INTERFACE
FOCUS

rsfs.royalsocietypublishing.org

Research

Cite this article: Powell R, Mariscal C. 2015 Convergent evolution as natural experiment: the tape of life reconsidered. *Interface Focus* **5**: 20150040.
http://dx.doi.org/10.1098/rsfs.2015.0040

One contribution of 12 to a theme issue 'Are there limits to evolution?'

Subject Areas:
astrobiology

Keywords:
contingency, convergence, exobiology, laws, experiment, Stephen Jay Gould

Author for correspondence:
Russell Powell
e-mail: rpowellesq@gmail.com

THE ROYAL SOCIETY
PUBLISHING

Convergent evolution as natural experiment: the tape of life reconsidered

Russell Powell[1] and Carlos Mariscal[2]

[1]Department of Philosophy, Boston University, Boston, MA 02215, USA
[2]Department of Biochemistry and Philosophy, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4R2

Stephen Jay Gould argued that replaying the 'tape of life' would result in radically different evolutionary outcomes. Recently, biologists and philosophers of science have paid increasing attention to the theoretical importance of convergent evolution—the independent origination of similar biological forms and functions—which many interpret as evidence against Gould's thesis. In this paper, we examine the evidentiary relevance of convergent evolution for the radical contingency debate. We show that under the right conditions, episodes of convergent evolution can constitute valid natural experiments that support inferences regarding the deep counterfactual stability of macroevolutionary outcomes. However, we argue that proponents of convergence have problematically lumped causally heterogeneous phenomena into a single evidentiary basket, in effect treating all convergent events as if they are of equivalent theoretical import. As a result, the 'critique from convergent evolution' fails to engage with key claims of the radical contingency thesis. To remedy this, we develop ways to break down the heterogeneous set of convergent events based on the nature of the generalizations they support. Adopting this more nuanced approach to convergent evolution allows us to differentiate iterated evolutionary outcomes that are probably common among alternative evolutionary histories and subject to law-like generalizations, from those that do little to undermine and may even support, the Gouldian view of life.

1. Replaying the tapes of life

In his book *Wonderful Life: The Burgess Shale and the Nature of History*, Gould [1] proposes a series of macroevolutionary thought experiments designed to probe the contingent nature of life's history. Each involves rewinding the 'tape of life' to a different period in the history of animal evolution and asking how its story would again unfurl. Gould's most crucial thought experiment speaks to the radical contingency of animal body plans that arose in the Cambrian, which include most of the animal phyla in existence today along with a variety of morphologically bizarre taxa that became extinct in the same period (p. 188, 227, 239). Gould argues that no biologist would have predicted the actual patterns of survivorship in the end-Cambrian extinctions on the basis of any plausibly relevant traits (e.g. degree of specialization, anatomical complexity, ecological prominence, etc.). From this epistemic vantage point, Gould draws a metaphysical conclusion, namely that early animal extinction patterns were essentially haphazard and counterfactually non-replicable. He further contends that owing to the path-dependent nature of macroevolution, these early stochastic sampling events dramatically shaped the future history of animal life on the Earth.

Gould reaches comparable conclusions in relation to more fine-grained taxonomic events, such as the seemingly improbable survival of lobe-finned fishes and its implications for the origins of tetrapods (p. 317, 318), as well as the extinction of the long-dominant non-avian dinosaurs, which allowed for the unlikely radiation of mammals. He concludes, 'any replay of the tape would lead evolution down a pathway radically different from the road actually taken' (p. 51). We will refer to this as the *radical contingency thesis* (RCT). Although Gould tended to focus on animal morphology, he argued that 'almost every interesting event of life's history falls into the realm of contingency' (p. 290)—remarks which suggest that the RCT was proposed not as a narrow claim about the evolution of animal body plans, but rather as a general thesis about the grand-scale organization of life on the Earth. For present purposes, we will assume this broader reading of the RCT.

There have been many challenges over the years to different elements of Gould's thesis. For example, many of the seemingly bizarre Cambrian taxa that inspired the RCT (Gould's *Wonderful Life* manuscript was originally titled *Homage to Opabinia* after one such taxonomic oddity [2]) have been recognized under modern evolutionary classification systems as 'stem taxa'—extinct basal lineages on the stems leading to the 'crown' groups represented by modern phyla.[1] Whereas Gould was fascinated by the Cambrian Problematica because of the unique combinations of features they possessed, cladistic analyses ignored these features and instead used shared derived characters to situate these taxa un-problematically in relation to modern phyla. In doing so, however, the cladistic reconstruction of Cambrian phylogeny simply bypassed the big theoretical questions that occupied Gould, such as in relation to patterns of early morphological disparity and extinction and their implications for the nature of the evolutionary process [4]. Even if it were the case that the early Cambrian fauna did not reflect the broad range of forms that Gould believed they did, this would not support the opposing view of life, namely the 'robust replicability thesis', which argues that macroevolutionary outcomes are *robust* (stable) across large differences in initial conditions. This is because the robust replicability view as it applies to the evolution of body plans requires a merit-based competition among early forms with lineage sorting based on functional superiority, and such a competition could not exist if only a small range of physically possible forms actually arose. In other words, the lack of an early great experimentation phase would only undermine the case for robust replicability, and would do little to undercut the RCT. For all of these reasons, the recent reclassification of the Cambrian fauna has left the contingency dispute unresolved.

Here, we focus on a more promising challenge to the RCT that appeals to the ubiquity of 'convergent evolution', or the ostensibly independent origination of similar biological forms and functions. Convergent evolution is taken by some researchers to support the view that a hypothetical replay of the tape of life on the Earth, and actual replays on the Earth-like planets elsewhere, would likely result in similar, predictable, outcomes. In this paper, we examine the evidentiary relevance of convergent evolution for the contingency debate. In §2, we consider where the RCT stands in relation to biological prediction, explanation, chance and laws, which in turn will help to clarify the types of evidence that might bear on its adjudication. In §3, we review the 'critique from convergent evolution', which views convergence as tantamount to experimental replication in the history of life, and we show that certain conceptual problems have prevented this critique from making crucial contact with Gould's core claims. We go on in §4 to determine the conditions under which episodes of convergent evolution can constitute valid natural experiments that support inferences regarding the deep robustness of evolutionary outcomes. We argue that proponents of convergence have problematically lumped causally heterogeneous phenomena into a single evidentiary basket, in effect treating all instances of convergent evolution as if they are of equivalent theoretical import. We attempt to remedy this in §5 by proposing ways to break down the heterogeneous class of convergent events based on the nature of the generalizations they support. Adopting this more nuanced approach to convergent evolution lays the groundwork for differentiating iterated evolutionary outcomes that are plausibly common among alternative evolutionary histories and subject to law-like generalizations, from those that do little to undermine, and may even support, the Gouldian view of life. This analysis sheds light on what convergence tells us about constraints on the history of life as it has unfolded on Earth, and how it might do in other parts of the universe.

2. Biological contingency, predictability, chance and laws

There are many non-equivalent definitions of evolutionary contingency on offer. For present purposes, we will consider an outcome merely *contingent* if its existence depends on the occurrence of an event that might not have occurred, even if the occurrence of this event was highly probable. For instance, we might say that the evolution of life on earth is contingent on the existence of nucleic acids, even though recent studies suggest that nucleic acids are readily created in the energy of asteroid impacts and hence are probably common in the universe. We will consider an outcome *radically contingent* when its existence depends on events occurring during a given evolutionary path that are unlikely to be replicated across the vast majority of alternative evolutionary histories. Radically contingent systems may be contrasted with *robustly convergent* systems in which many outcomes remain accessible from distant evolutionary trajectories (figure 1).

Some theorists have suggested that we think of the RCT as a claim about macroevolutionary path-dependence [5] or macroevolutionary stochasticity [6] *simplicter*. However, there may be systems that exhibit path-independence and predictability at finer grains of phylogenetic resolution (e.g. within orders, classes or phyla) but path-dependence and unpredictability at courser grains (e.g. across phyla or the whole of animal life), as depicted in figure 1*c*. Later (§3), we argue that such a mixed contingent/convergent system is consistent with, and indeed predicted by, the Gouldian view of life. For now, we note that such a mixed system is consistent with interpretations of Gouldian contingency as involving path dependency and/or stochasticity at the appropriate grain of resolution.

Outcomes are contingent (or radically contingent) only *with respect to a given initial set-up*—so, for example, an outcome might be robust with respect to complex multicellular eukaryote evolution even if the origin of eukaryotes is itself an improbable event. Gould's Cambrian thought experiment is structured in this manner, as it assumes the existence of complex multicellular organization among its initial conditions, and queries whether the evolution of specific body plans and their associated regularities are stable across replays of the multicellular tape of life. It follows that evolutionary outcomes will vary in their degree of evolutionary robustness, depending on the nature of the events on which they are contingent.

To complicate matters, not only do contingency and robustness admit of degrees, but each may also apply to many different evolutionary outcomes, some of which may turn out to be contingent while others robust. One

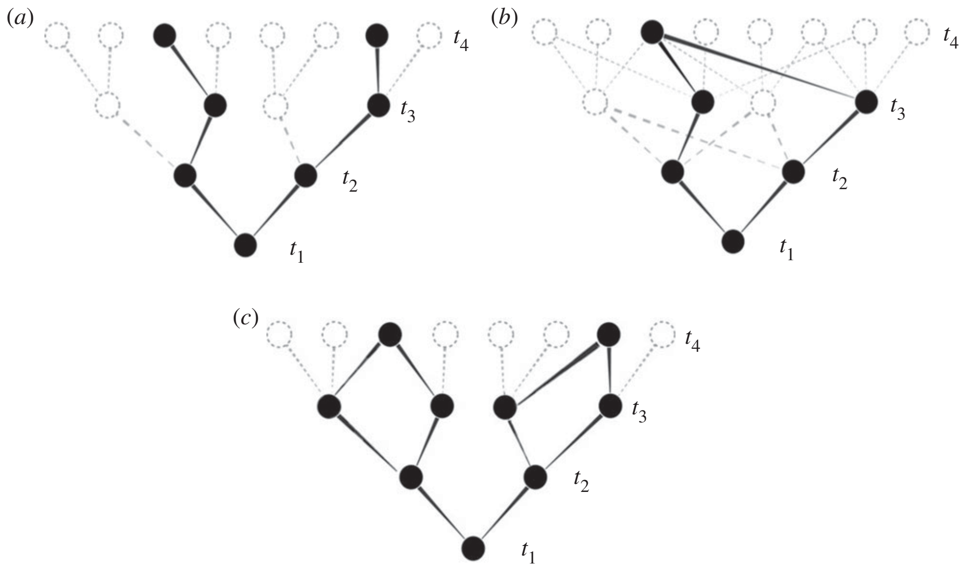

Figure 1. (*a*) A radically contingent system in which outcomes (morphological, functional or relational) depend on improbable events occurring during the evolutionary path under consideration. For example, at t_2, some paths are not taken (dashed lines) owing to non-replicable events, causing a number of evolutionary outcomes to become inaccessible at t_3. Conversely, (*b*) a robustly convergent system in which many outcomes are still accessible from even very distant evolutionary trajectories. For example, very divergent lineages at t_3 are still able to converge on the same evolutionary outcome at t_4. (*c*) A mixed contingent/convergent system that exhibits path-independence and predictability at finer grains of resolution (reflecting evolutionary iteration within the developmental parameters of orders, classes and phyla), but exhibiting path-dependence and unpredictability when viewed at courser grains (e.g. across phyla or the whole of animal life). Such mixed systems are consistent with and arguably predicted by a Gouldian view of life that accords theoretical primacy to internal developmental constraints. This underscores the problem with interpreting all evolutionary repetitions, and even certain impressive convergences, as counterexamples to the RCT.

methodological implication of this is that a handful of putative counterexamples can at best refute parts of the RCT, not the thesis as a whole. These methodological difficulties are not unique to the contingency dispute, as they confront many other ‘relative significance debates’ in evolutionary biology [7]. It is nonetheless meaningful to ask whether contingency is a dominant theme in macroevolution and to investigate the specific outcomes to which it applies.

Gould tended to illustrate the RCT in terms of ex ante unpredictability, an epistemic state he attributed to the idealized observer in his macroevolutionary thought experiments. However, linking contingency and unpredictability is misleading because it conflates metaphysical and epistemic claims, resulting in misinterpretations of Gould’s thesis (§3). Even if a late-Cretaceous observer could have predicted the K–Pg impact and the ensuing dinosaurian extinctions and mammalian radiations, and even if this idealized observer were capable of explaining the latter evolutionary outcomes by recourse to causes and laws of nature—these events would still be radically contingent. This is because the RCT is fundamentally a *metaphysical* (modal) thesis, not an *epistemic* one—it holds that certain macroevolutionary outcomes are highly sensitive to low probability events that are unlikely to be replicated across the vast majority of alternative histories of life. In Gould’s words: ‘Alter any early event, ever so slightly and without apparent importance at the time, and evolution cascades into a radically different channel’ ([1, p. 51] and [8, p. 1333]). It simply does not matter, for purposes of characterizing these dynamics, whether this metaphysics translates into unpredictability or undermines explanation.

Unfortunately, Gould abetted confusion on these points by contrasting the ‘contingent phenomenology’ of natural history with Laplacean determinism [8, p. 1333]. Yet if classic determinism is true, rewinding the tape of life is a trivial exercise, because the trajectory of life would unfold in precisely the same way. Determinism tells us nothing about the accidental or law-like nature of macroevolutionary outcomes. The RCT does not require that either objective macro-indeterminism or chaotic causal determinism obtain in macroevolution—it requires only that relevant outcomes are instable across the vast majority of nomically possible evolutionary worlds.

The RCT is best understood, therefore, as a universal biological claim about the sensitivity of large-scale evolutionary outcomes to initial conditions [2,4–11]. The claim is universal because it cites causal mechanisms applicable to all nomically possible histories of life. These include stochasticity, developmental entrenchment, mutational ordering, sequences of selection regimes and other processes that result in non-replicable, path-dependent and irreversible trajectories of macroevolution [12]. Thus, we think that contrary to the reading of some commentators [13], Gould’s macroevolutionary thought experiments were intended to draw lessons about the counterfactual stability of the evolution of life on the Earth-like worlds in general, not merely life as we know it on the Earth. Whether or not this was Gould’s intention, the theoretical framework of the RCT supports these broader implications.

Given this emphasis on nomic necessity, one might be tempted to think that the RCT debate boils down to whether there are biological laws describing the evolution of form and function—that if such laws are found, then Gould’s thesis will have been refuted. Indeed, Gould repeatedly attributed contingency in the history of life to the lack of invariant biological laws [8, p. 696, 1055, 1227, 1333]. Nevertheless, the RCT is not in tension with the existence of biological laws *tout court*. As we just saw, the RCT is itself premised on universal biological properties, and perhaps laws [14], that underwrite the stochastic, path-dependent nature of macroevolution—and it is consistent with many other biological laws (such as the zero force evolutionary law [15], which describes the tendency of biological systems to increase in diversity over time). What would pose a problem for the RCT is the existence of *strict deductive laws that describe the evolution of form and function*. Such generalizations would be universally projectable, support evolutionary counterfactuals and hold across alternative evolutionary histories.

Still, there are good reasons to divorce the contingency debate from disputes over the existence of laws in biology. Although philosophers of science have linked the supposed 'nomological vacuum' in biology to the metaphysics of evolutionary contingency [7], there are 'looser' accounts of lawhood that allow highly contingent antecedent conditions to be incorporated into law-like statements, and the existence of laws on such accounts may be consistent with the RCT. For instance, Sober [16] argues that generalizations formulated in terms of counterfactual conditionals with historically contingent antecedents can satisfy the desiderata for lawhood because, so formulated, they are exceptionless and universal. While Sober is right that the relations described by such conditionals are not *in themselves* radically contingent, if they contain antecedent conditions that are radically contingent, this will undermine the robustness of the outcomes they describe [17]. Given the diversity of philosophical thinking about lawhood, linking the radical contingency question to a verdict on biological laws is liable to confuse more than it is to illuminate.

It is better, we believe, to view the RCT as a series of claims about the frequency of certain outcomes across possible evolutionary histories. Low probability outcomes become probable given a sufficiently large population of chance set-ups. A key question for purposes of the contingency debate, therefore, is *whether a single history of life constitutes a sufficiently large population of chance setups so as to make certain low probability outcomes probable in any given history of life*. Gould appears to be making this sort of frequency claim when, for example, he discusses the contingency of human evolution:

> [I]f life started with all its models present, and constructed a later history from just a few survivors, then we face a disturbing possibility. Suppose that only a few will prevail, but all have an equal chance. The history of any surviving set is sensible, but each leads to a world thoroughly different from any other. If the human mind is a product of only one such set, then we may not be randomly evolved in the sense of coin flipping, but our origin is the product of massive historical contingency, and we would probably never arise again even if life's tape could be replayed a thousand times. [1, p. 233–234]

Gould thus argues that owing to the stochastic nature of certain key history of life-shaping events, many evolutionary outcomes would occur very infrequently across alternative evolutionary histories.

If the universe is infinite or nearly so, then there will be many runs of the tape of life that are very similar to our own. The question before us, therefore, is one of modal frequency. On this reading, the existence of law-like evolutionary outcomes with very low modal frequencies—that is, outcomes that obtain necessarily or with high probability but only under conditions that are astronomically uncommon in the universe—do not undermine the RCT or support the robust replicability view. This provides yet further reason to separate the contingency question from the question of laws in biology. In what follows, we consider the circumstances under which convergent evolution may be taken as evidence for the higher modal frequencies of certain evolutionary outcomes.

3. The critique from convergent evolution and its limitations

How might the contingency debate be adjudicated? The most straightforward way of evaluating the RCT would be to observe numerous independent histories of life. Each would begin from a wide range of initial conditions with respect to geophysical variables, the ordering of mutations, ecological relations, major evolutionary innovations and so on. With these data in hand, we would then search for regularities that allow us to infer frequency distributions and the macroevolutionary processes that underpin them. Unfortunately, we are currently privy to a single history of life, and thus we are unable to investigate the stability of large-scale evolutionary outcomes in this way. Gould's macroevolutionary thought experiments are designed to circumvent this '$n = 1$' problem [18], but such exercises of the imagination, even if intuitively compelling to some, are ultimately speculative and empirically inconclusive. Other approaches to biological contingency involve running controlled evolutionary experiments on microbes [19], observations of selection on animals in the wild [20] and evolutionary simulations [21], but such studies have limited or indeterminate generalizability to large-scale patterns in eukaryotic evolution that only manifest over immense geological timescales.[2]

Biologists and philosophers of science have begun paying increasing theoretical attention to convergent evolution, as it offers a promising avenue for evaluating the RCT.[3] Unlike the outcomes of controlled evolutionary experiments in the laboratory or observations of natural selection in the field, studies of convergent evolution can draw upon a voluminous database of natural history to make inferences about the robustness of evolutionary processes operating over vast timescales and across immense phylogenetic gaps. From the standpoint of the robust replicability view, we are not confronted with an $n = 1$ problem after all, because convergence amounts to natural experimental replication in the histories of life. Proceeding on this interpretation, some biologists have compiled expansive evidence bases of convergent evolution with the aim of debunking, or casting serious doubt on, Gould's thesis [24–26]. Philosophers of science have read convergent evolution in similar ways [27,28].

Convergence is widely regarded as evidence for adaptation [29]: the fact that both ichthyosaurs and dolphins evolved flippers in an aquatic environment strongly suggests their flippers are similarly functional. But convergence can also be read as evidence for three additional, more philosophically onerous claims that underpin the robust replicability view of life: first, that certain design problems are pervasive in the history of life; second, that the set of evolutionary solutions to pervasive design problems is highly circumscribed and third, that these restricted solutions are accessible to selection notwithstanding the 'internal' constraints of phylogeny. As we see (§4), drawing each of these inferences from existing data on convergent evolution can be highly problematic.

Before elaborating on these evidential issues, it is worth highlighting some conceptual problems that have prevented the critique from convergent evolution from engaging with key claims of the RCT. One such problem is that prominent critiques have often mischaracterized or attacked straw-man versions of Gould's thesis, which in turn has led them to misjudge the evidential value of convergence vis-à-vis the RCT. For example, George McGhee, in his leading review of convergent evolution, refutes several claims that he attributes to Gould, but which Gould almost certainly did not hold. One such refutation is 'the view that the evolutionary process is nonrepeating is demonstrably false' [24, p. 271]. Indeed, much of the philosophical attention to Gould's thesis has likewise focused on its implications for the non-repeatability and unpredictability of evolutionary outcomes [5,6,9,14,27].

However, there is ample room within the Gouldian view of life for a great deal of evolutionary iteration and predictability. By giving theoretical primacy to internal developmental constraints, the RCT, in fact, predicts certain kinds of repetition, namely repetitions that result from entrenched developmental systems that make certain adaptive outcomes more likely given their accessibility to selection.[4] To illustrate this phenomenon, Gould offers the repeated evolution in crustaceans of maxillipeds, feeding appendages that evolved from anterior walking legs, which was accomplished by the iterated selective deployment of homologous developmental pathways [8, p. 1134]. On the Gouldian view, evolutionary iterations such as these are only possible because of conserved developmental constraints and, crucially, these constraints could easily have been otherwise. This scenario is illustrated in schematic form by the system depicted in figure 1*c*, which exhibits path independence (convergence) at finer grains of phylogenetic resolution but path-dependence at courses grains.

In sum, the contingency dispute turns not on the existence of evolutionary repetitions *per se*, but on the *causes* of evolutionary iteration and whether they support the deep evolutionary robustness of the outcomes observed. What contingency theorists like Gould reject is the proposition that the driving forces behind convergence often transcend the contingently entrenched developmental plans of particular lineages. The key question, therefore, is whether instances of evolutionary iteration reflect this transcendence—and the problem is that whether they do, as we shall see in §4, is underdetermined by convergence data as it has thus far been collected and analysed.

Gould has also been criticized for advocating, as McGhee puts it, that 'evolution is entirely historically contingent, and thus unpredictable (and nonrepeating)' [24, p. 271]. Yet Gould states in no uncertain terms that he is *not* arguing that *all* of evolution is historically contingent and unpredictable:

> Am I really arguing that nothing about life's history could be predicted, or might follow directly from general laws of nature? Of course not; the question that we face is one of scale, or level of focus. Life exhibits a structure obedient to physical principles. We do not live amidst a chaos of historical circumstance…Much about the basic form of multicellular organisms must be constrained by rules of construction and good design…Invariant laws of nature impact the general forms and functions of organisms; they set the channels in which organic design must evolve. But the channels are so broad relative to the details that fascinate us! … When we set our focus upon the level of detail that regulates most common questions about the history of life, contingency dominates and the predictability of general form recedes to an irrelevant background…almost every interesting event of life's history falls into the realm of contingency.
>
> [1, p. 289–290]

Gould, it seems, would be happy to grant that (for example) the fusiform shape is an evolutionarily robust feature of complex multicellular aquatic life wherever it evolves. What he would deny is that the specific elements of body plans are themselves robust features of complex multicellular evolution wherever it evolves [8, p. 1212]. Ubiquitous convergence on fusiform morphology does little to detract from this conclusion.

One weak point in the above excerpt, however, is Gould's relative interest claim. Why should universal biomechanical constraints on the evolution of form not be as interesting to biologists as the quirky, more detailed outcomes of evolution? Gould offers no argument to support this claim. Indeed, as Haufe [14] contends, Gould's push for a nomothetic palaeontology shows that he was committed to the idea that a central goal of science, and indeed, palaeobiology, is to uncover spatio-temporally invariant laws. Whatever one makes of Gould's advocacy of historical narratives, it is hard to defend the claim that universal selective constraints on the evolution of form are *uninteresting*. In effect, Gould is making a rhetorical move similar to one that he, writing with Richard Lewontin, famously excoriated in the context of advancing non-selective explanations [30, p. 585]:

> In natural history, all possible things happen sometimes; you generally do not support your favoured phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival, but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an undogmatic and ecumenical chap.

Gould seems to be engaging in the very mode of argument for which he chastises adaptationists: he acknowledges there are some robustly replicable outcomes in evolution, but he relegates these to theoretically uninteresting phenomena in the history of life.

In short, both contingency theorists and convergence proponents intend to make universal biological claims about their favoured evolutionary dynamics—contingency in one case and convergence in the other—but they disagree as to the relative significance of these respective phenomena in any given history of life. With a clearer picture of the disagreement between contingency theorists and convergence proponents in mind, we will now go on to consider the circumstances in which evolutionary repetitions might serve as evidence against the RCT.

4. Evaluating the validity of natural experiments in convergent evolution

In what sense might convergent evolution constitute 'experimental' evidence that bears on the RCT? Although we normally think of scientific experiments as involving the manipulation of independent variables to assess their causal influence on dependent variables, 'to conduct an experiment', broadly construed, is simply to put one's self in an epistemic vantage point from which to make observations that affect our confidence in hypotheses about the causal, nomological or historical structure of the world. Given such a broad conception of experiment, it is not easy to draw a distinction between observational studies that make use of conditions already found in nature and experiments that involve the deliberate manipulation of nature, whether in terms of their procedures, epistemic goals, or confirmatory power [31].[5]

Nevertheless, natural experiments, which are common in the human sciences, are in certain ways more like proper manipulation-based experiments than they are like observational studies. This is because they are structured in ways that resemble a manipulationist experimental structure that might be designed by human agents were they tasked with running experiments to probe certain hypotheses or to make certain measurements. In natural experiments, researchers select samples that differ naturally in one or more independent variables but which are similar with respect to other variables, typically with the aim of corroborating a causal-historical hypothesis. Although natural experiments lack the control of laboratory and field experiments (thus risking internal validity), they have the advantage of allowing biologists to gather data across numerous taxa and habitats, reflecting vast timespans of evolution [34]. For instance, studies of plant and animal colonizations of isolated islands following the volcanic destruction of endemic ecosystems have been characterized as natural experiments—replays of the ecological tape from which biologists might infer relations that hold-up across disparate phylogenetic and environmental contexts. This breadth and depth of study, often impossible in laboratory and field experiments, permits inferences that are projectable to large-scale patterns in the history of life.

Like laboratory experiments, natural experiments can be contaminated, poorly structured and misleading. A crucial feature of natural evolutionary experiments is that they are isolated in ways that control for the relevant confounding variables.[6] The confounding variables that must be controlled in any given natural experiment are determined by the research questions being investigated or the hypotheses being tested. The key for present purposes, therefore, is to identify the validity criteria of natural experiments in convergent evolution *for purposes of assessing the RCT*. Evaluating the set-up of natural experiments in convergence for this specific evidentiary purpose requires at a minimum that we assess the 'independence' of observed replications by controlling for the influence of conserved developmental mechanisms. Failure to do so has resulted in flawed experimental set-ups, even in supposedly paradigmatic cases of natural experiments in convergent evolution.

For example, Morgan [35] touts John Beatty's [27] discussion of studies documenting the iterated evolution of Anolis lizard ecomorphs [36]—forms adapted to specialized microhabitats on isolated islands in the Caribbean—as paradigmatic natural evolutionary experiments that meet the requisite isolation conditions. Indeed, both Beatty [27] and Losos *et al*. [36] interpret these natural evolutionary experiments as contradicting Gould's thesis. However, *as tests of the RCT*, these experiments are invalid, because despite their geographical separation, the observed evolutionary systems are not isolated in crucial respects. In particular, the target systems begin with highly similar initial developmental conditions and thus fail to rule out the possibility that 'positive' internal constraints are responsible for the observed iterations.

A broader problem with evidential appeals to natural experiments in convergent evolution is that the sense of 'independence' that is operative in dominant definitions of convergence does not support some of the evidential uses to which convergence is put in the contingency debate. For example, on the standard 'taxic' account of convergence,[7] to satisfy the independence criterion, a similar trait in two lineages must not have been present in and continuously transmitted from their common ancestor [37]. A minority view [38] holds that satisfying the independence criterion requires that the genetic-developmental machinery causally responsible for a character resemblance must not have been continuously inherited from a common ancestor. Other authors wanting to distinguish iterations produced by shared developmental machinery have chosen to retain standard definitions of homology and instead distinguish between parallel and convergent homoplasies, where 'parallelism' is loosely defined as homoplasy that is underwritten by conserved developmental generators [10,11,39,40].

These different readings of the independence criterion have different implications for the evidentiary relevance of convergence. On the 'taxic' account, convergent evolution can ensue from both highly similar and highly disparate initial developmental conditions, resulting in a causally heterogeneous set of convergent events with differential implications for the evolutionary robustness of the regularities observed. The latter views, in contrast, do more to control for the 'internal' developmental determinants of iterated evolution, because they classify iterated traits produced by conserved generators as homologies or parallel homoplasies, rather than as cases of convergence simpliciter.

Because convergence proponents have tended to operate with the standard taxic framework for convergent evolution, their analyses have failed to distinguish natural experimental set-ups in convergent evolution that do not speak to the RCT from those that could genuinely undermine it. In essence, prominent studies of convergent evolution from which conclusions about the RCT are drawn have been working with tainted experimental set-ups. As a result, the critique from convergent evolution is unable to differentiate biological regularities that reflect deep truths about the living universe from those that are essentially accidental. We refer to this indiscriminate grouping of heterogeneous convergent events as the 'lumping problem'.

For instance, McGhee's [24] review of convergent evolution is a significant improvement over previous efforts because it includes extensive phylogenetic information; however, it fails to use this important data to control for the internal developmental determinants of iterated evolution. Consider the more than 20 instances of cantharophilous flowers evolving specialized morphologies to facilitate pollination by beetles (p. 121). The prediction 'where there are beetles, there will be cantharophilous flowers' may be law-like on looser accounts of biological laws, but if 'beetle' phenotypes and ecologies are highly contingent, then the outcomes to which this prediction refers will be highly restricted and add little to our understanding of the deep structure of evolution. The same goes for Conway Morris's [25, p. 211] discussion of iterations like ant mimicry, which has evolved more than 70 times in insects and spiders: if the 'ant' and broader 'arthropod' phenotype is radically contingent, then ant mimicry will not be projectable to alternative evolutionary histories. A compilation of such regularities may indicate the deep evolutionary robustness of very broad phenomena like coevolution and mimicry—but as we argued earlier, this is not the level of detail at which the contingency debate takes place.

The upshot is that the sheer *number of iterations*, without further analysis, tells us very little about the evolutionary robustness of the regularity they comprise. Dozens of convergences on flower and ant morphologies offer far weaker evidence against the RCT than do convergences that are fewer in number but that can be shown to arise from broad physico-chemical constraints on life. In short, the high frequency of an outcome in *this* evolutionary history, without further analysis, cannot be taken to imply a high frequency across *alternative* evolutionary histories. Unfortunately, prominent theoretical treatments of convergent evolution have tended to lump broad generalizations, like fusiform shapes, with narrower generalizations, like cantharophilous flowers, as if the total body of evidence spoke against the RCT. Yet as we have shown, the fact that structures or functions recur in evolution is not always, or even often, evidence that they are universal or probable across alternative histories of life. The data do not speak for itself, because the causes of evolutionary iteration are underdetermined. This failure to differentiate among convergent events may be due, in part, to the common misconception that the RCT predicts the wholesale lack of repetition in evolution (§3). Whatever its motivation, lumping is harmful to the critique from convergent evolution because it obscures the strongest evidence against the RCT. In compiling a body of undifferentiated data on convergence, the whole is less than the sum of its parts.

As we saw earlier, the crux of Gould's disagreement with convergence proponents turns not on whether evolutionary iteration is important, but *why* it is important, and what it signifies about the causal structure of the biological world. Let us call it 'Gouldian repetition' when evolutionary iteration results from selection acting on radically contingent conserved developmental substrates or their sequelae. Let us refer to all other iterations as cases of 'true convergence'.[8] The key is to identify meaningful cases of true convergence, and then to assess the relative significance of such events in the overarching history of life.[9]

Iterations that constitute Gouldian repetitions do not support the 'critique from convergent evolution' because they fail to serve as evidence for three key premises of the robust replicability view (§3). First, they fail to show that certain design problems are pervasive in the history of life, because the design problems that prompt Gouldian repetitions may be shaped by the internal developmental parameters of the clade and thus may be highly restricted in scope. Second, Gouldian repetitions fail to show that the set of solutions to these design problems is highly circumscribed by forces other than conserved developmental constraints. Third, they fail to show that these restricted solutions are accessible to selection notwithstanding the internal constraints of phylogeny, because internal constraints are integral to defining and explaining Gouldian repetitions.

We should not assume, however, that all iterated regularities underwritten by developmental biases constitute Gouldian repetitions. For instance, the hammerhead ribozyme, a catalytic RNA that undergoes self-cleavage, has evolved convergently many times, quite possibly as a result of biases that are intrinsic to genotype–phenotype maps [45]. Thus, we should not assume that deep replicability in macroevolution hinges entirely on the power of selection, even though convergence is often taken to be evidence of selection's power.

This discussion raises a thorny problem: how can we determine empirically whether a given iteration is a case of Gouldian repetition or a case of true convergence? One workable approach would be to infer a degree of developmental independence that is proportional to phylogenetic distance. Let us call this the 'taxonomy heuristic'. The taxonomy heuristic helps explain why a small number of convergences across kingdoms, phyla and classes tend to indicate deeper counterfactual stability than, say, numerous convergences confined to families, genera and species. The trouble with such a heuristic, however, is that phylogenetic distance is not always a good indicator of developmental difference when it comes to evolutionary iteration. It is now well established that convergence between higher taxonomic groupings can be caused by the activation of 'deep homologues', developmental substrates conserved across vast phylogenetic distances. We must therefore develop more precise ways of assessing the independence of convergent episodes and not rely exclusively on the blunt instrument of phylogenetic distance (§5).

A further challenge relates to quantifying the *degree* of convergence, which one might think crucial to the contingency debate. As with counting the number of repetitions, however, merely quantifying the degree of convergence tells us little about the evolutionary robustness of the regularity observed. Weaker degrees of convergence that transcend the body plans of disparate clades (such as viviparity or filter-feeding or the fusiform shape) offer stronger evidence of deep robustness than do higher degrees of convergence within shared developmental constraints (such as Anolis lizard or dolphinoid ecomorphology).

5. Towards a taxonomy of convergence

We have argued that instances of convergent evolution can, under certain conditions, constitute natural experiments that bear on the RCT. However, we have also shown that the causes of iterated evolution are underdetermined in ways that affect the validity of the experimental set-ups. Lumping together causally heterogeneous convergent events has detracted from what might otherwise be a persuasive case against the RCT. To remedy this, we will distinguish biological iterations along several dimensions that have implications for their evidentiary relevance to the contingency debate. We will argue that convergent regularities exhibiting high levels of what we term *specificity*, *independence* and *scope* constitute the strongest evidence against the RCT. These categories apply to all types of evolutionary outcomes (not merely those surrounding the evolution of animal body plans), and thus they offer a way of evaluating the RCT read broadly as a thesis about the whole of life.

5.1. Specificity

It is widely appreciated that convergence and homology are relative to hierarchical level. For instance, a structure can be convergent at the level of morphology while being homologous at the level of tissues, proteins or genes. Many of the described cases of convergence are 'superficial' in the sense that the details of their structure and, especially, their underlying developmental architecture, can be dramatically divergent [46]. Contingency may dominate the molecular particulars of development while convergence reins at the level of general form. We might therefore want to distinguish convergences that occur on multiple levels simultaneously from those that occur at only a single level [47]. As we shall see, however, the fact that a specific structure was realized independently by completely different molecular-developmental pathways bespeaks its evolutionary robustness—and the more specific that structure is, the more it contradicts the RCT's claim that contingent history dominates at the level of details 'that have always defined the guts and soul of biology' [8, p. 1212].

Convergence is relative not only to hierarchical level, but also to the degree of *specificity* at which a given trait is delineated.[10] The 'specificity' of an iteration here refers to the number of shared character dimensions at a given level of biological organization—e.g. genes, proteins, cells, morphology, etc.—a measure that applies equally to homologies as it does to convergences. Generally speaking, one trait is more specific than another when it occupies a higher 'level' in a nested multiple realization base: e.g. blubber is more specific than insulation, and AFP type II is more specific than ice-binding protein. In this case, 'levels' refer to the degree of abstraction with which a trait is described, not to levels of biological organization. Insulation and blubber are traits at the same organizational level, but the former is less specific than the latter because it implicates fewer character dimensions. All else equal, the less specific a convergent regularity, the more likely it is to be evolutionarily robust.

Consider a sensory modality like hearing. The tympanal organ or 'ear' in insects (a vibrating drum-like structure) and its neuronal accouterments have evolved more than 20 times in the bodies, wings and legs of various insects to detect pressure waves generated by sound. A structurally similar ear has evolved in vertebrates using radically different developmental machinery—an instance of cross-phyla convergence on a highly physically constrained function that indicates the deep evolutionary robustness of this specific structural trait. However, if 'hearing' is defined more broadly to include not only drum-like structures, but any specialized capacity to detect sound vibrations in a substrate and to translate these stimuli into neuronal impulses used to generate adaptive behaviour, then the trait will be less structurally specific but far more evolutionarily robust. In this way, the multiple realizability of function can underwrite the evolutionary robustness of functions even while it undercuts the law-like replicability of form [48].

Generality, however, can come with a cost: vague regularities are typically less scientifically useful and interesting. Compare the following predictions about thermoregulation: 'organisms will evolve a covering with the property of thermal conservation' and 'organisms will evolve strategies for maintaining viable internal temperatures'. The former description admits of a limited set of possible realizations, whereas the latter, though highly evolutionarily robust, is so vague it might be realized by an indefinite disjunction of structural or functional configurations. In contrast, the anti-freeze protein AFP type II is a molecule with a very specific structure that has evolved a handful of times for the function of lowering the freezing temperature to prevent cell rupture [24, p. 190].[11] Because numerous proteins have ice-binding properties, however, 'anti-freeze' is a multiply realizable property that has been achieved through selection on different genes and molecular mechanisms, with a diverse range of proteins evolving in groups as disparate as animals, plants, fungi and prokaryotes [51].

A key point of weakness in the critique from convergent evolution relates to ambiguities regarding the specificity of the trait under consideration. Take, for instance, McGhee's formulation of the following regularity: 'If any large, fast-swimming organisms exist in the oceans of Jupiter's moon Europa. . .I predict with confidence that they will have streamlined, fusiform bodies; that is, they will look very similar to a porpoise, an ichthyosaur, a swordfish, or a shark' [24, p. 272]. The ambiguity here lies with the phrase 'look very similar'. If McGhee is making the weaker, less interesting claim that aquatic extraterrestrial 'animals' would have a fusiform shape, then, as we have seen, Gould would be unlikely to disagree. Yet at finer grains of resolution, variation between such course-grained convergences can be enormous [13,46]. If, on the other hand, McGhee is making the stronger and more interesting claim that such organisms would resemble sharks or porpoises in features *other than their fusiform morphology*, then we run into a problem of evidence. Squid have spindle shapes, but in all other respects (save for their camera-type eye) bear little resemblance to a shark or a porpoise. Because we see only 'dolphinoid' convergence *within the vertebrate body plan*—i.e. in Mesozoic marine reptiles, cetaceans and perhaps fish—this tells us little about the evolutionary robustness of these more specific outcomes across the whole of complex multicellular life. The same is true, for example, of convergence between marsupial and placental mammals, and within placentals, on a range of ecomorphologies throughout the Cenozoic [52]. Such observations do not clearly contradict Gould's thesis, because they may be caused by entrenched developmental constraints driving iterated evolution within clades.

This analysis of specificity helps to resolve key points of disagreement in the contingency discussion. Recall Gould's claim that universal generalizations about form and function do not speak to the details that fascinate biologists. We might reasonably interpret the RCT as claiming that only highly non-specific biological regularities about the evolution of form and function will be evolutionarily robust. Advocates of the robust replicability view, on the other hand, argue that some cases of convergence support robust regularities with fairly high levels of specificity.

Yet there remains an unsatisfactory vagueness in all of this. Because the RCT and its opposing thesis have been underspecified, it is difficult, in some cases, to determine at what degree of specificity convergence begins to contradict the RCT. We can say that iterated traits as specific as the camera-type eye are among the strongest (though still not definitive) cases against the RCT (Box 1), whereas others as vague as 'sensory apparatus' leave the RCT unchecked. Still others, such as urticating hair, fall into a greyer evidential zone. Yet the key to this discussion is not which evolutionary outcomes contradict the RCT, but which are robust in relevant ways and which are not. We can begin to answer this question by specifying the outcome in question and then considering the evidence for and nature of its evolutionary replicability.

5.2. Independence

The lumping problem as described above (§4) stems in part from the failure to take into account the independence of observed evolutionary iterations. By saying that evolutionary iterations vary in terms of their *independence*, we mean they vary in the degree of conserved molecular-developmental generators that they implicate. Ceteris paribus, a small number of

Box 1. Case study: image-forming eyes.

When faced with a particular convergent event, an analysis of its specificity, independence and scope can help us to assess its evidentiary relevance for the contingency debate. An example of a convergent regularity that arguably exhibits high levels of all of these features is the iterated evolution of image-forming eyes.

Specificity. The image-forming eye includes several distinct eye types—variations on camera and compound configurations—each of which is highly structurally specific, including particular cornea, lens and retina configurations. Both major eye types exhibit wide phylogenetic distribution, originating numerous times in distant animal phyla [53]. Camera-type eyes, for example, have evolved not only in vertebrates, arthropods, molluscs and cnidarians, but even in microbial eukaryotes [54].

Scope. Image-forming eyes are likely to have broad scope given the ubiquity of the light stimulus, the universal laws of optics, and the availability of substrates that can readily be coopted for optical functions. For example, a diverse set of crystallin proteins have been used for lens transparency, and the photoreceptor pigment rhodopsin is functionally optimized to trigger an electrophysiological cascade in response to individual quanta of light, suggesting that optimal solutions to this highly constrained design problem are accessible to selection.

Independence. The complex morphological arrangements characteristic of the major eye types are polyphyletic, and they are produced by diverse molecular developmental pathways in the phylogenetically disparate lineages in which they are found. However, there are two interrelated facts that might lead one to think these impressive iterations might be Gouldian repetitions. First, there is a growing consensus that the two cell types involved in photoreception were present in the ancestor of all bilaterians [55] and, on some accounts, descended from a single photoreceptive cell that originated once in early cyanobacteria and was subsequently taken up by eukaryotic cells in endosymbiotic events [56]. Second, all known eyes in animals are produced by certain conserved molecular sequences that are relied upon in development—such as deep homologues like ‘Pax-6’, which serve as ‘master control’ genes that trigger eye morphogenesis in groups as diverse as vertebrates and arthropods. Do these genetic and cell-type homologies prevent us from classifying the iterated evolution of complex eyes as cases of true convergence? There are several reasons why we believe they may not. First, as noted above, the weight of the evidence suggests that the ancestral bilaterian did not have complex eyes, even if it possessed Pax-6 and both types of photoreceptors [57], and thus describing ‘eyes’ as monophyletic would be misleading. Second, the presence of conserved molecular developmental homologues, such as Pax-6, does not in itself (contra Gould [8]) undermine the independence of such iterated outcomes for purposes of assessing the RCT. To convert such cases into Gouldian repetitions, it must be shown that the features subject to convergence result from developmental constraints that could easily have been otherwise. While it is true that Pax-6 is not functionally constrained and thus its conservation betrays the signature of historical contingency (like other upstream components of development that are refractory to selective modification), as gene manipulation studies show, Pax-6 is not causally responsible for the *specific* morphological features on which convergence judgements rest [10,11], and thus there is no reason to believe that Pax-6 plays a constraining role that would undermine the independence of complex eye evolution. Third, even if the photoreceptor cell evolved only once in the known history of life, the extent to which this fact undercuts the evolutionary robustness of complex eye evolution depends on whether the origination of photoreception is itself a radically contingent event—i.e. whether it is unlikely to be replicated across the vast majority of alternative histories of life. We cannot simply infer from the fact of singular origin that a trait is radically contingent, given the possibility of strong functional constraints (as with rhodopsin, above). At present, the most we can say is that the evolutionary replicability of basic photoreception remains an open question, although the replicability of complex eyes is robust if the existence of photoreception is taken among the relevant initial conditions. So even if complex eye evolution is not universally biologically projectable, it is certainly in tension with the RCT insofar as the latter is framed as a more limited thesis about what we can expect from replays of the tape of *multicellular* life on the *Earth*.

Evolutionary significance. The evolution of eyes was of great evolutionary significance. Eye-bearing species comprise a huge proportion of known animal diversity [53] and likely kickstarted active predation in the Cambrian [51], which in turn led to a cascade of strategic coevolutionary responses including (quite plausibly) the evolution of skeletons in numerous groups as well as the convergent evolution of competing eyes and associated expansions of nervous and musculoskeletal systems [57].

Multiple realizability. In the nested multiple realizability base, camera and compound eye configurations each represent more specific structural realizations of the vision adaptation, which in turn is a more specific realization of image formation. Image formation is greater in scope because it includes not only vision but also certain non-visual sensory modalities, such echolocation and electrolocation, which have also been the subject of evolutionary iterations. Due to constraints of the physical laws, there may be only a few types of waveform energy that organisms can use—such as light, electromagnetism and sound—to form complex, real-time topographic images of their environment. The few biological structures that are capable of harnessing this energy, and the brains necessary to process the information it provides, have been realized repeatedly in the history of life.

Evidential summary. This case supports key premises of the robust replicability view, namely that certain design problems are (i) pervasive in the history of life, (ii) solved by a limited number of structurally specific solutions, and (iii) accessible to selection notwithstanding the internal constraints of phylogeny.

highly independent evolutionary repetitions offers stronger evidence against the RCT than does a large number of non-independent repetitions. Independence is most directly assessed by laboratory investigation, such as through gene 'knock-out' studies that determine the specific causal role of the developmental mechanisms involved in the production of a given homoplasy, in conjunction with phylogenetic reconstructions of those mechanisms. Although a full accounting of independence requires causal and phylogenetic analyses of developmental mechanisms, where this information is unavailable, phylogenetic distance can be used as a tentative proxy by which to infer levels of independence.

According to the 'taxonomy heuristic' (§4), iterations with a narrow phylogenetic distribution (i.e. iterations limited to lower-level taxa) are, ceteris paribus, more likely to be Gouldian repetitions than cases of true convergence. Consider, for example, 'hummingbird flowers'—odourless, brick-red, trumpet-shaped flowers rich in nectar that have evolved to attract nectar-feeding birds nearly 20 times [24, p. 124]. Because this iteration is confined to low-level taxa, it gives us little confidence in the deep counterfactual stability or projectibility of the outcomes involved. In some cases, the specific conserved genetic homologues involved in a given phylogenetically narrow iteration have been identified. This is true for the iterated evolution of stickleback fish morphologies (see below), and will likely soon be true for the evolution of Anolis ecomorphologies whose conserved developmental underpinnings are now coming to light. On the other hand, the phylogenetic distribution of an iterated outcome like viviparity (the development of embryos within the body of the female) is wide, having evolved over twice as many times as hummingbird flowers in far more distant lineages [24, p. 86] using very disparate developmental machinery. Other iterated traits with fairly high levels of specificity are found in clades as developmentally disparate as plants and animals: an example is urticating hair, barbed bristles connected to cells that produce poisonous or acrid fluids, which are designed to break off and lodge in the skin of predators.

The concept of independence as it applies to convergent regularities poses its own set of difficulties, however. In addition to the limitations of using phylogenetic distance as a proxy for developmental distance (discussed in §4), there is also the problem that we must be able to determine which aspects of development are the relevant causes of a given homoplasy, because some homologous molecular-developmental mechanisms will nearly always exist. One way of doing this would be to identify 'causally specific' developmental homologues implicated in convergent regularities [11]. On this approach, the less casually specific the conserved developmental causes of a given iteration, the more independent that iteration would be. As others have pointed out [40], however, there is an ambiguity with regard to what constitutes the same or sufficiently similar developmental underpinnings, as this could refer to the same gene mutations, to mutations to the same gene region, and so on. These conceptual problems must be ironed out if we are to fully understand the evidentiary relevance of convergence to the contingency debate.

5.3. Scope

Convergent regularities also vary in terms of their *scope*, or the ubiquity of the conditions under which they obtain. If convergent regularities are limited by conditions that are astronomically uncommon in the universe and only accidentally obtain on the Earth, then these regularities will have a very narrow scope. This, in turn, will restrict their ability to serve as evidence for the robust replicability thesis and against the RCT, where these theses are interpreted as contradicting claims about the modal frequency of certain evolutionary outcomes (§2).

The limiting conditions for convergent regularities include (i) 'external' ecological factors, such as features of the abiotic environment like the drag of water or the presence of light, and features of the biotic environment like the existence of companion species with which to coevolve, and (ii) 'internal' developmental factors relating to the evolvability of the trait in question. Each of these conditions is expected to be applicable to different extents over the course of the Earth's evolutionary history and to other potential life worlds. Some limiting conditions, like the presence of light or water, are ubiquitous and thus do little to restrict the scope (and hence projectibility) of convergent regularities. Other limiting conditions are extremely rare or accidentally confined to narrow subsets of life on the Earth, thus resulting in a highly restricted scope that precludes exobiological projectibility.

The scope of a convergent regularity is only as wide as its most limiting condition, be it external or internal to the lineage in question. Because we are not currently capable of identifying all of the conditions that underpin a given regularity, in assessing scope the most limiting conditions should be our focus. For example, some convergent regularities, like the cantharophilous flowers discussed above, depend on the existence of morphologically specific pollinator lineages such as beetles—a form that so far as we know is a radically contingent outcome of life on this planet. So the repeated evolution of beetle-pollinated flowers, by virtue of this highly contingent ecological prerequisite, does not support exobiological projectibility, or even stability across deep Earthly rewinds of the tape of life. Contrast this case with, say, convergence in vessels to transport water, which evolved many fewer times than beetle-pollinating morphology but which, given the invariant physics of water transport, are universally projectible. Similarly, the biomechanics of fusiform shapes require the existence of life in a fluid environment and the internal structure necessary to maintain a sturdy shape—conditions that are likely widespread in environments conducive to the evolution of complex multicellular life.

Internal factors that limit the scope of convergent regularities are equally important, because they explain why Gouldian repetitions are neither projectible to other life worlds nor stable across deep replays of the known tape of life on the Earth—even though they underwrite impressive degrees of replicability across 'shallower' rewinds. For example, given constraints of the mammalian body plan and certain recurring ecological conditions, a range of specific mammalian forms may be highly replicable over a 55 million year period of evolution. Yet due to their internal developmental limiting conditions, these regularities may not be highly robust features of life on the Earth, let alone of other potential life worlds.

Nevertheless, Gouldian repetitions are to some degree projectible, and the extent of their projectibility depends in part on the phylogenetic prevalence of their conserved developmental mechanisms. For example, the iterated evolution of reduced pelvic armour in stickleback fish in isolated glacial lakes was accomplished via the selective deactivation of a common genetic homologue (the Pitx1 gene, which controls pelvic fin development) [58], indicating the stability of these forms

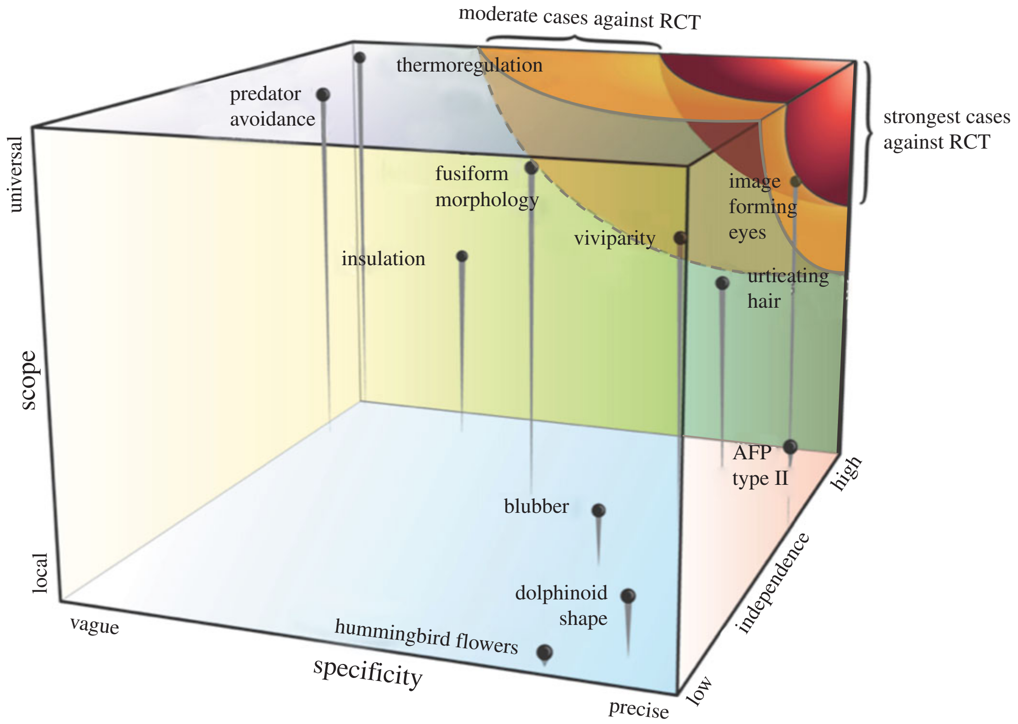

Figure 2. Here we illustrate three key dimensions of robustness with respect to which evolutionary regularities may be described. Each evolutionary iteration (labelled) is approximately located in a three-dimensional space along axes reflecting its degree of specificity, independence and scope, illustrating the extent to which it represents evidence against the RCT conceived either as a general thesis or as a particular thesis about specific evolutionary outcomes (with the strongest such cases found in the darkest shaded region). Some iterated regularities, such as hummingbird flowers or predator avoidance, do little to contradict the RCT, whereas others, such as image-forming eyes, constitute much stronger, though still not decisive, evidence against it. (Online version in colour.)

across very 'shallow' rewinds of the tape of life. These outcomes are robust only when the initial set-up includes the specific developmental parameters of this particular family of fish, which easily could have been otherwise. Mammalian ecomorphologies may be more evolutionarily robust as indicated by their somewhat deeper phylogenetic distribution, but they will still be of limited projectibility, even on the Earth. In contrast, the convergent evolution of salinity tolerance across the archeal and bacterial domains reflects a deeper rewind of the tape of life, controlling for internal limiting conditions and thus indicating a high degree of exobiological projectibility—though this outcome is arguably of sufficiently low specificity to be consistent with the RCT.

Couched in terms of scope, a key problem with the critique from convergent evolution is that it has failed to distinguish iterations of plausibly universal scope from those with more narrow limiting conditions. For instance, the evolution of fusiform shapes among mobile macrobes in aquatic environments has greater scope than that of dolphinoid shapes, given the additional body plan limiting conditions that underpin the latter regularity. By paying careful attention to the interrelated aspects of specificity, independence and scope, we can address the lumping problem and begin to distinguish convergences that speak to the RCT from those that do not (figure 2).

6. Conclusion

We have shown that many convergent regularities do not contradict the RCT, either because they are too vague (such as mimicry, thermoregulation or the fusiform shape), or because they depend on radically contingent internal constraints that make them Gouldian repetitions (such as Anolis, stickleback and mammalian ecomorphologies). We have sketched a taxonomy of natural experiments in convergent evolution with an eye toward identifying iterations that are maximally robust: outcomes that are expected to occur in most or all replays of the tape of life. Nevertheless, some worries remain.

One worry is that even if we were able to document an impressive number of true convergences that exhibit great specificity, independence and scope, there are clearly many more cases of non-convergence given similar initial conditions, suggesting that such outcomes are readily defeasible. For instance, the vast majority of lineages with eyespots (approx. 70) have never evolved an eye with a distinct lens, which suggests that non-convergence is far more common than convergence when it comes to the evolution of eyes. This, in turn, might be taken to undercut the law-like necessity of this particular macroevolutionary outcome. However, this conclusion does not follow, for as we saw earlier (§2), true convergence can be framed as a low-probability probabilistic phenomenon (covered, e.g. by statistical laws). Moreover, it is unclear how to determine which conditions must be similar in order to identify instances of non-convergence. The worry here is that any given outcome could be disqualified as a candidate for non-convergence on the grounds that the natural experiment was not similar in the right ways.

In any case, the ratio of convergent/non-convergent events given similar initial conditions is orthogonal to the task at hand, which is to evaluate the RCT. As we have seen (§2), the RCT does not claim certain evolutionary outcomes will be rare *in each history of life*; rather, it holds that there exist many important outcomes at particular levels of description that are *unlikely to be* (*re*)*produced at all* in thousands of histories of life. So even a handful of true convergences in a sea of non-convergence in a single history of life can provide powerful evidence against a strong version of the RCT—namely, the thesis that no important and sufficiently specific evolutionary outcomes are robustly replicable—as well as against *specific* contingency theses concerning particular outcomes.

Of course, we can never be certain about the robustness of an iterated regularity without an extraterrestrial dataset, because we cannot rule out the possibility that what is thought to be a case of true convergence is in fact the result of unknown hidden variables that make it a Gouldian repetition. Nevertheless, we have shown that there are strong candidates for true convergence that are of great scientific interest and importance in the history of life. In so doing, we hope to have laid the groundwork for a more effective use of convergent evolution as evidence in evaluating Gould's provocative thesis.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

Acknowledgements. We thank Austin Booth, Robert Brandon, Tyler Brunet, Adrian Currie, Ford Doolittle, Doug Erwin, George McGhee, Dan McShea, Gordon McOuat, Letitia Meynell, Louise Roth, audiences at the University of Cambridge, National Evolutionary Synthesis Center and Washington University-Saint Louis, as well as three anonymous referees, for comments on earlier drafts of this manuscript. Powell is grateful to John Templeton Foundation grant no. 43160 for financial support of this research.

Endnotes

[1] To pick but one example, the famously bizarre *Hallucigenia* has been identified as a stem onychophoran (velvet worm), a group closely related to arthropods [3].

[2] Our general focus on eukaryotic evolution is partially due to historical interest (this was Gould's focus), and partly due to the fact that prokaryotic evolution may be overwhelmed by lateral gene transfer [22], making convergent evolution hard to assess in the prokaryotic history of life [23].

[3] Although most discussions of convergent evolution focus on morphology, the concept is broad enough to encompass similarities in molecular and developmental structure, as well as functional and relational resemblances, irrespective of their particular structural realization.

[4] In his last and most comprehensive monograph, Gould contends that 'homologous developmental pathways can also be employed … as active facilitators of homoplastic adaptations that might otherwise be very difficult, if not impossible, to construct in such strikingly similar form from such different starting points across such immense phyletic gaps' [8, p. 1122–1123].

[5] As Okasha [32] points out, many classic experiments in physics, such as the crucial tests of general relativity, are essentially observational and do not involve human intervention in nature. At the same time, few studies in the historical sciences, such as palaeontology and geology, are purely observational, because they often involve controlled, systematic searches, guided by a wealth of background theory (such as plate tectonics and common descent), employ increasingly refined methods of data collection (such as dating and reconstructing fossils) and are carried out with the aim of testing hypotheses (such as the claim that a bolide impact triggered the end-Cretaceous extinction). Nor is there a clear epistemic asymmetry between observational and manipulational studies in terms of their tendency to generate knowledge about the causal structure of the biological world. In fact, Brandon [33] shows that controlled studies in evolutionary biology, which create highly artificial conditions in the laboratory, are often weakly projectable to the natural living world. Laboratory or field manipulation is therefore not a prerequisite for making justified inferences about evolutionary history.

[6] The concept of 'natural experiment' that we employ here corresponds, roughly, to what Morgan [35] calls 'nature's experiments'.

[7] There are several other major accounts of homology on offer in the literature, such as developmental and transformational accounts (see reference [37] for a review). We focus on the taxic account here because not only is it the dominant account, it is the account most often relied upon by convergence proponents in their critiques of the RCT.

[8] We introduce this distinction, rather than relying on the contested convergence-parallelism demarcation, because some accounts of parallelism would exclude many Gouldian repetitions, and because the theoretical motivations for delineating parallelism extend beyond the contingency debate.

[9] 'Developmental constraint' remains a contested concept in biology [41–43], and precisely how conserved developmental mechanisms that are relevant to a given iteration can be distinguished from broader developmental homology poses conceptual problems that remain unsolved [11,39,40,44]. These conceptual problems must be sorted out, however, if convergence is to serve as evidence against the RCT.

[10] Currie [29] reserves the term 'specificity' for the level of detail in a generalization and 'grain' as a measure of the similarity between existing traits (or environments).

[11] Although many examples of functional convergence lack specificity, in some cases functional traits can be specified with high precision—such as various jaw functions related to prey capture in fish [49], and specific cleaning symbioses in crustaceans, fish and birds [50].

References

1. Gould SJ. 1990 *Wonderful life: the burgess shale and the nature of history*. New York, NY: W.W. Norton & Co.
2. Knoll AH. 2003 *Life on a young planet: the first three billion years of evolution on earth*. Princeton, NJ: Princeton University Press.
3. Yang J, Ortega-Hernández J, Gerber S, Butterfield NJ, Hou J-B, Lan T, Zhang X. 2015 A superarmored lobopodian from the Cambrian of China and early disparity in the evolution of Onychophora. *Proc. Natl Acad. Sci. USA* **112**, 8678–8683. (doi:10.1073/pnas.1505596112)
4. Brysse K. 2008 From weird wonders to stem lineages: the second reclassification of the Burgess Shale fauna. *Stud. Hist. Phil. Biol. Biomed. Sci.* **39**, 298–313. (doi:10.1016/j.shpsc.2008.06.004)
5. Desjardins E. 2011 Historicity and experimental evolution. *Biol. Philos.* **26**, 339–364. (doi:10.1007/s10539-011-9256-4)
6. Turner D. 2011 Gould's replay revisited. *Biol. Philos.* **26**, 65–79. (doi:10.1007/s10539-010-9228-0)
7. Beatty J. 1995 The evolutionary contingency thesis. In *Concepts, theories and rationality in the biological sciences* (eds JG Lennox, G Wolters), pp. 45–81. Konstanz, Germany: University of Konstanz Press. and Pittsburgh, Penn.: University of Pittsburgh Press.
8. Gould SJ. 2002 *The structure of evolutionary theory*. Cambridge, MA: The Belknap Press of Harvard University Press.
9. Inkpen R, Turner D. 2012 The topography of historical contingency. *J. Philos. Hist.* **6**, 1–19. (doi:10.1163/187226312X625573)
10. Powell R. 2007 Is convergence more than an analogy? Homoplasy and its implications for macroevolutionary predictability. *Biol. Philos.* **22**, 565–578. (doi:10.1007/s10539-006-9057-3)
11. Powell R. 2012 Convergent evolution and the limits of natural selection. *Eur. J. Philos. Sci.* **2**, 355–373. (doi:10.1007/s13194-012-0047-9)
12. Beatty J, Desjardins E. 2009 Natural selection and history. *Biol. Philos.* **24**, 231–246. (doi:10.1007/s10539-008-9149-3)
13. Currie A. 2012 Convergence, contingency & morphospace. *Biol. Philos.* **27**, 583–593. (doi:10.1007/s10539-012-9319-1)
14. Haufe C. 2015 Gould's laws. *Philos. Sci.* **82**, 1–20. (doi:10.1086/678979)
15. Brandon RN, McShea D. 2011 *Biology's first law: the tendency for diversity and complexity to increase in evolutionary systems*. Chicago, IL: University of Chicago Press.
16. Sober E. 1997 Two outbreaks of lawlessness in recent philosophy of biology. *Philos. Sci.* **64**, S458–S467. (doi:10.1086/392622)

17. Mitchell SD. 1997 Pragmatic laws. *Philos. Sci.* **64**, S468–S479. (doi:10.1086/392623)
18. Sterelny K, Griffiths PE. 1999 *Sex & death: an introduction to philosophy of biology*. Chicago, IL: University of Chicago Press.
19. Travisano M, Mongold JA, Bennett AF, Lenski RE. 1995 Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* **267**, 87–90. (doi:10.1126/science.7809610)
20. Grant PR, Grant R. 2002 Unpredictable evolution in a 30-year study of Darwin's finches. *Science* **296**, 707–711. (doi:10.1126/science.1070315)
21. Goldberg EE, Igić B. 2008 On Phylogenetic Tests of Irreversible Evolution. *Evolution* **62**, 2727–2741. (doi:10.1111/j.1558-5646.2008.00505.x)
22. Doolittle WF. 1999 Phylogenetic classification and the universal tree. *Science* **284**, 2124–2128. (doi:10.1126/science.284.5423.2124)
23. Mariscal C, Doolittle WF. 2015 Eukaryotes first: how could that be? *Phil. Trans. R. Soc. B* **370**, 20140322. (doi:10.1098/rstb.2014.0322)
24. McGhee G. 2013 *Convergent evolution: limited forms most beautiful*. Cambridge, MA: MIT Press.
25. Conway Morris S. 2003 *Life's solution: Inevitable Humans in a Lonely Universe*. Cambridge, UK: Cambridge University Press.
26. Vermeij GJ. 2006 *Evolution and escalation: an ecology history of life*. Princeton, NJ: Princeton University Press.
27. Beatty J. 2006 Replaying life's tape. *J. Philos.* **103**, 336–362. (doi:10.5840/jphil2006103716)
28. Dennett DC. 1995 *Darwin's dangerous idea: evolution and the meaning of life*. New York, NY: Simon & Schuster.
29. Currie AM. 2013 Convergence as evidence. *Br. J. Philos. Sci.* **64**, 763–786. (doi:10.1093/bjps/axs027)
30. Gould SJ, Lewontin RC. 1979 The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**, 581–598. (doi:10.1098/rspb.1979.0086)
31. Bernard C. 1957.1927 *An introduction to the study of experimental medicine*. New York, NY: Dover Publications.
32. Okasha S. 2011 Experiment, observation and the confirmation of laws. *Analysis* **71**, 222–232. (doi:10.1093/analys/anr014)
33. Brandon RN. 1997 Does biology have laws? The experimental evidence. *Philos. Sci.* **64**, S444–S457. (doi:10.1086/392621)
34. Diamond JM. 1986 Overview: laboratory experiments, field experiments, and natural experiments. In *Community ecology* (eds J Diamond, TJ Case), pp. 3–22. New York, NY: Harper and Row.
35. Morgan MS. 2013 Nature's experiments and natural experiments in the social sciences. *Phil. Soc. Sci.* **43**, 341–357. (doi:10.1177/0048393113489100)
36. Losos JB, Jackman TR, Larson A, De Queiroz K, Rodriguez-Schettino L. 1998 Contingency and determinism in replicated adaptive radiations of Island Lizards. *Science* **279**, 2115–2118. (doi:10.1126/science.279.5359.2115)
37. Wiley EO, Lieberman BS. 2011 *Phylogenetics: theory and practice of phylogenetic systematics*. Hoboken, NJ: Wiley-Blackwell.
38. Ramsey G, Peterson A. 2012 Sameness in biology. *Philos. Sci.* **79**, 255–275. (doi:10.1086/664744)
39. Currie AM. 2014 Venomous dinosaurs and rear-fanged snakes: Homology and homoplasy characterized. *Erkenntnis* **79**, 701–727. (doi:10.1007/s10670-013-9533-5)
40. Pearce T. 2011 Convergence and parallelism in evolution: a Neo-Gouldian account. *Br. J. Philos. Sci.* **63**, 429–448. (doi:10.1093/bjps/axr046)
41. Wagner GP. 1988 The influence of variation and of developmental constraints on the rate of multivariate phenotypic evolution. *J. Evol. Biol.* **1**, 45–66. (doi:10.1046/j.1420-9101.1988.1010045.x)
42. Amundson R. 1994 Two concepts of constraint: adaptationism and the challenge from developmental biology. *Philos. Sci.* **61**, 556–578. (doi:10.1086/289822)
43. Sansom R. 2003 Constraining the adaptationism debate. *Biol. Philos.* **18**, 493–512. (doi:10.1023/A:1025581622161)
44. Arendt J, Reznick D. 2007 Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32. (doi:10.1016/j.tree.2007.09.011)
45. Greenbury SF, Johnston IG, Louis AA, Ahnert SE. 2014 A tractable genotype–phenotype map modelling the self-assembly of protein quaternary structure. *J. R. Soc. Interface* **11**, 20140249. (doi:10.1098/rsif.2014.0249)
46. Griffiths PE. 2007 Evo-devo meets the mind: towards a developmental evolutionary psychology. In *Integrating evolution and development* (eds R Brandon, R Sansom). Cambridge, UK: Cambridge University Press.
47. Donley JM, Sepulveda CA, Konstantinidis P, Gemballa S, Shadwick RE. 2004 Convergent evolution in mechanical design of lamnid sharks and tunas. *Nature* **429**, 61–65. (doi:10.1038/nature02435)
48. Rosenberg A. 1985 *Structure of biological science*. Cambridge, UK: Cambridge University Press.
49. Wainwright PC, Alfaro ME, Bolnick DI, Hulsey CD. 2005 Many-to-one mapping of form to function: a general principle in organismal design? *Integr. Comp. Biol.* **45**, 256–262. (doi:10.1093/icb/45.2.256)
50. Poulin R, Grutter AS. 1996 Cleaning symbioses: proximate and adaptive explanations. *Bioscience* **46**, 512–517. (doi:10.2307/1312929)
51. Ewart KV, Lin Q, Hew CL. 1999 Structure, function and evolution of antifreeze proteins. *Cell. Mol. Life Sci.* **55**, 271–283. (doi:10.1007/s000180050289)
52. Martin LD, Meehan TJ. 2005 Extinction may not be forever. *Naturwissenschaften* **92**, 1–19. (doi:10.1007/s00114-004-0586-9)
53. Land MF, Fernald RD. 1992 The evolution of eyes. *Annu. Rev. Neurosci.* **15**, 1–29. (doi:10.1146/annurev.ne.15.030192.000245)
54. Leander BS. 2008 A hierarchical view of convergent evolution in microbial eukaryotes. *J. Eukaryotic Microbiol.* **55**, 59–68. (doi:10.1111/j.1550-7408.2008.00308.x)
55. Arendt D. 2003 Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* **47**, 563–571.
56. Gehring WJ. 2004 Historical perspective on the development and evolution of eyes and photoreceptors. *Int. J. Dev. Biol.* **48**, 707–717. (doi:10.1387/ijdb.041900wg)
57. Erwin DH, Valentine JW. 2013 *The Cambrian explosion: the construction of animal biodiversity*. Greenwood Village, CO: Roberts and Co.
58. Foster SA, Baker JA. 2004 Evolution in parallel: new insights from a classic system. *Trends Ecol. Evol.* **19**, 456–459. (doi:10.1016/j.tree.2004.07.004)