PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2010) **365**, 679–690
doi:10.1098/rstb.2009.0262

Idealization in evolutionary developmental investigation: a tension between phenotypic plasticity and normal stages

Alan C. Love*

Department of Philosophy, Minnesota Center for Philosophy of Science, University of Minnesota, 831 Heller Hall, 271 19th Avenue South, Minneapolis, MN 55455, USA

Idealization is a reasoning strategy that biologists use to describe, model and explain that purposefully departs from features known to be present in nature. Similar to other strategies of scientific reasoning, idealization combines distinctive strengths alongside of latent weaknesses. The study of ontogeny in model organisms is usually executed by establishing a set of normal stages for embryonic development, which enables researchers in different laboratory contexts to have standardized comparisons of experimental results. Normal stages are a form of idealization because they intentionally ignore known variation in development, including variation associated with phenotypic plasticity (e.g. via strict control of environmental variables). This is a tension between the phenomenon of plasticity and the practice of staging that has consequences for evolutionary developmental investigation because variation is conceptually removed as a part of rendering model organisms experimentally tractable. Two compensatory tactics for mitigating these consequences are discussed: employing a diversity of model organisms and adopting alternative periodizations.

Keywords: idealization; model organisms; normal stages; periodization; phenotypic plasticity; variation

1. ANALYSING REASONING STRATEGIES IN BIOLOGY

In a recent discussion about the possible advantages of incorporating aspects of systems biology into evolutionary developmental biology, Koentges (2008) remarked: 'there might be some initial disappointment that nature neither constructed its regulatory circuits with an engineer's intelligence nor used Occam's razor, whereas we must use both to describe it' (p. 663). While Koentges is appropriately optimistic about future investigative endeavours, his comment hints at an important conceptual point; we must be mindful of the differences between our descriptive or explanatory resources and the biological phenomena we attempt to descriptively capture and systematically explain. But the mindfulness is not merely a caution—sometimes our descriptive and explanatory resources are successful precisely because they intentionally ignore aspects of natural phenomena or use a variety of approximation techniques.

Idealization is one type of reasoning strategy that scientists use to describe, model and explain that purposefully departs from features known to be present in nature. For example, the interior space of a cell is often depicted as relatively empty even though intracellular space is known to be crowded (Ellis 2001); the variable of cellular volume takes on a value that is known to be false (i.e. relatively empty). Other well-known examples from physical science include frictionless planes and ideal gases. Idealizations involve knowingly ignoring variation in properties or excluding particular values for variables, in a variety of different ways, for descriptive and explanatory purposes (Jones 2005; Weisberg 2007). All reasoning strategies, including idealization, combine distinctive strengths alongside of latent weaknesses. For example, decomposing a system into its constituents to understand the features manifested by the system promotes a powerful dissection of the causal interactions of the localized constituents, while simultaneously downplaying interactions with elements external to the system (Wimsatt 1980; Bechtel & Richardson 1993). Studying the characteristic strengths and weaknesses of idealizations used in scientific investigation represents a form of 'reasoning explication', which calls for the philosophical reconstruction and evaluation of reasoning with the aim of engaging methodological (modelling or data gathering) and epistemological (explanatory evaluation or data interpretation) aspects of ongoing research (Love 2008*c*).

The goal of this paper is to scrutinize a particular form of idealization found in some evolutionary developmental (evo-devo) investigations that impinges on our understanding of phenotypic plasticity. The study of ontogeny in model organisms is usually executed by establishing a set of normal stages for embryonic development, which enables researchers in different laboratory contexts to have standardized comparisons of experimental results (Hopwood 2005, 2007). These normal stages are a form of idealization because they intentionally ignore kinds of

*aclove@umn.edu

One contribution of 12 to a Theme Issue 'From polyphenism to complex metazoan life cycles'.

variation in development, including variation associated with phenotypic plasticity (e.g. via strict control of environmental variables). This is a tension between the *phenomenon* of developmental plasticity and the *practice* of developmental staging. The tension has consequences for evo-devo investigation because specific kinds of variation in developmental features that might be relevant to evolution are downplayed in the process of rendering ontogeny experimentally tractable. After reviewing the phenomenon of plasticity and the practice of staging (§2), I focus on the tension and its consequences for evo-devo (§3). Two tactics for mitigating these consequences are discussed, both of which are already in use by biologists: employing a diversity of model organisms, and adopting alternative periodizations of developmental processes (§4). Both tactics are costly and controversial, but biologists are in a better position to prosecute further empirical inquiry when these trade-offs are explicitly characterized.

2. NATURAL PHENOMENA AND SCIENTIFIC PRACTICES

(a) *Phenotypic plasticity and evolution*

Phenotypic plasticity is a ubiquitous biological phenomenon. It involves the capacity of a particular genotype to generate phenotypic variation, often in the guise of qualitatively distinct phenotypes, in response to differential environmental cues (Pigliucci 2001; DeWitt & Scheiner 2004; Kaplan 2008; Gilbert & Epel 2009). One familiar example is seasonal caterpillar morphs that depend on different nutritional sources (Greene 1989). Some of the relevant environmental variables include temperature, nutrition, pressure/gravity, light, predators or stressful conditions and population density effects such as the presence or absence of conspecifics (Gilbert & Epel 2009). The reaction norm is a summary of the range of phenotypes, whether quantitatively or qualitatively varying, exhibited by organisms of a given genotype for different environmental conditions. When the reaction norm exhibits discontinuous variation or bivalent phenotypes (rather than quantitative, continuous variation), it is often labelled a polyphenism.

Phenotypic plasticity has been of recurring interest to biological researchers (Pigliucci 2001, ch. 3; Sarkar 2004) and controversial in evolutionary theory (Via *et al.* 1995). Extensive study of phenotypic plasticity has occurred in the context of quantitative genetic methods and phenotypic selection analyses, where the extent of plasticity in natural populations has been clearly demonstrated and operational measures delineated for its detection (Scheiner 1993; Pigliucci 2001; DeWitt & Scheiner 2004). Other aspects of plasticity that require different investigative methods include the sources of plasticity during ontogeny, the molecular genetic mechanisms that encourage (or discourage) plasticity and the kinds of mapping functions that exist between genotype and phenotype (Pigliucci 2001; cf. Kirschner & Gerhart 2005, ch. 5). In the present context, the focus is not on the extent of plasticity or how selection operates on it, but rather the origin of phenotypic variation during *and after* ontogeny: where does plasticity emerge from? How do molecular genetic mechanisms produce (or reduce) plasticity? What genotype–phenotype mapping functions are prevalent or rare?

To turn away from selection processes might initially seem to remove us from the context of *evolutionary* developmental biology and return us to developmental biology *simpliciter*. But evo-devo researchers study a variety of problems, each of which has a different structure and relates to other evolutionary problems in distinct ways (Love in press). Some of the more prominent problems, such as evolvability or the origin of novelties, are primarily about how variation is developmentally generated and what types of variation are possible (or not) in the history of life (Gerhart & Kirschner 1997; Frankino & Raff 2004; Kirschner & Gerhart 2005). The problem of explaining the origin of novelties revolves around understanding how special kinds of variation that are outside the range of current developmental production were generated at distinct phylogenetic junctures (Love 2006*a*, 2008*a*). Questions about phenotypic plasticity related to the production of variation fall within this province and thus are germane to evo-devo investigation (Pigliucci 2001, pp. 207–214).

We can observe this more concretely by looking at an account of how novelties originate that explicitly appeals to developmental plasticity: phenotypic accommodation (West-Eberhard 2003, ch. 7; cf. Moczek 2008). West-Eberhard argues that an environmental change can alter development such that a novel phenotype appears ('plasticity'), which then can be incorporated into the existing flexibility of ontogeny ('accommodation'). Given that the environmental change would alter the ontogeny of multiple members of a population, the new variant has a mechanism for spreading through the population. Finally, genetic variation existing within the population allows natural selection to fix the trait ('assimilation') so that it can be produced without the environmental trigger in subsequent generations (cf. Suzuki & Nijhout 2006).

In order to evaluate this model properly, we need to be able to alter development through manipulation of environmental variables and observe how a novel phenotype can be established within the existing plasticity of an organism (Kirschner & Gerhart 2005, ch. 5). This manipulation would ideally allow for the identification of patterns of variation through the reliable replication of particular alterations of developmental processes. Thus, we need to measure variation and document specific patterns within different environmental regimes (which assumes variation in ontogeny is actually being measured in any one regime). Without measuring the variation across different environmental regimes, you cannot observe phenotypic plasticity (except by accident in rare cases; Cudilo *et al.* 2007). These measurements are required to document the degree of plasticity for a particular trait, as well as any patterns of plasticity that emerge, such as correlated traits or qualitatively distinct morphs (Pigliucci 2001). If the environment is a 'normal agent' in the ontogenetic process of generating phenotypes (Gilbert & Epel 2009), then an evaluation of the significance of phenotypic plasticity for

evolution (e.g. phenotypic accommodation models of the origin of novelty) requires answers to questions about where plasticity emerges, how molecular genetic mechanisms are involved in the plasticity and what genotype–phenotype relations obtain.

(b) *Developmental staging*

The study of ontogeny in model organisms is often executed by establishing a set of numbered or named stages for 'normal' embryonic development that allow researchers in different laboratory contexts to obtain standardized experimental results that can be compared and contrasted (Hopwood 2005, 2007). They are critical to large communities of developmental biologists working on well-established models, such as chick (Hamburger & Hamilton 1951), *Xenopus* (Nieuwkoop & Faber 1956) and zebrafish (Kimmel *et al.* 1995): 'Embryological research is now unimaginable without such standard series' (Hopwood 2005, p. 239). The developmental trajectory from fertilized zygote to fully formed adult is broken down into distinct temporal periods by reference to the occurrence of major events such as fertilization, gastrulation or metamorphosis (Minelli 2003, ch. 4). The trajectory of development need not be understood as beginning with fertilization and ending with a morphologically or sexually mature adult (cf. Minelli 2009), but this represents a widely used conception in ontogenetic studies.

These developmental stages compose a 'periodization' that intentionally ignores variation associated with phenotypic plasticity. Animals and plants are raised under stable environmental conditions so that stages can be reproduced in different laboratory settings and variation is often viewed as 'noise' that must be reduced or eliminated if one is to understand how development works (Frankino & Raff 2004). This practice also encourages the selection of model organisms that exhibit less plasticity, such as nutritional polyphenisms (Bolker 1995). The laboratory domestication of a model organism also may reduce the amount or type of observable phenotypic variation (cf. Gu *et al.* 2005), but I ignore this type of effect here in part because laboratory domestication can increase variation (e.g. via inbreeding).

Despite attempts to reduce variation by controlling environmental factors, some of it always remains and is displayed by the fact that absolute chronology is not a reliable measure of time in ontogeny, and neither is the initiation or completion of its different parts (Mabee *et al.* 2000; Sheil & Greencbaum 2005). Developmental stages allow this recalcitrant variation to be effectively ignored by judgements of embryological typicality (e.g. Hamburger & Hamilton 1951). Normal stages also involve assumptions about the causal connections between different processes across sequences of stages (Alberch 1985; Minelli 2003, ch. 4). Once these stages have been constructed, it is possible to use them as a visual standard against which to recognize and describe variation as a deviation from the norm. But, more typically, variation ignored in the construction of these stages is also ignored or treated as noise in the routine consultation of the stages in day-to-day research contexts (Frankino & Raff 2004).

Let us observe a particular set of normal stages in more detail: the periodization of zebrafish embryogenesis (Kimmel *et al.* 1995). These idealized stages, which are keyed to morphological features of a live embryo that are easily identifiable by standard observation methods (naked eye and microscope), represent 'typical' development and their pictorial representation (either in a drawing or in a photograph) reinforces this typicality by only having one or two iconic representations. Some embryos arrive at the 64-cell stage before 2 h, some after and most relatively close to 2 h at the stable temperature of 28.5°C in filtered sea water at the concentration of approximately 5–10 embryos per millilitre. In some cases, this variation is noted: 'Frequently the 32 blastomeres of this stage are present in a 4×8 array, but other regular patterns, as well as irregular ones involving one or more of the blastomeres, also occur' (Kimmel *et al.* 1995, p. 263). The rationale for eliminating the variability is explicit:

> A staging series is a tool that provides accuracy in developmental studies. This is because different embryos, even together within a single clutch, develop at slightly different rates. We have seen asynchrony appearing in the development of zebrafish . . . embryos fertilized simultaneously *in vitro* . . . and incubated at an optimal temperature without crowding . . . Comparisons reveal more of this variability among embryos from different clutches than from within a single clutch. Genetic uniformity may alleviate but does not eliminate this problem; even embryos of a clonal strain . . . develop asynchronously.
>
> (Kimmel *et al.* 1995, p. 253)

Notice that asynchrony is most exaggerated in those conditions that undergird the comparison and contrast of experimental findings across different laboratory contexts. Sometimes, this asynchrony makes it difficult to follow cellular phenomena: 'The cleavage furrows that bring the 64-cell stage to an end generally occur so irregularly that with few exceptions one cannot after this time deduce a blastomere's cellular ancestry from its position' (Kimmel *et al.* 1995, p. 267). Although it is not described as such, asynchrony might be a form of phenotypic plasticity with respect to the duration of developmental events that could be delineated by systematically varying environmental conditions.

Normal stages fulfil a number of goals related to descriptive and explanatory endeavours that developmental biologists engage in. First, normal stages yield a way to measure experimental replication: 'a staging series provides a good way to ensure reproducibility' (Kimmel *et al.* 1995, p. 268). Second, they enable consistent and unambiguous communication among researchers, especially if stages are founded on commonly observable morphological features. Third, normal stages facilitate accurate predictions of developmental phenomena: 'Staging by somite number more accurately predicts where these neurons will be in their development than does staging by elapsed time after fertilization' (Kimmel *et al.* 1995, p. 253). Finally, they aid in making comparisons or generalizations across species: 'the descriptor '18-somite embryo' has more meaning

than '18-hour-old embryo',' particularly in cross-species comparisons' (Kimmel *et al.* 1995, p. 254).

In the midst of these goals, drawbacks also appear. Key morphological indicators sometimes overlap: 'In some embryos, the (yolk syncytial layer) forms over the course of two cycles, sometimes beginning a stage earlier and other times a stage later' (Kimmel *et al.* 1995, p. 267). Terminology that is useful for one purpose may be misleading for another. 'We define per cent-epiboly to mean the fraction of the yolk cell that the blastoderm covers; per cent-coverage would be a more precise term for what we mean to say, but per cent-epiboly immediately focuses on the process and is in common usage' (Kimmel *et al.* 1995, p. 268). Terms also can be misleading in cross-species comparisons: 'epiblast and hypoblast, are also used to describe layers of the avian embryonic blastoderm, but the layers so named seem to be altogether different in these two kinds of vertebrate embryos' (Kimmel *et al.* 1995, p. 268). Manipulation of the embryo for continued observation can have a causal impact on ontogeny: 'Late in the pharyngula period the embryo . . . swims away in response to touch; this can be prevented by anesthesia . . . Repeated anesthesia and rinsing appears to slightly but significantly retard subsequent development' (Kimmel *et al.* 1995, p. 254). Avoiding variability in stage indicators can encourage overlooking the significance of this variation or at least provide a reason to favour its minimization: 'time of hatching is not useful as a staging index for the zebrafish . . . because individuals within a single developing clutch hatch sporadically during the whole 3rd day of development (at standard temperature), and occasionally later' (Kimmel *et al.* 1995, p. 298).

Reviewing some of the details of zebrafish normal stages helps to drive home a point about the practice that we want to assess in relation to the phenomenon of phenotypic plasticity. There are good reasons for adopting normal stages to periodize model organism ontogeny, and these help explain why their continued use results in empirical success. At the same time, there are ways in which the practice of staging involves idealizations—decisions to ignore particular kinds of variation. These decisions yield distinct advantages but also engender biases that require scrutiny. This is especially important given that, similar to other standard (successful) practices in science, normal stages are often taken for granted, which means their biasing effects are neglected (Wimsatt 1980). Thus, the advantages of zebrafish normal stages can overshadow their drawbacks (e.g. systematically underestimating the extent of variation in a population), some of which are directly relevant to evolutionary questions.

3. OF SCIENTIFIC REASONING: ADVANTAGES AND A TENSION

(a) *Advantages of idealization*

Periodizations are central to many areas of biology (Griesemer 1996; Winther 2006), and thus 'temporal framework choices' are aspects of scientific reasoning that can be subject to reasoning explication (Love 2008*c*). Since developmental stages involve ignoring types of known variation, there is good reason to treat them as a form of idealization—representations of developmental phenomena based on concrete observational features and measurement techniques that intentionally set aside variations in specific parameters to depict a non-abstract typical case for various descriptive and explanatory purposes (Jones 2005; Weisberg 2007). Normal stage idealizations of ontogeny allow for a classification of developmental events that exhibit virtues, such as comprehensiveness of the periodization, suitably sized stages, reasonably homogeneous stages, reasonably sharp boundaries between stages and stability under different investigative conditions (Dupré 2001). The zebrafish normal stages accomplish these goals through a willingness to treat particular forms of variation as less relevant to explaining (because invariant features are more explanatory) or in need of explanation (because variation is akin to noise). Minimizing variation through typicality considerations produces 'reasonably' homogeneous periods and clearly defined boundaries, which encourages more precise explanations within particular disciplinary approaches (Griesemer 1996), especially those involving experimental manipulations. The success of a periodization is not a function of the eventual ability to relax its idealizations: 'it is apparent that the successful use of models does not involve refinements to a unique idealized representation of some phenomenon or group of properties, but rather a proliferation of structures, each of which is used for different purposes' (Morrison 2005, p. 169). As a general rule, periodizations are not slowly corrected so that they become less idealized; instead, new periodizations are constructed and used alongside the existing ones because different idealizations involve different judgements of typicality that serve diverse descriptive and explanatory aims.

Idealizations also can facilitate abstraction and generalization, both of which are a part of extrapolating findings from the specific investigative context of a model organism to other domains (Love 2008*b*,*c*, 2009; Steel 2008). Abstraction operates by omitting concrete particulars and is used for making comparisons and contrasts over different degrees and kinds of exclusion. Generalization refers to the range of application for scientific claims (wider scope usually being preferable) and is related to how we conceptualize laws (Mitchell 2000). Idealizations that disregard variability can yield generalizations that cover all ontogenies within a species: 'We observed groups of about a dozen embryos developing together to obtain the data, and here ignore variability among the embryos' (Kimmel *et al.* 1995, p. 270). Zebrafish researchers graphically depicted several generalizations of this type: (i) idealized blastomere number as a function of time after fertilization (p. 262), (ii) idealized rate of advance of the blastoderm margin over the yolk cell during epiboly (p. 270), and (iii) idealized rate of somitogenesis (p. 278). Idealizations that set aside values for variables, such as the mode of neural tube formation, also can facilitate abstractions, such as the vertebrate pharyngula. Sometimes, a generalization is achieved by abstraction because the exclusion of details facilitates extending the scope of application for a claim. Isolating generalizations of different

scopes about phenotypic plasticity are required to evaluate its evolutionary significance.

These advantages, in conjunction with those seen directly in the zebrafish normal stages (e.g. prediction, communication and replicability), are what encourage idealization as a reasoning strategy in scientific investigations. The past successes of normal stage idealizations in terms of precision, prediction or generalization with respect to specific explanatory goals also offer a guide to their future use. Advantages of idealization compose criteria of adequacy for the continued utilization of the same strategies. Desiderata of this nature appear unambiguously in the historical origins of normal stage periodizations: 'There was demand because the more elaborate experiments became, the more necessary it was to standardize stages of operation and of assay, within a single experiment, through an experimental series, and to establish a 'common language' between laboratories' (Hopwood 2005, p. 275).

Thus, the virtues and limits of idealizations can be assessed on the basis of how well they contribute to the explanatory goals of the research community in which they are routinely deployed (Weisberg 2007; Love 2009). Additionally, these advantages articulate with the benefits of working with model organisms: short generation times, rapid development, preferable life history properties (transparent embryos, developmental canalization and small adult size), ease of manipulation with molecular experimental tools (including having a genomic database) and other practical factors such as cost of maintenance (Bolker 1995). The characteristics of model organisms combine with the advantages of idealized normal stages to generate a picture of ontogeny that is abstract enough to compare with other model organisms and identify substantive generalizations about the causal mechanisms operating in developmental processes.

(b) *A tension between developmental staging and phenotypic plasticity*

Having highlighted the advantages of idealization in investigating the ontogeny of model organisms, it is incumbent upon us to look for ways in which this type of idealization may have weaknesses or blindspots. An easy place to begin observing these trade-offs is with respect to the desirable properties of model organisms. Short generation times and rapid development are tightly correlated with insensitivity to environmental conditions ('developmental canalization') through various mechanisms such as prepatterning, maternal loading or cell lineage determinism (Bolker 1995). Consequentially, most model organisms are poorly suited to informing us about how environmental effects modulate or combine with genetic or other 'internal' factors in development. The advantages that accrue from the experimental investigation of development with model organisms are accompanied by a key drawback: the relative difficulty to discover information about mechanisms underlying reaction norms (Bolker 1995). This drawback is not directly related to the practice of developmental staging, as is evidenced by the fact that some model organisms show phenotypic plasticity (e.g. Pigliucci 2002). But the positive reinforcement between criteria for model organism selection and the choices made in creating developmental stages is a signal that these practices can involve idealizations that mutually reinforce one another and thereby affect what variation is observable (Robert 2004).

We can assemble the materials from our discussion thus far and explicitly demonstrate a *tension* between the practice of developmental staging and the phenomenon of plasticity.

(1) Variation due to phenotypic plasticity is a normal feature of ontogeny.
(2) The developmental staging of model organisms intentionally downplays variation in ontogeny associated with the effects of environmental variables (e.g. phenotypic plasticity) by strictly limiting the range of values for environmental variables and by removing variation in characters used to establish the comprehensive periodization.
(3) Therefore, using model organisms with specified developmental stages will make it difficult, if not impossible, to observe patterns of variation due to phenotypic plasticity.

Notice that the tension in (3) is obtained even if the focus is not on evolutionary questions. If the primary goal is to understand ontogeny, then the tension remains and questions about the developmental significance of phenotypic plasticity are left open (Robert 2004; Love 2008*b*). Incubating embryos at different temperatures can retard or accelerate overall developmental rates that are relevant for constructing normal stages (Kimmel *et al.* 1995). Thus, even though zebrafish embryos do not appear to display malformations when temporarily incubated at temperatures within an 8°C range (25–33°C), and the overall rates are approximately linear (as is the case for the standard incubation temperature of 28.5°C), the significance of phenotypic plasticity with respect to fluctuating temperatures, sustained non-standard or more extreme temperatures, and the developmental rate or sequence for select processes (rather than overall rate or sequence) remains unknown or sparsely documented (Mabee *et al.* 2000).

Zebrafish investigators did recognize some of these issues explicitly: 'Comparisons of embryos raised at different temperatures should be made with caution, because there is no assurance that all features of the embryo coordinately change their rates of development when the temperature is changed' (Kimmel *et al.* 1995, p. 260). But even the experiments used to probe the effects of temperature were subject to their own idealizations, such as making a decision about when a particular stage was reached based on the majority of embryos in a given group having reached that stage (thereby eliminating temporal variation). This kind of recalcitrant variation in embryonic development and how it is (or is not) represented visually in idealized normal stages has been an issue of debate for more than 100 years (Hopwood 2005, 2007). The fact that phenotypic plasticity phenomena are sparsely documented could be a spur

to investigate it more thoroughly, but the opposite conclusion is often drawn: absence of evidence is taken to imply evidence of its developmental insignificance.

The tension is exacerbated for researchers interested in evolutionary questions because the documentation of patterns of variation is precisely what is required to gauge the evolutionary significance of phenotypic plasticity (Robert 2004; Love 2006*b*). If it is claimed that types of variation arising from phenotypic plasticity are important for understanding evolutionary processes, then practices of developmental staging in model organisms can retard our ability to make either a positive or negative assessment. The first step in West-Eberhard's (2003) model of novelty origination via phenotypic accommodation will not be subject to experiment unless we shift our attention to systematic deviations from the characteristics used to construct the normal stages. Only then can we observe an environmental change that can alter development such that a novel phenotype appears. In practice, developmental staging will encourage a negative assessment of the evolutionary significance of phenotypic plasticity because the variation will not be manifested and documented, and therefore is unlikely to be reckoned as substantive. Even if we decide to treat zebrafish as a model for understanding developmental mechanisms relevant to the origin of vertebrate jaws (Kimmel *et al.* 2001; cf. Metscher & Ahlberg 1999), the evolutionary significance of phenotypic plasticity is unlikely to play a role in our explanatory framework (Shigetani *et al.* 2005). Phenotypic variation related to jaw structures (mandibular arches or jaw cartilages) is not noted in the Pharyngula and Hatching periods of the normal stages—only whether these features make their appearance, which itself involves a typicality judgement (i.e. an idealization). Although there is good evidence of phenotypic plasticity in jaw muscles (Hoh 2002) and structure (Meyer 1987; Wimberger 1991), whether these forms of plasticity were present and played a role in the origin of jaw features is methodologically difficult to assess and largely ignored.

Idealizations involving normal stages discourage a robust experimental probing of phenotypic plasticity, which is a tension between the phenomenon of plasticity and the practice of developmental staging. There are epistemic costs to staging that must be set alongside of the characteristic strengths obtained through this form of idealization. The consequences of this tension for evo-devo investigation are twofold. First, the most powerful experimental systems for studying development are set up to minimize variation that may be critical to comprehending how evolutionary processes occur in nature. Second, if evo-devo investigations revolve around a character that was assessed for typicality to underwrite the temporal partitions that we call stages (e.g. epiboly in the case of zebrafish gastrulation), then much of the variation in this character was conceptually removed as a part of rendering the model organism experimentally tractable. It is worth observing this second situation via a specific example (see also Love 2009).

In a paper on the evolution of development in arthropods, Minelli and colleagues argued that the standard periodization for post-embryonic ontogeny in arthropods is sometimes a barrier to evolutionary analyses (Minelli *et al.* 2006). The conventional periodization is in terms of instars (molt-to-molt intervals) subsequent to hatching, which are then grouped into stages (larva, pupa and imago for insects). Although this periodization has a legitimate function in describing aspects of the ontogeny of particular arthropod taxa, and has been a key ingredient of successful reasoning about ontogeny in past investigations (e.g. copulatory structure origination during development), a molt–molt periodization is problematic when used for a different explanatory purpose—understanding molt-timing evolution and the origin of holometaboly. The lability of the characters used to define the intervals raises concerns about the biasing effects of idealization, especially in regard to how well one can generalize across different arthropod taxa. Drawing attention to this variation can be seen as a challenge to the purported advantages obtained via idealization, such as stability under different investigative conditions (Dupré 2001).

The tension between the phenomenon of phenotypic plasticity and the practice of developmental staging is now documented. When normal stages are constructed for model organism embryogenesis, idealizations that involve ignoring or minimizing variation are used. One of the ways this is accomplished is through strict control of environmental variables, the very same variables whose values might be involved in generating phenotypic plasticity. Much of the remaining variation is removed in order to generate a comprehensive periodization with reasonably homogeneous stages and operationally sharp boundaries between stages. On top of this, model organisms are selected because of features that correlate with embryological trajectories that are insensitive to environmental variables (e.g. short development time). Taken together, phenotypic plasticity becomes effectively minimized, if not wholly effaced, from the study of ontogeny. It is not surprising that plasticity seems relatively unimportant if it is insubstantially manifested. Therefore, any attempted evo-devo inquiry into the significance of phenotypic plasticity, such as for models of the origin of novelty, will be frustrated systematically. Claims about the importance of phenotypic plasticity will appear speculative because of the fact that variation due to environmental variables is rarely observed or relatively minimal in laboratory model organisms. Thus, the advantages of idealization obtained for studies of ontogeny are accompanied by a blindness to variation that might be relevant to how evolutionary processes occur in nature and, more specifically, hinder evo-devo investigations of characters whose variation was ignored in order to produce the typicality assessments that undergird a normal stage periodization.

4. COMPENSATORY TACTICS: EASING THE TENSION

The identification of drawbacks that accompany strategies of idealization used to study development and their impact on assessing the evolutionary significance

of phenotypic plasticity should not leave us paralysed. They invite us to consider ways to address the liabilities identified (Love 2006*b*). In this sense, reasoning explication can be relevant to ongoing scientific practice in terms of data gathering, model building and explanatory evaluation (Love 2008*c*). Recalling the previous argument to display the tension between the practice of staging and the phenomenon of plasticity (§3*a*), we can construct a principled perspective on how to address these liabilities by adding three further premises.

(4) Reasoning strategies involving idealization, such as (2), are necessary for the successful prosecution of biological investigations of ontogeny.
(5) Therefore, compensatory tactics should be chosen in such a way as to specifically redress the blindspots arising from the kind of idealizations used.
(6) Given (1)–(3), compensatory tactics must be related to the effects of ignoring variation due to phenotypic plasticity that result from the developmental staging of model organisms.

Point (6) offers general guidance in isolating relevant compensatory tactics. What practices will promote observations of variation due to phenotypic plasticity that is typically ignored when developmental stages are constructed for model organisms? There are at least two tactics corresponding to (6) that ease the tension between the practice of developmental staging and the phenomenon of phenotypic plasticity: the employment of diverse model organisms and the adoption of alternate periodizations.

(a) *Employing a diversity of model organisms*

One compensatory tactic that combats the idealization-related liabilities arising from the use of normal stage periodizations is to study phenotypic plasticity in non-standard models. Variation often will be observable in these models because they do not have comprehensive normal stages, and therefore do not exhibit the liabilities associated with the idealization strategy. In turn, researchers are sensitized to the ways in which these kinds of variation are being muted in the study of standard models. Stages can be used then as visual standards to identify variation as deviations from the norm and thereby characterize patterns of variability. Although experimental protocols use similar molecular methods to investigate the ontogeny of non-standard models, it is clear that unified normal stages are often absent (Crotty & Gann 2009). Organisms that do not have large communities built around them are less likely to have had their embryonic development formally staged, and thus the effects of idealization on phenotypic plasticity would not be operative. A good example is seen in evo-devo studies of dung beetle horn morphology, where some of the most persuasive evidence for the evolutionary significance of phenotypic plasticity has emerged (Emlen 2000; Moczek & Nagy 2005; Moczek 2008). It can be argued that the absence (or relative unimportance) of normal stages for all of ontogeny in this non-standard model facilitated these discoveries. A concentration on intra- and interspecific variation in specific morphological features (horns) of dung beetle congenerics means that normal stages for one particular species were not constructed using the eruption time or other characteristics of horn ontogeny. As a consequence, variability in these features was not minimized and remains salient in the context of experimental observations of development.

At this point, it becomes necessary to delimit the range of meanings associated with the label 'model organism'. One recent discussion highlights two distinct facets: 'A model organism must exemplify some key general biological problem that can be solved relatively easily with it and that will turn out to have the same answer for more important but experimentally less tractable organisms' (Slack 2009, pp. 1674–1675). Call the first component a 'problem focus' and the second a 'representation requirement'. The former is connected with the famous Krogh principle: 'For a large number of problems there will be some animal of choice or a few such animals on which it can be most conveniently studied' (Krogh 1929, p. 202). But Slack's problem focus includes qualifications; biological problems must be 'key' and 'general'. Krogh-principle model organisms do not always meet these qualifications, as some problems are more lineage specific or encourage a proliferation of models (Krebs 1975), such as in human disease and biomedical research. Discussions of the Krogh principle have emphasized difficulties with the representation requirement: 'generalisations from one species to another must necessarily be more restricted at higher and more complex levels' (Krebs & Krebs 1980, p. 380). The Krogh principle also operates when models are chosen for lineage-specific evolutionary questions, such as using lampreys to study the origin of vertebrate jaws (e.g. Kuratani *et al.* 2002). Many model organisms are introduced for these reasons (Collins *et al.* 2007; Jenner & Wills 2007; Milinkovitch & Tzika 2007; Crotty & Gann 2009), although care is required to avoid problematic assumptions about representation (Jenner 2006).

In contrast, the models that are subject to the practice of normal staging in the study of ontogeny are different. These general model organisms are not tailored to address one specific problem but rather a very broad range of key questions that are presumed to be less lineage specific. The rationale for their use is general representativeness, which is supposed to justify the extrapolation of findings (Bolker 1995), and the availability of powerful experimental tools that can address functional issues (Sommer 2009). By definition, the set of general model organisms always will be relatively small.

We can mark this difference in meaning by distinguishing 'Krogh-principle' model organisms from general model organisms. This leads us to a more precise formulation of the compensatory tactic. Employing diverse model organisms is a relevant compensatory tactic for the effects of ignoring variation due to phenotypic plasticity resulting from developmental staging because: (i) variation due to phenotypic plasticity will be more readily observed in Krogh-principle model organisms because they do not have comprehensive normal stages and therefore do not exhibit the variation suppression liabilities

associated with the idealization strategy and (ii) the observation of this kind of variation highlights its absence in the study of general model organisms and prompts a different use of stages as standards to identify patterns of variation that deviate from the idealized norm. Thus, there is good reason to continuously employ diverse model organisms as a compensatory tactic in this case and not settle on circumscribed, consensus list (Milinkovitch & Tzika 2007; Sommer 2009).

(b) *Using alternative periodizations*

A second compensatory tactic is the adoption of alternative periodizations (cf. Minelli 2003, pp. 57ff). This involves choosing different characters to construct new temporal partitions, thereby facilitating the observation of variation with respect to characteristics previously stabilized in the normal stage periodization. These alternative periodizations often divide a subset of developmental events according to processes or landmarks that differ from the principles used to construct the normal stages. As a consequence, periodizations may not map one–one onto the existing normal stages, especially if they encompass events beyond a fertilization to sexually mature adult trajectory (Minelli 2009). This lack of isomorphism between periodizations also will be manifested if different measures of time are used, whether sequence (event ordering) or duration (succession of defined intervals), and whether sequences or durations are measured relative to one another or against an external standard, such as absolute chronology (Reiss 2003; cf. Colbert & Rowe 2008). These incompatibilities prevent the alternative periodizations from being assimilated into a single, overarching staging scheme. In all of these cases, idealization is involved and therefore each new periodization is subject to the liabilities of ignoring kinds of variation. But the new periodizations require choosing different characters to stabilize and typify when defining their temporal partitions, which means different kinds of variation will be exposed than were previously observable. Once these new kinds of variation are uncovered, their evolutionary significance can be genuinely assessed. The price of detecting the overlooked variation is ignoring variation in *other* features. Alternative periodizations are useful not because they are more comprehensive but because they involve different idealizations that overlap and crosscut the temporal partitions in the standard normal staging of embryogenesis.

Evo-devo investigations already exhibit these approaches in different guises. For example, in order to explore wing developmental evolution more precisely, Reed *et al.* (2007) constructed a staging for final-instar wing disc ontogeny in the common buckeye butterfly. This was pursued because the timing of larval events is relatively dissociated from wing disc development, and thus temporal measures of larval ontogeny do not correlate tightly with the developmental state of the wing discs. Another reason was that the researchers were interested in the phenomenon of phenotypic plasticity: 'Some cryptic variation, however, might manifest developmentally or physiologically, but simply not have an effect on phenotypes that is obvious or accessible to investigators' (Reed *et al.* 2007, p. 2). Consonant with the requirements for assessing the evolutionary significance of phenotypic plasticity (§2*b*), the first step is actually measuring it: 'few direct measurements of intraspecific developmental variation have been made' (Reed *et al.* 2007, p. 2). Substantial heterochronic variability has been uncovered in this and other cases (Mabee *et al.* 2000; Sheil & Greenbaum 2005; Colbert & Rowe 2008).

In the case of molt–molt intervals (§3*b*), the researchers also advocate alternative periodizations: 'Paying due attention to the time course of different aspects of development without roughly conflating all of them into the usual periodization of stages and instars punctuated by molts offers a good chance for asking interesting questions on the evolution of arthropod development' (Minelli *et al.* 2006, p. 381). The motivation for this is not wholly due to temporal variation, as alternative spatial decompositions can be pursued for features whose variation has been idealized, such as 'segments' (Minelli & Fusco 2004; cf. Wimsatt 2007, ch. 9). Often, alternative spatial and temporal decompositions are jointly pursued, such as in the case of treating pluteus larval skeletogenesis as distinct from larval arm morphogenesis in indirect developing echinoids (Love *et al.* 2007; Love 2008*c*).

Each new set of temporal partitions can be used as a methodological tool both within and across disciplinary approaches for dissecting complex biological phenomena (Griesemer 1996). Because these periodizations serve a methodological role, and therefore proliferate as needed rather than become unified into a single representation (Morrison 2005), this leaves new issues to be tackled, such as how we relate or coordinate different periodizations within and across disciplinary approaches. A new staging system of egg embryogenesis in parasitic trematodes (Jurberg *et al.* 2009) was motivated by the staging of a free-living flatworm (Hartenstein & Ehlers 2000), and then correlated with preexisting stages for egg embryogenesis (some of which were put forward more than 50 years ago). A staging of the mouse limb was spurred by the need to standardize comparisons of *in situ* hybridization results and then was compared with chick limb staging and the normal stages for mouse embryogenesis (Wanek *et al.* 1989). Another more comprehensive staging system has been put forward as a tool for making comparative assessments of vertebrate ontogeny (Werneburg 2009). The 'plastochron index' is a mathematical formula that was created as a way to stage shoot development in plants to mitigate variability with respect to chronological age and partition time using characters that can be easily scored (Erickson & Michelini 1957). It is also a reminder of how alternative periodizations with new idealizations also carry liabilities in terms of ignoring variation due to phenotypic plasticity: 'this technique served to circumvent the environmental variability of the material to a considerable extent' (p. 297). The plastochron index can be coordinated with generalized growth stages for plants (Lancashire *et al.* 1991) or specific developmental stages for *Arabidopsis* flowering (Smyth *et al.* 1990).

Adopting alternative periodizations is a relevant compensatory tactic for the effects of ignoring variation due to phenotypic plasticity that result from the developmental staging of general model organisms because it involves new idealizations that are different from those used in normal stages. This facilitates the discovery of previously unseen variation due to phenotypic plasticity (Reed *et al.* 2007). The need to coordinate alternative periodizations also provides an impetus to shift staging perspectives because the new sets of temporal partitions coexist with older ones, each retaining methodological autonomy. Idealizations are not eliminated but rather played off of one another. The intersection of these alternative temporal periodizations is one route to a more empirically adequate account of biological phenomena (Wimsatt 1987).

(c) *Objections: costs and controversies*

Some natural objections and worries emerge in response to our discussion thus far. For example, how can developmental stages be a barrier to observing phenotypic plasticity and assessing its significance when researchers seem to have done precisely that (and repeatedly)? Here it is necessary to emphasize the difference between establishing the existence of phenotypic plasticity and understanding the mechanisms underlying its production. It is the latter that has been in view (where plasticity emerges, how molecular genetic mechanisms are involved in the plasticity and what genotype–phenotype relations obtain) because of its pertinence to evo-devo investigation. This means we should distinguish between types of significance; understanding the significance of phenotypic plasticity for responding to selection is different from understanding its significance for how a genotype maps to a phenotype.

In relation to this first question we cannot forget the logistical difficulties of studying these developmental aspects of phenotypic plasticity—setting up different environments to observe the variation, having separate environments with sufficient numbers of organisms for statistical analysis and dealing with the confounding effect of non-additive interactions, which multiplies the number of experimental replicates required (Scheiner 1993; Pigliucci 2001; Kaplan 2008). These logistical hurdles have an impact on how much we can understand about different kinds of significance. Thus, even if in principle we can answer these questions of significance, it does not mean in practice that the achievement will be accomplished easily.

Another objection involves questions about whether the compensatory tactics are really necessary, if not downright wasteful. Do we really need to document variation resulting from phenotypic plasticity to comprehend the evolutionary significance of development, especially if it requires new model organisms that are closely related to existing ones (Slack 2009)? If we employ a diversity of model organisms, do we decrease our ability to investigate and experimentally manipulate ontogeny with functional tools (Sommer 2009)? Can we afford a diversity of model organisms, especially when they are difficult to rear in the laboratory or must be harvested in exotic locales (Slack 2009)? Should not we plumb the depths of general model organisms with our increasingly sophisticated post-genomic apparatus of molecular methods?

Working from general models and a preferred periodization facilitates the transfer of empirical results to subsequent generations of researchers and students. It is a form of good stewardship as well because enormous resources were expended to produce normal stages that can serve a large scientific community. Because normal stages for embryogenesis are widely used and empirically successful, there may be concerns about disrupting the advantages gained from their utilization. In fact, these standard periodizations arose a century ago out of a frustration experienced by researchers with too many (seemingly) arbitrary periodizations.

> Many set up stages on the basis of whatever single organ they happened to be studying—organs that did not necessary track the main developmental events—and aggregated features observed in different embryos into stages. But variations in the development of organs within species made arbitrary staging unreliable and the addition of new material difficult. Only from systematic descriptions of the development of many organs within a series of individual embryos . . . could effective comparisons be made.
>
> (Hopwood 2005, p. 253)

New attempts at systematic staging of vertebrate embryos are still motivated by a similar rationale (Werneburg 2009).

These intertwined worries and concerns are genuine; there is no clear-cut answer that will satisfy all parties. The compensatory tactics of employing a diversity of model organisms and adopting alternative periodizations may be conceptually appropriate for addressing how the practice of developmental staging has an impact on the detection of phenotypic plasticity, but this does not magically remove associated costs or controversy. The advantages of a single, comprehensive periodization for a general model organism (e.g. zebrafish normal stages) must be weighed in light of the advantages of alternative, process-specific periodizations or the use of Krogh-principle model organisms. Establishing new temporal partitions for a general model organism's ontogeny may help mitigate the consequences of idealization, but it also will be a burden on human and material resources. The cost of engaging in these practices will encourage claims that diverse model organisms and alternative periodizations are unnecessary given the existence of general model organisms with their embryogenesis already staged. This is one reason why we must assess advantages and liabilities with respect to a specific phenomenon (e.g. phenotypic plasticity) and for a specific idealization practice (e.g. developmental staging). Change the phenomenon or practice and new trade-offs will be manifested, and new questions will arise about adequate resources and funding. The achievement is that we are now openly scrutinizing these practices in relation to the phenomenon of interest, recognizing both advantages and drawbacks involved in the idealizations used, thereby putting the

research community in a stronger position with respect to reasoning towards systematic descriptions and comprehensive explanations of the biological phenomena in view.

5. CONCLUDING REMARKS

Although there might be some initial disappointment that ontogeny's temporal events do not unfold according to standard, invariant units of time, there are clear advantages for developmental biologists when using normal stages to describe them. The differences between our epistemological resources and the biological phenomena we desire to describe, model and explain is both a cause of empirical success and a source of liabilities with respect to detecting and evaluating other phenomena. It should not surprise us that compensatory tactics come alloyed with costs and controversies; trade-offs in reasoning always co-travel with financial and methodological trade-offs. The proliferation of idealizations, whether in terms of diverse model organisms or alternative periodizations, never will be met with universal acclaim. But the goal of offering a philosophical reconstruction and evaluation of scientific reasoning (reasoning explication; see Love 2008*c*) in the context of these idealizations and associated compensatory tactics is now clear. Because the developmental staging of general model organisms downplays variation in ontogeny resulting from phenotypic plasticity, compensatory tactics are required to observe this variation. Employing diverse Krogh-principle model organisms and adopting alternative periodizations are appropriate compensatory tactics because they specifically redress the blindspots related to how the idealization strategy of developmental staging in general model organisms diminishes variation due to phenotypic plasticity.

Other compensatory tactics that address the liabilities arising from the routine use of normal stages to study ontogeny might be pursued. In addition to systematically manipulating environmental variables and comparing patterns of variation with the expectation established by normal stages, another tactic would be to relax the strict control of environmental variables and isolate variation that consequently emerges. This might yield more information about what is *not* variable, either in process or outcome, during ontogeny (Richter *et al.* 2009). Adopting this tactic would involve choosing how to relax control on different environmental factors and tracking the new trade-offs that result from following this experimental path.

The particular tension isolated and discussed here is not the only issue one might worry about in the context of evo-devo investigation. Phenotypic plasticity might be masked in part through developmental canalization (Frankino & Raff 2004; Kaplan 2008). This is an interpretive tension between two different types of phenomena and includes deciding whether gene expression noise should be viewed as something to be buffered (canalization) or variation that can be actively used (plasticity) (Rao *et al.* 2002). Similarly, two different kinds of practices used in the study of ontogeny might yield different conflicts, such as embryo fixation and *in situ* hybridization. To recognize these possibilities is to learn from reasoning explication—to scientifically benefit from the philosophical reconstruction and evaluation of reasoning strategies—because the awareness of these possibilities comes from a portrayal of their methodological prospects and pitfalls, as well as available compensatory tactics.

Although compensatory tactics always will be attended by cost and controversy, reasoning explication leaves ongoing research in biology with additional methodological and epistemological resources. The increasing documentation of unexpected and dramatic intraspecific variation in ontogeny (e.g. Mabee *et al.* 2000; Sheil & Greenbaum 2005), and new methods for analysing and interpreting it (Colbert & Rowe 2008), suggests that the regular utilization of these compensatory tactics (and others) will generate a variety of new empirical results, which in turn will deliver more robust answers to questions about the evolutionary significance of phenotypic plasticity. This is a reminder that the liabilities associated with specific strategies of scientific reasoning such as idealization are not incompatible with, and sometimes conducive to, the methodological and epistemological advantages derived from their use.

Ingo Brigandt, Chris DiTeresi, Giuseppe Fusco, James Lennox, Alessandro Minelli, Ric Otte and Massimo Pigliucci provided constructive feedback and criticism on an earlier version of this manuscript. Their timely help was invaluable and gave me numerous resources to improve the argument, regardless of whether they agreed and despite the fact that I could not accommodate all of their suggestions. I also would like to thank Giuseppe Fusco and Alessandro Minelli for inviting me to make a philosophical contribution to this special issue and their editorial initiative throughout the process.

REFERENCES

Alberch, P. 1985 Problems with the interpretation of developmental sequences. *Syst. Zool.* **34**, 46–58. (doi:10.2307/2413344)

Bechtel, W. & Richardson, R. 1993 *Discovering complexity: decomposition and localization as strategies in scientific research*. Princeton, NJ: Princeton University Press.

Bolker, J. A. 1995 Model systems in developmental biology. *BioEssays* **17**, 451–455. (doi:10.1002/bies.950170513)

Colbert, M. W. & Rowe, T. 2008 Ontogenetic sequence analysis: using parsimony to characterize developmental sequences and sequence polymorphism. *J. Exp. Zool. (Mol. Dev. Evol.)* **310B**, 398–416.

Collins, J. P., Gilbert, S. F., Laubichler, M. D. & Müller, G. B. 2007 Modeling in EvoDevo: how to integrate development, evolution, and ecology. In *Modeling biology* (eds M. D. Laubichler & G. B. Müller), pp. 355–378. Cambridge, MA: MIT Press.

Crotty, D. A. & Gann, A. (eds) 2009 *Emerging model organisms: a laboratory manual*, vol. 1. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Cudilo, E., Al Naemi, H., Marmorstein, L. & Baldwin, A. 2007 Knockout mice: is it just genetics? Effect of enriched housing on fibulin-4(+/−) mice. *PLoS ONE* **2**, e229. (doi:10.1371/journal.pone.0000229)

DeWitt, T. J. & Scheiner, S. M. (eds) 2004 *Phenotypic plasticity: functional and conceptual approaches*. New York, NY: Oxford University Press.

Dupré, J. 2001 In defence of classification. *Stud. Hist. Phil. Biol. Biomed. Sci.* **32**, 203–219.

Ellis, R. J. 2001 Macromolecular crowding: obvious but underappreciated. *Trends Biochem. Sci.* **26**, 597–604. (doi:10.1016/S0968-0004(01)01938-7)

Emlen, D. J. 2000 Integrating development with evolution: a case study with beetle horns. *BioScience* **50**, 403–418. (doi:10.1641/0006-3568(2000)050[0403:IDWEAC]2.0.CO;2)

Erickson, R. O. & Michelini, F. J. 1957 The plastochron index. *Am. J. Bot.* **44**, 297–305. (doi:10.2307/2438380)

Frankino, W. A. & Raff, R. A. 2004 Evolutionary importance and pattern of phenotypic plasticity. In *Phenotypic plasticity: functional and conceptual approaches* (eds T. J. DeWitt & S. M. Scheiner), pp. 64–81. New York, NY: Oxford University Press.

Gerhart, J. & Kirschner, M. 1997 *Cells, embryos, and evolution: towards a cellular and developmental understanding of phenotypic variation and evolutionary adaptability.* Malden, MA: Blackwell Science, Inc.

Gilbert, S. F. & Epel, D. 2009 *Ecological developmental biology: integrating epigenetics, medicine, and evolution*. Sunderland, MA: Sinauer.

Greene, E. 1989 A diet-induced developmental polymorphism in a caterpillar. *Science* **243**, 643–646. (doi:10.1126/science.243.4891.643)

Griesemer, J. R. 1996 Periodization and models in historical biology. In *New perspectives on the history of life* (eds M. T. Ghiselin & G. Pinna), pp. 19–30. San Francisco, CA: California Academy of Sciences.

Gu, Z., David, L., Petrov, D., Jones, T., Davis, R. W. & Steinmetz, L. M. 2005 Elevated evolutionary rates in the laboratory strain of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **102**, 1092–1097. (doi:10.1073/pnas.0409159102)

Hamburger, V. & Hamilton, H. L. 1951 A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92. (doi:10.1002/jmor.1050880104)

Hartenstein, V. & Ehlers, U. 2000 The embryonic development of the rhabdocoel flatworm *Mesostoma lingua* (Abildgaard, 1789). *Dev. Genes Evol.* **210**, 399–415. (doi:10.1007/s004270000085)

Hoh, J. F. Y. 2002 'Superfast' or masticatory myosin and the evolution of jaw-closing muscles of vertebrates. *J. Exp. Biol.* **205**, 2203–2210.

Hopwood, N. 2005 Visual standards and disciplinary change: normal plates, tables and stages in embryology. *Hist. Sci.* **43**, 239–303.

Hopwood, N. 2007 A history of normal plates, tables and stages in vertebrate embryology. *Int. J. Dev. Biol.* **51**, 1–26. (doi:10.1387/ijdb.062189nh)

Jenner, R. A. 2006 Unburdening evo-devo: ancestral attractions, model organisms, and basal baloney. *Dev. Genes Evol.* **216**, 385–394. (doi:10.1007/s00427-006-0084-5)

Jenner, R. A. & Wills, M. A. 2007 The choice of model organisms in evo-devo. *Nat. Rev. Genet.* **8**, 311–319. (doi:10.1038/nrg2062)

Jones, M. R. 2005 Idealization and abstraction: a framework. In *Idealization XII: correcting the model. Idealization and abstraction in the sciences* (*Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 86*) (eds M. R. Jones & N. Cartwright), pp. 173–217. Amsterdam, The Netherlands: Rodopi.

Jurberg, A. *et al.* 2009 The embryonic development of *Schistosoma mansoni* eggs: proposal for a new staging system. *Dev. Genes Evol.* **219**, 219–234. (doi:10.1007/s00427-009-0285-9)

Kaplan, J. M. 2008 Phenotypic plasticity and reaction norms. In *A companion to the philosophy of biology* (eds S. Sarkar & A. Plutynski), pp. 205–222. Malden, MA: Blackwell Publishers.

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B. & Schilling, T. F. 1995 Stages of embryonic development of the zebrafish. *Dev. Dyn.* **203**, 253–310.

Kimmel, C. B., Miller, C. T. & Keynes, R. J. 2001 Neural crest patterning and the evolution of the jaw. *J. Anat.* **199**, 105–119. (doi:10.1046/j.1469-7580.2001.19910105.x)

Kirschner, M. W. & Gerhart, J. C. 2005 *The plausibility of life: resolving Darwin's dilemma*. New Haven, CT: Yale University Press.

Koentges, G. 2008 Evolution of anatomy and gene control. *Nature* **451**, 658–663. (doi:10.1038/451658a)

Krebs, H. A. 1975 The August Krogh principle: 'for many problems there is an animal on which it can be most conveniently studied'. *J. Exp. Zool.* **194**, 221–226. (doi:10.1002/jez.1401940115)

Krebs, H. A. & Krebs, J. R. 1980 The 'August Krogh principle'. *Comp. Biochem. Phys.* **67B**, 379–380.

Krogh, A. 1929 The progress of physiology. *Science* **70**, 200–204. (doi:10.1126/science.70.1809.200)

Kuratani, S., Kuraku, S. & Murakami, Y. 2002 Lamprey as an evo-devo model: lessons from comparative embryology and molecular phylogenetics. *Genesis* **34**, 175–183. (doi:10.1002/gene.10142)

Lancashire, P. D., Bleiholder, H., Boom, T. V. D., Langeluddeke, P., Stauss, R., Weber, E. & Witzenberger, A. 1991 A uniform decimal code for growth stages of crops and weeds. *Ann. Appl. Biol.* **119**, 561–601. (doi:10.1111/j.1744-7348.1991.tb04895.x)

Love, A. C. 2006*a* Evolutionary morphology and evo-devo: hierarchy and novelty. *Theory Biosci.* **124**, 317–333. (doi:10.1016/j.thbio.2005.11.006)

Love, A. C. 2006*b* Taking development seriously: who, what, where, why, when, how? *Biol. Phil.* **21**, 575–589. (doi:10.1007/s10539-005-7587-8)

Love, A. C. 2008*a* Explaining evolutionary innovation and novelty: criteria of explanatory adequacy and epistemological prerequisites. *Phil. Sci.* **75**, 874–886. (doi:10.1086/594531)

Love, A. C. 2008*b* Explaining the ontogeny of form: philosophical issues. In *The Blackwell companion to philosophy of biology* (eds A. Plutynski & S. Sarkar), pp. 223–247. Malden, MA: Blackwell Publishers.

Love, A. C. 2008*c* From philosophy to science (to natural philosophy): evolutionary developmental perspectives. *Q. Rev. Biol.* **83**, 65–76. (doi:10.1086/529564)

Love, A. C. 2009 Typology reconfigured: from the metaphysics of essentialism to the epistemology of representation. *Acta Biotheor.* **57**, 51–75. (doi:10.1007/s10441-008-9059-4)

Love, A. C. In press. Rethinking the structure of evolutionary theory for an extended synthesis. In *Toward an extended evolutionary synthesis* (eds M. Pigliucci & G. B. Müller). Cambridge, MA: MIT Press.

Love, A. C., Andrews, M. E. & Raff, R. A. 2007 Gene expression patterns in a novel animal appendage: the sea urchin pluteus arm. *Evol. Dev.* **9**, 51–68.

Mabee, P. M., Olmstead, K. L. & Cubbage, C. C. 2000 An experimental study of intraspecific variation, developmental timing, and heterochrony in fishes. *Evolution* **54**, 2091–2106. (doi:10.1111/j.0014-3820.2000.tb01252.x)

Metscher, B. D. & Ahlberg, P. E. 1999 Zebrafish in context: uses of a laboratory model in comparative studies. *Dev. Biol.* **210**, 1–14. (doi:10.1006/dbio.1999.9230)

Meyer, A. 1987 Phenotypic plasticity and heterochrony in *Cichlasoma managuense* (Pisces, Cichlidae) and their implications for speciation in cichlid fishes. *Evolution* **41**, 1357–1369. (doi:10.2307/2409100)

Milinkovitch, M. & Tzika, A. 2007 Escaping the mouse trap: the selection of new evo-devo species. *J. Exp. Zool.* (*Mol. Dev. Evol.*) **308B**, 337–346.

Minelli, A. 2003 *The development of animal form: ontogeny, morphology, and evolution*. Cambridge, UK: Cambridge University Press.

Minelli, A. 2009 *Perspectives in animal phylogeny and evolution*. New York, NY: Oxford University Press.

Minelli, A. & Fusco, G. 2004 Evo-devo perspectives on segmentation: model organisms, and beyond. *Trends Ecol. Evol.* **19**, 423–429. (doi:10.1016/j.tree.2004.06.007)

Minelli, A., Brena, C., Deflorian, G., Maruzzo, D. & Fusco, G. 2006 From embryo to adult–beyond the conventional periodization of arthropod development. *Dev. Genes Evol.* **216**, 373–383. (doi:10.1007/s00427-006-0075-6)

Mitchell, S. D. 2000 Dimensions of scientific law. *Phil. Sci.* **67**, 242–265. (doi:10.1086/392774)

Moczek, A. 2008 On the origins of novelty in development and evolution. *BioEssays* **30**, 432–447. (doi:10.1002/bies.20754)

Moczek, A. P. & Nagy, L. M. 2005 Diverse developmental mechanisms contribute to different levels of diversity in horned beetles. *Evol. Dev.* **7**, 175–185. (doi:10.1111/j.1525-142X.2005.05020.x)

Morrison, M. 2005 Approximating the real: the role of idealizations in physical theory. In *Idealization XII: correcting the model. Idealization and abstraction in the sciences (Poznań Studies in the Philosophy of the Sciences and the Humanities, vol. 86)* (eds M. R. Jones & N. Cartwright), pp. 145–172. Amsterdam, The Netherlands: Rodopi.

Nieuwkoop, P. D. & Faber, J. (eds) 1956 *Normal table of Xenopus laevis (Daudin): a systematical and chronological survey of the development from fertilized egg till the end of metamorphosis*. Amsterdam, The Netherlands: North Holland.

Pigliucci, M. 2001 *Phenotypic plasticity: beyond nature and nurture*. Baltimore, MD: The Johns Hopkins University Press.

Pigliucci, M. 2002 Touchy and bushy: phenotypic plasticity and integration in response to wind stimulation in *Arabidopsis thaliana*. *Int. J. Plant Sci.* **163**, 399–408. (doi:10.1086/339158)

Rao, C. V., Wolf, D. M. & Arkin, A. P. 2002 Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237. (doi:10.1038/nature01258)

Reed, R. D., Chen, P.-H. & Nijhout, H. F. 2007 Cryptic variation in butterfly eyespot development: the importance of sample size in gene expression studies. *Evol. Dev.* **9**, 2–9.

Reiss, J. O. 2003 Time. In *Keywords and concepts in evolutionary developmental biology* (eds B. K. Hall & W. M. Olson), pp. 359–368. Cambridge, MA: Harvard University Press.

Richter, S., Garner, J. & Würbel, H. 2009 Environmental standardization: cure or cause of poor reproducibility in animal experiments? *Nat. Methods* **6**, 257–261. (doi:10.1038/nmeth.1312)

Robert, J. S. 2004 *Embryology, epigenesis, and evolution: taking development seriously*, New York, NY: Cambridge University Press.

Sarkar, S. 2004 From the *Reaktionsnorm* to the evolution of adaptive plasticity: a historical sketch, 1909–1999. In *Phenotypic plasticity: functional and conceptual approaches* (eds T. J. DeWitt & S. M. Scheiner), pp. 10–30. New York, NY: Oxford University Press.

Scheiner, S. M. 1993 Genetics and evolution of phenotypic plasticity. *Ann. Rev. Ecol. Syst.* **24**, 35–68. (doi:10.1146/annurev.es.24.110193.000343)

Sheil, C. A. & Greenbaum, E. 2005 Reconsideration of skeletal development of *Chelydra serpentina* (Reptilia: Testudinata: Chelydridae): evidence for intraspecific variation. *J. Zool.* **265**, 235–267. (doi:10.1017/S0952836904006296)

Shigetani, Y., Sugahara, F. & Kuratani, S. 2005 A new evolutionary scenario for the vertebrate jaw. *BioEssays* **27**, 331–338. (doi:10.1002/bies.20182)

Slack, J. M. W. 2009 Emerging market organisms. *Science* **323**, 1674–1675. (doi:10.1126/science.1171948)

Smyth, D. R., Bowman, J. L. & Meyerowitz, E. M. 1990 Early flower development in *Arabidopsis*. *Plant Cell* **2**, 755–767. (doi:10.1105/tpc.2.8.755)

Sommer, R. J. 2009 The future of evo-devo: model systems and evolutionary theory. *Nat. Rev. Genet.* **10**, 416–422. (doi:10.1038/nrg2567)

Steel, D. P. 2008 *Across the boundaries: extrapolation in biology and social science*. New York, NY: Oxford University Press.

Suzuki, Y. & Nijhout, H. F. 2006 Evolution of a polyphenism by genetic accommodation. *Science* **311**, 650–652. (doi:10.1126/science.1118888)

Via, S., Gomulkiewicz, R., De Jong, G., Scheiner, S. M., Schlichting, C. D. & Van Tienderen, P. H. 1995 Adaptive phenotypic plasticity: consensus and controversy. *Trends Ecol. Evol.* **10**, 212–217. (doi:10.1016/S0169-5347(00)89061-8)

Wanek, N., Muneoka, K., Holler-dinsmore, G., Burton, R. & Bryant, S. V. 1989 A staging system for mouse limb development. *J. Exp. Zool.* **249**, 41–49. (doi:10.1002/jez.1402490109)

Weisberg, M. 2007 Three kinds of idealization. *J. Phil.* **104**, 639–659.

Werneburg, I. 2009 A standard system to study vertebrate embryos. *PLoS ONE* **4**, e5887. (doi:10.1371/journal.pone.0005887)

West-Eberhard, M. J. 2003 *Developmental plasticity and evolution*. New York, NY: Oxford University Press.

Wimberger, P. H. 1991 Plasticity of jaw and skull morphology in the neotropical cichlids *Geophagus brasiliensis* and *G. steindachneri*. *Evolution* **45**, 1545–1563. (doi:10.2307/2409778)

Wimsatt, W. C. 1980 Reductionistic research strategies and their biases in the units of selection controversy. In *Scientific discovery: case studies* (ed. T. Nickles), pp. 213–259. Boston, MA: D. Reidel Publishing Company.

Wimsatt, W. C. 1987 False models as means to truer theories. In *Neutral models in biology* (eds M. H. Nitecki & A. Hoffman), pp. 23–55. New York, NY: Oxford University Press.

Wimsatt, W. C. 2007 *Re-engineering philosophy for limited beings: piecewise approximations to reality*. Cambridge, MA: Harvard University Press.

Winther, R. G. 2006 Parts and theories in compositional biology. *Biol. Phil.* **21**, 471–499. (doi:10.1007/s10539-005-9002-x)