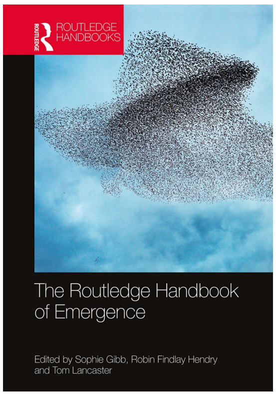

The Routledge Handbook of Emergence

Sophie Gibb, Robin Findlay Hendry, Tom Lancaster

Reduction

Publication details
https://www.routledgehandbooks.com/doi/10.4324/9781315675213-5
John Bickle
Published online on: 22 Mar 2019

4
REDUCTION

John Bickle

Why include an entry on reduction in a handbook on emergence? Historically, reduction has been the antithesis of emergence. Their antagonism was noted by Ernest Nagel in his classic chapter on theory reduction (1961) (much more on which below), which ends with a detailed discussion of emergence. We can often learn interesting things about a concept by investigating its contradictory.

I will restrict my discussion to reduction as characterized in philosophy of science. Even so restricted, "reduction" is vastly equivocal. Nevertheless, all varieties share "nothing-but"-ism: a reduction shows that the reduced kind (whatever kind it might be) is thereby "nothing but" the reducing; no reduced content is left out or over. This is the feature that sets reduction in opposition to emergence.

Term/sentence reduction

Term or sentence reductionism was popular through the mid-20th century. It was one of two "dogmas" Quine (1951) urged fellow empiricists to abandon. Its brief popularity in philosophy of mind rested on attempts to translate psychological terms or sentences into "topic-neutral" synonyms, which in turn were hypothesized to be contingently identifiable with physical terms or sentences: those of future neuroscience, according to central state materialists (Smart 1959). The proposed definability of psychological expressions into "topic-neutral" synonyms, for example, "something is going on in me which . . .," and the predicted future contingent identities of those "somethings" with physical (e.g., brain) goings-on, captured reductionism's "nothing-but"-ism.

This approach to reduction has zero popularity now. Yet it is worth dwelling on here briefly, because the most philosophically sophisticated advocate of term/sentence reduction, Rudolph Carnap, held a much more nuanced view, as far back as the mid-1930s, than is typically realized. Logical positivism is widely acknowledged, and roundly criticized, for advocating a definitional reductionism based on a "verificationist" theory of meaning. Carnap (1932) held such a view. However, by his 1936 work, he realized the limits of verificationism. There he shifted explicitly to confirmation, a much weaker notion, adopted an explicitly pragmatic interpretation of observation languages for science . . . and then gave up on definability. Deeply theoretical terms from physics weren't the only problem; simple disposition terms like "soluable" "cannot be defined by means of the terms by which conditions and reactions are described" (1936, 440). Reduction by definability, for the most prosaic of sciences, had to go.

Instead of being defined by observation terms, a new scientific term, say, Q_3, can be introduced initially by a "reduction pair":

(R1) $Q_1 \supset (Q_2 \supset Q_3)$
(R2) $Q_4 \supset (Q_5 \supset {\sim}Q_3)$,

where Q_1 and Q_4 describe experimental conditions, and Q_2 and Q_5 experimental results (1936. 441). A single reduction pair (typically) does not establish the "complete" meaning of a new term. It only does so for cases where one of the test conditions, $Q_1 \vee Q_4$, is fulfilled; for all other cases Q_3 remains indeterminate, "meaningless" in Carnap's (1936) parlance. But scientists can further diminish this "region of indeterminateness" by adding additional reductive pairs, to specify additional conditions and related experimental results that determine whether Q_3 or ${\sim}Q_3$ applies. (The most straightforward example is a term for a physical magnitude that can be determined by different methods.) The term will then have meaning for the disjunction of all the experimental conditions in its set of reduction pairs, and this set expands with each new "decision" about its usage (1936, 448–449).

Carnap's target in his 1936 work are those who insist on reduction via definability. The problem is that "a definition determines the meaning of a new term once and for all," a practice "not in accordance with the intentions of the scientist concerning the use of the predicate" (1936, 449). For the indeterminate cases not (yet) governed by a reduction pair, Carnap insists, "the scientist wishes to leave these questions open until the results of further investigations suggest the statement of a new reduction pair" (1936, 449). Reduction pairs thus constitute a "partial definition of meaning only and can therefore not be replaced by a definition" (1936, 449–450). This shift raises a fascinating puzzle for even the brief history of analytic philosophy. W.V.O. Quine at least acknowledges that Carnap had "long since" given up on the "radical reductionism" of term/sentence definability, but then famously goes on in the remainder of that essay to level his "Duhemian holism" challenge to "dogmatic" empiricists, Carnap included. Yet seventeen years prior Carnap had written: "Science is a system of statements based on direct experience, and controlled by experimental verification. Verification in science is not, however, of single statements but of the entire system or subsystem of such statements" (1934). It is difficult to find a more straightforward statement of "Duhemian holism." It is also worth noting that Carnap's appeals in his 1936 work to "scientists' intentions" suggest that he had given up on "rational reconstructions" as early as then. It can be baffling how credit for influential ideas gets distorted over even short histories in philosophy.

Aside from the historical interest just noted, why is all this worth repeating? Because Carnap's nuanced account of term introduction via sets of reduction pairs may not be so widely implausible about actual scientific practice than logical positivism is so often accused of being. Morton and Bickle (2005) argue that the scientific term, "memory consolidation," looks to have been introduced and developed in present-day molecular neurobiology in a fashion resembling Carnap's reduction pairs.

Theory reduction

We turn next to the most prominent account of the reduction relation, theory reduction. The classic account is Ernest Nagel's (1961, chapter 11); Schaffner (2012) provides the best recent overview of Nagel's account, its numerous proposed revisions and replacements, and its remarkable resiliency. For Nagel, reduction was *deduction*: of the sentences constituting the reduced theory, from those of the reducing serving as premises. Logical derivability captured reductionism's "nothing-but"-ism: no content of the reduced theory remains beyond the scope of the reducing.

In many actual scientific cases, the reducing theories correct the reduced. Real planets do not travel in ellipses; real bodies do not fall with uniform vertical acceleration over any finite interval near the surface of the earth. Kepler's theory of planetary motion and Galileo's mechanics are false vis-à-vis Newton's mechanics, yet these are textbook scientific reductions. How can a false theory validly be derived from a true one (presumably, at the time of the reduction)? Nagel captured this feature by adding various limiting assumptions and boundary conditions on the applicability of the reducing theory, possibly counterfactual, to the premises of the derivation. He also noted that many scientific reductions are "heterogeneous": the reduced theory contains terms not part of the descriptive vocabulary of the reducing. "Heat" and "pressure" of classical equilibrium thermodynamics, for example, nowhere occur in statistical mechanics and the kinetic/corpuscular theory of matter. For Nagel, various "conditions of connectability" must also be included among the premises of the derivation. These "bridge laws" or "correspondence rules," as they soon came to be called, had to have at least the logical strength of conditionals, with sentences containing terms of the reducing theory as antecedents, and others containing terms of the reduced theory which did not occur in the reducing as consequents. Unfortunately, in his famous illustration, the reduction of the ideal gas laws of classical thermodynamics to statistical mechanics, these connections are stated as biconditionals. This perhaps rationalizes the otherwise puzzling fact that biconditional conditions of cross-theoretic connectability were so often assumed to be required on Nagel's account. He explicitly states that only conditionals are required, and the logic of first-order derivability clearly only requires conditionals. This misunderstanding misinformed some anti-reductionist "multiple realization" arguments in philosophy of mind for decades. If "[brain sentence]⊃[psychological sentence]" is all the cross-theory connectability that a Nagel reduction (deduction) of psychology via neuroscience requires, it matters not at all that [psychological sentence] is true of other creatures lacking brains. Nagel discussed some difficulties interpreting these conditions as either conventional definitions or synthetic statements. Their status remained a constant source of discussion within the Nagel-inspired reduction literature.

Nagel's work (1961) was not the first account of intertheoretic reduction to appear. Schaffner (2012) provides a helpful history of Nagel's account, tracing it back a dozen years prior to 1961. Other philosophers were working on intertheoretic reduction, too. John Kemeny and Paul Oppenheim (1956) offered a logically weaker account, in which the reducing theory explains all the "observational data" that the reduced theory explains, and typically more. The reducing theory's increased explanatory scope captures reductionism's "nothing-but"-ism; the reduced theory thereby plays no ineliminable explanatory role. Additionally, for Kemeny and Oppenheim, the reducing theory must be at least as "well systematized" as the reduced. Systematization balances simplicity and explanatory power. It is a measure of how well the increased complexity of the reducing theory is compensated for by its increased explanatory power. A simpler theory thus can be reduced to a more complex one if the more complex theory is significantly stronger in explanatory power by enough to offset its increased complexity. Unfortunately, Kemeny and Oppenheim leave systematicity intuitive. Their account was quickly rejected as being logically too weak: presumably, it counts as reductions cases which are not. Despite this popular challenge, Theruer and Bickle (2013) argue that some apparent reductions in the recent neurobiological field of "molecular and cellular cognition" appear to follow a modified Kemeny–Oppenheim pattern.

Patrick Suppes (1967) suggested an isomorphic "sameness of structure" condition on intertheoretic reduction within a broader framework which conceived of theories as ordered sets of models. Building explicitly on Suppes' suggestion, Balzer, Moulines, and Sneed (1987) specified additional set-theoretic conditions, including analogues of Nagel's connectability and derivability conditions, and "ontological reduction links" between components of reducing and reduced models, such that every "confirmed empirical application" of the reduced theory is related to some model of the reducing. Logically stronger than mere isomorphism, these conditions expressed a model-theoretic analog of derivability, and thereby captured reductionism's characteristic "nothing-but" feature. With great formal ingenuity, Balzer, Moulines, and Sneed (1987, chapter VI) reconstruct historical cases of scientific reductions using their set-theoretic apparatus.

Paul Feyerabend (1962) leveled one of the first responses to Nagel's work (1961). Denying that any scientific examples actually meet Nagel's connectability and derivability conditions, Feyerabend proposed a radical "ontological replacement" account of reduction: the ontology of the reducing theory replaces without remainder that of its "incommensurable" reduced analog in all contexts of scientific usage, including the observational. Most philosophers of science rejected Feyerabend's radical alternative, but both Kenneth Schaffner (1967, 1992) and Clifford Hooker (1981) saw reason to accommodate some of his views. For Schaffner, reduction remains deduction, with the reducing theory serving as premises, but not of the reduced theory, but rather, of a *corrected version* of the reduced theory. The deduced structure is still in the theoretical language of the reduced theory (perhaps altered by corrections necessitated by the reducing), and so something akin to Nagelian "bridge laws" remain required in heterogeneous reductions. Schaffer dubbed these "reduction functions." His "General Reduction Paradigm" (later renamed the "General Reduction-Replacement Paradigm") also insisted that an "analog relation" held between the actual reduced theory and the corrected version actually deduced in the reduction. He never specified this analog relation in any formal detail. Over the years Schaffner has increasingly characterized his account as more closely aligned with Nagel's (see most recently Schaffner (2012)).

Hooker (1981) accommodated some Feyerabendian features also while still keeping reduction as deduction, but replacing the actual reduced theory with its *image*, already formulated within the descriptive vocabulary of the reducing theory. The constructed image is *explanatorily equipotent* to the actual reduced theory, with the strength of logical derivation and this explanatory equipotence capturing reductionism's "nothing-but"-ism. In cases of reduction which correct, and perhaps even falsify significantly, the actual reduced theory, various counterfactual limiting assumptions and boundary conditions must also be added as premises from which this image is derived. But since the derived image is already within the vocabulary of the reducing theory, no "conditions of connectability" or "reduction functions" are required to affect its derivation. The number and counterfactual extent of the limiting assumptions and boundary conditions needed to derive the explanatory equipotent image of the reduced theory suggests a measure of the relative "smoothness" of a given intertheoretic reduction, although Hooker admits that he leaves this notion in a frustratingly intuitive state (1981, 223–224). Still, working with a number of actual scientific examples displaying varying amounts of Feyerbendian "incommensurability," Hooker suggests that a given case's (intuitive) location on the "smoothness/bumpiness of intertheoretic reduction" spectrum nicely matches its location on the "ontological retention/replacement" spectrum for the reduced theory's postulated kinds. So we learn something about cross-theoretic ontology – identity, to revision, to outright elimination – when we investigate a given intertheoretic reduction for its relative "smoothness."

Hooker's account influenced Paul and Patricia Churchland's neuroscience-based eliminative materialism (P.M. Churchland 1985; P.S. Churchland 1986). Bickle (1998) exploited Hooker's insights within a "semantic" (model-theoretic) account of theory structure, borrowing resources from Balzer, Moulines, and Sneed (1987). He offers a detailed model of the isomorphic location of scientific cases across the intertheretic reduction and ontological consequences spectra; develops a detailed Hooker-inspired account of the reduction of the classical gas laws of equilibrium thermodynamics to statistical mechanics and the kinetic theory of gases, including the derivation of an image of the gas laws within the reducing theory; develops a quasi-formal account of "smoothness of reduction" (based on set cardinality); develops in detail a response to the multiple realizability challenge to reduction inspired by Hooker's remarks in his work (1981, Part III); and argues for a revisionary ontological outcome for the propositional attitudes of folk psychology vis-à-vis a then-current model of synaptic plasticity in cellular neurobiology. Endicott (1998) subjects Hooker's, the Churchlands', and Bickle's argument to searching evaluation and criticism. Schaffner (2012) recently concurs with Endicott's criticism.

Well into the 1990s, "reduction" in serious philosophy typically meant intertheretic reduction, and that usually meant Nagel's account. But over the past twenty years some new accounts have developed, and we close with some of those.

Functional reduction

One alternative to intertheoretic reduction emerged from consciousness studies: functional reduction. Joseph Levine (1993) was an early advocate of an "explanatory gap" between the subjective qualitative features of conscious experiences, that "what-it-is-like-ness" (Thomas Nagel 1974) of, for example, the vivid redness of the visual experience of a polished fire truck or the taste of a ripe peach, and the physical features of the object experienced and the conscious experiencer. Levine (1993) defended this gap by way of an account of "reductive explanation," one quickly adopted by fellow "qualiaphiles" such as David Chalmers (1996) and even Jaegwon Kim (2005). Reduction, according to Levine, is a two-stage process. "Stage 1 involves the (relatively? quasi?) *a priori* process of working the concept of the property to be reduced 'into shape' for reduction by identifying the causal role for which we are seeking the underlying mechanisms" (1993, 132). This is to "functionalize" the concept – hence the name of this account. "Stage 2 involves the empirical work of discovering just what those underlying mechanisms are" (1993, 132), that is, doing the science to discover which kinds in the natural world play, or at least approximate, that functional role.

Levine fleshes out this skeleton with an example: the reduction of water to collections of H_2O molecules in liquid state. "Our very concept of water is of a substance that plays such-and-such a causal role" (1993, 131): for example, boils at 212°F at sea level. Empirical-cum-theoretical investigations of collections of H_2O molecules, including their varying speeds and locations near the liquid's surface, intermolecular attractive forces between them, and their increasing kinetic energy with increased heat and its relation to atmospheric pressure, explain the 212°F boiling point. Those are the empirically discovered features of the world that realize this aspect of water's full functional profile. That the behavior of the physical aggregates matches these causes and effects "without remainder" captures reductionism's "nothing-but"-ism. If there is "nothing more" to being water than its possessing that complete set of causes and effects, and if the empirically discovered behavior of collections of H_2O molecules in liquid state realize that full functional profile, then water is "nothing but" H_2O molecules in liquid state.

With so much serious work on intertheretic reduction already in existence, why would qualiaphiles bother to articulate another account? Conveniently, at least for the anti-reductionists among them, adopting functional reduction yielded a straightforward argument for the *ir*reducibility of qualia. A central conclusion of qualiaphiles is that qualia cannot be functionalized: cannot be "fully elaborated without remainder" into a profile of causes and effects. Such treatments inevitably "leave out" their qualitative features, the "what-it-is-like-ness" of these conscious experiences, as thought experiments involving inverted spectra, absent qualia, what Mary the future neuroscientist doesn't know, and philosophical zombies purport to show. Chalmers (1996) is a key source for many of these famous puzzles, but the entire consciousness studies literature wallows in them and their seemingly endless variations. But according to functional reduction, if no functionalization of a concept is possible . . . then neither is any (functional) reduction of it! If Stage 1 of a reduction is a nonstarter, then no Stage 2 empirical work will find the mechanisms which exhaust the functional profile, which exhausts that concept's content – because no functional profile does. Viola! Qualia are irreducible to physical mechanisms!

The problem with functional reduction, at least in philosophy of science, is that it is as much a philosopher's fiction as the famous thought experiments purporting to show that qualia can't be functionalized. Levine's (1993) "scientific" example, which Chalmers (1996), Kim (2005), and others adopt, draws neither from contemporary science nor its recent history – as advocates of intertheoretic reduction attempted. The case study comes instead from elementary-school science education; it's an example we use to instruct children about our basic scientific world view. One may be excused for questioning whether real reductionism in real scientific practice circa now resembles an analysis derived from cartoon "science."

This counter to functional reduction (first offered in Bickle 2012) amounts to a challenge: take the Methods section from a published experimental report appearing in any reputable current science journal – *Cell, Science, Nature, Proceedings of the National Academy of Sciences* – and show how the experiments described there can be interpreted reasonably as "functionalizing" some concept, even partially. Help yourself to the Supplementary Materials now routinely published online with all major scientific publications. Prediction: you will not succeed. That is because "functional reduction" is no part of actual current scientific practice.

Explanatory, metascientific, and mechanistic reduction

Reduction in philosophy of science has experienced a recent renaissance. This final section documents just three examples. Michael Strevens' "explanatory reduction" is particularly apt to include in a volume on emergence. He aims to provide an account that is genuinely reductionist, but that also accommodates concerns that have motivated anti-reductionists, including emergentists. He builds his account explicitly on a doctrine of ontological physicalism: everything that exists is made up entirely of physical stuff, and everything that occurs does so because of the effects of physical laws. From this he concludes that all events, no matter how high level (above that of fundamental physics) or abstract, can be derived from fundamental physical facts and laws. Some of these derivations, though not all and in fact very few, are also explanations: namely, those successful derivations of the explanandum limited exclusively to the components of the causal network that "make a difference" to its occurrence. This is Strevens' broader "kairetic" account of explanation, detailed most extensively in his 2011 work; its name is drawn from the Greek "kairos," meaning "decisive point." Strevens even provides a "recipe" for finding these reductive explanations, via a process of abstracting away from the details of a "complete" specification of the causal network generating the explanandum until a derivation of the event remains, but no further abstracting away of causal details will maintain that logical relation. Explanatory reductionism is then one further conjunct: for any higher-level phenomenon, there is at least one derivation of it from the most fundamental physical level that is (kairetic) explanatory, and that one is its best explanation. The general reductionist appeal to derivation is familiar: it was the logical relation Nagel (1961) required between reduced and reducing theories. Here, however, "derivation" is part of the apparatus for representing the operation of a causal process, not merely a relation between theories qua syntactic structures. It is also the key relation in Strevens' account that captures reductionism's "nothing-but"-ism.

How then does Strevens save anti-reductionist motivations, like the methodological autonomy of the special sciences? In his 2016 work responding to Phillip Kitcher's famous appeal to the autonomy of classical genetics from molecular biology, Strevens distinguishes between two senses of explanatory irrelevance. "Objectively" irrelevant causal details make no difference to the occurrence of the explanandum. "Contextually" irrelevant causal details can be, and often are, objectively relevant to a complete explanation of it. But for practical purposes of pursuing particular scientific endeavors, these details can be, and often must be, ignored: "black boxed," as Strevens puts it, for the efficient special-scientific investigation into the explanandum to proceed. This scientific division of labor can be either "compartmental," at a single level of investigation, or "stratified," across levels. And while kairetic explanatory reductionism requires that eventually all the objectively relevant difference-making causal details must get included in our best (fundamental physical) explanation, the scientific practices involved in achieving such explanations recognize a demand for the contextual autonomy of special-scientific pursuits.

Strevens (2016) remains uncommitted about "whether it is worth our explanatory while" actually to trace particular lines of implementation down to their fundamental physics (karietic) explanation. The ensuing reductive explanation will be better than any which stops partway down, but practically speaking achieving it may be too expensive . . . or too boring. These practicalities depend on the specific case at hand. But what about entire "reductionistic" scientific endeavors that seek explanations of phenomena down many recognized levels at once, in a "single bound"? The autonomy Strevens (2016) seeks to salvage within his reductionism might be no part of such scientific endeavors; these sciences might violate even his practical norms of "contextual irrelevance." His shift to context and pragmatics affords him wiggle room; perhaps "compartmentalizing" norms typically operative in single-level special sciences get suspended in these "single-bound" endeavors.[1] It is worth reflecting on such cases, however, because a different sense of reduction has been built directly upon them.

If our overarching goal in investigating reduction is to understand a prominent (yet hardly universal) feature of scientific practice, that goal suggests a strategy. Find some field that both its practitioners and scientists working in related fields label "reductionist," and investigate practices – experimental, explanatory – specific to it. Conduct this analysis as bereft of philosophical/epistemological assumptions about "what reduction has to do or be" as one can render oneself. The result should be a "metascientific" account of actual scientific reductionism in actual scientific practice. With such an account in hand, one can follow up investigating whether other scientific fields routinely labeled "reductionistic" employ similar practices and whether the account delivers what philosophers have wanted (or targeted) in "an account of reduction."

John Bickle (2006, 2009, 2012) has sought such an account. The first step is choosing a scientific field to investigate. Bickle chooses "molecular and cellular cognition" (MCC), a field that raises the worry suggested earlier for Strevens' account. MCC started in the early 1990s with the application of gene targeting techniques from developmental molecular biology to neurobiology and has expanded to encompass more recent developments such as optogenetics (Bickle 2016). The Molecular and Cellular Cognition Society (www.molcellcog.org) now has more than 2300 members worldwide. Its proponents contrast their approaches and goals explicitly with those of "cognitive neuroscience" in their focus on cellular and molecular mechanisms of cognitive functions and their use of animal models (http://molcellcog.org/index.asp?page=about). But MCC is hardly some fringe field in current neuroscience. Its centrality to the discipline for nearly two decades is apparent in a quote from the introductory chapter to one of the principal textbooks in the field, the edition from fifteen years ago (since supplanted by a fifth edition):

> This book . . . describes how neural science is attempting to link molecules to mind – how proteins responsible for the activities of individual nerve cells are related to the complexity of neural processes. Today it is possible to link the molecular dynamics of

> individual nerve cells to representations of perceptual and motor acts in the brain and to relate these internal mechanisms to observable behavior.
>
> *(Kandel, Schwartz, and Jessel 2000, 3–4; my emphases)*

These "links" are nothing less than "single-bound" mind-to-molecular pathway reductions, introduced fifteen years ago as "textbook" neuroscience. Learning and memory, including "declarative" or "explicit" forms, constitute MCC's most notable achievements to date (Sweatt 2009; Silva, Landreth, and Bickle 2014).

What does a metascientific analysis of landmark MCC results reveal actual scientific reductionism to be? First, it is part of the more general search for causal connections between neuroscientific kinds. Three kinds of experiments are crucial to establish any causal hypothesis A → B with a strong degree of scientific confidence. *Negative interventions* manipulate the hypothesized cause (A) by decreasing its probability or intensity and measure changes in the hypothesized effect (B). For hypothesized excitatory causes, under widely accepted conditions of experimental control, a decrease in the hypothesized effect shows that the hypothesized cause is necessary for the effect. (For hypothesized inhibitory causes, the change to the effect will be exactly the opposite, i.e., an increase. Similarly for positive intervention experiments, to follow.) But even the most scrupulously controlled negative intervention experiments can never show the sufficiency of the hypothesized cause. *Positive interventions*, especially ones integrated with successful negative interventions, can. Here experimenters manipulate the hypothesized cause (A) by increasing its probability or intensity and measure changes to the hypothesized effect (B). Increases in the effect show the sufficiency of the hypothesized cause. Now, however, the problem of experimental artifact looms, since in most positive interventions the hypothesized cause is increased beyond its normal biological limits. The integration of results from a third kind of experiment is required: *non-intervention experiments*, in which the hypothesized cause (A) is not manipulated experimentally, but merely measured in correlation with the occurrence of the effect (B) in as biologically realistic circumstances that permit the required measurements. For any specific causal hypothesis, scientific confidence that it has been established experimentally increases with the successful integration of results from all three types of experiments.[2]

So far this metascientific analysis of landmark MCC experimental results does not distinguish the field from causal-mechanistic sciences more broadly, not all of which are reductionist. What additional features makes MCC reductionist? The first concerns particular experimental practices. Specific cognitive functions are operationalized into behavioral protocols and measures for use with animal models, typically rodents. The molecular-biological and gene targeting experimental tools then routinely employed to manipulate the hypothesized causes manipulate components of the biological system generating the operationalized behavioral measures – and most often components of the behaving system's components (of its components): specific proteins involved in intra- and inter-cellular signaling pathways. These are the hypothesized mechanisms of the experimentally operationalized cognitive functions under direct experimental test. The second reductionist feature is another form of multiexperiment integration, mediation analysis (Silva, Landreth, and Bickle 2014). Even when a neurobiological causal hypothesis has been established by the integration of all three negative manipulation, positive manipulation, and nonintervention experiments, MCC investigations do not stop. In particular, MCC scientists next investigate causes mediating the established connection A → B: those causes by which A causes B. To do this, MCC scientists inevitably "look down" into the components that constitute A and B. This is what their experimental tools of choice, drawn from molecular biology and increasingly biochemistry, permit. The now quarter-century track record of this approach far exceeds anything that less reductionistic fields of science can offer, at least for the causal mechanisms of learning and memory.

Or so holds metascientific, "ruthless" reductionism. The underlying picture is straightforward, even if the MCC scientific details can be daunting. Real reduction in real scientific practice is a matter of intervening experimentally into increasingly lower levels of biological organization, from specific neurons now down to specific configured proteins in intra- and inter-neuronal signaling pathways, then tracking the effects of these interventions on the system's behavior in experimental protocols well accepted as indicators of specific cognitive functions. When these experiments are individually successful *and* successfully integrated, a reduction is claimed: of the specific cognitive function directly to the lowest level of biological organization for which the full set of integrated interventionist experimental results has been achieved. This is explanation "in a single bound," of cognition directly to its cellular or molecular mechanisms. Nowadays, at least for many aspects of learning and memory, the operative mechanistic level is specific genes and proteins. This account wears reductionism's "nothing-but"-ism on its sleeve. The cognitive function is the cellular/molecular activity experimentally intervened into and integrated successfully, delivered via the system's neuroanatomy to the behavioral periphery, to drive muscle contractions against the skeletal frame and produce the behaviors that operationalize the specific cognitive function for experimental investigation. Ruthless reductionism indeed!

Both Strevens' and Bickle's talk of causes and mechanisms will ring familiar to recent philosophers of science. The last fifteen years has seen the rise and increasing dominance of a "new mechanism" (Machamer, Darden, and Craver 2000, with a precursor in Bechtel and Richardson (1993) and before that numerous papers of William Wimsatt's throughout the 1970s and 1980s). New mechanists are not univocal about reduction, but one view has been developed by William Bechtel (2009). Bechtel's "mechanistic reduction" draws directly on the basic new mechanist account of explanation. To explain the action of some system S's Ψing, one specifies its mechanism: the individual components of S $x_1, \ldots, x_n$, those components' activities $\phi_1, \ldots, \phi_m$, and the causal organization of the x_is' $\phi_i$ings which produces S's Ψing (Craver 2007, 7, Fig. 1.1). Bechtel illustrates his account using ruthless reductionists' favorite case studies: the consolidation of long-term memory. But he does not find reduction "in a single-bound" lurking here, from behavior to molecular pathways. Instead, he finds "nested mechanisms within mechanisms," organized like Russian nesting dolls. The molecular pathways intervened into in MCC experiments are not mechanisms of the behaving animal (e.g., the rat navigating the Morris water maze). Rather, those molecular activities are the mechanisms of specific cells in the rat's hippocampus and cortex inducing long-term potentiation (LTP), a form of activity-driven synaptic plasticity (enhancement). And those specific neurons inducing LTP are a mechanism of place cell spatial map formation across the rat's hippocampus. And place cell spatial map activity is the mechanism of the rat navigating the water maze. A complete mechanistic explanation requires specifying the full mechanism for each component of the system at the next level down. Animal experimental psychology and cognitive neuroscience attend to the mechanisms at higher levels of organization, with cellular and molecular biology the lower levels. No level is explanatorily privileged; all are required for a full mechanistic reduction of the cognitive function. The ruthless reductionist's alternative account, of seeking explanation "in a single bound," risks various methodological errors that cognitive neuroscientists routinely impress upon their MCC colleagues (Bechtel 2009; see also Craver 2007, chapter 7). Nevertheless reductionism's "nothing-but"-ism still obtains: the cognitive function is the full nested collection of mechanisms within mechanisms, with nothing "left over."

This dispute between ruthless and mechanistic reductionists appears resolvable: Which account gets the actual science right? That debate continues, with new case studies continually on offer. But two lines of possible rapprochement have also emerged. Perhaps mechanistic reduction correctly describes the vision of cognitive/systems neuroscience, while ruthless reductionism that of cellular and molecular neurobiology? Bechtel (2009) first suggested this rapprochement, describing the history of neuroscience's mid-20th-century split into the Society for Neuroscience and the cognitive/systems/neuropsychology camps. Neuroscience has been a vastly interdisciplinary enterprise since its inception. Why assume that it speaks with one voice about what reduction is?

Second, new mechanists David Kaplan and Carl Craver (2011) have addressed a challenge raised by "dynamic systems theorists," that purely mechanistic explanations in neuroscience must be supplemented with another form of explanation. Dynamicists insist that most neural systems involve thousands of parts or more, organized and interacting in myriad complex ways, with multiple redundancy and feedback. As this complexity increases, it becomes increasingly implausible to decompose the system into its discrete, interacting parts. New concepts must appear in explanations, including systems variables and order parameters, and the appropriate explanatory tools are turning out increasingly to be ones from nonlinear dynamical systems theory. These explanations violate the strictures of mechanistic explanation.

Kaplan and Craver's (2011) response is radical. To the extent that dynamical systems "explanations" violate the principles of correct mechanistic explanations, such "explanations" aren't genuine explanations at all, but rather just descriptions of the phenomenon to be (mechanistically) explained. They examine in some detail a landmark case of dynamicist neuroscience, the Haken, Kelso, and Bunz (1985) model of human bimanual finger-movement coordination. They point out that these modelers only intended for their dynamical systems model to be a mathematically compact description of the temporal evolution of a purely behavioral dependent variable. None of the model's variables or parameters were interpreted to map onto components or operations of any hypothetical mechanism generating the behavioral data; none of the mathematical relations or dependencies between variables in the model were intended to map onto hypothesized causal interactions between components or activities of any mechanisms. Furthermore, subsequent to developing and publishing their dynamicist model, Kaplan and Craver point out that these modelers themselves began to investigate how the behavioral regularities their model described might be produced by neural motor system components, activities, and organization. This suggests that the modelers themselves were taking their dynamicist model as a heuristic and sought to move towards a "how-possibly," and ultimately a "how-actually," mechanistic explanation.

Kaplan and Craver's (2011) last argument is reminiscent of Silva, Landreth, and Bickle's appeal to mediation analysis in MCC, to "looking down" to components of established causal hypotheses to find the causes mediating them. Mediation Analysis is part of what makes MCC (metascientifically/ruthlessly) reductionistic. So in addition to the challenge Kaplan and Craver (2011) raise for dynamicists, a comparison of one of their arguments with Mediation Analysis yields a significant potential comparison between an explicit kind of reductionism and a new mechanism. It doesn't resolve the dispute about explanations in a "single bound," from behavior directly to molecular pathways, versus nested hierarchies of mechanisms within mechanisms, but it does reveal more clearly what is at dispute. It is the scientific status of "higher-level" causal mechanisms when lower-level ones have been established experimentally with strong scientific confidence. Ruthless reductionists then eschew the higher level as mechanistic; mechanistic reductionists continue to embrace them as such.

Notes

1 Strevens (personal communication) suggested this reply, along with numerous helpful comments that improved my discussion of his views.
2 The terminology for these three kinds of experiments is from Silva, Landreth, and Bickle (2014). Sweatt (2009) refers to them, respectively, as "blockages," "mimicries," and "correlations." There are subtle differences between Silva, Landreth, and Bickle's "positive interventions" and Sweatt's "mimicries," but there is not space to explore those here.

References

Balzer, W., Moulines, C.-U., and Sneed, J. (1987). *An Architectonic for Science.* Dordrecht: Reidel.

Bechtel, W. (2009). "Molecules, systems, and behavior: Another view of memory consolidation." In Bickle, J. (ed.), *The Oxford Handbook of Philosophy and Neuroscience.* New York: Oxford University Press, 3–40.

Bechtel, W. and Richardson, R. (1993). *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research.* Princeton, NJ: Princeton University Press.

Bickle, J. (1998). *Psychoneural Reduction: The New Wave.* Cambridge, MA: MIT Press.

Bickle, J. (2006). "Reducing mind to molecular pathways: Explicating the reductionism implicit in current cellular and molecular neuroscience." *Synthese* 151 (3): 411–434.

Bickle, J. (2009). "Real reductionism in real neuroscience: Metascience, not philosophy of science (and certainly not metaphysics!)." In Hohwy, J. and Kallestrup, J. (eds.), *Being Reduced: New Essays on Reduction, Explanation, and Causation.* New York: Oxford University Press, 34–51.

Bickle, J. (2012). "A brief history of neuroscience's actual influences on mind-brain reductionism." In Gozzano, S. and Hill, C. (eds.), *New Perspectives on Type Identity Theory.* Cambridge: Cambridge University Press, 88–109.

Bickle, J. (2016). "Revolutions in neuroscience: Tool development." *Frontiers I Systems Neuroscience* (March): http://journal.frontiersin.org/article/10.3389/fnsys.2016.00024/full

Carnap, R. (1932). "The elimination of metaphysics through the logical analysis of language." (translation by Arthur Pap), *Erkenntnis* 2: 60–81.

Carnap, R. (1934). *The Unity of Science.* London: K. Paul, Trench, Trubner and Co. Ltd.

Carnap, R. (1936). "Testability and meaning." *Philosophy of Science* 3 (4): 419–471.

Chalmers, D. (1996). *The Conscious Mind.* New York: Oxford University Press.

Churchland, P.M. (1985). "Reduction, qualia, and the direct introspection of brain states." *Journal of Philosophy* 82 (1): 8–28.

Churchland, P.S. (1986). *Neurophilosophy.* Cambridge, MA: MIT Press.

Craver, C.F. (2007). *Explaining the Brain.* New York: Oxford University Press.

Endicott, R. (1998). "Collapse of the new wave." *Journal of Philosophy* 95 (2): 53–72.

Feyerabend, P.K. (1962). "Explanation, reduction and empiricism." In Feigl, H. and Maxwell, G. (eds.), *Minnesota Studies in Philosophy of Science* 3, Minneapolis: University of Minnesota Press, 28–96.

Haken, H., Kelso, S., and Bunz, H. (1985). "A theoretical model of phase transitions in human hand movements." *Biological Cybernetics* 51 (5): 347–356.

Hooker, C.A. (1981). "Towards a general theory of reduction, Part I: Historical and scientific setting, Part II: Identity in reduction, Part III: Cross-categorial reduction." *Dialogue* 20: 38–59, 201–236, 496–529.

Kandel, E.R., Schwartz, J., and Jessel, T. (eds.) (2000). *Principles of Neural Science*, 4th Ed. New York: McGraw-Hill.

Kaplan, D.M. and Craver, C.F. (2011). "The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective." *Philosophy of Science* 78 (4): 601–627.

Kemeny, J. and Oppenheim, P. (1956). "On reduction." *Philosophical Studies* 7 (1–2): 6–19.

Kim, J. (2005). *Physicalism, or Something Near Enough.* Princeton, NJ: Princeton University Press.

Levine, J. (1993). "On leaving out what it is like." In Davis, M. and Humphreys, G.W. (eds.), *Consciousness: Psychological and Philosophical Essays.* London: Blackwell, 121–136.

Machamer, P., Darden, L., and Craver, C.F. (2000). "Thinking about mechanisms." *Philosophy of Science* 67 (1): 1–25.

Morton, A.L. and Bickle, J. (2005). "Re-examining logical positivism: Testability and meaning in contemporary neuroscience." *Journal of Contemporary Philosophy* 25: 3–11.

Nagel, E. (1961). *The Structure of Science.* New York: Harcourt, Brace, and World.

Nagel, T. (1974). "What is it like to be a bat?" *Philosophical Review* 83 (4): 435–450.

Quine, W.V.O. (1951). "Two dogmas of empiricism." *Philosophical Review* 60 (1): 20–43.

Schaffner, K. (1967). "Approaches to reduction." *Philosophy of Science* 34 (2): 137–147.

Schaffner, K. (1992). "Philosophy of medicine." In Salmon, M., Earman, J., Glymour, C., Lennox, J., Machamer, P., McGuire, J., Salmon, W., and Schaffner, K. (eds.), *Introduction to the Philosophy of Science.* Englewood Cliffs, NJ: Prentice Hall, 310–344.

Schaffner, K. (2012). "Ernest Nagel and reduction." *Journal of Philosophy* 109 (8–9): 534–565.

Silva, A.J., Landreth, A., and Bickle, J. (2014). *Engineering the Next Revolution in Neuroscience.* New York: Oxford University Press.

Smart, J.J.C. (1959). "Sensations and brain processes." *Philosophical Review* 68: 141–156.

Strevens, M. (2011). *Depth: An Account of Scientific Explanation.* Cambridge, MA: Harvard University Press.

Strevens, M. (2016). "Special science autonomy and the division of labor." In Couch, M. and Pfeifer, J. (eds.), *The Philosophy of Phillip Kitcher.* New York: Oxford University Press.
Suppes, P. (1967). "What is a scientific theory?" In Morgenbesser, S. (ed.), *Philosophy of Science Today.* New York: Basic Books, 55–67.
Sweatt, D. (2009). *Mechanisms of Memory*, 2nd Ed. San Diego: Academic Press.
Theruer, K. and Bickle, J. (2013). "What's old is new again: Kemeny-Oppenheim reduction at work in current molecular neuroscience." *Philosophia Scientia* 17 (2): 89–113.