. But it is a well-known phenomenon that sometimes, the only practically feasible way to establish a thesis is to first establish a logically stronger thesis, and then deduce the former as a corollary. This might be the situation with Chalmers' scrutability theses.

So far, my discussion suggests a return to a Carnapian definitional project. Still, I wish to conclude with two reasons to be cautious.

First, definitions raise their own technical questions. They are typically conceived as quantified biconditionals, perhaps with a certain modal force. Moreover, they are standardly taken to license substitution in every context. It is this feature which ensures that the definitional account satisfies the chaining constraint. However, the relevant quantified biconditionals may not be available even in infinitary languages, unless there is an upper bound on the size of scenarios.[33] Even if they are available, they may not license substitution in every context unless a certain global assumption

[31] At least there would not be a direct threat. There may still be an indirect one. Unlike in the case of vagueness, there would be little reason to think that there was higher-order indeterminacy. If there is not, claims of the form "indet p" will be determinately true (Chalmers 2012, pp. 33–34). But it is not clear whether those indeterminacy claims are a priori knowable.

[32] I will not spell out the notion of definition formally here, or discuss these assumptions.

[33] Pertinent issues are discussed in Leuenberger (2018).

about the language holds. In Carnap's case, that assumption was extensionality. But these assumptions are themselves very problematic, and it is not clear how they can be justified.[34]

Second, definitions raise broader philosophical questions. In the introductory section, I suggested that rationalism differs from empiricism by interpreting truths of metaphysics and mathematics in a non-deflationary way. I also noted that such definitional reduction suggests a deflationary reading of the relevant sentences. It thus remains to be seen whether a successful universal reductionist project can remain in the spirit of rationalism.

7 Conclusion

The problem of cross-family sentences that I have raised prompts us to pay close attention to the question of what the relata of reduction are. Are they such things as expressions, concepts or properties, or rather things like sentences, propositions or facts? The latter form an infinite and in some sense open-ended domain, making it difficult to argue for a conclusion about everything that belongs to it. On this horn of the dilemma, what I have called "the exhaustion constraint" fails to be satisfied. The domain of expressions, in contrast, is arguably finite, and we can hope to exhaust it by tackling its members one by one. However, the problem of cross-family sentences shows that exhausting the domain of expressions is not in general enough to exhaust the domain of sentences made up from those expressions. It suffices only if certain chaining and accumulation principles hold for the relation of reduction. But the relevant relation between classes of expressions may well fail to satisfy the chaining constraint, leading to the second horn of my dilemma for someone arguing for universal reductionism.

While our candidate bases may change over time, universal reductionism itself will not lose its allure. Philosophers of a certain intellectual temperament will keep trying to vindicate versions of the view. My aim here has been to articulate design constraints for their arguments.

Acknowledgements Thanks to audiences in London, Aberdeen, Florence, Geneva, and Glasgow. I am particularly grateful to my commentators on two of these occasions, Jonathan Simon and Robert Michels, as well as to Philipp Blum, David Chalmers, Stephan Krämer, Gabriel Rabin, Bruno Whittle, and to anonymous referees. This work was supported by the Arts and Humanities Research Council [grant number AH/M009610/1].

[34] In the preface to the second edition, Carnap expresses dissatisfaction with his attempt at justifying extensionality in $$ 43–45 of Carnap (1928a).

References

Ayer, A. J. (2001). *Language, truth and logic*. London: Penguin.
Bennett, K. (2003). Why the exclusion problem is intractable, and how, just maybe, to tract it. *Noûs*, *37*, 471–497.
Bennett, K. (2004). Global supervenience and dependence. *Philosophy and Phenomenological Research*, *68*(3), 501–529.
Bennett, K. (2008). Exclusion again. In J. Hohwy & J. Kallestrup (Eds.), *Being reduced* (pp. 280–305). Oxford: Oxford University Press.
Bliss, R., & Trogdon, K. (2016). Metaphysical grounding. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (Winter 2016 ed.). Stanford: Stanford University.
Carnap, R. (1928a). *Der logische Aufbau der Welt*. Leipzig: Felix Meiner Verlag. Translated into English as *The logical structure of the world*. Cambridge University Press, 1967.
Carnap, R. (1928b). *Scheinprobleme in der Philosophie* (pp. 301–343). Berlin: Benary. In translation of Carnap (1928a).
Chalmers, D. J. (1996). *The conscious mind*. Oxford: Oxford University Press.
Chalmers, D. J. (2012). *Constructing the world*. Oxford: Oxford University Press.
Chalmers, D. J., & Jackson, F. (2001). Conceptual analysis and reductive explanation. *The Philosophical Review*, *110*, 315–361.
Eddington, A. (1938). *The philosophy of physical science*. Cambridge: Cambridge University Press.
Jackson, F. (1998). *From metaphysics to ethics*. Oxford: Oxford University Press.
Jackson, F. C. (1977). *Perception. A representative theory*. Cambridge: Cambridge University Press.
Leuenberger, S. (2009). What is global supervenience? *Synthese*, *170*, 115–129.
Leuenberger, S. (2018). Global supervenience without reducibility. *The Journal of Philosophy*, *115*, 389–422.
Lewis, D. (1986). Introduction. In *Philosophical papers* (Vol. 2). Oxford: Oxford University Press.
List, C., & Menzies, P. (2009). Non-reductive physicalism and the limits of the exclusion principle. *The Journal of Philosophy*, *106*, 475–502.
Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Minnesota Studies in the Philosophy of Science* (pp. 3–36). Minneapolis: University of Minnesota Press.
Shagrir, O. (2002). Global supervenience, coincident entities and anti-individualism. *Philosophical Studies*, *109*, 171–196.
Shoemaker, S. (1981). Varieties of functionalism. *Philosophical Topics*, *12*, 93–119.
Sider, T. (1999). Global supervenience and identity across times and worlds. *Philosophy and Phenomenological Research*, *59*(59), 913–937.
Stalnaker, R. C. (1996). Varieties of supervenience. *Philosophical Perspectives*, *10*, 221–241. Reprinted, with new appendices, in Stalnaker 2003, pp. 86–108.
Stalnaker, R. C. (2003). *Ways a world might be. Metaphysical and anti-metaphysical essays*. Oxford: Oxford University Press.
Williamson, T. (2001). Ethics, supervenience and ramsey sentences. *Philosophy and Phenomenological Research*, *62*, 625–30.