Brit. J. Phil. Sci. **70** (2019), 127–152

Underdetermination and Evidence in the Developmental Plasticity Debate

Karen Kovaka

ABSTRACT

I identify a controversial hypothesis in evolutionary biology called the plasticity-first hypothesis. I argue that the plasticity-first hypothesis is underdetermined and that the most popular means of studying the plasticity-first hypothesis are insufficient to confirm or disconfirm it. I offer a strategy for overcoming this problem. Researchers need to develop a richer middle-range theory of plasticity-first evolution that allows them to identify distinctive empirical traces of the hypothesis. They can then use those traces to discriminate between rival explanations of evolutionary episodes. The best tools for developing such a middle-range theory are experimental evolution and formal modelling.

1 Introduction

It is a striking fact of the biological world that phenotypic expression may be, and often is, influenced by an organism's developmental environment.

doi:10.1093/bjps/axx038

Advance Access published on August 30, 2017

This means that two genetically identical organisms raised in different environments can have different physical characteristics, use different strategies to avoid predators and find food, and even be of different sexes. Biologists increasingly recognize the importance of this phenomenon, which they call developmental plasticity (Pigliucci [2001]), but hypotheses about its role in evolution are controversial.

Perhaps the most controversial hypothesis is that many important phenotypic novelties owe their origins to developmental plasticity, because plasticity allows new phenotypes to emerge prior to genetic mutations or recombination (West-Eberhard [2003]). This is the plasticity-first hypothesis. Biologists have long debated both its plausibility and importance (for example, Orr [1999]; Pigliucci [2007]; Gilbert and Epel [2009]; Jablonka and Raz [2009]; Futuyma [2011]; Dickins and Rahman [2012]; Moczek [2015]), in part because it informs a broader controversy about the adequacy of central components of evolutionary theory (Laland *et al.* [2015], p. 2). Philosophers, too, are interested in whether the plasticity-first hypothesis poses a challenge to biological orthodoxy. But where they have examined its theoretical implications, I am interested in an epistemic question: what evidence do we need to settle the longstanding debate about the hypothesis?[1]

I argue that the plasticity-first hypothesis is transiently underdetermined (Sklar [1975]), that is, the presently available data does not confirm nor disconfirm the plasticity-first hypothesis because it does not discriminate between the plasticity-first hypothesis and its theoretical rivals. Moreover, ancestral–descendent comparisons, which are the most popular means of studying the plasticity-first hypothesis, do not on their own generate the evidence needed to confirm or disconfirm the hypothesis. Together, these two facts explain why even though the number of empirical studies about plasticity has grown substantially in the last decade (Forsman [2015]), the debate about the plasticity-first hypothesis has reached a stalemate.

I offer a strategy for overcoming this underdetermination problem. Researchers need to develop a richer middle-range theory (Binford [1982]; Jeffares [2008]) of plasticity-first evolution, one that allows them to identify distinctive empirical traces of the hypothesis. Then they can search for those traces and use them to discriminate between rival explanations of evolutionary episodes. The best tools for developing that middle-range theory are to be found in experimental evolution and formal modelling, not ancestral–descendant comparisons.

My epistemic analysis also informs the theoretical side of the controversy over the plasticity-first hypothesis. There is a deflationary interpretation of the

[1] Examples include an edited volume about the evolutionary significance of the Baldwin effect (Weber and Depew [2003]), articles about the theoretical integration of evolution and development (Sterelny [2000]; Amundson [2005]; Love [2006]), and Kaplan ([2008]) on non-genetic inheritance and developmental plasticity.

debate on which describing a particular evolutionary process as either an instance of plasticity-first or gene-first evolution is simply a matter of taste (Wagner [2011], p. 182). This deflationary view depends on the claim that there is no difference in the evolutionary patterns generated by plasticity-first and gene-first mechanisms, and thus the distinction between the two is explanatorily inert. The research strategy I suggest addresses this concern about explanatory inertia by prioritizing the discovery of empirical differences between the two kinds of mechanisms. Confirming the plasticity-first hypothesis and demonstrating its scientific interest are thus related issues, and resolving them requires a shift in evidence-collecting methods and priorities.

2 The Plasticity-First Hypothesis and Its Rivals

I'll begin by introducing the plasticity-first hypothesis, its theoretical rivals, and the controversy between them. The plasticity-first hypothesis proposes that because organisms are developmentally plastic (that is, sensitive to environmental inputs), a process of environmental induction followed by genetic assimilation is an evolutionarily significant mechanism for the emergence of phenotypic novelties (West-Eberhard [2003]; Moczek *et al.* [2011]).[2]

It's easiest to understand this hypothesis in light of an example. Sticklebacks are small fish that inhabit oceans, lakes, streams, and estuaries throughout the northern hemisphere. Stickleback from different habitats have different mouth shapes. Experiments have shown that many of these differences are due to environmental factors rather than genetic ones. For instance, if you capture juvenile fish from a stream habitat and feed them zooplankton rather than bloodworms, they will develop mouth phenotypes typical of lake-dwelling fish (Lucek *et al.* [2014]). Dietary changes can also induce fish from the surface of a lake to develop mouths typical of fish that live on lake bottoms (Wund *et al.* [2008]). Thus, sticklebacks are developmentally plastic for mouth shape.

When developmentally plastic individuals encounter new environmental conditions, they sometimes develop new phenotypes in response (Waddington [1953], [1956]), where a new phenotype is simply one that differs qualitatively or quantitatively from the other phenotypes that a genotype has produced in the

[2] 'Phenotypic novelty' is a term of art with a narrower meaning than 'new phenotype'. To be considered a novelty, a phenotype must differ from its predecessors in some important way, but the biological literature is still divided on what is required for something to count as a novelty (for a review, see Wagner and Lynch [2010]). My conception follows Pigliucci's ([2008], p. 890), according to which novelties are 'new traits or behaviours, or novel combinations of new traits or behaviours, arising during the evolution of a lineage, and that perform a new function within the ecology of that lineage'. The plasticity-first hypothesis is not committed to any one conception of novelty. Instead, the hypothesis claims that plasticity-first evolution is part of the explanation for the origin of some traits that will count as novelties on whatever conception(s) biologists adopt.

population's recent history.[3] This phenomenon is called environmental induction: the appearance (and recurrence) of a new developmental variant that occurs when some new environmental input affects a pre-existing responsive phenotype, causing a phenotypic change or reorganization (modified from West-Eberhard [2003], p. 140). Of course, all traits are developmentally dependent on both genetic and environmental inputs, but the concept of environmental induction is useful to biologists because it foregrounds the difference-making role that the environment plays in some cases of development that it does not play in others.

Environmental conditions do not only generate phenotypic variation. If they are intergenerationally stable, they may allow for the transmission of phenotypic variation by inducing it anew in each generation, as parents expose their offspring to developmental environments similar to those in which they themselves matured. Suppose the inducing conditions are stable across generations and the induced phenotypic variant is adaptive. An example might be a population of stream-dwelling sticklebacks that migrate to a lake and whose mouth phenotypes respond plastically to prey availability. Then natural selection will prefer the induced phenotype. The genotypes that are capable of producing the adaptive variant will become more frequent, while the genotypes that are not will dwindle.[4] This is genetic evolution (change in allele frequencies across generations due to natural selection), but it does not require new genetic variants.[5]

Evolution by selection of an environmentally induced phenotype may not seem powerful because the adaptive variants maintained by environmental induction are fragile. If, for example, the stickleback population leaves the lake and migrates back to a stream, the new mouth phenotype will disappear as environmental conditions change. Insofar as evolutionary biology aims to explain the evolution of complex traits that depend on the slow accumulation of phenotypic changes over long periods of time and varied environmental conditions, we may be sceptical of how relevant these traits can be.

There is, however, a further process called genetic assimilation, which can reduce the dependency of environmentally induced traits on their initial inducing

[3] Not all new phenotypes count as true phenotypic novelties, but they may still serve as the foundation for the evolution of future novelties.

[4] Here I make the realistic assumption that there is genetic variation for plasticity in the population. If all members of the population are equally plastic, then, of course, natural selection will not occur.

[5] The term for this kind of evolution by selection of a phenotypic variant is genetic accommodation. Though both environmentally and genetically induced variants can be genetically accommodated (Crispo [2007]), here I am interested only in genetic accommodation of environmentally induced traits.

conditions.[6] Genetic assimilation occurs when changes in the genetic basis of an induced trait make the trait more adaptive, but less plastic. The more adaptive genetic variants have a selective advantage, so they spread throughout the population, and the genetic system gains more control over the environmentally induced trait. Eventually, the trait will develop even in the absence of the original inducing conditions, meaning it is no longer environmentally induced. Once this transfer of developmental control from the environment to the genome occurs, the trait is more likely to figure in future episodes of cumulative evolution.

A final component of the plasticity-first hypothesis that needs definition is evolutionary significance. Advocates of the hypothesis are vague about the meaning of this term, but as I understand it, their primary concern is the explanatory significance of the plasticity-first hypothesis, meaning that the plasticity-first hypothesis provides part of the answer to a central problem in evolutionary theory, namely, the problem of how phenotypic novelties emerge and spread.[7] If plasticity-first evolution is explanatorily significant, then a satisfactory resolution of the problem of novelty will invoke it as a key novelty-generating mechanism. I am following the literature here by speaking of one general problem of novelty in evolutionary theory (Muller and Wagner [1991]), but it might also be productive to approach the issue more narrowly, asking about the role of plasticity-first evolution in the appearance of novelties in particular clades.

How do we determine if plasticity-first evolution is a key novelty-generating mechanism, either in general or in particular clades? Frequency is one consideration, but presumably plasticity-first evolution can be evolutionarily significant even if it is somewhat rare. For example, if plasticity-first evolution has produced high-profile phenotypic novelties such as limbs (Standen *et al.* [2014]), bipedalism (Pigliucci [2008]), types of social learning (Sterelny [2012]), or CAM photosynthesis (West-Eberhard *et al.* [2011]), then these are reasons to think it meets the significance threshold. So, evidence for the significance of the plasticity-first hypothesis can be evidence about either its frequency or its role in the emergence of particular high-profile novelties.

Judgements of explanatory significance depend in part on philosophical considerations about the nature of explanations, particularly historical explanations, but the issue is also beset by empirical challenges. These stem from the fact that it is difficult to discriminate between plasticity-first evolution and alternative explanations for the emergence of novelties. There are three of these alternative explanations: novelty from coding mutations, novelty from non-coding (regulatory) mutations, and novelty from

[6] A related process is called the Baldwin effect, but the Baldwin effect picks out cases in which plasticity helps organisms survive in a new environment without being followed by plasticity loss or further adaptation of the plastic trait(s) (Crispo [2007]).

[7] This conception of explanatory significance is drawn from Godfrey-Smith's ([2001]) analysis of explanatory adaptationism and Orzack and Sober's ([1994]) I-adaptationism thesis.

recombination, which is the reshuffling of genetic material on or between chromosomes. Unlike plasticity-first evolution, in which phenotypes change before genes do, these mechanisms all involve gene-first changes. It is difficult to determine which of these mechanisms is responsible for any particular phenotypic novelty because researchers have to make inferences about past events on the basis of presently available data, but both phenotype-first and gene-first mechanisms produce the same evolutionary outcomes: phenotypic novelties and genetic adaptations. Researchers need additional evidence to confirm or disconfirm the plasticity-first hypothesis, but there is no consensus on what data would be sufficient for this purpose.

3 Evidential Disagreements

The scientific debate about the plasticity-first hypothesis is dominated by two issues: the quality of direct versus indirect evidence, and the proper placement of the burden of proof. In this Section, I show that in focusing on these issues, the debate has neglected substantive questions about what data would qualify as confirmatory evidence for plasticity-first evolution and how to obtain such evidence. I also identify two evidential standards that need to be met in order for the plasticity-first hypothesis to be confirmed: First, the evidence must be discriminatory, that is, for at least some novelties, it must favour the plasticity-first explanation over the alternative gene-first explanations. Second, the evidence must be relevant to the significance claim embedded in the hypothesis.

3.1 Direct and indirect evidence

Advocates of the plasticity-first hypothesis characterize the evidence for their position as indirect, rather than direct (West-Eberhard [2003]; Pigliucci and Murren [2003]). They argue that plasticity-first evolution, like many other evolutionary mechanisms, is difficult to observe directly, and thus it is unreasonable to ask (as some sceptics do) for direct evidence for the hypothesis before accepting it. Instead, researchers can test the hypothesis by collecting indirect evidence. This indirect evidence usually consists of comparisons between ancestral and descendant populations (studies that compare characteristics of a descendant population to those of an ancestral one). Scientists less friendly to the plasticity-first hypothesis criticize the reliance on indirect evidence (de Jong and Crozier [2003]; de Jong [2005]). In fact, there is a long tradition of scepticism about plasticity-first evolution that continues despite detailed reviews showcasing the latest empirical work on the topic (Wund [2012]; Schlichting and Wund [2014]).

For all their disagreements, sceptics and advocates do agree that the distinction between direct and indirect evidence generates the controversy about the evidential status of the plasticity-first hypothesis. Advocates accept indirect evidence as confirmatory, sceptics do not, and both sides are aware of this. But what exactly is direct evidence? That is far from clear. In fact, there are as many as four different and plausible interpretations of the direct–indirect distinction.

The first and most straightforward interpretation of the distinction is that obtaining direct evidence means actually observing the hypothesized event happen in the wild (though additional laboratory experiments may be needed to supplement these observations). All other evidence, by contrast, is indirect. Call this the distinction between direct and indirect observation.[8] It is tempting to read the sceptics' demand for direct evidence as demands for direct observation, especially calls for a 'crucial laboratory experiment' (de Jong and Crozier [2003], p. 17) to confirm the hypothesis. Direct observation is not, however, the typical standard against which evidence for an evolutionary hypothesis is judged, so it is perhaps unfair to interpret sceptics as calling for it. Inferences about the occurrence of natural selection in the wild are rarely direct in this sense (Endler [1986]), nor is research on speciation (Wund [2012]). Though there are rare exceptions (Grant and Grant [2009]), biologists almost never witness speciation events from beginning to end. If the sceptics really are appealing to direct observation in their criticisms of plasticity-first evolution, then they are holding the hypothesis to an unreasonably high standard.

A second interpretation of the sceptics' calls for direct evidence is that they want evidence for the occurrence of plasticity-first evolution in particular cases to discriminate between plasticity-first evolution and the alternative mechanisms that can also produce new phenotypes. On this interpretation, direct evidence is discriminatory: to be considered direct evidence, a data set must favour the plasticity-first explanation for the emergence of a particular novelty over the alternative gene-first mechanisms. Indirect evidence, by contrast, is data that are consistent with plasticity-first evolution, but not discriminatory. On this construal, demanding direct evidence for the plasticity-first hypothesis is completely reasonable. In fact, from the perspective of confirmation theory, data that are not direct in this sense are not evidence at all. When data fail to discriminate, they fail to provide evidence, and the result is a kind of underdetermination problem called contrast failure (Forber [2009]). If this is what sceptics mean when they call for 'clear empirical evidence'

[8] I am offering this as a plausible interpretation of the direct–indirect evidence distinction as it arises in the debate about plasticity-first evolution, not as an endorsement of the idea that laboratory observations are not direct observations. In Section 5 I will discuss the role of laboratory observations, particularly LNS, in resolving this debate.

yes, but if it is not discriminatory for the plasticity first hp, it is also not discriminatory for the alternative hp. this means that the gene first hp also lacks evidence

(Santos *et al.* [2015], p. 128) and 'direct support' (de Jong [2005]), then they are simply denying that the advocates have offered discriminatory evidence for the plasticity-first hypothesis.

Yet another way of drawing the distinction between direct and indirect evidence is to say that direct evidence for the plasticity-first hypothesis must be relevant to the significance claim embedded in the hypothesis. If evidence merely raises the probability that plasticity-first evolution occurred in this or that particular case, but does not speak to overall significance, it is indirect. This conception of direct evidence allows sceptics to concede that some evidence for plasticity-first evolution is discriminatory, but still deny that this evidence is significance-relevant because it does not tell us how frequent plasticity-first evolution is or whether it is responsible for high-profile evolutionary novelties. Direct evidence as significance-relevant evidence is a plausible interpretation of at least some of the plasticity-first sceptics; see, for example, Wray *et al.*'s call to 'strengthen the evidence for [the] importance' (Laland *et al.* [2014], p. 164) of phenotypic plasticity.

Though interpreting the sceptics as calling for either discriminatory or significance-relevant evidence is both charitable and plausible, it is clear that neither of these is the distinction advocates have in mind when they talk about direct and indirect evidence. Advocates of the plasticity-first hypothesis do not make the claim that the hypothesis should be adopted on the basis of evidence that lacks discriminatory power or fails to address the significance of plasticity-first evolution. Rather, their claim is that indirect evidence can be both discriminatory and significance-relevant.

Advocates use the distinction between direct and indirect evidence in two different ways. First, they sometimes mean that direct evidence is observationally direct, as in Pigliucci and Murren's ([2003], p. 1462) reference to catching genetic assimilation 'in the act'. Second, they sometimes interpret calls for direct evidence as calls to demonstrate each element of the larger plasticity-first hypothesis within one model system or, even more stringently, in a single study. According to Schlichting and Wund ([2014]), such integrated evidence is desirable, but not a requirement on confirmatory evidence. Wund ([2012], p. 6) argues that asking for such a demonstration in one study is a 'flawed expectation', because, like the observational directness requirement, it holds the plasticity-first hypothesis to a higher standard than that used for analogous hypotheses. What advocates are sometimes defending when they speak about indirect evidence, then, is that it is possible to obtain discriminatory evidence in a piecemeal fashion by splitting the larger plasticity-first process into smaller units and then investigating these units across different experiments and model systems.

These four different ways of distinguishing between direct and indirect evidence are blurred together in the plasticity-first debate. Advocates talk about

direct evidence both in terms of direct observation and integration. These may also be the conceptions of directness that sceptics have in mind, but if they are, then important questions about the extent to which the data on plasticity-first evolution are discriminatory and significance-relevant are ignored. Alternatively, if sceptics are arguing that evidence for plasticity-first evolution needs to be discriminatory and/or significance-relevant, then they are indeed raising important questions about confirmation, but the two sides are talking past one another. In either case, the debate has yet to produce a clear discussion of whether the data on plasticity-first evolution manage to be discriminatory or significance-relevant. And such a discussion is critical to understanding the relationship between the plasticity-first hypothesis and the evidence for it.

3.2 Burden of proof

One way in which advocates have tried to address worries about discriminatory evidence is by claiming that there is often evidential parity between plasticity-first hypothesis and its theoretical rivals; that is, that phenotype-first mechanisms and gene-first mechanisms enjoy comparable levels of evidential support in many particular cases. This burden of proof argument concedes that we often do not know whether a population's evolutionary history involved genetic assimilation, but it claims that we also do not know whether and how often that population has evolved novelties by gene-first mechanisms (Pigliucci *et al.* [2006]). Biologists have long assumed that the gene-first mechanisms account for most if not all phenotypic novelties, but only because there was no alternative hypothesis that fit the available evidence. But now the plasticity-first hypothesis presents just such an alternative. As a result, we have learned that some data once thought to support the occurrence and significance of gene-first evolutionary mechanisms are not actually fine-grained enough to discriminate between phenotype-first and gene-first explanations.

According to the burden of proof argument, when sceptics talk as though the evidence for gene-first mechanisms greatly outweighs the evidence for phenotype-first mechanisms, they are overlooking the fact that the mere construction of a plausible rival hypothesis can create an underdetermination problem, independent of the strength of the evidence for that rival hypothesis. It is important not to overstate this burden of proof claim, however. Unless advocates of plasticity-first evolution mean to make the radical assertion that there is no evidence for gene-first mechanisms that is not subject to a plasticity-first interpretation, the most the argument can do is lead us to adjust our priors concerning the plausibility of plasticity-first evolution, not deliver evidential parity across the board.

In any case, the burden of proof argument is a double-edged sword. If, as advocates generally concede, gene-first and phenotype-first explanations are often underdetermined relative to the present evidence, the argument does not give us reason to accept the plasticity-first hypothesis. At best, we should suspend judgement about the relative importance of gene-first and phenotype-first mechanisms because we do not have evidence that can help us determine which mechanism was active in particular evolutionary episodes, much less give us information about the relative frequency and significance of these mechanisms. It is therefore important to face the issues of discriminatory and significance-relevant evidence head-on. It is to these issues that I turn next. There are substantive questions about confirmation and the plasticity-first hypothesis that the debate surrounding the hypothesis has not addressed. How can they be resolved?

4 The Importance of Middle-Range Theory

In this Section I argue that to be in a position to say what data would be discriminatory and significance-relevant, researchers need a richer middle-range theory of plasticity-first evolution, one that allows them to identify its distinctive empirical traces.[9] A middle-range theory is a theory of the relationship between a process of interest and the observable evidence it produces (Binford [1982]; Jeffares [2008]). Such a theory allows scientists to guard against false positives and false negatives, and to address worries about the degradation of evidential traces over time.

First, I revisit the possibility of confirming the hypothesis by observing it in a natural population. I have already claimed that this sense of direct evidence is an unreasonably high evidential standard, and that it is a mistake to say that only observations of natural populations count as direct. Here, I show that such evidence is also unable to confirm the plasticity-first hypothesis because it so rarely yields discriminatory evidence. Then I show that for similar reasons, ancestral–descendant comparisons, the most popular means of studying plasticity-first evolution, are not sufficient to confirm the hypothesis as long as our middle-range theory remains as rudimentary as it currently is.

4.1 Direct observation

Sceptics and advocates agree that the ideal way to gather evidence about plasticity-first evolution would be to watch it happen, and though demanding this ideal is unreasonable, there is still value in thinking about what

[9] These traces need not be unique, just distinct enough to allow us to discriminate between the plasticity-first hypothesis and the competing alternative explanations for the origin of evolutionary novelties.

researchers would have to observe in the ideal case in order to conclude that plasticity-first evolution was occurring. To observe environmental induction and genetic assimilation in real time, biologists would first need to identify a population undergoing rapid evolution in the wild. In such a population, there are seven observations that, taken together, would establish that a novel trait evolved by plasticity-first evolution rather than some competing alternative mechanism.

(1) Emergence of a novel phenotype: Some members of the population develop a phenotype (such as smaller eyes or narrower pectoral bones) that was not formerly present in the population and which performs a new function in the ecology of the lineage.

(2) Presence of an inducer: The population experiences a novel and intergenerationally stable environmental condition such as a change in salinity, conductivity, temperature, or nutrient availability.

(3) Causal link between inducer and novel phenotype: The inducer is the difference-maker between the novel phenotype emerging or not emerging.

(4) Adaptiveness of the novel phenotype: The novel phenotype has a fitness benefit in the novel environment.

(5) Genetic basis of the novel phenotype: The same genetic variants that underpin the novel phenotype were associated with a different phenotype before the introduction of the inducer.

(6) Spread of the novel phenotype: The adaptive phenotype must spread throughout the population.

(7) Subsequent selection on the novel phenotype: Once (or as) the novel phenotype spreads, changes in its genetic basis that further improve its form, function, or regulation are selected, leading to the persistence of the novel phenotype even in the absence of the original inducer.

Pursuing such direct confirmation of the plasticity-first hypothesis has a number of disadvantages. First, biologists must have the good fortune to catch a natural population in the act of rapid evolution, as well as the ability to identify in advance what trait to measure. As challenging as this seems, it may not be impossible (Moczek [2007]). Second, provided biologists can pass this first hurdle, the ensuing research would be difficult, time-consuming, and expensive. Third, it is a forward-looking approach, and cannot tell us about evolutionary episodes that have already happened. If direct observation were the only evidence-gathering option, then we would not be able to answer questions about plasticity-first evolution versus some alternative in particular

historical cases. Finally, one or two or even ten direct observations do not amount to confirmation of the significance claim embedded in the plasticity-first hypothesis except in the unlikely event that the observation is of a high-profile novelty.[10]

To demonstrate evolutionary significance, we have to be able to generalize beyond a few observations and make inferences about how often and under what conditions plasticity-first evolution occurs. Making these inferences requires more information than the mere fact of direct observations. Thus, even if successful, direct observation is not significance-relevant and so does not confirm the plasticity-first hypothesis. Direct observations might be very powerful evidence if combined with additional data generated by other methodological approaches, but they do not, in isolation, have much confirmatory power.

The conceit of the idealized data set also highlights the difficulties of using techniques other than direct observation to confirm the plasticity-first hypothesis. The primary difficulty is that induction and assimilation occur quickly (probably over several tens of generations), and such transient processes do not leave stable and easily detectable empirical traces in their wake. Most of the evidence of past evolution that is readily accessible tells us only that both genes and phenotypes change over time, but not the order in which those changes occur.

In the case of direct observation, researchers' ability to collect dynamic rather than static data (Lewontin [2002]; Forber [2009]) could provide insight into the ordering of genetic change and phenotypic change. The challenge for other evidence-gathering techniques is to devise approximations of dynamic data from static data. This, of course, is a problem common to the historical sciences rather than a unique issue for the plasticity-first hypothesis. The general solution to the problem requires researchers to search out physical traces left by past events. Their goal is to find a signature: a trace or set of traces that uniquely picks out one of the mechanisms under consideration. Sometimes, a single trace, or 'smoking gun', may serve to discriminate between alternatives, though more often, multiple independent traces must converge in order to rule out one alternative and confirm another (Cleland [2002]; Forber and Griffith [2011]). Of course, there is no guarantee that every mechanism or process of interest has a signature (Turner [2007]), but it often happens that methodological and technical advances uncover confirming traces that scientists previously believed were inaccessible (Currie [2014]; Turner [2016]).

How can researchers identify the signature of a process like plasticity-first evolution? This is where middle-range theory becomes important. When the

[10] I say 'unlikely' because novelties such as the tetrapod limb take many, many human lifetimes to evolve.

distinctive empirical traces of a process are not apparent, researchers must invest in developing a theory about the relationship between the process they are interested in and the observable evidence the process produces. Once armed with the appropriate middle-range theory, researchers can confirm plasticity-first evolution in particular cases without directly observing it. They can determine which data are actually discriminatory, and then design studies of evolutionary episodes that seek out the relevant signature. When they find the signature, they can rule out alternative explanations and conclude that plasticity-first evolution has occurred.

Even more importantly, the signature of plasticity-first evolution can assist researchers in the project of making generalizations about evolutionary significance. They can use the signature to compile a database of individual cases of plasticity-first evolution, as well as confirmed cases of gene-first novelties, and then use this database to support inferences about the frequency of plasticity-first evolution, about what if any distinctive modes and tempos of evolution it produces, and about the probability that particular high-profile novelties emerged due to plasticity-first evolution. This kind of information is necessary to respond to sceptics who worry that the distinction between plasticity-first evolution and gene-first evolution is explanatorily inert.

4.2 Ancestral–descendant comparisons

Unfortunately, research on plasticity-first evolution is not focused on developing middle-range theory nor characterizing a signature. Advocates have identified one empirical trace that under certain conditions provides discriminatory evidence for plasticity-first evolution, and they focus instead on carrying out ancestral–descendant comparisons (studies that compare characteristics of a descendant population to those of an ancestral one) that provide this kind of evidence.[11] Researchers may compare genomic information, when it is available, but most often they test for plasticity in the development of particular traits. When (i) a descendant population has phenotypically diverged from its ancestors in ways that appear to be adaptive and (ii) the ancestral population shows plasticity for the trait or traits that have evolved in the descendants, researchers take this as evidence that induced plastic responses in the ancestors drove subsequent evolution in their descendants.[12]

[11] Of course, researchers cannot look at the literal ancestral populations (because they are dead), but must identify extant populations that are reasonable analogues of the true ancestors. Identifying such proxy ancestral populations is a general problem in evolutionary biology rather than a problem that is particular to the plasticity-first hypothesis, so I will not take it up in this article.

[12] The full set of data that researchers look for is a bit more detailed (see Levis and Pfennig [2016]), but these are the key findings.

But the inferences one can draw from these comparative studies are more limited than advocates admit. There are two reasons. First, many ancestral–descendant comparisons are not discriminatory because they provide evidence about either environmental induction or genetic assimilation, but not both. Second, those studies that are discriminatory are too rare to support inferences about evolutionary significance.

In order to rule out a gene-first explanation for a particular novelty, a study needs to show that both environmental induction and genetic assimilation occurred in the same population. It is not sufficient to show the independent occurrence of just one or the other, because the independent occurrence of either of these processes is compatible with a gene-first explanation for phenotypic novelty.

Consider the stickleback example from Section 2. Recently diverged lake and stream stickleback populations have adaptive differences in mouth shape, and you can experimentally induce the lake phenotype in stream-dwelling fish (and vice versa). This case provides compelling evidence for adaptive plasticity in both stickleback populations. But it does not allow us to make inferences about plasticity-first evolution in these populations. The newer mouth phenotype may be the precursor to a true evolutionary novelty, or it may not. At present, the descendant stickleback population is as plastic as the ancestral one, meaning that genetic assimilation has not occurred. We have no evidence that any genetic changes have occurred to make the adaptive new phenotype more stable, that the induced response is on its way to becoming constitutive, that the stickleback population(s) will become less plastic with respect to this trait in the future, or that the new phenotype will figure in their subsequent evolution.

A second example is Carol Lee's work on marine copepod invasions of freshwater habitats, work that supports conclusions about genetic assimilation but not environmental induction. In order to invade a freshwater habitat, these tiny marine crustaceans must evolve new ion regulation mechanisms. Lee *et al.* ([2011]) studied the role of two enzymes (V-type H ATPase and Na/K ATPase) involved in copepod ion transportation and showed that copepods from marine habitats in the Atlantic Ocean and Gulf of Mexico rapidly evolved increased plasticity for enzyme function when exposed to freshwater conditions. This increased plasticity was adaptive because it improved the copepods' ability to survive in freshwater. Further, copepod populations that have already made the transition from seawater to freshwater show less plasticity for enzyme function than Lee's experimental populations, supporting the conclusion that these invading populations have genetically assimilated a previous plastic response. But it's an open question whether plasticity in the original invaders existed before the invasion, or if a regulatory mutation following the invasion increased plasticity (a gene-first mechanism).

If anything, the study supports the latter possibility, since Lee's experimental population evolved increased plasticity for ion regulation after exposure to novel salinity levels rather than demonstrating plasticity immediately.

Advocates cite these examples, and others like them, as evidence for the plasticity-first hypothesis. In a sense, they are correct. Ancestral–descendant comparisons support the plasticity-first hypothesis by deepening our understanding of how and when developmental plasticity, environmental induction, and genetic assimilation occur. By demonstrating individual components of the hypothesis, these studies also lend plausibility to the claim that the entire process occurs in nature sometimes. They help to build a circumstantial case for plasticity-first evolution. But circumstantial evidence isn't good enough, because the plasticity-first hypothesis does not merely claim that plasticity-first evolution probably occurs in nature sometimes. It claims that plasticity-first evolution is an evolutionarily significant novelty-generating mechanism. To support that claim, we need to be able to connect the plasticity-first process to particular novelties, and that requires us to be able to discriminate between gene-first and phenotype-first explanations in particular cases (some researchers do recognize this explicitly, see Levis and Pfennig [2016]).

The second problem with ancestral–descendant comparisons is that even when they are discriminatory, there are very few candidate model systems that can support conclusions about evolutionary significance, and these model systems have special characteristics from which it is difficult to generalize. Ancestral-derived comparisons require recently diverged population pairs in which the descendants have adaptively diverged from the ancestors. Adaptive radiations (for example, African cichlid fish and stickleback) and recent invasions (for example, copepods and tiger snakes) are good sources for such pairs, but there are many additional features that populations need to have that further narrows down the set of good candidates for an integrated demonstration of plasticity-first evolution (Levis and Pfennig [2016]). If we want to be able to make inferences about older evolutionary novelties or about novelties that did not originate in the context of invasions or adaptive radiations, special cases are of limited value. Even if researchers can demonstrate plasticity-first evolution in each of these model systems, they still need a way to leverage these demonstrations into sources of evidence about other evolutionary episodes to which our access is more limited. So there is an important gap between discriminatory evidence and significance-relevant evidence. The best way to bridge this gap is to search for additional empirical traces of plasticity-first evolution. If we identify additional traces, we can expand the set of confirmed cases of plasticity-first evolution that forms the basis for significance judgements. That's why it is important to enrich the middle-range theory of plasticity-first evolution. We need to build up our understanding of how the process occurs and the kinds of marks it leaves

on the world so that we can identify tokens of the process when we encounter them and tell tokens of different processes apart.

There are a couple of ways to enrich the middle-range theory and find more traces. One approach, which would not require much of a shift from researchers' current methodology, is to continue doing ancestral–descendant comparisons, but to look for additional traces in model systems where it is already established that plasticity-first evolution has occurred. These additional traces might then be identifiable in other natural populations, even ones that are not ideal for conducting ancestral-derived comparisons. Researchers have not yet tried this strategy, perhaps because there are still no uncontroversial demonstrations of plasticity-first evolution in a natural population (Levis and Pfennig [2016]).

Other approaches to characterizing the signature of plasticity-first evolution depart more significantly from existing methodological norms. In fact, they go against truisms about quality of evidence espoused by many researchers. The approaches I have in mind are experimental evolution and formal modelling.

5 Adjusting Methodological Norms

Researchers recognize that formal models and experimental evolution can contribute to the study of plasticity-first evolution, but they also grant epistemic priority to data from natural populations. According to Schlichting and Wund ([2014], pp. 665–6), 'Ancestral–descendant or sister taxon comparisons provide the strongest evidence that genetic accommodation is frequent in nature'. They also write that 'Although experimental evolution studies provide definitive evidence for demonstrating both the possibility and mechanisms of genetic accommodation, evidence from natural populations is preferable for indicating the prevalence of this process in nature' ([2014], pp. 660–1). Similarly, Levis and Pfennig ([2016], p. 3) tell us that 'studying the plasticity-first hypothesis in lab populations of rapidly evolving organisms would be worthwhile but would not clarify whether plasticity has contributed to adaptation in any natural population'.

These biologists are correct that we must ultimately refer to data from natural populations to assess the historical evolutionary significance of plasticity-first evolution. But in drawing this conclusion, they discount the critical role of other methodological techniques in determining what kind of data from natural populations they should look for. This discounting shows up not only in their explicit statements about evidence, but in their concrete research recommendations as well. And it echoes the deep-seated yet dubious idea that certain kinds of experiments enjoy some in-principle epistemic privilege relative to other investigative techniques (Parke [2014]).

Such discounting is a mistake because though the literature on plasticity offers some clues about what additional traces of plasticity-first evolution might be, the area is largely uncharted territory. One suggestion is that the time scale of plasticity-first evolution is much shorter than that of gene-first mechanisms (Lande [2009]). If so, researchers may be able to make inferences about which process is more likely in cases where they have information about the rate of evolution. Authors also occasionally reference a possible genomic signature of plasticity-first evolution that distinguishes it from gene-first processes. Characterizing such a signature would allow for inferences and generalizations that go beyond those licensed by ancestral-derived comparisons.

The ease of suggesting and imagining possible signatures of plasticity-first evolution raises the question of how these possibilities can be tested. Researchers are, after all, in a paradoxical position. In order to identify distinctive characteristics of plasticity-first evolution, they need clear cases of plasticity-first evolution to observe and manipulate. But in order to identify such cases, they need some prior knowledge about its distinctive characteristics.

This chicken and egg problem arises because it is almost prohibitively difficult to identify cases of plasticity-first evolution in nature. But there is another option: researchers can construct cases of plasticity-first evolution for themselves. Experimental evolution and formal modelling confer this ability. Not only do they allow researchers to construct cases of plasticity-first evolution, they make it possible to observe plasticity-first processes in real time, not just once, but many times over. Thus, these techniques allow for more direct observation and manipulation of plasticity-first evolution than comparative studies do. They are better suited for the tasks of developing middle-range theory and characterizing the signature of plasticity-first evolution than studies of natural populations.

To illustrate the value of the formal modelling and experimental evolution, I will discuss the case of selective sweeps, in which these techniques are already helping to elucidate the signatures of evolutionary processes. Then I will consider how an extension of this approach could contribute to the study of plasticity-first evolution. The growing literature on how to assess the plasticity-first hypothesis has not seriously entertained the kind of strategy I am advancing here, but the limitations of comparative studies demand that we combine them with different approaches.

5.1 Formal modelling and experimental evolution

Before introducing the selective sweep example, I will review the modelling and experimental techniques I have in mind. Both traditional quantitative genetic models and agent-based simulations are useful for investigating the signature

of plasticity-first evolution. Quantitative genetic models can generate empirically testable predictions about the differences between gene-first and phenotype-first processes, predictions that can guide future experimental set-ups and parameter choices. For instance, the prediction that plasticity-first evolution is more rapid than gene-first evolution comes from the quantitative genetic literature (Lande [2009]; Frank [2011]).

Simulations, by contrast, allow biologists to set up both phenotype-first and genotype-first evolutionary scenarios, run them thousands of times, and then look for interesting differences in outcome between the two kinds of scenarios. If they can be externally validated, these differences will be diagnostic of plasticity-first evolution in natural populations. An intriguing example of this kind of strategy comes from Draghi and Whitlock ([2012]), who simulated evolution of a gene-network model in three different of environments. They use the model to investigate the genetic basis of traits that have evolved as plastic responses to environmental variation, but it's an approach that could also be used to investigate plasticity-first evolution and identify features of plasticity-first evolution that differ from gene-first evolution.

Experimental evolution is 'research in which populations are studied across multiple generations under defined and reproducible conditions, whether in the laboratory or in nature' (Garland and Rose [2009], p. 3). This broad definition encompasses a range of experimental techniques as varied as artificial selection, laboratory natural selection (LNS), habitat alteration, and monitoring invasive species. What sets these approaches apart is their ability to generate dynamic rather than static data about a population's response to selection (Parke [unpublished]).

Of these experimental techniques, artificial selection—which involves breeding populations in a laboratory setting and selecting for a particular trait in each generation—has played an outsized role in research on plasticity-first evolution. Some of the earliest empirical investigations of the hypothesis were artificial selection experiments. Waddington ([1953], [1956]) reduced the activity of the Hsp90 protein (Rutherford and Lindquist [1998]) by exposing fruit fly larvae to heat shock, which induced a new phenotype: wings without cross-veined patterns. After several generations of artificial selection for this phenotype, it developed consistently even without the heat shock treatment that originally induced it.[13]

Today, Waddington's study is regarded as the classic proof of possibility of genetic assimilation, but the artificial selection method suffers from two limitations that consign results such as Waddington's to second-best status in the eyes of many researchers. First, the novel phenotype Waddington induced is not adaptive. Second, it is not clear whether the inducing condition, heat

[13] See (Suzuki and Nijhout [2006], [2008]) for additional examples of artificial selection studies.

shock, is one that a natural population of fruit flies would ever encounter. These two limitations highlight what many see as a more general drawback of artificial selection experiments: the difficulty of using them to make inferences about selective processes in nature (Rohner *et al.* [2013]).

LNS is an approach that circumvents these limitations. In LNS, 'the experimenter divides replicate lines among two or more environmental treatments and examines how the experimental stocks respond over time' (Fuller *et al.* [2005], p. 391). The key difference between artificial selection and LNS is that the experimenter does not choose which individuals will reproduce in each generation. As a result, any phenotypic novelties that emerge over the course of the experiment are adaptive responses to the experimental population's environmental treatment. If these environmental treatments are reasonable approximations of selection pressures the experimental population might encounter outside the laboratory, then the limitations of artificial selection experiments are not problems for LNS.

LNS should be more common because it is an ideal method for characterizing the signature of plasticity.[14] Once researchers have evolved a novel phenotype and then confirmed the mechanism by which it occurred (gene-first or phenotype-first), they can go back and re-examine data from each time-step of the process in search of distinctive, identifying patterns. They can even contrast cases in which different mechanisms predominate and look for divergences between them.

5.2 An example: selective sweeps

These techniques—quantitative genetic models, simulations, and LNS—have all contributed to attempts to find empirical differences between adaptation from standing genetic variation and adaptation from new (*de novo*) mutations. This project is related to plasticity-first research, but the contrast between standing variation and *de novo* mutations does not perfectly overlap the one between gene-first and phenotype-first mechanisms. *De novo* mutation includes both coding and non-coding mutations, while standing variation includes cases of recombination, environmental induction, and selection of phenotypes that already exist in a population at low frequency (so, not novelties at all). Thus, when biologists ask about the relative frequency of adaptation from standing genetic variation, they are asking about something importantly different from plasticity-first evolution. Still, the techniques that are helping to identify the first kind of signature may also help characterize the second.

Biologists try to distinguish between adaptation from standing variation and *de novo* mutation by looking for differences in what are called selective

[14] For one rare case, see (Lachapelle *et al.* [2015]).

sweeps. Selective sweeps occur when the alleles that are located near an adaptive allele under natural selection also get selected for. The resulting reduction in genetic variation surrounding the allele under direct selection is called a selective sweep. Selective sweeps are both weaker and narrower if selection is for standing genetic variation rather than a *de novo* mutation because a *de novo* mutation is selected for from the moment it appears, while standing variation is neutral for a time before it becomes adaptive in a new environment. Previously neutral alleles can reach intermediate frequency in a population without ever being under direct selection. Once they are under direct selection, they sweep through a population more quickly than a *de novo* mutation would, and this is what creates the weaker and narrower (or, soft, as opposed to hard) sweep pattern (Barrett and Schluter [2008]).

We have this prediction about selective sweeps because of formal modelling. And though looking for hard and soft sweeps is far from a perfect method for distinguishing between *de novo* adaptations and adaptation from standing variation (Teshima *et al.* [2006]), it is a helpful tool, and one that improves over time as modelling approaches become more sophisticated (for an example of such increased sophistication, see Peter *et al.* [2012]).

The other technique that is improving biologists' ability to distinguish between these two sources of adaptation and provide information about their relative frequency is LNS. Depending on their choice of experimental population they can study either hard sweeps (in asexual populations) or soft sweeps (in sexual populations) in isolation (Burke [2012]). There are also some model systems that allow for comparative analyses of both types of sweeps, and these support, among other things, a much larger role for adaptation from standing variation than from *de novo* variation (Burke *et al.* [2014]).

This selective sweep example is of dual relevance to the plasticity-first hypothesis. First, it is a case where formal modelling and experimental evolution are doing precisely the kind of work that is of paramount importance for plasticity-first research. It shows that the methodological strategy I am defending can be successful. Second, the idea that selective sweeps or other genomic patterns may be diagnostic of plasticity-first evolution is barely addressed in the literature (Gibson and Dworkin [2004]). As a result, no one knows if there is a difference between gene-first and phenotype-first adaptation that parallels the one between standing variation and *de novo* mutation. But this question should absolutely be explored. If there is a signature of phenotype-first evolution that is analogous to that of standing genetic variation, discovering it would transform the debate about the evolutionary significance of plasticity. Even if genomic data alone are not sufficient to pick out phenotype-first adaptations, they may be able to do so in combination with other kinds of data, such as information about a fixed allele's selective environment.

6 Conclusion

The existing methodological strategy in plasticity-first research involves combining comparative studies of different populations that demonstrate different pieces of the plasticity-first hypothesis. Together, these studies are supposed to approximate the ideal data set much like a mosaic might approximate a photograph. I have argued that this strategy has not proven effective for generating discriminatory evidence. Worse, it is insufficient as a stand-alone strategy for confirming or disconfirming the plasticity-first hypothesis in the future. We should be concerned about this state of affairs because the plasticity-first hypothesis stands at the centre of one of the great questions facing evolutionary biology today: the extent to which the modern synthesis needs to be revised or expanded to accommodate developmental and environmental influences on evolution.

I have argued that in addition to piecing together a mosaic of studies that address the diverse components of plasticity-first evolution, researchers should go in for diversity of a further kind—methodological diversity that makes greater use of the full set of tools available to evolutionary biologists. This means directing resources toward formal modelling and experimental evolution studies. Once researchers use these tools to learn more about the signature of plasticity-first evolution, they may find that the data they have already collected have more discriminatory power than they thought, or they may find that they need to look for very different data. But the priority has to be getting researchers to a point where they can identify discriminatory evidence when they see it.

It is also possible that the plasticity-first hypothesis will not turn out to have a distinct signature. There is no guarantee we can find evidence that bears on every empirical question that interests us, though the historical sciences continue to surprise the wider scientific community with their ingenuity (Currie [2014]; Turner [2016]). Regardless of whether there is a signature to be found, however, experimental evolution and formal models are where sceptics and advocates of plasticity-first evolution alike should continue their dispute about the quality of the evidence for the hypothesis. These are the research areas that will allow us to judge the depth of the underdetermination problem facing the plasticity-first hypothesis and whether distinguishing between plasticity-first and gene-first mechanisms of evolution is a matter of preference or genuine scientific interest.

Acknowledgements

I am very grateful to Adrian Currie, Alex Guerrero, Tim Linksvayer, Mary Morgan, Emily Parke, Brian Reese, Carlos Santana, Elliott Sober, Quayshawn

Spencer, Michael Weisberg, Matt Wund, and two anonymous reviewers for their constructive comments on earlier drafts of the manuscript. I also thank the members of the Penn Philosophy of Science Reading Group and the audience at the 2016 Philosophy of Biology at Madison Workshop for their encouragement and helpful suggestions.

Department of Philosophy
University of Pennsylvania
Philadelphia, PA, USA
kkovaka@sas.upenn.edu

References

Amundson, R. [2005]: *The Changing Role of the Embryo in Evolutionary Thought: Roots of Evo-Devo*, Cambridge: Cambridge University Press.

Barrett, R. D. H. and Schluter, D. [2008]: 'Adaptation from Standing Genetic Variation', *Trends in Ecology and Evolution*, **23**, pp. 38–44.

Binford, L. R. [1982]: 'Objectivity–Explanation–Archaeology', in C. Renfrew, M. J. Rowlands and B. A. Segraves (*eds*), *Theory and Explanation in Archaeology*, New York: Academic Press, pp. 125–38.

Burke, M. K. [2012]: 'How Does Adaptation Sweep through the Genome? Insights from Long-term Selection Experiments', *Proceedings of the Royal Society B*, **279**, pp. 5029–38.

Burke, M. K., Liti, G. and Long, A. D. [2014]: 'Standing Genetic Variation Drives Repeatable Experimental Evolution in Outcrossing Populations of *Saccharomyces cerevisiae*', *Molecular Biology and Evolution*, **3**, pp. 3228–39.

Cleland, C. E. [2002]: 'Methodological and Epistemic Differences between Historical Science and Experimental Science', *Philosophy of Science*, **69**, pp. 447–51.

Crispo, E. [2007]: 'The Baldwin Effect and Genetic Assimilation: Revisiting Two Mechanisms of Evolutionary Change Mediated by Phenotypic Plasticity', *Evolution*, **61**, pp. 2469–79.

Currie, A. [2014]: 'Marsupial Lions and Methodological Omnivory: Function, Success, and Reconstruction in Paleobiology', *Biology and Philosophy*, **30**, pp. 187–209.

de Jong, G. [2005]: 'Evolution of Phenotypic Plasticity: Patterns of Plasticity and the Emergence of Ecotypes', *New Phytologist*, **166**, pp. 101–18.

de Jong, G. and Crozier, R. H. [2003]: 'A Flexible Theory of Evolution', *Nature*, **424**, pp. 16–17.

Dickins, T. E. and Rahman, Q. [2012]: 'The Extended Evolutionary Synthesis and the Role of Soft Inheritance in Evolution', *Proceedings of the Royal Society B*, **278**, pp. 1721–7.

Draghi, J. A. and Whitlock, M. C. [2012]: 'Phenotypic Plasticity Facilitates Mutational Variance, Genetic Variance, and Evolvability along the Major Axis of Environmental Variation', *Evolution*, **66**, pp. 2891–902.

Endler, J. A. [1986]: *Natural Selection in the Wild*, Princeton, NJ: Princeton University Press.

Forber, P. [2009]: 'Spandrels and a Pervasive Problem of Evidence', *Biology and Philosophy*, **24**, pp. 247–66.

Forber, P. and Griffith, E. [2011]: 'Historical Reconstruction: Gaining Epistemic Access to the Deep Past', *Philosophy and Theory in Biology*, **3**, available at <dx.doi.org/10.3998/ptb.6959004.0003.003>.

Forsman, A. [2015]: 'Rethinking Phenotypic Plasticity and Its Consequences for Individuals, Populations, and Species', *Heredity*, **115**, pp. 276–84.

Frank, S. A. [2011]: 'Natural Selection, II: Developmental Variability and Evolutionary Rate', *Journal of Evolutionary Biology*, **24**, pp. 2310–20.

Fuller, R. C., Baer, C. F. and Travis, J. [2005]: 'How and When Selection Experiments Might Actually Be Useful', *Integrative and Comparative Biology*, **45**, pp. 391–404.

Futuyma, D. J. [2011]: 'Expand or Revise? The Evolutionary Synthesis Today', *The Quarterly Review of Biology*, **86**, pp. 203–8.

Garland, T. and Rose, M. R. [2009]: *Experimental Evolution*, Berkeley, CA: University of California Press.

Gibson, G. and Dworkin, I. [2004]: 'Uncovering Cryptic Genetic Variation', *Nature Reviews Genetics*, **5**, pp. 681–90.

Gilbert, S. and Epel, D. [2009]: *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution*, Sunderland, MA: Sinauer Associates.

Godfrey-Smith, P. [2001]: 'Three Kinds of Adaptationism', in S. H. Orzack and E. Sober (*eds*), *Adaptationism and Optimality*, Cambridge: Cambridge University Press, pp. 335–57.

Grant, P. R. and Grant, B. R. [2009]: 'The Secondary Contact Phase of Allopatric Speciation in Darwin's Finches', *Proceedings of the National Academy of Sciences*, **106**, pp. 20141–8.

Jablonka, E. and Raz, G. [2009]: 'Transgenerational Epigenetic Inheritance: Prevalence, Mechanisms, and Implications for the Study of Heredity and Evolution', *The Quarterly Review of Biology*, **84**, pp. 131–76.

Jeffares, B. [2008]: 'Testing Times: Regularities in the Historical Sciences', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **39**, pp. 469–75.

Kaplan, J. [2008]: 'Evolutionary Innovations and Developmental Resources: From Stability to Variation and Back Again', *Philosophy of Science*, **75**, pp. 861–73.

Lachapelle, J., Bell, G. and Colegrave, N. [2015]: 'Experimental Adaptation to Marine Conditions by a Freshwater Alga', *Evolution*, **69**, pp. 2662–75.

Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E. and Odling-Smee, J. [2015]: 'The Extended Evolutionary Synthesis: Its Structure, Assumptions, and Predictions', *Proceedings of the Royal Society B*, **282**, p. 20151019.

Laland, K., Uller, T., Feldman, M., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E., Odling-Smee, J., Wray, G., Hoekstra, H., Futuyma, D., Lenski, R., Mackay, T., Schluter, D. and Strassmann, J. E. [2014]: 'Does Evolutionary Theory Need a Rethink?', *Nature*, **514**, pp. 161–4.

Lande, R. [2009]: 'Adaptation to an Extraordinary Environment by Evolution of Phenotypic Plasticity and Genetic Assimilation', *Journal of Evolutionary Biology*, **22**, pp. 1435–46.

Lee, C. E., Kiergaard, M., Gelembiuk, G. W., Eads, B. D. and Posavi, M. [2011]: 'Pumping Ions: Rapid Parallel Evolution of Ionic Regulation Following Habitat Invasions', *Evolution*, **65**, pp. 2229–44.

Levis, N. A. and Pfennig, D. W. [2016]: 'Evaluating "Plasticity-First" Evolution in Nature: Key Criteria and Empirical Approaches', *Trends in Ecology and Evolution*, **31**, pp. 563–74.

Lewontin, R. C. [2002]: 'Directions in Evolutionary Biology', *Annual Review of Genetics*, **36**, pp. 1–18.

Love, A. C. [2006]: 'Evolutionary Morphology and Evo-Devo: Hierarchy and Novelty', *Theory in Biosciences*, **124**, pp. 317–33.

Lucek, K., Sivasundar, A. and Seehausen, O. [2014]: 'Disentangling the Role of Phenotypic Plasticity and Genetic Divergence in Contemporary Ecotype Formation during a Biological Invasion', *Evolution*, **68**, pp. 2619–32.

Moczek, A. P. [2007]: 'Developmental Capacitance, Genetic Accommodation, and Adaptive Evolution', *Evolution and Development*, **9**, pp. 299–305.

Moczek, A. P. [2015]: 'Developmental Plasticity and Evolution—Quo Vadis?', *Heredity*, **115**, pp. 302–5.

Moczek, A. P., Sultan, S., Foster, S., Ledón-Rettig, C., Dworkin, I., Nijhout, H. F., Abouheif, E. and Pfennig, D. W. [2011]: 'The Role of Developmental Plasticity in Evolutionary Innovation', *Proceedings of the Royal Society of London B*, **278**, pp. 2705–13.

Muller, G. B. and Wagner, G. P. [1991]: 'Novelty in Evolution: Restructuring the Concept', *Annual Review of Ecology and Systematics*, **22**, pp. 229–56.

Orr, H. A. [1999]: 'An Evolutionary Dead End?', *Science*, **285**, pp. 343–4.

Orzack, S. H. and Sober, E. [1994]: 'Optimality Models and the Test of Adaptationism', *The American Naturalist*, **143**, pp. 361–80.

Parke, E. C. [2014]: 'Experiments, Simulations, and Epistemic Privilege', *Philosophy of Science*, **81**, pp. 516–36.

Parke, E. [unpublished]: 'Tapes of Life on Ice: Philosophical Lessons from Experimental Evolution'.

Peter, B. M., Huerta-Sanchez, E. and Nielsen, R. [2012]: 'Distinguishing between Selective Sweeps from Standing Variation and from a *de novo* Mutation', *PLOS Genetics*, **8**, p. e1003011.

Pigliucci, M. [2001]: *Phenotypic Plasticity: Beyond Nature and Nurture*, Baltimore, MD: Johns Hopkins University Press.

Pigliucci, M. [2007]: 'Do We Need an Extended Evolutionary Synthesis?', *Evolution*, **61**, pp. 2743–9.

Pigliucci, M. [2008]: 'What, If Anything, Is an Evolutionary Novelty?', *Philosophy of Science*, **75**, pp. 887–98.

Pigliucci, M. and Murren, C. J. [2003]: 'Perspective: Genetic Assimilation and a Possible Evolutionary Paradox: Can Macroevolution Sometimes Be so Fast as to Pass Us By?', *Evolution*, **57**, pp. 1455–64.

Pigliucci, M., Murren, C. J. and Schlichting, C. D. [2006]: 'Phenotypic Plasticity and Evolution by Genetic Assimilation', *Journal of Experimental Biology*, **209**, pp. 2362–7.

Rohner, N., Jarosz, D. F., Kowalko, J. E., Yoshizawa, M., Jeffery, W. R., Borowsky, R. L., Lindquist, S. and Tabin, C. J. [2013]: 'Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish', *Science*, **342**, pp. 1372–5.

Rutherford, S. L. and Lindquist, S. [1998]: 'Hsp90 as a Capacitor for Morphological Evolution', *Nature*, **396**, pp. 336–42.

Santos, M., Szathmáry, E. and Fontanari, J. F. [2015]: 'Phenotypic Plasticity, the Baldwin Effect, and the Speeding up of Evolution: The Computational Roots of an Illusion', *Journal of Theoretical Biology*, **371**, pp. 127–36.

Schlichting, C. D. and Wund, M. A. [2014]: 'Phenotypic Plasticity and Epigenetic Marking: An Assessment of Evidence for Genetic Accommodation', *Evolution*, **68**, pp. 656–72.

Sklar, L. [1975]: 'Methodological Conservatism', *Philosophical Review*, **84**, pp. 374–400.

Standen, E. M., Du, T. Y. and Larsson, H. C. E. [2014]: 'Developmental Plasticity and the Origin of Tetrapods', *Nature*, **513**, pp. 54–8.

Sterelny, K. [2000]: 'Development, Evolution, and Adaptation', *Philosophy of Science*, **67**, pp. S369–87.

Sterelny, K. [2012]: *The Evolved Apprentice*, Cambridge, MA: MIT Press.

Suzuki, Y. and Nijhout, H. F. [2006]: 'Evolution of a Polyphenism by Genetic Accommodation', *Science*, **311**, pp. 650–2.

Suzuki, Y. and Nijhout, H. F. [2008]: 'Genetic Basis of Adaptive Evolution of a Polyphenism by Genetic Accommodation', *Journal of Evolutionary Biology*, **21**, pp. 57–66.

Teshima, K. M., Coop, G. and Przeworski, M. [2006]: 'How Reliable Are Empirical Genomic Scans for Selective Sweeps?', *Genome Research*, **16**, pp. 702–12.

Turner, D. [2007]: *Making Prehistory: Historical Science and the Scientific Realism Debate*, Cambridge: Cambridge University Press.

Turner, D. [2016]: 'A Second Look at the Colors of the Dinosaurs', *Studies in History and Philosophy of Science A*, **55**, pp. 60–8.

Waddington, C. H. [1953]: 'Genetic Assimilation of an Acquired Character', *Evolution*, **7**, pp. 118–26.

Waddington, C. H. [1956]: 'Assimilation of the Bithorax Phenotype', *Evolution*, **10**, pp. 1–13. 'Genetic

Wagner, A. [2011]: *The Origins of Evolutionary Innovations: A Theory of Transformative Change in Living Systems*, Oxford: Oxford University Press.

Wagner, G. P. and Lynch, V. J. [2010]: 'Evolutionary Novelties', *Current Biology*, **20**, pp. R48–52.

Weber, B. H. and Depew, D. J. [2003]: *Evolution and Learning: The Baldwin Effect Reconsidered*, Cambridge, MA: MIT Press.

West-Eberhard, M. J. [2003]: *Developmental Plasticity and Evolution*, Oxford: Oxford University Press.

West-Eberhard, M. J., Smith, J. A. C. and Winter, K. [2011]: 'Photosynthesis, Reorganized', *Science*, **332**, pp. 311–12.

Wund, M. A. [2012]: 'Assessing the Impacts of Phenotypic Plasticity on Evolution', *Integrative and Comparative Biology*, **52**, pp. 5–15.

Wund, M. A., Baker, J. A., Clancy, B., Golub, J. L., Foster, S. A., Gibson, A. E. G. and Whitlock, E. M. C. [2008]: 'A Test of the "Flexible Stem" Model of Evolution: Ancestral Plasticity, Genetic Accommodation, and Morphological Divergence in the Threespine Stickleback Radiation', *The American Naturalist*, **172**, pp. 449–62.