



Okasha, S. (2018). The strategy of endogenization in evolutionary biology. *Synthese*, 2018, 1-23. Advance online publication. <https://doi.org/10.1007/s11229-018-1832-6>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1007/s11229-018-1832-6](https://doi.org/10.1007/s11229-018-1832-6)

[Link to publication record on the Bristol Research Portal](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer at <https://link.springer.com/article/10.1007/s11229-018-1832-6> . Please refer to any applicable terms of use of the publisher.

University of Bristol – Bristol Research Portal

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/brp-terms/>

The strategy of endogenization in evolutionary biology

Samir Okasha¹ 

Received: 5 December 2017 / Accepted: 24 May 2018
© The Author(s) 2018

Abstract Evolutionary biology is striking for its ability to explain a large and diverse range of empirical phenomena on the basis of a few general theoretical principles. This article offers a philosophical perspective on the way that evolutionary biology has come to achieve such impressive generality, by focusing on “the strategy of endogenization”. This strategy involves devising evolutionary explanations for biological features that were originally part of the background conditions, or scaffolding, against which such explanations take place. Where successful, the strategy moves biology a step closer to the ideal of explaining as much as possible from evolutionary first principles. The strategy of endogenization is illustrated through a series of biological examples, historical and recent, and its philosophical implications are explored.

Keywords Evolutionary biology · Reductionism · First principles · Strategy of endogenization

1 Introduction

One striking feature of evolutionary biology is its ability to explain a diverse range of empirical phenomena on the basis of a few general theoretical principles. The basic Darwinian argument, that natural selection acting on heritable variation will lead organisms to evolve adaptations to their environment, turned out to be able to explain quite a lot, as Darwin anticipated. The integration of Darwin’s principles with Mendelian genetics in the early years of the twentieth century, and the ensuing “modern evolutionary synthesis”, further increased the explanatory power of evolutionary

✉ Samir Okasha
samir.okasha@bristol.ac.uk

¹ Department of Philosophy, University of Bristol, Bristol BS6 6JL, UK

biology. Today opinions differ about whether the synthesis is still adequate, and about which if any additional principles should be added to its toolkit.¹ But whatever one's view on this, it seems clear that compared with most of the biosciences, which are typically data-rich but theory-poor, evolutionary biology comes closer to the philosopher's ideal of subsuming many particular facts under as small a theoretical umbrella as possible. In this respect, though not others, it is perhaps the branch of biology that is most like physics.

My aim in this article is to offer a philosophical perspective on the historical trajectory of evolutionary biology, by focusing on what I call the “strategy of endogenization”. This refers to a particular way in which the generality of evolutionary theory has been increased over time. It involves devising evolutionary explanations for biological features that were originally part of the background conditions, or scaffolding, against which such explanations took place. Where successful, the features in question cease to be part of the background and are brought within the fold of evolutionary theory, or endogenized. One example is fair meiosis, or Mendelian segregation. Most population-genetic models of evolution take for granted that each gene in a diploid organism has a 50% chance of being transmitted to any gamete; this fact plays a key role in the models' workings, and it has important evolutionary consequences. However as was eventually realized, there is a question as to *why* meiosis is usually fair. When evolutionists tried to answer this question, they thereby endogenized fair meiosis—it moved from background assumption to something that receives an evolutionary explanation in its own right. Another example is hierarchical organization. That biological entities form a nested hierarchy (gene, cell, organism, group, etc.) is a presupposition of much evolutionary analysis. But in recent years there has been a move to endogenize this hierarchy, by explaining how it evolved in the first place. Many other examples are discussed below.

The strategy of endogenization is a useful concept for understanding how evolutionary biology has increased its explanatory power over time, for three reasons. Firstly, it highlights a common methodological thread running through diverse scientific developments. Indeed a number of the “big themes” in recent evolutionary discussions, such as evolvability, major transitions, and niche construction, involve an element of endogenization. Secondly, the endogenization concept is relevant for philosophical debates about how best to characterize the “essence” of evolution by natural selection. The point here is that some candidates, e.g. the “replicator–interactor” framework, while purporting to offer a fully general analysis of Darwinian evolution, in fact rely on evolved features that, in a fuller analysis, should be endogenized. Thirdly, endogenization is directly relevant to the theme of this special issue: first principles in science. For each time that a background feature is endogenized, this moves biology a step closer towards the ideal of explaining as much as possible from evolutionary first principles.

To appreciate this third point, note the pervasive tension within evolutionary biology between the generality of the core Darwinian principles, which can be stated abstractly, and the day-to-day applications of those principles, which are to biolog-

¹ See the debate in *Nature* entitled “Does evolutionary theory need a re-think?” (Laland et al. 2014).

ical systems with numerous contingent features, e.g. DNA-based storage of genetic information, sexual reproduction, and germ-soma separation, to pick just three. The tension is heightened because these features often function as enabling conditions, explicit or implicit, in the specific explanations that biologists construct. By using the strategy of endogenization, these features can sometimes be brought within the scope of evolutionary explanation, thus reducing the number of primitives that evolutionary biology has to assume. In this way, endogenization serves to implement a form of reductionism, as the core Darwinian principles come to assume a larger and larger explanatory burden.

The structure of this paper is as follows. Section 2 explains the strategy of endogenization in more detail, distinguishing it from other ways in which the scope of evolutionary theory has been increased. Section 3 explores six successful applications of the endogenization strategy, to a range of biological phenomena. Section 4 considers the limits of endogenization, while Sect. 5 relates endogenization to debates about the “essence” of Darwinian evolution. Section 6 discusses first principles and reductionism. Section 7 concludes.

2 What is the strategy of endogenization?

I borrow the term “endogenization” from economics, where a distinction is customarily drawn between endogenous and exogenous variables in a theoretical model. The former are ones whose values are determined by the model in question, while the latter are treated as given. Thus for example, the aggregate demand for a commodity depends on both the commodity’s price and on consumers’ income; but in the elementary model of supply and demand, price is an endogenous variable while income is exogenous. This is because the model itself tells us what the equilibrium price will be, but says nothing about what consumers’ income will be. Of course, there must be *some* story about what determines income levels, but it lies outside the purview of the model in question. Thus the endogenous/exogenous distinction is inherently model-relative. In some cases it is possible to endogenize a variable that was previously treated as exogenous, by devising a more general model.

I claim that in the historical development of evolutionary biology, we see a successive endogenization of variables that previous theorists had treated as exogenous. As a result, the scope of evolutionary biology has been successively expanded, as evolutionary explanations were devised for biological features that were previously part of the unexplained background against which such explanations were constructed. These features are diverse in character. They include: (i) aspects of genetics, e.g. fair meiosis, dominance, linkage; (ii) aspects of the inheritance system, e.g. high-fidelity replication, a single-celled bottleneck, non-inheritance of acquired characters; (iii) aspects of reproduction, e.g. sex, recombination, gamete dimorphism; (iv) developmental features, e.g. modularity, germ-line sequestration, the genotype–phenotype map; (v) population-level features, e.g. genetic variation, dispersal, population subdivision; (vi) large-scale features of biological systems, e.g. hierarchical organization.

The motivations for pursuing endogenization are various. In some cases it stems from the realization that a particular feature is not universal, or exhibits previously

unsuspected variation. For example, the discovery that rates of mutation and recombination vary between genetic loci and between species prompted biologists to search for adaptive explanations of these rates. Similarly, the discovery that sexual reproduction sometimes involves fusion of equal-sized gametes prompted biologists to wonder why gamete dimorphism—the more usual arrangement—had evolved. In other cases, though a given feature is known not to be universal, the drive to endogenize arises from the realization that existing evolutionary biology tacitly makes assumptions which limit its applicability to biological systems lacking the feature in question. Thus for example, Buss (1988) makes a persuasive case that neo-Darwinian evolutionary theory really only works well for non-clonal multi-celled organisms which sequester their germ line in early ontogeny, attributes which many taxa lack. Buss's argument played a major role in prompting biologists to treat "individuality" as something to be explained, rather than assumed.

Endogenization comes in two varieties, depending on the sort of evolutionary explanation that the biological feature in question receives. In some cases the feature is explained in terms of the fitness benefit that it confers on an individual organism (or possibly, a whole group or species). Thus for example, Eshel's famous explanation for why meiosis is usually fair identifies a selective advantage that will accrue, in certain circumstances, to an organism containing a modifier gene that restores fair Mendelian segregation at other genetic loci at which it has broken down (Eshel 1985). In other cases, the feature is explained as an unselected side-effect. Thus for example, one explanation for why most species contain so much standing genetic variation is that it is a side-effect of adaptation to a heterogeneous environment. So here, a biological feature—within-species genetic variation—is accounted for in evolutionary terms, but not because of any adaptive benefit that it itself brings.

To better understand what the strategy of endogenization involves, it helps to distinguish it from other ways in which evolutionary biology has increased its explanatory scope over time. Firstly, endogenization is not merely the extension of evolutionary theory to new classes of phenomena (though it may involve this). Examples of the latter include the application of evolutionary principles to animal behaviour in the mid twentieth century, which led to the field of behavioural ecology, and to the human mind in the late twentieth century, which led to the field of evolutionary psychology. Such extensions are of course important, but they do not count as examples of endogenization as I am using the term. For the phenomena in question were not exogenous to begin with—they were not part of the background conditions against which other evolutionary explanations had been constructed. Rather, they were things about which previous evolutionary theory simply did not have much to say.

Secondly, endogenization isn't just a matter of making more complicated or realistic models. This occurred, for example, when evolutionary biologists incorporated environmental stochasticity into their models, rather than treating the environment as fixed as had been traditionally been done. This makes quite a big difference, demonstrating that the assumption of environmental fixity was critical to many of the explanations that evolutionary biology had devised.² But the increase in generality that resulted

² See Yoshimura and Clark (1993) ch.1 for a historical survey of the attempt to grapple with stochastic environments.

from incorporating environmental stochasticity was not a case of endogenization, for it did not involve giving an *evolutionary* explanation of something that had previously been taken for granted.

Finally, endogenization isn't the same as filling in a black box. It is common in science to black-box some questions in order to make progress with others. For example, Darwin black-boxed heredity in *The Origin*—he invoked the fact of parent-offspring resemblance without knowing how it came about. Similarly, the early neo-Darwinians black-boxed ontogenetic development—they assumed that the genes in an embryo affect the adult phenotype but without knowing how. In both cases, the black box was filled in as science progressed, by advances in transmission genetics and developmental biology respectively. This is not endogenization, since although parent-offspring resemblance and a systematic link between embryo genotype and adult phenotype were (crucial) background assumptions in the explanations that Darwin and the neo-Darwinians offered, the explanation of why these assumptions are valid was not itself evolutionary, but mechanistic.

The strategy of endogenization can be illustrated by a comparison of J. B. S. Haldane with R. A. Fisher, two of the major evolutionary theorists of the early twentieth century. Haldane and Fisher both pursued the integration of Darwinian evolution with Mendelian genetics, but in different ways.³ Haldane's approach was to use the known facts about Mendelian inheritance to construct models of the evolutionary process. In a series of ten papers entitled "A mathematical theory of natural and artificial selection", Haldane explored the evolutionary consequences of a range of possible assumptions concerning mating pattern, dominance/recessiveness, selection intensity and more. However Fisher went one step further, offering evolutionary explanations for the existence of dominance and linkage, rather than taking them as given. Fisher (1932) described Haldane's approach as "deductive" and his own as "inductive". While Haldane used known genetic facts as parameters in his evolutionary models, Fisher's view was that "genetics supplies the facts as to living things as they now are, facts which ... have an evolutionary history and may be capable of an evolutionary explanation ... facts which are not immutable laws of the workings of things but which might have been different had evolutionary history taken a different course" (1932 p. 165). This is a neat expression of the philosophy behind the endogenization strategy.

With these preliminaries in place, we are in a position to explore the strategy of endogenization in more detail. How often does the strategy succeed? By what means can the strategy be implemented? What are its logical limits? Does endogenization teach us anything about the conceptual structure of evolutionary theory? Is a failure to endogenize always a shortcoming, or can it be justified by the necessity of holding some factors constant in order to study others? I probe these questions through a series of biological examples.

³ This contrast between Haldane and Fisher is discussed by Edwards (2011), to whom I am indebted in this paragraph.

3 Endogenization: some success stories

3.1 The origin of variation

One important historical example of endogenization concerns the treatment of *variation* in evolutionary biology. Darwin himself emphasized that variation was an essential ingredient of evolution by natural selection—if organisms do not vary, there is nothing to select between. From his field observations, Darwin knew that abundant variability was the norm in natural populations, but he did not understand where it came from, nor why it was maintained. Indeed he famously worried that sexual mixing would quickly render a population homogenous. This worry was defused with the advent of Mendelism, which showed that sexual mixing has no inherent tendency to destroy genetic variation, given the particulate nature of inheritance (i.e. the fact that genes are inherited as discrete particles).⁴ But a deeper worry persisted, which is that natural selection itself will typically *destroy* variation. The selective preservation of some variants and the elimination of others will reduce the total variance, genotypic and phenotypic, in a population, other things being equal. So paradoxically, Darwinian evolution eats up the very variation on which its continued operation depends. This paradox is heightened by the fact that repeated rounds of selection are required to produce the complex adaptations found in nature, as Darwin himself realized.

Given this situation, understanding the origins and maintenance of intra-specific variation quickly became an urgent task for evolutionary biology. The urgency increased in the 1960s and 70s with the discovery of extensive genetic polymorphisms in natural populations, thanks to novel experimental techniques such as protein electrophoresis. Today we understand that the ultimate source of new genetic variation is mutation; while sex and recombination lead to the production of offspring with novel combinations of genes. We have a detailed quantitative understanding of the effects of mutation, sex and recombination on the genetic variation in a population; and some understanding of how this genetic variation translates into phenotypic variation. Also, we have a variety of candidate explanations for why natural selection may sometimes preserve genetic variation rather than destroying it. These include overdominance (or heterozygote superiority), frequency-dependent selection, spatial and temporal variability in selection coefficients, and the long-term advantage to a species of being able to adapt quickly to environmental change. The net result is that between Darwin's day and today, variation has been endogenized: it has moved from being an unexplained presupposition of evolutionary theory to something which receives an evolutionary explanation in its own right.

This brief summary of how variation was endogenized does not do justice to the intellectual effort involved. Understanding the evolutionary forces that affect mutation, sex and recombination is a difficult task, that continues today. In a 1937 paper, the geneticist A. H. Sturtevant asked why the mutation rate does not evolve to zero, given that most mutations are harmful. He concluded that “no answer seems possible

⁴ This is shown by the Hardy–Weinberg equilibrium, which shows that in the absence of mutation and selection, random mating will lead the frequency distribution of diploid genotypes to reach a stable equilibrium.

at present, other than the surmise that the nature of genes does not permit such a reduction” (quoted in Sniegowski et al. 2000, p. 1064). Today we can say much more than this. Thanks to a combination of theoretical and experimental work, we know that the mutation rate can evolve, and does vary widely between taxa; that in certain circumstances, modifier genes that raise the rate of mutation (at other loci) can be selected; and that the mutation rate can typically be reduced, but only at a cost.

As regards sex and recombination, the situation is similar. Early attempts to explain these phenomena included Weismann (1886), who suggested that sex is valuable because it produces “individual differences” for natural selection to act on; and Fisher (1930) and Muller (1932), who argued that recombination is advantageous because it allows beneficial mutations that arise separately to be brought together in a single individual, thus speeding up the rate of adaptation. These ideas are still taken seriously today, but the field has moved on. There exist a variety of putative explanations of the adaptive advantage of sex, including DNA repair, parasite resistance, and producing offspring that are genetically heterogenous; and detailed mathematical models exploring the costs and benefits of recombination in different circumstances.⁵ Despite considerable empirical data, and experimental tests that attempt to separate the theories, no final consensus on why sex and recombination evolved has been reached. Thus despite the great advances made since Darwin’s day, the endogenization of variability represents unfinished business for evolutionary biology.

This example raises one important issue. Theories on the evolution of mutation, sex and recombination fall into two camps. Some posit a short-term advantage to the individual, while others posit a long-term advantage to the lineage or species. Thus creating a store of genetic variability to buffer against environmental deterioration is a long-term advantage of sex; while aiding DNA repair, or producing a genetically variable offspring brood, is a short-term advantage. This distinction has sometimes been obscured (in part because some early theorists discussed what are really individual-level advantages in terms of their effect on population mean fitness), but it is crucial. In general, positing individual-level advantage is preferable, given how natural selection works. Positing a lineage-level advantage is only valid if the evolutionary mechanism one has in mind is lineage-level selection, i.e. survival of some lineages and extinction of others. This does occur, but it is less common than individual-level selection and unlikely to explain the origin of complex adaptations (though may explain their maintenance.) This point is by now widely accepted, though it was not always so. Thus the route that the endogenization took, in this case, involved an increasing shift in favour of short-term explanations.

3.2 Males and females

A second example of the endogenization strategy concerns anisogamy, or gamete dimorphism. This describes species like our own in which there are two sexes: females who large produce gametes and males who produce small gametes. This has impor-

⁵ See Hartfield and Keightley (2012) for a fairly recent review of theories on the evolution of sex and recombination.

tant evolutionary consequences, since the optimal mating strategy for a male and a female will typically differ, given that small gametes can be produced in much greater number. Darwin himself realised this, in his pioneering discussion of “selection in relation to sex”, and it lies at the heart of modern sexual selection theory. In effect, this body of theory deals with the downstream evolutionary consequences of anisogamy for male/female differences, both morphological and behavioural. But there is a further question, which sexual selection theory does not answer, about why there exist two different sexes in the first place. (Note that this is not the same as asking why reproduction is sexual rather than asexual.) Though anisogamy is the rule in plants and animals, it is not the ancestral state, and many microbial and fungal species that reproduce sexually are isogamous (though often they have distinct mating types.) Recognition of this fact led biologists to consider how anisogamy might have evolved initially, that is, to endogenize it.

This problem was originally studied by Kalmus in the 1930s but came to fruition in the 1960s and continues to be studied today. The well-known work of Parker et al. (1972) provides one possible answer. They argued that there is a trade-off between the number and size of gametes that an organism can produce, and that larger gametes are generally fitter than smaller one, i.e. have a survival advantage. Given these assumptions, disruptive selection can quite easily lead, from an isogamous starting point, to distinct male and female castes, with different specializations. In effect, males specialize in making lots of low-fitness gametes, while females specialize in making a few high-fitness ones. On this theory, therefore, the inherent trade-off between gamete size and number implies a direct selective advantage for anisogamy; thus potentially explaining the existence of males and females.

This theory enjoys empirical support but does not fit all the known facts, so a number of rival explanations have been developed too.⁶ One is that anisogamy evolved in order to prevent intra-genomic conflict. Anisogamy ensures that all of the cytoplasmic elements in a fertilized zygote, such as mitochondria, derive from the mother—the male contributes only nuclear genes. This may well be advantageous, since if cytoplasmic elements derived from both parents, then conflict between these elements would be inevitable, as they would be genetically diverse, so would each promote its own interests at the expense of the organism. Anisogamy is thus the organism’s way of preempting such internal conflict, according to this explanation.

As in the case of sex and recombination, debate over the merits of the rival theories continues, so understanding the evolution of anisogamy is a work-in-progress. But the point to note is how, as evolutionary biology progressed, a biological feature went from being a presupposition of other evolutionary explanations, indeed the premise on which the whole field of sexual selection theory was built, to being something of which an evolutionary explanation was itself sought. This is a paradigm illustration of how endogenization can deepen the explanatory reach of evolutionary theory.

⁶ See Billiard et al. (2010) for a useful survey of the alternative theories.

3.3 Altruism and population structure

A third example of endogenization comes from social evolution theory. As is well-known, a major preoccupation in this field is altruistic or pro-social behaviour, that is, behaviour which reduces an organism's fitness but enhances that of others. Altruism is *prima facie* puzzling, as Darwin himself saw, since an organism which behaves altruistically will apparently be at a selective disadvantage. The basic solution to this puzzle is that for altruism to evolve by natural selection, the beneficiaries of the altruistic actions must have a statistical tendency to be altruists themselves. Thus non-random assortment is needed, i.e. altruists must tend to associate with each other. Such assortment can arise because of kinship, or because of population sub-division, or both. If organisms tend to interact with their kin, or with other members of their social group, this can generate the statistical association between altruist and recipient that is the precondition for altruism to evolve. Precursors of this idea can be found in Darwin (1879), Fisher (1930) and Haldane (1932), but it was first spelled out clearly by Hamilton (1964); it lies at the heart of the explanations for the evolution of altruism given by the modern theories of kin and multi-level selection.

In retrospect, we can see an interesting trend in social evolution theory over the last 40 years. Early treatments tended simply to posit the population structure, or pattern of assortative interaction, necessary for altruism to evolve. Thus in Hamilton's original theoretical work, it is simply assumed that social behaviour occurs between relatives; his "coefficient of relatedness", denoted r , which measures how closely altruist and recipient are related, is exogenous in his theoretical models. Something similar is true of early work in the group selection tradition, which simply starts from the assumption that populations are sub-divided into groups, such as colonies or demes, which are genetically different. This is not necessarily a problem, and could be justified on empirical grounds, particularly given that altruism does indeed appear to have evolved. But there is a deeper question about how the necessary population structure came into being in the first place. Social evolution theorists came to realize this, and thus tried to endogenize these features by constructing more complicated models.

A recent paper by Powers et al. (2011) illustrates this point nicely. They argue that many extant explanations of cooperation (or altruism) are incomplete, as they fail to account for the origin of the population structures that they invoke as premises in their explanation. To redress this, the authors construct a model in which cooperation and population structure co-evolve, and show that this makes a substantial difference to the ease with which cooperation can evolve. In a similar vein, other social evolutionists have devised models in which dispersal co-evolves with altruism; the point here is that how far an organism disperses from its parents determines how likely it is to interact with relatives, and thus whether the resulting population structure will be conducive to the evolution of altruistic behaviour or not.

This endogenizing trend is also reflected in how the coefficient of relatedness r has come to be treated in social evolution theory. Though Hamilton originally defined r in genealogical terms, as the probability that donor and recipient share genes that are "identical by descent", what really matters is statistical association between social partners, whether due to kinship or not (as Hamilton himself noted in a later 1975 paper). Modern treatments thus typically define r directly as a measure of statistical

association. This shift in the meaning of r is well-known, but it has one implication that is less well-known, namely that the value of r may change as the population evolves. Thus in a full evolutionary analysis, r needs to be treated as a dynamic variable, not a fixed parameter, which is how precisely how some recent models of social evolution treat the relatedness coefficient.⁷

In this example, the motivation to endogenize is somewhat different from in other examples. It reflects a realization that early explanations for the evolution of altruism were “too easy”, in that they had unwittingly helped themselves to something that was equally in need of explanation as altruism itself, namely the population structures that permit it to flourish. Thus it is not an abstract commitment to reductionism, nor to extending the scope of evolutionary theory, that drives the endogenization in this case, but rather the need to explain the evolution of altruism without begging the question.

3.4 Niche construction

A fourth example of endogenization concerns niche construction, currently a hot topic in evolutionary biology. This refers to the fact that organisms’ activities often modify the environment, of both their own and other species. The origins of this concept lie in Lewontin’s observation that in principle, there are two ways in which an adaptive fit between organism and environment can come about (Lewontin 1983). The organism can evolve adaptations to the environment via natural selection, or it can modify the environment within its own lifetime to suits its needs, as for example when a beaver builds a dam. Lewontin argued that traditional neo-Darwinism recognized only the first of these routes to achieving organism–environment fit, neglecting the second. This reflected a broader tendency in evolutionary biology to treat “the environment” as a given rather than as something that can itself undergo evolutionary modification.

Though in a sense the point that Lewontin was making is obvious—biologists have never denied that beavers build dams, after all—it is true that a lot is often packed into “the environment” in evolutionary explanations, and rather little said about how the environment came to be the way it is. This might be justified on pragmatic grounds: one cannot explain everything at once. However, in their influential book on niche construction, Odling-Smee et al. (2003) argue persuasively that organism-induced modifications of the environment cannot safely be ignored in evolutionary biology, even as a pragmatic expedient. They construct theoretical models to show that such modifications can generate novel evolutionary dynamics. Their flagship model is based on two-locus population genetics. One locus, with alleles E and e , affects an organismic behaviour which itself alters the level of a critical resource \mathbf{R} in the environment. The second locus, with alleles A and a , has an effect on organismic viability that depends on the amount of \mathbf{R} . Thus as niche-constructing behaviour evolves at the first locus, the environment is modified, and this in turn alters selection pressures at the second locus. Odling-Smee et al. adduce empirical evidence, from diverse taxa, for thinking that niche construction of this sort actually occurs. And they argue that in our own

⁷ See van Baalen and Rand (1998) for an example of such a model.

species, cultural niche construction can dramatically alter the rate and direction of human evolution.

Though the empirical significance of niche construction, and the extent to which the concept marks a break with traditional neo-Darwinism, are both matters of dispute, the basic logic of Odling-Smee et al.'s argument is surely correct. I suggest that their work be understood as a particular application of the strategy of endogenization, where the feature being endogenized is the selective environment itself. In effect, Lewontin's critique was that classical evolutionary models treat the environment as exogenous when really it should be endogenized. Niche construction theory addresses Lewontin's point by describing how the environment co-evolves with the organisms in it. This relaxes an idealization that is implicit in many classical models, and thus increases the generality of evolutionary theory.

In Sect. 2, we argued that endogenization can be of two different sorts, depending on whether the feature in question is explained as an adaptation or a side-effect. Interestingly, Odling-Smee et al.'s concept of niche construction covers both of these cases. They define the concept broadly, to include any organism-induced modification of the environment at all, whether of their own niche or of others. Thus beavers' building dams and spiders' spinning webs count as niche construction, but so too do micro-organisms causing oxygen to accrue in the atmosphere over millennia.⁸ In the former examples, the organism-induced modifications are beneficial for the organisms in question, or their offspring; while in the latter, they are side-effects, positive or negative, on the selective environment encountered by other species, possibly at a much later date. Thus niche construction involves endogenization of both sorts.

3.5 Hierarchical organization

A fifth example concerns hierarchical organization. It is a familiar observation that the entities biologists study are hierarchically arranged. How exactly "the" biological hierarchy should be characterized (and whether there is only one), is not entirely obvious, but a typical account goes something like "gene–chromosome–cell–multicelled organism–colony–population–species", where adjacent entities stand in a part-whole relation. The existence of hierarchical organization is not of course a recent discovery, nor one that evolutionary biology has ignored. Indeed, it lies at the heart of the "levels-of-selection" question in biology, which goes right back to Darwin. That question is about the level(s) of the hierarchy at which the process of natural selection occurs, and at which adaptations are found. However there has been a subtle shift in the way that hierarchical organization is conceptualized, in discussions of the levels-of-selection, which relates directly to the strategy of endogenization.

To see this, consider the way that the levels-of-selection question is traditionally set up, for example by Lewontin (1970). Firstly, it is noted that evolution by natural selection will operate on any entities satisfying three conditions: variability, resulting fitness differences, and heritability (or parent-offspring resemblance). Secondly, it is

⁸ Okasha (2005a) argues that that these two sorts of examples should not be lumped together under a single label (niche-construction), since their evolutionary logic is quite different.

noted that these conditions can be satisfied by entities at more than one hierarchical level, e.g. genes, organisms or groups; so in principle, selection and adaptation can occur at multiple levels. That is, it is the joint facts of hierarchical organization, plus the abstractness of the conditions required for Darwinian evolution, that generate the levels-of-selection question, on this view.

The problem with setting the issue up this way, as Griesemer (2000) insightfully notes, is that it takes the existence of the biological hierarchy for granted, as if hierarchical organization were simply an exogenously given fact about the organic world. But of course the biological hierarchy is itself the *product* of evolution, not something that was there at the dawn of life. Even cells, entities fairly low down the hierarchy as usually described, are highly complex, evolved entities; and eukaryotic cells, we now know, were originally formed by the merger of two prokaryotic cell lineages. Now presumably, there is an evolutionary story to be told about how this came about; and similarly for entities at other hierarchical levels, e.g. multi-celled organisms. So ideally, we would like an evolutionary theory that explains how the biological hierarchy came into existence, rather than taking it for granted.

Over the last twenty-five years or so, evolutionary biology has risen to this challenge. The burgeoning literature on “major transitions in evolution”, also known as “evolutionary transitions in individuality”, is in large part an attempt to grapple with the origins of hierarchical organization.⁹ As characterized by Maynard Smith and Szathmáry (1995), such transitions occur when a number of free-living entities, originally capable of surviving and reproducing alone, become aggregated into a higher-level unit, giving rise to a new hierarchical level; eventually, the lower-level entities lose the ability to survive and reproduce alone, and become parts of a whole. Examples include the transition from solitary replicators to networks of replicators enclosed in compartments, from independent genes to chromosomes, from unicellular to multicellular organisms, and from solitary organisms to colonies. The challenge is to understand these transitions in Darwinian terms. Why was it advantageous for the lower-level entities to sacrifice their individuality, cooperate with one another, and form a larger corporate body? And how could such an arrangement, once it arose, be evolutionarily stable?

The traditional levels-of-selection question reappears in the context of the evolutionary transitions, but in a subtly transformed way (cf. Okasha 2005b). In the traditional setting, the levels question was “synchronic”: it was about selection and adaptation at pre-existing hierarchical levels. In the evolutionary transitions literature, the levels question becomes “diachronic”: it is about the evolution of the hierarchy in the first place. One way to understand this change is to think of it as an application of the strategy of endogenization, where the feature that is being endogenized is hierarchical organization itself. Traditionally evolutionary biology treated this feature as exogenous, part of the background against which other explanations were constructed; but as the science progressed, and the evolutionary transitions discussion took off, the feature came to receive an evolutionary explanation of its own. This perspective, I

⁹ Key early works in this tradition include Buss (1988), Maynard Smith and Szathmáry (1995) and Michod (1999). For more recent discussion, see Sterelny and Calcott (2011), Bourke (2011), and Szathmáry (2015).

suggest, helps us to understand the (much-disputed) relationship between the older and the newer discussions of the levels-of-selection question.

This example has two interesting features. Firstly, it points to one way that the endogenization strategy can be implemented, namely by re-deploying theoretical tools originally designed for a different purpose. Biologists were readily able to understand the evolutionary transitions in Darwinian terms, because the principles of social evolution theory, which were originally designed to explain social behaviour in animals, turned out to have much broader applicability. The major themes of social evolution theory—cooperation, conflict, relatedness, division-of-labour—are relevant to each of the transitions, and thus provided a ready theoretical framework with which to endogenize hierarchical organization.

Secondly, this example shows how a lack of endogenization can sometimes be a source of scientific error. In the earlier discussions, in which the biological hierarchy was taken as given, many biologists were inclined to dismiss the levels-of-selection debate as a storm in teacup—arguing that in practice, selection on individual organisms is the most important factor in evolution, whatever about other theoretical possibilities.¹⁰ But as Michod (1999) stresses, multicellular organisms did not come from nowhere; they evolved from aggregations of single cells, so in a sense are just highly cooperative cell groups. Thus levels of selection other than that of “the individual organism” must have existed in the past, whether or not they still operate today. Once the hierarchy is endogenized, the argument that individual selection “is all that matters in practice” is seen to be unsustainable.

3.6 The genotype–phenotype map

Our sixth example concerns the genotype–phenotype (G–P) map, that is, the way in which changes in an organism’s genotype are translated into changes in its phenotype. The importance of this map for Darwinian evolution has long been recognized. Thus in an influential paper, Lewontin (1978) argued that in order for organisms to evolve adaptations by cumulative natural selection, a condition called “quasi-independence” must be met. This means that it is possible to modify one organismic trait without too much of an effect on its other traits; in other words, pleiotropic effects should not be too pervasive.¹¹ Were this not so, natural selection would not be able to gradually hone a trait for a particular function, for the probability of the necessary adaptive variants arising by mutation would simply be too low. Since many organisms have evolved a close adaptive fit to the environment, along multiple trait dimensions, this suggests that the G–P map of many species does in fact satisfy quasi-independence, at least approximately.

Lewontin’s argument is insightful, and makes a point that is surely correct. However with the wisdom of hindsight there is something unsatisfactory about his discussion. For while he argues persuasively that quasi-independence is a hidden presupposition

¹⁰ Thus for example C. H. Waddington described the original levels-of-selection debate as “a rather foolish controversy” (quoted in Maynard Smith 1976 p. 277).

¹¹ Pleiotropy means that a given gene affects more than one phenotypic trait.

of Darwinian explanations, he does not consider why the presupposition holds good. That is, Lewontin's analysis tells us what the G–P map must be like in order that natural selection be capable of producing complex adaptations, but does not address the question of how the map got to be like that in the first place. Is it simply a lucky accident, or did it itself evolve from an ancestral state in which it was different? Lewontin does not tell us. Ideally, we would like our evolutionary theory to endogenize the G–P map, rather than to treat it as a given.

Again, recent evolutionary biology has risen to this challenge, in the literature on the evolution of evolvability.¹² Evolvability refers to the capacity of a lineage to undergo sustained adaptive change over time. In a pioneering discussion, Wagner and Altenberg (1996) identified a number of conditions on the G–P map that are necessary for evolvability, which incorporate but go beyond Lewontin's quasi-independence. They include: that mutations generate a continual supply of adaptive variants; that multiple simultaneous mutations are not needed to improve a trait; and that traits can be modified independently of each other. In general, these conditions are facilitated by an organism's genetic architecture being *modular*. This means that many genes within the genome have phenotypic effects that are confined to a limited number of developmental pathways, or modules, so do not affect the whole organism's phenotype at once.

Wagner and Altenberg stress, surely correctly, that evolvability is not inevitable: it is easy to conceive of genetic architectures, or G–P maps, that would frustrate the capacity of a lineage to undergo sustained evolutionary change. More controversially, they suggest that evolvability should be thought of as itself an adaptation, on the grounds that the G–P map "is under genetic control" (p. 969). They adduce evidence for thinking that mutations at one genetic locus can affect the way in which genotypic variation maps onto phenotypic variation at other loci, e.g. by increasing the degree of developmental canalization, that is, buffering the organism's phenotype against genetic and environmental perturbations. If Wagner and Altenberg are right, this suggests that the G–P map can be endogenized in the first way, i.e. that it admits of an adaptive explanation in its own right.

An alternative view is that evolvability, and/or the G–P map that underlies it, should be thought of as a byproduct rather than a direct target of selection, i.e. that it can only be endogenized the second way (Pigliucci 2008). This view may seem preferable given the widely-agreed point that natural selection cannot anticipate the future. Surely the capacity of a lineage to undergo future evolution is not the sort of thing that can be directly selected for in the present? Though this is correct, note that selection at the lineage level is a possibility, at the time-scales relevant for the evolution of the G–P map. That is, maybe lineages whose G–P maps did not lead them to be sufficiently evolvable went extinct, as they were unable to adapt when the environment changed, leaving behind only lineages with G–P maps which facilitate evolvability?

The underlying issue here—side-effect versus individual-level benefit versus lineage-level benefit—is a matter of ongoing debate in the literature on the evolution of evolvability; here is not the place to attempt a resolution. Whichever of these

¹² See for example Wagner and Altenberg (1996), Hansen (2006), and Pigliucci (2008).

views turns out to be correct, the point to note, again, is how a biological feature went from being an implicit presupposition of most other evolutionary explanations to receiving one of its own.

4 The limits of endogenization

The foregoing examples show that the strategy of endogenization is fairly pervasive in evolutionary biology; it is a thread that runs through a number of important developments in the field. This raises an immediate question. How far can the strategy be pushed? Are there cases where it cannot be applied?

One limit arises because not all biological features have an evolutionary explanation (whether as adaptations or side-effects). For they may be inevitable consequences of other scientific laws, so never have been different in the first place. In certain cases, such features may nonetheless be an important part of the background against which evolutionary explanations are constructed. To illustrate, consider life-history theory, the branch of evolutionary biology which studies how organisms apportion their reproductive effort over their lifetime. Models in life-history theory invariably assume a negative trade-off between survival and reproduction—the better an organism is at one, the less good it will be at the other.¹³ This trade-off is fundamental to how the models work, so certainly counts as a background assumption that has shaped other evolutionary explanations; indeed if survival and reproduction did *not* exhibit a trade-off, the consequences for life-history evolution would be quite dramatic. But it would not be possible to apply the endogenization strategy here. The existence of this trade-off is not an evolved feature of modern organisms; rather its universality reflects a physical constraint, namely that organisms cannot expend energy on reproduction and on trying to survive at the same time.

Though this is true, it is debatable how much of a limit to the endogenization strategy it really represents. For even in this example, the particular *shape* of the survival–reproduction trade-off, e.g. whether it is convex or concave, is something that can and does evolve; and this shape plays a crucial role in the predictions of life-history theory. (A convex trade-off favours iteroparity, i.e. many reproductive episodes in an organism’s lifetime, while a concave trade-off favours semelparity, i.e. reproducing just once.) Now the convexity or concavity of the survival–reproduction trade-off depends on the mortality and fertility schedule of the species in question, and these features do of course respond to natural selection, hence can change over time. So in short, although the existence of the trade-off is not a candidate for endogenization, its precise nature, or form, certainly is.

More generally, the importance or otherwise of structural constraints, that derive from physical or biophysical laws, in constraining the organic forms that can evolve is an old debate in evolutionary biology, that has re-surfaced repeatedly; here is not the place to attempt a resolution.¹⁴ Certainly if such constraints are pervasive, it follows

¹³ See Stearns (1992) for an introduction to life-history theory.

¹⁴ See Hansen (2015) for a good survey of the role of constraints in evolution, including discussion of the “structuralist” opposition to neo-Darwinism.

that the set of biological features that do not admit of evolutionary explanation, because they could not have been otherwise (in some fairly robust sense of “could”) will be quite large. However even so, the endogenization strategy will often play an indirect role in helping us to understand constrained features. For distinguishing such features from universal features of modern organisms that owe their universality not to constraints but to stabilizing selection, and thus *do* admit of an evolutionary explanation, is not easy. (The genetic code is a possible example of the latter: alternative codes are not prohibited by the laws of nature, but stabilizing selection has worked against them.) Indeed it may only be possible to draw this distinction by trying and failing to construct an evolutionary explanation for a given feature; in which case the attempt to endogenize, even if unsuccessful, will play an important role.

A quite different sort of limit arises simply because evolution by natural selection must have begun somewhere. As work on chemical or pre-biotic evolution has taught us, a type of Darwinian process was likely at work on earth long before anything distinctively biological had arisen, involving simple molecules with the ability to self-replicate. Given this ability, the conditions necessary for a Darwinian process to occur—variation, differences in replication rate, and heritability—were satisfied, so an evolutionary process could get underway, eventually giving rise to something like modern RNA. Inevitably our knowledge of how this happened is partial, but it is clear that the very first self-replicating molecules must have arisen by non-Darwinian means. For this to happen, the necessary chemical building blocks must have been present in the primeval soup; their presence thus represents a feature that, as a matter of logic, could not be endogenized by evolutionary theory.

This point is clearly correct, however its practical import is fairly small. For although abiogenesis and pre-biotic evolution are important topics, relatively little is known about them, and they are not at the forefront of modern evolutionary research. That Darwinian evolution must have had a beginning is perfectly true, and represents a logical limit on the endogenization strategy; however this is quite compatible with the main thesis of this essay, namely that the strategy represents one important way in which evolutionary theory has successively expanded its explanatory reach over time.

5 The “core” of Darwinian evolution

The strategy of endogenization is made possible by the fact that the core Darwinian principles are abstract in nature, so potentially applicable very widely. This is what permits diverse biological features, including ones that were originally part of the unexplained background, to be subsumed under the evolutionary umbrella. However a minor irony in fact lurks here. For although the abstractness of the core Darwinian principles is a widely acknowledged point, there is a lack of agreement on how exactly these principles should be formulated. Moreover, some candidate formulations appeal to biological features which are in fact the *products* of evolution, and so in a fully general analysis should themselves be endogenized. Ironically, then, what makes the strategy of endogenization possible is the fact that the core Darwinian principles are abstract; and yet those very principles are often formulated in a way that fails the test of endogenization.

To illustrate, consider the famous “replicator–interactor” conceptualization of evolution by natural selection, due to Dawkins (1982) and Hull (1980). According to these authors, Darwinian evolution involves repeated rounds of replication followed by environmental interaction, so involves entities of two different sorts: replicators and interactors.¹⁵ Replicators are defined as entities of which copies are made, and which “pass on their structure intact” from one generation to the next; they are characterized by the attributes of “longevity” and “copying fidelity”. Interactors are defined as entities that “interact as cohesive wholes with their environment”, in a way that causes replication to be differential. Both Dawkins and Hull intend this to be a fully general analysis of Darwinian evolution, in principle applicable to biological systems of various sorts, and perhaps also to non-biological (e.g. cultural) evolution. However, there is a mismatch between the desired generality and the way that their key terms are defined.

The point is a simple one. The longevity and copying fidelity of replicators (such as genes), and the cohesiveness of interactors (such as organisms) are highly evolved properties, themselves the product of many rounds of cumulative natural selection.¹⁶ The very first replicators must have had extremely *poor* copying fidelity, and lacked longevity; indeed a major challenge in pre-biotic evolution is to understand how in the absence of error correcting enzymes, replicating molecules more than a few base pairs long could have persisted, given the mutation rate. (This is known as the Eigen paradox.¹⁷) Similarly, the earliest multicellular organisms must have been highly *non-cohesive* entities, owing to competition between their constituent cell lineages (Buss 1988; Michod 1999). As we know from recent work on intra-organismic conflict, the cohesion and unity of modern organisms, to the extent that it exists, is an evolutionary achievement, not the ancestral condition, and it is constantly in danger of breaking down. This points to a limitation of the Dawkins/Hull framework. For if we wish to understand how copying fidelity and cohesiveness evolved in the first place, we cannot build these requirements into the very concepts used to formulate the Darwinian theory.

What exactly does this show? One moral is as follows. In so far as the aim is to formulate the Darwinian principles in a way that is maximally general, the Dawkins/Hull formulation is inferior to certain alternative formulations that have been proposed in the literature.¹⁸ These include: the widely-used “heritable variation in fitness” analysis of Lewontin (1970); the closely related analysis of Maynard Smith (1988) according to which “multiplication, variation and heredity” are the key ingredients; the “Price equation” approach of Price et al. (1970), which says that trait-fitness covariance is key; and the formulation due to Griesemer (2000), in which the key notion is that of a “reproducer”, rather than a replicator. Though these formulations differ somewhat in purpose, they all purport to identify the essence, or conceptual core, of Darwinian evolution. But unlike the Dawkins/Hull formulation, then do so without relying, at

¹⁵ Dawkins uses the term “vehicle” in place of Hull’s “interactor”; the meaning is essentially the same.

¹⁶ This point is emphasized by Griesemer (2000).

¹⁷ The paradox is named after Eigen (1971). See Maynard Smith and Szathmáry (1995) for a good discussion.

¹⁸ This point has been argued for on independent grounds by Okasha (2006) chapter 1, and Godfrey-Smith (2009).

least overtly, on any evolved features of modern organisms, thus passing the test of endogenization.

Now it might be argued that this is unfair to Dawkins and Hull, on the grounds that their framework does not aim to capture every conceivable case of evolution by natural selection, but rather to characterize the particular sort of evolution that is capable of leading to adaptive complexity. Conceivably it may be true that without high-fidelity replicators and cohesive interactors, the Darwinian process would never have been able to produce the “interesting” organismic adaptations that we find in nature; in which case treating these features as essential conditions for Darwinian evolution would make a certain amount of sense, despite the features themselves having an evolutionary history. In short, the lack of endogenization is certainly a problem if the aim is to devise a scheme that captures all possible instances of Darwinian evolution; but it is less of a problem if the aim is to capture the subset capable of producing interesting adaptations.

Whether something like this is true to Dawkins’ and Hull’s original intentions, and those of other “replicator-first” theorists, is not a question that can be addressed here. However it raises a more general issue, of how much in evolutionary biology can be explained in terms of the core Darwinian principles themselves—however exactly they are formulated—as opposed to those principles taken in conjunction with additional biological features. I look briefly at this issue next.

6 On reduction to first principles

Let us return to the theme of this special issue: first principles in science. The prevalence of the endogenization strategy in evolutionary biology, and its numerous successes, may *seem* to support a strong form of reductionism, according to which all or almost all biological features can be explained on the basis of fundamental evolutionary principles. For each successful endogenization means that features that were originally part of the unexplained background are brought under the evolutionary umbrella, thus extending the explanatory scope of the core Darwinian principles. So endogenization may seem to fit with the reductionist vision championed by Dennett (1995), who describes evolution by natural selection as a “universal acid” that cuts through everything in its path.

Though tempting, this conclusion does not necessarily follow. For there is a less reductionist alternative, that is also compatible with the prevalence and success of the endogenization strategy. The alternative is this. It is not the core Darwinian principles themselves that bear the explanatory burden in evolutionary biology, but rather those principles as they operate in specific biological settings, in the presence of additional contingent biological features. There are many candidates for what these additional features are, some of which we have discussed already. They include: high-fidelity replication (Dawkins 1976); digital information storage (Maynard Smith and Szathmáry 1995); particulate rather than blending inheritance (Fisher 1930); quasi-independence (Lewontin 1978); and gene regulation by allosteric proteins (Monod 1971). Each of these features has been argued to be essential, for Darwinian evolution to have produced the particular phenomena, such as adaptive complexity, that we find

in the biosphere. Thus for example, Fisher argues that without particulate inheritance, evolution would quickly grind to a halt unless the mutation rate were improbably high; Maynard Smith and Szathmary argue that without digital information storage, the number of possible life forms would be too small to permit anything very interesting to evolve; while Dawkins and Lewontin argue that high-fidelity replication and quasi-independence, respectively, are pre-requisites for the evolution of complex adaptations.

Now as we have stressed, many if not all of these features have themselves evolved, or at least depend on biological structures that have evolved. Thus they are candidates for endogenization. But the key point is this. Suppose we grant that one of these features is essential to explaining adaptive complexity, as the above authors argue. Suppose it then turns out that the feature in question can be given an evolutionary explanation of its own, or endogenized. This may seem to invite the reductionist conclusion that in fact, adaptive complexity *can* be explained in terms of the core Darwinian principles alone, contrary to our first supposition. But in fact this does not follow. That an evolutionary explanation of some phenomenon invokes a given background feature; and that that feature can itself be given (some other) evolutionary explanation, does not imply that the initial phenomenon can be explained in terms of evolutionary principles alone, without appeal to the background feature.

This point reflects a sort of “non-monotonicity” of explanation. If theory T , in conjunction with background condition Y , explains phenomenon X ; and if theory T can itself explain why condition Y holds; it does not follow that T alone can explain X . (Let T be the theory of evolution by natural selection, X the existence of adaptive complexity, and Y a feature such as high-fidelity replication.) Obviously this would follow if we replaced “explains” with “entails” (or “logically implies”); but philosophers of science know well that entailment is neither necessary nor sufficient for scientific explanation. If instead we identify explaining something with raising its probability,¹⁹ then the failure of the above principle is immediate, for the conditions **(I)** $P(X|T&Y) > P(X)$ and **(II)** $P(Y|T) > P(Y)$ do not jointly imply **(III)** $P(X|T) > P(X)$.

To help see why this is, note that **(I)** is equivalent to $P(X&T&Y)/P(T&Y) > P(X)$. Now let us suppose, as will be plausible in certain cases, that $P(Y|X&T) = 1$, that is, the probability of background condition Y obtaining, conditional on the truth of theory T and the occurrence of phenomenon X , is 1; which is equivalent to $P(X&T&Y) = P(X&T)$. Substituting into **(I)**, this gives $P(X&T)/P(T&Y) > P(X)$. Now this is compatible with the falsity of **(III)**, so long as $P(T&Y)$ is sufficiently smaller than $P(T)$ (since **(III)** is equivalent to $P(X&T)/P(T) > P(X)$). And $P(T&Y) \ll P(T)$ is compatible with condition **(II)**, so long as $P(Y)$ is sufficiently small. More generally, whenever we have $P(Y|X&T) \approx 1$ and $P(T&Y) \ll P(T)$, then so long as $P(Y)$ is sufficiently small, conditions **(I)** and **(II)** will be true but **(III)** will be false.

¹⁹ The idea that enhancement of prior probability, aka “positive relevance”, rather than entailment, is the key to the concept of scientific explanation is a theme in Salmon’s critique of Hempel’s theory on explanation. See for example Salmon (1989).

This is all rather abstract, but our illustrative example may in fact constitute a case in point. Suppose that the prior probability of organisms exhibiting high-fidelity replication (Y) is very low; this seems reasonable, given that most conceivable replication systems are not high-fidelity. Suppose further that the probability of high-fidelity replication, conditional on the truth of the theory of evolution (T) and the existence of adaptive complexity (X), is extremely high; this is also plausible, if Dawkins is right that evolution could *only* have produced adaptive complexity in the presence of high-fidelity replication. Suppose finally that the prior probability that the theory of evolution is true is substantially greater than the probability that the theory is true *and* that high-fidelity replication exists ($T \& Y$); this too is plausible, since by assumption high-fidelity replication is improbable, and the theory of evolution itself, *sans* the existence of adaptive complexity, arguably does not make it probable in absolute terms (though it does increase its probability). If all of this is right, then we have that $P(Y)$ is very low, $P(Y|X \& T) \approx 1$ and $P(T \& Y) \ll P(T)$; so by the argument of the last paragraph, conditions (I) and (II) are true but (III) is not. If we are happy to equate “explains” with “raises the prior probability of”, this means that the theory of evolution and the assumption of high-fidelity replication together explain adaptive complexity, the theory of evolution itself explains high-fidelity replication, but the theory of evolution alone does *not* explain adaptive complexity.

(Note that in practice, when evolutionary theory endogenizes some background feature Y , it will probably make use of further background assumptions Z , i.e. it is not T alone but rather $T \& Z$ that will explain Y ; which then raises the question of whether Z can itself be endogenized, and so-on. (What explains the evolution of high-fidelity replication is not the core Darwinian principles alone, but those principles plus further biological assumptions.) But the point of the foregoing paragraph is that even if this process eventually bottoms out, and a given background feature *is* explained by T alone, it still does not follow that T alone can explain all the phenomena that T plus that background feature can explain, so long as “explains” is understood in a way that does not imply “entails”).

This analysis can illuminate one strand in the current debate over niche construction.²⁰ Proponents of niche construction argue that it is an “evolutionary process in its own right”, by which they mean that it plays an essential role in evolutionary explanation. One of the central explananda of evolutionary biology—the striking “fit” of organism to environment—cannot be explained by natural selection alone, they claim; niche construction is essential too. Opponents respond that organisms’ ability to successfully niche-construct is itself an evolved attribute, and thus in principle must admit of an explanation in selective terms, i.e. natural selection has led organisms to acquire the ability to modify their niches. Reconstructed this way, the debate can be seen to instantiate the above pattern. The proponents argue that to explain a particular phenomenon, traditional Darwinian evolution is not enough—an additional explanatory factor is needed too; the opponents respond that since that background factor is itself evolved, then at root, Darwinian evolution *can* do all the explanatory work, so no fundamental modification of the theory is needed. Our analysis suggests that, in principle,

²⁰ See the for-and-against arguments in Scott-Phillips et al. (2014). Thanks to an anonymous referee for suggesting this example.

either camp in this debate may be right. Nothing in the logic of scientific explanation shows that the opponents' reductionist story is compulsory, but nothing precludes it either.

The upshot, therefore, is that recognizing the prevalence of the strategy of endogenization in evolutionary biology does not automatically push us to a Dennett-style "universal acid" position, in which everything is explained by the core Darwinian principles themselves (i.e. variation, fitness differences, and heredity). It is equally compatible with an alternative, milder form of reductionism, according to which it is not the core Darwinian principles alone, but rather those principles in conjunction with additional contingent biological features, that do the explanatory work. That these additional features have themselves evolved, so are candidates for endogenization, does not force us to the extreme reductionist alternative, though it does not preclude it either. Clearly, the substantive issue here is whether it is true that contingent biological features, in addition to the core Darwinian principles themselves, are essential to the explanations that evolutionary theory give, and if so what those features are. However, a proper treatment of that question is a task for another day.

7 Conclusion

The theory of evolution occupies a unique position in biology, explaining and unifying a vast and diverse body of phenomena. This is hardly a new observation, as Dobzhansky's oft-repeated dictum "nothing in biology makes sense except in the light of evolution" shows (Dobzhansky 1973). The aim of this article has been to identify one interesting way in which the explanatory scope of evolutionary theory has increased over time, namely by the endogenization of features that were originally presuppositions of other evolutionary explanations. This concept of endogenization provides a useful perspective on the recent history of evolutionary biology, helping to reveal commonalities between diverse scientific developments. And the concept is directly relevant to two important issues in the philosophy of evolutionary biology: the question about how best to formulate the core Darwinian principles, and the question of reduction to first principles.

Acknowledgements Versions of this paper were presented at the Philosophy of Biology in the UK conference in Oxford in 2012, and the First Principles in Science conference in Munich in 2016. Thanks to the audience members on both occasions for discussion, and to two anonymous referees for thoughtful comments.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Billiard, S., López-Villavicencio, M., Devier, B., Hood, M. E., Fairhead, C., & Giraud, T. (2010). Having sex, yes, but with whom? Inferences from fungi on the evolution of anisogamy and mating types. *Biological Reviews*, 86(2), 421–442.

- Bourke, A. F. G. (2011). *Principles of social evolution*. Oxford: Oxford University Press.
- Buss, L. (1988). *The evolution of individuality*. Princeton: Princeton University Press.
- Darwin, C. (1879). *On the origin of species by means of natural selection*. London: John Murray.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- Dawkins, R. (1982). *The extended phenotype*. Oxford: Oxford University Press.
- Dennett, D. (1995). *Darwin's dangerous idea*. London: Penguin.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *The American biology teacher*, 35(3), 125–29. <https://doi.org/10.2307/4444260>.
- Edwards, A. W. F. (2011). Mathematizing Darwin. *Behavioural Ecology and Sociobiology*, 65(3), 421–30.
- Eigen, M. (1971). Self-organization of matter and evolution of biological macromolecules. *Naturwissenschaften*, 58(10), 465–523.
- Eshel, I. (1985). Evolutionary genetic stability of Mendelian segregation and the role of free recombination in the chromosomal system. *American Naturalist*, 125, 412–420.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Fisher, R. A. (1932). The evolutionary modification of genetical phenomena. In *Proceedings of the 6th international congress of genetics I* (pp. 165–72).
- Godfrey-Smith, P. (2009). *Darwinian populations and natural selection*. Oxford: Oxford University Press.
- Griesemer, J. (2000). The units of evolutionary transition. *Selection*, 1, 67–80.
- Haldane, J. B. S. (1932). *The causes of evolution*. London: Macmillan.
- Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, 7, 1–52. [https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4).
- Hamilton, W. D. (1975). Innate social aptitudes of man: an approach from evolutionary genetics. In R. Fox (Ed.), *Biosocial anthropology* (pp. 133–55). New York: Wiley.
- Hansen, T. F. (2006). The evolution of genetic architecture. *Annual Review of Ecology and Systematics*, 37, 123–157.
- Hansen, T. F. (2015). *Evolutionary constraints*. Retrieved 21 Nov. 2017, from <http://www.oxfordbibliographies.com/view/document/obo-9780199941728/obo-9780199941728-0061.xml>.
- Hartfield, M., & Keightley, P. D. (2012). Current hypotheses for the evolution of sex and recombination. *Integrative Zoology*, 7(2), 1–52. [https://doi.org/10.1016/0022-5193\(64\)90038-4](https://doi.org/10.1016/0022-5193(64)90038-4).
- Hull, D. (1980). Individuality and selection. *Annual Review of Ecology and Systematics*, 11, 311–32.
- Laland, K., et al. (2014). Does evolutionary theory need a re-think? *Nature*, 514(7521), 161–4. <https://doi.org/10.1038/514161a>.
- Lewontin, R. C. (1970). The units of selection. *Annual Review of Ecology and Systematics*, 1, 1–18.
- Lewontin, R. C. (1978). Adaptation. *Scientific American*, 239(3), 212–30.
- Lewontin, R. C. (1983). The organism as the subject and object of evolution. *Scientia*, 118, 65–82.
- Maynard Smith, J. M. (1976). Group selection. *Quarterly Review of Biology*, 51(2), 277–83.
- Maynard Smith, J. M. (1988). Evolutionary progress and levels of selection. In M. H. Nitecki (Ed.), *Evolutionary Progress* (pp. 219–30). Chicago, IL: University of Chicago Press.
- Maynard Smith, J. M., & Szathmáry, E. (1995). *The major transitions in evolution*. Oxford: Oxford University Press.
- Michod, R. (1999). *Darwinian dynamics: Evolutionary transitions in fitness and individuality*. Princeton: Princeton University Press.
- Monod, J. (1971). *Chance and necessity*. New York: Knopf Press.
- Muller, H. J. (1932). Some genetic aspects of sex. *American Naturalist*, 66(703), 118–38. <https://doi.org/10.1086/280418>.
- Odling-Smee, F. J., Laland, K. N., & Feldman, M. W. (2003). *Niche construction: The neglected process in evolution*. Princeton: Princeton University Press.
- Okasha, S. (2005a). On niche construction and extended evolutionary theory. *Biology and Philosophy*, 20, 1–10.
- Okasha, S. (2005b). Multi-level selection and the major transitions in evolution. *Philosophy of Science*, 72, 1013–25.
- Okasha, S. (2006). *Evolution and the levels of selection*. Oxford: Oxford University Press.
- Parker, G. A., Baker, R. R., & Smith, V. G. F. (1972). The origin and evolution of gamete dimorphism and the male–female phenomenon. *Journal of Theoretical Biology*, 36(3), 529–53. [https://doi.org/10.1016/0022-5193\(72\)90007-0](https://doi.org/10.1016/0022-5193(72)90007-0).
- Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics*, 9, 75–82. <https://doi.org/10.1038/nrg2278>.

- Powers, S. T., Penn, A. S., Watson, R. A., et al. (2011). The concurrent evolution of cooperation and the population structures that support it. *Evolution*, 65(6), 1527–43.
- Price, G. R., et al. (1970). Selection and covariance. *Nature*, 227, 520–1.
- Salmon, W. (1989). *Four decades of scientific explanation*. Minneapolis: University of Minnesota Press.
- Scott-Phillips, T. C., et al. (2014). The niche construction perspective: A critical appraisal. *Evolution*, 68(5), 1231–43.
- Sniegowski, P., et al. (2000). The evolution of mutation rates: separating causes from consequences. *Bioessays*, 22(12), 1057–66.
- Stearns, S. (1992). *The evolution of life histories*. Oxford: Oxford University Press.
- Sterelny, K., & Calcott, B. (Eds.). (2011). *The major transitions in evolution revisited*. Cambridge, MA: MIT Press.
- Sturtevant, A. H. (1937). Essays on evolution I. On the effects of selection on mutation rate. *Quarterly Review of Biology*, 12, 467–77.
- Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. *Proceedings of the National Academy of Sciences USA*, 112(33), 10104–11.
- van Baalen, M., & Rand, D. (1998). The unit of selection in viscous populations and the evolution of altruism. *Journal of Theoretical Biology*, 21(193(4)), 631–48.
- Wagner, G. P., & Altenberg, L. (1996). Complex adaptations and the evolution of evolvability. *Evolution*, 50(3), 631–48.
- Weismann, A. (1886) [1889]. The significance of sexual reproduction on the theory of natural selection (S. Scholand, Vol. 1, pp. 251–332, Trans.). *Essays upon Heredity and Kindred Biological Problems*. Oxford: Clarendon Press.
- Yoshimura, J., & Clark, C. W. (1993). *Adaptation in stochastic environments*. Berlin: Springer.