bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

# Seeing the future: a better way to model and test for adaptive developmental plasticity

Anup Malani, Stacy Rosenbaum, Susan C. Alberts, & Elizabeth A. Archie*

February 11, 2022

**Abstract**

Early life conditions can have profound effects on individual health, longevity, and biological fitness. Two classes of hypotheses are used to explain the evolutionary origins of these effects: developmental constraints (DC) hypotheses, which focus on the deleterious effects of low-quality early-life environments, and predictive adaptive response (PAR) models, which focus on organisms' predictions about their adult environment, phenotypic adaptations based on that prediction, and the deleterious consequences of incorrect predictions. Despite their popularity, these ideas remain poorly defined. To remedy this, we provide mathematical definitions for DC, PARs, and related concepts, and develop statistical tests derived from these definitions. We use simulations to demonstrate that PARs are more readily detected by tests based on quadratic regressions than by tests based on more commonly used interaction regression models. Specifically, quadratic regression-based tests on simulated data yield 90.7% sensitivity and 71.5% specificity in detecting PARs, while interaction-based tests yield sensitivity and specificity roughly equal to chance. We demonstrate that the poor performance of interaction models stems from two problems: first, they are mathematically incapable of detecting a central prediction of PAR, and second, they conceptually conflate PARs with DC. Our results emphasize the value of formal statistical modeling to reveal the theoretical underpinnings of verbal and visual models, and their importance for helping resolve conflicting and ambiguous results in this field of research. We conclude by providing recommendations for how researchers can make use of explicit definitions and properly-aligned visualizations and statistical tests to make progress in this important research area.

*Malani: University of Chicago Law School and National Bureau of Economic Research; Rosenbaum: University of Michigan Department of Anthropology; Alberts: Duke University Departments of Biology and Evolutionary Anthropology; Archie: University of Notre Dame Department of Biological Sciences. Malani and Rosenbaum contributed equally to this work. Correspondence: rosenbas@umich.edu.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

# Introduction

Early life adversity is associated with a wide variety of negative outcomes in adulthood, including poor health, short lifespans, and low biological fitness. This phenomenon has generated interest from many academic fields, including public health [58, 41], medicine [15, 45], psychology and psychiatry [12, 50, 29], biology [9, 53], sociology [28, 18], economics [10, 49], and ecology and evolution [23, 40]. The widespread observation of early life effects, which have been documented in species ranging from plants to insects to humans [16, 3], has led to two evolutionary theories that connect early life experiences to later life outcomes (e.g. adult health, disease risk, and longevity, to name but a few; reviewed in [31, 34, 35]. The first theory, the developmental constraints hypothesis (also known as the silver spoon hypothesis [38, 43]), posits that harsh conditions in early life lead to adaptive trade-offs during development. These trade-offs promote survival during development, but come at the expense of later-life phenotypic quality, health, and disease risk [60, 35]. The second, the predictive adaptive response hypothesis, proposes that organisms use their early life conditions to predict the quality of their adult environment and adjust their developmental trade-offs accordingly. [24, 43, 46]. Under this model, incorrect "guesses" about the future lead to poor health outcomes during adulthood. Organisms may predict their external environment (commonly referred to as an external PAR (ePAR), or the quality of their own somatic state (an internal PAR (iPAR) [48, 46, 6, 22]).

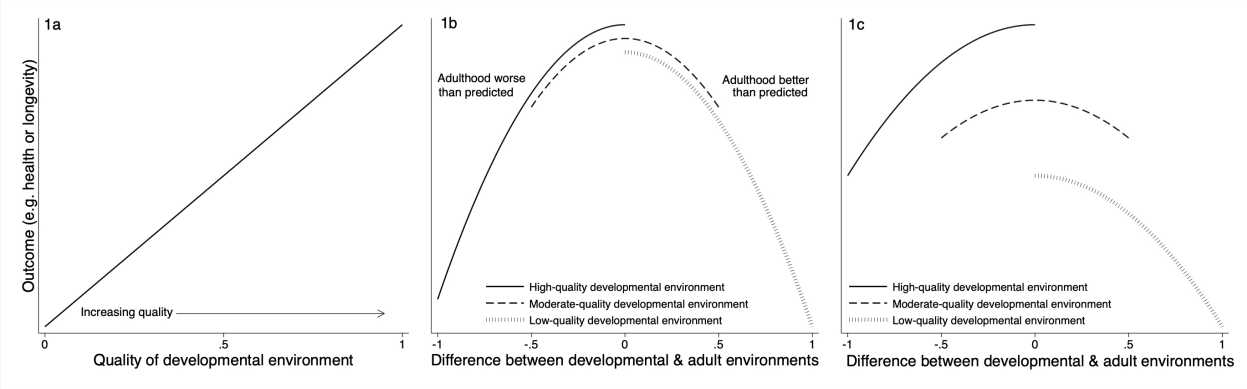
The developmental constraints (DC) and predictive adaptive response (PAR) hypotheses address how two different phenomena affect adult outcomes. The DC hypothesis primarily makes predictions about the effect of an organism's starting point (i.e., the quality of their developmental environment) on adult outcomes (Figure 1a). Specifically, it proposes that a poor-quality developmental environment will lead to worse outcomes in adulthood [43]. In contrast, PAR hypotheses are primarily about the effects of *change* in the environment (more specifically, a correct or incorrect prediction about this change) on adult outcomes (Figure 1b). Though there are several flavors of predictive hypotheses [34], all propose that organisms use cues available during development to make predictions about some feature of adulthood (e.g., what their adult environment will be like, or what their own somatic state will be). The worse the prediction, the worse an organism's outcomes in adulthood will be [4, 24].

The two theories are not mutually exclusive because adult outcomes can simultaneously be determined by the quality of the developmental environmental and how well organisms predict some feature of adulthood [20, 5]. For example, depending on the size of the effects of developmental constraints, organisms who start off in low-quality environments may always fare worse than those who start in high-quality environments, even if both also fare worse if there is a bigger mismatch between their developmental and adult environments (Figure 1c). In addition, higher-quality adult environments may also yield better outcomes, independent of any effects of developmental environment or how well an organism predicted their adult environment [44, 52]. This is called the adult environmental quality hypothesis (AEQ) [43].

Currently, most empirical studies on humans and other mammals support the DC hypothesis and not PAR hypotheses [reviewed in 34]. However, it is difficult to know what conclusions to draw from this pattern, in part because the literature often does not clearly delineate when these and related hypotheses (e.g., AEQ) do and do not generate distinguishable predictions. Further, we contend that the literature is increasingly reliant on empirical tests that conflate, rather than separate, the two phenomena. To remedy these problems, we provide mathematical definitions for the DC, PAR, and AEQ models. We then explicate several conceptual issues that constrain any empirical tests of these hypotheses, as well as problems that are specific to statistical tests that rely on interaction effects (e.g., interactions between developmental and adult environments

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 1: Illustration of the two dominant evolutionary theories for the observed relationship between developmental environments and adult outcomes.




Notes. **1a:** The primary prediction of the developmental constraints (DC) hypothesis is that, controlling for adult environmental quality, low-quality developmental environments will result in worse adult outcomes (e.g., health or longevity) than high-quality ones. **1b:** The primary prediction of the most common version of the predictive adaptive response (PAR) hypothesis is that, controlling for adult environmental quality, the worse the match between an organism's developmental environment and its adult environment the worse its adult outcomes will be. The center of the x-axis (0) represents a perfect match between developmental and adult environment. The left-hand side (-1) would represent an organism who predicted the highest-possible quality adult environment (based on their developmental environment) but ended up in the lowest-possible quality environment. The right-hand side (1) would represent an organism who predicted the lowest-possible quality adult environment (based on their developmental environment) but ended up in the highest-possible quality environment. Organisms who have a moderate-quality developmental environment can find that their adult environment is either worse or better than their development-based prediction (represented by -0.5 and 0.5, respectively), but are limited in how wrong their prediction can be in either direction. That is, if environmental quality is bounded between 0 (worst) and 1 (best) as in Figure 1a and an organism predicts 0.5, then the possible range of prediction errors is -0.5 to 0.5. **1c:** Both the DC and PAR hypotheses can simultaneously be true. Organisms who experience low-quality developmental environments can have worse outcomes relative to those who experienced high-quality developmental environments, and organisms can fare worse the greater the difference between their developmental and adult environments.

or developmental environments and phenotypes). Finally, we outline new empirical tests that are consistent with the predictions derived from our mathematical models. We use simulations to compare the sensitivity and specificity of these new tests to tests that are often used in the existing literature. We conclude by providing guidance on when and how researchers can implement these tests using their own data.

## Mathematical definitions and predictions

The DC and PAR theories have suffered from ambiguity about the precise phenomena under consideration. Here we solve this problem by mapping descriptions of the DC and PAR theories, along with necessary related concepts, to mathematical definitions and theoretical predictions. The goal of this exercise is to highlight and make explicit assumptions that are often left implicit. For convenience, our definitions and predictions are summarized in Table 1. To simplify our exposition, we assume the functions that map environment, adaptation errors, or prediction errors to outcomes are differentiable. Our analysis can easily be extended to non-differentiable functions.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 1: Formal definitions and derived predictions of theories for the relationship between the quality of developmental environments and adult outcomes.

| Theory | Definition | Observable Variation | Prediction or Test (Eqn. No.) |
|---|---|---|---|
| Developmental Constraints (DC) | $\frac{\partial y_1}{\partial e_0} > 0$ | $(y_1, e_0)$ | $\frac{\partial y_1}{\partial e_0} > 0$ (1) |
| Predictive Adaptive Response (PAR) | $\mathrm{E}(e_1) = e_0$ $\frac{\partial p}{\partial \mathrm{E}(e_1)} < 0$ $\frac{\partial^2 y_1}{\partial p \partial e_1} < 0$ | | |
| Phenotypic adaptation test | | $(e_0, p)$ | $\frac{\partial p}{\partial e_0} < 0$ (5) |
| Environmental mismatch test | | $(y_1, \lvert\Delta e\rvert)$ | $\frac{\partial y_1}{\partial \lvert\Delta e\rvert} < 0$ (7) |
| Developmental Adaptive Response (DAR) | $\frac{\partial p}{\partial e_0} < 0$ $\frac{\partial^2 y_1}{\partial p \partial e_1} < 0$ | (see PAR) | (see PAR) |
| Adult Environmental Quality (AEQ)* | $\frac{\partial y_1}{\partial e_1} > 0$ | $(y_1, e_1)$ | $\frac{\partial y_1}{\partial e_1} > 0$ (11) |

Notes. $y_0$= developmental outcome; $y_1$=outcome in adulthood; $e_0$=developmental environment; $e_1$=adult environment; $\mathrm{E}(e_1)$=adult environment the organism expects; $\Delta e$=difference between developmental and adult environments. Definitions assume a differentiable function that maps from environment or discrepancy between environment and adapted-to environment to outcomes. Theories are described in detail in the "Mathematical definitions and predictions" section.

## Developmental constraints

The DC hypothesis proposes that low-quality developmental environments will lead to worse outcomes in adulthood, relative to high-quality developmental environments [43, 25]. There are many examples of this in the real world. For example, *Drosophila* who experience nutritional stress during development are smaller and have lower egg viability in adulthood than those who do not [30].

DC is described by a simple causal chain:

(1) Experience low-quality developmental environment[1]
↓
(2) Exhibit poor adult health/fitness outcomes.

Mathematically, this phenomenon can be represented as:

$$y_1 = f(e_0) \text{ where } \partial y_1 / \partial e_0 > 0 \tag{1}$$

where $y_1$ is adult outcomes (e.g., health or longevity), $f$ is a decreasing function, and $e_0$ is developmental environment. Higher values of $y_1$ and $e_0$ indicate better outcomes or higher-quality environments, respectively. DC is verifiable with observable data, so it can be directly tested.

[1] Although DC typically focuses on external developmental environments, the hypothesis could theoretically be extended to internal states, mirroring the distinction that is drawn between external and internal PARs. However, we will focus on DC application to external environments.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## PAR-related theories

In addition to defining the PAR hypothesis, our goals in this section are to 1) connect PAR to a related theory, the mismatch hypothesis; 2) define the theoretical (as opposed to empirical) tests for PAR; 3) describe variants of PAR; and 4) outline the relationship between PAR and the often empirically indistinguishable developmental adaptive response (DAR) hypothesis.

In its most general form, the PAR hypothesis proposes that organisms adapt to (i.e., adopt phenotypes optimized for [66]) their expected adult environment. A widely-cited example occurs in meadow voles. Young voles who mature in colder temperatures are born with thicker coats than those who mature in warmer ones, even if their uterine environments are identical temperatures [36]. This is an example of a predictive adaptive response because the benefit of the adaptation—i.e., the thick coat phenotype—primarily accrues in adulthood.

A related theory, the mismatch hypothesis, is ubiquitous in the evolutionary health literature[2] [21, 22, 57, 19]. It posits that the more closely the environment at life-history stage $k$ matches the environment to which the organism is adapted at stage $k$, the better off an organism's outcomes at stage $k$ will be. An example of this occurs in *Bicyclus anynana* butterflies, who have higher survival rates if the temperature and rainfall of their adult environments is similar to the temperature and rainfall of their developmental environments [8].

The causal chain that motivates both the PAR and mismatch hypotheses is expressed as a set of four steps, which we will repeatedly refer to as we develop our mathematical argument:

(1) Experience low-quality developmental environment
↓
(2) Predict low-quality environment in adulthood
↓
(3) Adopt phenotypic adaption to low-quality adult environment
↓
(4) Exhibit best health/fitness outcomes in low-quality adult environment.

While this description focuses on low-quality environments, it can be reversed for high-quality environments. In each step, internal state can be substituted for external environment[3]; later, this modification will enable us to distinguish internal and external PAR. Until then, we will focus on the external environment to simplify exposition.

Each step in the above chain is functional. The mechanism of PAR is captured by steps 2 to 3: the organism's prediction about its adult environment causes it to develop phenotypic adaptations to that environment. Step 4 is required to complete the logic of PAR: the organism adopts the phenotypic adaptation to a low-quality adult environment because doing so improves health and fitness outcomes in adulthood, relative to not adopting the phenotype. The purpose of step 1 is to enable taking PAR to data. It is difficult to measure organisms' predictions. Step 1 posits a model for formulation of an organism's predictions which has an input—developmental environment—that is observable.

The steps outlined above can be mapped to mathematical expressions. At its core, PAR is

[2]Though it will be used here in a developmental mismatch context, it is generalizable to evolutionary mismatch [22]: conditions during some historical time period would replace the developmental environment, and some point in an organism's lifespan would replace the adult environment.

[3]The internal PAR hypothesis proposes that organisms use their external environment during development to predict their somatic state in adulthood [48, 46], but there is little reason to believe that internal state during development could not also be used to predict internal state in adulthood. Somatic state could theoretically be substituted for external environment either at $e_1$ only, or at $e_0$ and $e_1$.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

defined by two propositions. The first captures steps 2 to 3:

$$p = f(E(e_1)), \text{ where } \partial f/\partial E(e_1) < 0 \tag{2}$$

where $p$ is a phenotypic adaptation to a low-quality environment and $e_1$ is the quality adult environment. This equation says that predicting a higher-quality environment reduces the degree of adaptation to a low-quality environment (or stops it altogether, if the phenotypic adjustment is binary). The second proposition captures steps 3 to 4:

$$y_1 = g(p, e_1), \text{ where } \partial^2 y_1/\partial p \partial e_1 < 0, \tag{3}$$

i.e., adapting to a low-quality adult environment yields worse outcomes as adult environmental quality improves. Together, these two propositions formalize steps 2 to 4.

In order to make PAR testable, we must also formalize step 1, i.e., the assumption that the organism predicts its adult environment will be the same as its developmental environment:

$$E(e_1) = e_0 \tag{4}$$

This claim is not essential to defining PAR. However, we include it because, in contrast to the organism's prediction, the developmental environment is observable. This makes the PAR hypothesis falsifiable.

### Tests for PAR

There are two basic strategies for testing PAR: phenotypic adaptation tests and mismatch tests. Both of these strategies are implied by our mathematical models, and are used in the literature.

The first strategy—phenotypic adaptation tests—focuses on the *mechanism* of PAR, i.e., the mechanism that produces an organism's phenotypic adjustment (e.g., the molecular basis of wing spot pattern development in *Bicyclus anynana* butterflies when they live in wet versus dry environments [8]). We refer to the phenotypic adjustment as "the mechanism" because this is how organisms get from predictions to presumptively better outcomes. To understand the specific tests implied by this phenotype-focused strategy, we can collapse the three equations that capture PAR (Eqs. 2, 3, and 4) above into two different expressions. The first expression is obtained by plugging steps 1 to 2 of our causal chain (4) into steps 2 to 3 (2) to obtain

$$p = f(e_0), \text{ where } \partial f/\partial e_0 < 0 \tag{5}$$

This equation says that low-quality developmental environments cause phenotypic adaptations to low-quality adult environments. The second expression is simply steps 3 to 4, (i.e., (3)), which says that a phenotypic adaptation to a low-quality environment improves adult outcomes in that environment. Each of these two expressions implies a separate test for PAR that makes use of data on phenotypic adaptations.

The second strategy for testing for PAR—mismatch tests—ignores information about the mechanism and relates to the mismatch hypothesis. Specifically, it plugs the first expression above (5) into steps 3 to 4 (3) to connect developmental and adult environment, without reference to any specific phenotypic "choice:"

$$\begin{gathered} y_1 = g(f(e_0), e_1), \text{ where} \\ \frac{\partial^2 y_1}{(-\partial e_0)\partial e_1} = \frac{\partial^2 y_1}{(-\partial p)\partial e_1}\frac{\partial p}{\partial e_0} < 0 \end{gathered} \tag{6}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

This equation gives both the definition and the test of the mismatch hypothesis. To see why, it helps to evaluate the function at the point where $e_0 = e_1$. From here, a concomitant decrease in developmental environmental quality and an increase in adult environmental quality (or vice versa) increases the mismatch between developmental and adult environments, leading to worse outcomes.[4] The aforementioned definition of the mismatch hypothesis is awkward because it relies on a cross-derivative and perhaps restrictions on $(e_0, e_1)$ (from note 4). Therefore, we simplify the mismatch hypothesis as follows:

$$\partial y_1 / \partial |e_1 - e_0| < 0 \tag{7}$$

Now, the hypothesis states that adult outcomes decline as the gap between developmental and adult environments increases.

The first strategy—tests of phenotypic adaptation—is a stepping stone to the second, so we presented it first. For further treatment of this first strategy, refer to the appendix. We now focus exclusively on the second strategy—mismatch tests—because it is commonly invoked in the literature.

### Variants of PAR

Although the above descriptions of PARs and the mismatch hypothesis align with the most common variants found in the literature, one can define alternative versions that are also cohesive. Here we discuss one prominent variant, the internal PAR model. In the appendix we discuss two other versions that could easily be defined: (i) a version that predicts adult environments will be different than developmental environments, and (ii) a version that posits that overly optimistic and overly pessimistic predictions have differential effects.

The internal PAR model distinguishes between the external environment and internal state of the organism [65, 48, 46, 64]. One can capture this version of PAR with steps 1-4 of our causal chain by replacing either $e_1$ only or both $e_0$ and $e_1$ (note 17) with an organism's internal state in the relevant life stages [46]. The testable predictions remain as in (5) or (7) so long as an organism uses its developmental environment (or its developmental somatic state) to predict its future somatic state. A major challenge to distinguishing internal from external PAR, however, is that when internal states are influenced by external environments—the premise upon which the internal PAR theory is based—observed differences in the former may simply reflect unobserved differences in the latter.

In the literature, tests for internal PAR have focused on phenotypic adaptations, while tests for external PAR have focused on environmental mismatch [46]. However, given that the definitions of the concepts are the same except for re-defining one or both of the $e$ terms (note 17), this distinction has no theoretical justification. External PAR can be tested using a phenotypic adaptation strategy ((5), which we discuss in the appendix), and internal PAR could be tested using mismatch (7).

### Developmental adaptive response

The developmental adaptive response (DAR) model is theoretically distinct—but often not empirically distinguishable—from PAR. DAR posits that organisms adapt to their developmental environ-

[4] A technical caveat: opposite sign changes can decrease the mismatch if one starts from where $e_0 > e_1$. Therefore, the mismatch hypothesis may require the assumption that $e_0 \leq e_1$ to be a valid test of PAR. When that is not true, mismatch may be true but needs a theoretical justification aside from PAR. This is related to the limited environments in which one can test for PAR listed in Table 2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

ment[5]. This theory captures a common alternative explanation for observations that mismatched developmental and adult environments may lead to poor outcomes, namely that an organism may adapt to its developmental environment not because that environment is its best prediction of its adult environment, but because its goal was simply to adapt to its developmental environment [4, 54, 34]. One example of a DAR comes from water fleas, who acquire a protective "helmet" when they receive signals that they will be in a predator-rich environment [7]. This is considered a DAR, rather than a PAR, because the fleas benefit immediately from their phenotypic adjustment [24]. DARs are not merely distinct from PARs, but depending on circumstance, can be mutually exclusive. However, DARs can generate the same prediction as PARs, which makes distinguishing the theories difficult.

The causal chain that motivates DAR is described by the following three steps:

(1) Experience low-quality developmental environment
↓
(2) Adopt phenotypic adaption to low-quality developmental environment
↓
(3) Exhibit best health/fitness outcomes if adult environment is low-quality.

The only difference between DAR and PAR is the removal of the prediction in the causal chain for PAR (i.e., step 2 in that chain). The organism is adapting to the developmental environment, which is directly observable, and thus no prediction is necessary. As with PAR, we could replace environment with internal state to define an internal variant of DAR.

Two mathematical formulas define DAR. Steps 1 and 2 of the DAR causal chain are captured by

$$p = f(e_0), \text{ where } \partial f/\partial e_0 < 0. \tag{8}$$

That is, phenotypic changes are a direct response to the developmental environment. The same is true for PAR, except that the developmental environment matters *indirectly* because it is used to predict the adult environment. Steps 2 and 3 for DAR imply

$$y_1 = g(p, e_1), \text{ where } \partial^2 y_1/\partial p \partial e_1 < 0. \tag{9}$$

That is, phenotypic adaptations to a low-quality developmental environment harm adult outcomes when the adult environment is high- rather than low-quality. This claim is the same as steps 3 and 4 of the PAR causal chain.

The similarity between the equations for DAR (8, 9) and PAR (2, 3) highlight the difficulty of distinguishing DAR from PAR. DAR's first proposition is identical to the first test for PAR in (5). If we plug the first DAR equation into the second, the result,

$$\begin{gathered} y_1 = g(f(e_0), e_1), \text{ where} \\ \frac{\partial^2 y_1}{(-\partial e_0)\partial e_1} = \frac{\partial^2 y_1}{(-\partial p)\partial e_1}\frac{\partial p}{\partial e_0} < 0 \end{gathered} \tag{10}$$

is identical to the mismatch hypothesis (6) that flows from PAR.

There are, however, some circumstances in which it is possible to distinguish DAR and PAR. If the phenotypic adaptation for DAR is different than for PAR, i.e., one examines an adaptation

[5]The term immediate adaptive response is in use elsewhere in the literature [e.g., 5, 69]. Because it is often used as a synonym for developmental constraints, we deliberately avoid using it here.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

that is only useful in early life, then one can test for DAR but not PAR. Likewise, an adaption only useful during adult life can be used to test for PAR but not DAR. An example of this would be wing polyphenisms in locusts. Wing phenotype varies between adult morphs depending on the developmental environments larvae are exposed to, but the wing adaptations—which are not expressed until adulthood—clearly have no benefit during the larval phase [2, 51, 59, or for a more general treatment of such polymorphisms, see [67]]. Note that this approach to distinguishing DAR and PAR requires the use of a phenotypic adaptation test. The mismatch test yields the same prediction for DAR and PAR because it does not examine phenotypic adaptations.

Because a primary goal of this paper is to address the conceptual framing of the literature on DC and PARs, hereafter we will refer to the prediction that a mis-fit between an adaptation and adult environment leads to worse adult outcomes as PAR rather than DAR, even though the two are empirically indistinguishable with a mismatch test.

### Adult environmental quality hypothesis

Finally, the AEQ hypothesis says that a higher-quality adult environment will result in better adult outcomes, i.e.,

$$\partial y_1/\partial e_1 > 0. \tag{11}$$

## Conceptual issues in testing models

Three conceptual issues are central to any empirical tests of the DC, PAR, and AEQ hypotheses: the non-independence of $e_0$, $e_1$, and $\Delta e$; overlap in empirical predictions; and the non-mutually exclusive nature of the theories. Here we explicate each of these issues in turn.

### Issue 1: Non-independence of $\mathbf{e_0}$, $\mathbf{e_1}$, and $\mathbf{\Delta e}$

First, it is impossible to independently vary early life and adult environments in a way that allows us to conduct independent tests of DC, PAR, and AEQ. An experimental paradigm would vary $e_0$ to test the effects of developmental environments on adult outcomes (i.e., DC), vary $e_1 - e_0$ ($=\Delta e$) to test the effects of the match between developmental and adult environments on adult outcomes (i.e., PAR), and vary $e_1$ to test the effects of adult environmental quality on adult outcomes (i.e., AEQ). However, one cannot vary $e_0$ or $e_1$ without simultaneously varying $\Delta e$. For example, if we hold $e_0$ constant but increase $e_1$, one cannot determine if an observed change in outcomes is associated with an increase in $e_1$ or a increase in the change in environment $\Delta e$ because an increase in adult environment mechanically necessitates an increase in the change in environment. This means that if one is testing for DC and for PARs, one cannot usually test the AEQ hypothesis[6].

To deal with this problem[7] in a way that does not arbitrarily prioritize the ability to test one hypothesis over another, researchers will benefit from specifying a data generating process [13]; that is, researchers should explicitly state the relationships between early life and adult environments and differences between the two, and decide which of these is/are exogenous to the theoretical model (i.e., independent of the model's structure), and which are endogenous to the model (i.e., a byproduct of the exogenous variation, not independent of the exogenous component). Specifying

[6]The inability to independently vary the triggers for the three hypotheses usually precludes testing all three, but there is one narrow exception that capitalizes on differences in predictions generated by symmetric versus asymmetric PARs. It is discussed in the appendix (Figure S1).

[7]Statisticians and economists call this an identification problem.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 2: Predicted relative health or fitness outcomes for organisms experiencing various early-life and adult environmental quality combinations, depending on whether the outcomes are generated by developmental constraints (DC), predictive adaptive responses (PARs), or both DC and PARs acting together.

| Environment | Only DC is true | Only PAR is true | Both DC & PAR are true | Implication |
|---|---|---|---|---|
| HL vs HH | HL = HH | HL < HH | HL < HH | If HL $\geq$ HH, can reject PAR & both |
| HL vs LL | HL > LL | HL < LL | Ambiguous | If HL >(<) LL, DC more (less) powerful than PAR |
| HL vs LH | HL > LL | HL = LH | HL > LL | If HL $\leq$ LL, can reject DC & both |
| HH vs LL | HH > LL | HH = LL | HH > LL | If HH $\leq$ LL, can reject DC & both |
| HH vs LH | HH > LH | HH > LH | HH > LH | If HH $\leq$ LH, can reject all theories |
| LL vs LH | LL = LH | LL > LH | LL > LH | If LL $\leq$ LH, can reject PAR & both |

Notes. H = high-quality environment, L = low-quality environment. The first letter in every pair represents the organism's developmental environment, while the second represents their adult environment. E.g., row one compares individuals who had high-quality developmental but low-quality adult environments (HL) to those who had high-quality environments for their whole lives (HH) (the first column). It then explicates the predictions that follow if only DC were true (second column), only PARs were true (third column), or both were true (fourth column). < and > signs refer to which organism would have a worse or better outcome in adulthood under the given hypothesis, respectively.

this data generating process is important because only hypotheses that concern the exogenous variation are testable. The data generating process that best represents the biological phenomenon of interest here is that "nature" (defined very broadly as environment, genetics, or some combination of the two) determines $e_0$ and $\Delta e$, and that these set points collectively generate $e_1$ via the formula $e_1 = e_0 + \Delta e$. This specification allows us to test DC and PAR, but not AEQ. An alternative assumption—that $e_1$ is set by "nature," but that $\Delta e$ is simply a byproduct—seems less defensible, because time moves linearly: $e_0$ and $\Delta e$ come prior to $e_1$, so it is unlikely $e_1$ causes either $e_0$ or $\Delta e$. Researchers could assume a different data generating process, but for any data generation process, only two of the three hypotheses will be testable.

## Issue 2: Overlapping predictions

A second difficulty is that the DC, PAR, and AEQ hypotheses only generate divergent predictions under specific circumstances (Table 2). Even if we only examine 2 of the 3 hypotheses, e.g., DC and PAR, only certain types of variation in developmental and adult environments allow us to distinguish one theory from the other (Table 2). Specifically, PAR *is not testable* without variation in both the developmental *and* the adult environment [34, 26, 54]. There is no test that can distinguish PAR from DC unless some subjects experience matching conditions and some experience non-matching conditions.

Relatively few observational studies can meet this challenging requirement, especially studies of human subjects. For example, the canonical studies of the Dutch Hunger Winter [55] and Leningrad Siege [61] only contain information on the conditions in the fifth and second rows of Table 2, respectively (though the degree to which adult environments matched developmental environments in Leningrad is open to interpretation [4]). One implication of this is that far fewer human data sets than animal data sets are suitable for distinguishing PAR from DC because it is relatively rare to have access to subjects in all the necessary developmental and adult conditions [but see, e.g., 26, 27, 47, 33, 57, for exceptions].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

### Issue 3: Non-mutually exclusive theories

A third difficulty is that the DC, PAR and AEQ theories are not mutually exclusive. Any combination of them could be true at the same time. An implication of this fact is that univariate models without higher-order polynomial terms (e.g., interacting and squared variables) may not be able to distinguish which theories are true. Consider the following simple, empirical model of outcomes:

$$y_1 = F(|e_1 - e_0|) = \beta|e_1 - e_0| + e. \tag{12}$$

This is an empirical model in the sense that it seeks to correlate adult outcomes with the mismatch between early life and adult environments. We differentiate this model from the theoretical models in the previous section by using capital letters to define the function. Suppose, however, that both DC and PAR are true. Specifying the empirical model as above does not simply control for or eliminate the influence of DC just by excluding a separate $e_0$ term. Therefore, the estimated coefficient will suffer omitted variable bias and not equal the partial derivative of adult outcomes with respect to mismatch ($\partial y_1/\partial|e_1 - e_0|$):

$$\mathrm{E}(\hat{\beta}) = \beta + \gamma_0\delta_{0d}, \tag{13}$$

where $\gamma_0$ is equal to $\partial y_1/\partial e_0$ and $\delta_{0d}$ is the coefficient from a regression of $|e_1 - e_0|$ on $e_0$. A similar problem afflicts any empirical specification that does not allow all plausible models to be true.

## Issues with a popular visualization and testing strategy

One popular empirical approach for implementing the mismatched environment[8] test for PARs relies on estimating a regression with interaction effects between developmental and adult environmental quality[9] [e.g. 14, 27, 26, 32, 11, 63, 33, 52, 42], i.e.,

$$y_1 = \beta + \beta_0 e_0 + \beta_1 e_1 + \beta_{01} e_0 e_1 + u, \tag{14}$$

where $u$ is a regression error term. We will hereafter refer to this equation as the interaction regression.

The test for DC with such a model focuses on the coefficient on $e_0$, while the test for a PAR focuses on the coefficient on the interaction term $\beta_{01}$. This strategy is in part motivated by conceptual visualizations in the literature like the one depicted in Figure 2. These visualizations plot adult outcomes against adult environmental quality separately for organisms who experienced different developmental environments [e.g., 43, 52]. Failing to reject the hypothesis that $\beta_{01} = 0$ is interpreted as evidence for DC (Figure 2a), while rejecting the hypothesis that $\beta_{01} = 0$ is interpreted as preliminary evidence for a PAR.

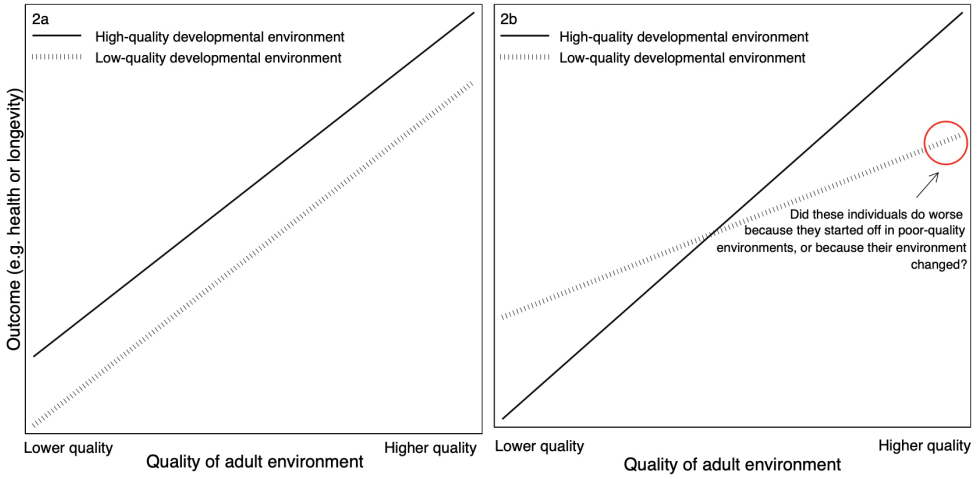
Because it is difficult to map the sign of the interaction term to a positive or negative effect of a change in environment, interpretation of $\beta_{01}$ involves visualizing the interaction effect. Specifically,

[8] The literature also sometimes uses the term mismatch to refer to a mismatch between actual adult environment and the phenotype chosen based on expected adult environment [68]. We are referring to the mismatch between developmental environment and adult environment, consistent with our definition in the Mathematical Definitions and Predictions section.

[9] We will primarily refer to the structure of models used to evaluate external PAR (aka informational adaptive developmental plasticity [46]). An internal PAR (aka somatic-state based adaptive developmental plasticity) test may interact the developmental environment with the presence of a phenotypic feature that is thought to be adaptive [46]. The conceptual overlap between these hypotheses is discussed in the Mathematical Definitions and Predictions section, and the specifics of tests that rely on environment/phenotype interactions in the appendix.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Figure 2: Commonly-used depictions of empirical evidence for the developmental constraints (DC, 2a) and predictive adaptive response (PAR, 2b) hypotheses.




Notes. **2a** As adult environmental quality improves, outcomes for those who experienced both high and low-quality developmental environments improves. However, those who started in low-quality developmental environments always fare worse than peers who started in high-quality environments. **2b** In low-quality adult environments, organisms who experienced "matching" (i.e. similarly poor) environments during development and in adulthood have better outcomes; in high-quality adult environments, those who experienced high-quality developmental environments fare better. These depictions manipulate both the starting point (i.e., the variable developmental constraints theory is concerned with) and how well the developmental and adult environments match (i.e., the variable predictive adaptive response theory is concerned with) simultaneously, making it difficult to distinguish the two. Moreover, the x-axis is the variable that the adult environmental quality hypothesis (AEQ) tests, i.e., that adult environmental quality impacts adult outcomes, regardless of developmental environment or how well the developmental and adult environments match.

if the line for organisms with high-quality developmental environments crosses from below the line for organisms with low-quality developmental environments (as in Figure 2b), this pattern is interpreted as evidence for a PAR, because organisms whose developmental and adult environments "match" do better than those in a "non-match" condition[10].

It is true that Figure 2b is consistent with predictions derived from the PAR hypothesis: organisms in low-quality adult environments who experienced low-quality developmental environments do better than those that had high-quality developmental environments and vice versa. It is also consistent with the AEQ hypothesis (11), since all organisms do better as their adult environmental quality improves.

However, any visualization that has the same axes as Figure 2 suffers from a fundamental problem: it is manipulating two variables simultaneously—developmental environment (implicitly, DC) and match/mismatch (PARs)—making it difficult to determine which variable is responsible for the observed effect. For example, on the right-hand side of Figure 2 (where the quality of the adult environment is high), the organisms represented by the dashed line started off in a low-quality environment, which might account for their worse outcomes relative to organisms represented by the solid line. However, they are *also* in an adult environment that does not match their developmental environment (i.e., they experienced change). The organisms represented by the solid line had a high-quality developmental environment, but they are also in an adult environment that matches

[10]If the low-quality developmental environment line merely has a lower slope than the high-quality developmental line, the plot can be interpreted as evidence for a PAR: the lines do not need to cross, and indeed may be likely not to if developmental constraints are also at work [e.g., 43].

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

their developmental environment (i.e., they did not experience change). It is not clear whether the individuals who grew up in high-quality environments have better outcomes because they started their lives on top or because their developmental and adult environments match. While this is discussed in greater detail in the appendix, simply comparing individuals from low and high-quality developmental environments in low-quality adult environments (the left-hand side of the x-axis in Figure 2) still does not cleanly separate the effects of starting point from the effects of mismatch, because it is still manipulating two different variables while embedding information about a third (adult environmental quality). Fundamentally, the visualization is flattening a three (or higher) dimensional problem into two dimensions, and in doing so, it conflates the two primary variables of interest.

Using adult environmental quality (i.e., the variable of interest in the AEQ hypothesis and not the main variables of interest when testing DC or PARs) as the x-axis variable generates a variety of interpretability problems and provides no clear way to distinguish between the theories of interest. In contrast, fewer mental gymnastics are required when, in the case of DC, the x-axis is the effect of the developmental environment (Figure 1a), or in the case of PARs, the x-axis reflects how well the developmental and adult environments match (Figure 1b & c)[11]. Graphs like Figure 1a-c allow us to separately, visually identify DC and/or PARs, as well as to separate how a negative change in environment might be different than a positive one (see also [32] for a related visualization technique).

In addition to the interpretability issues raised by Figure 2, using an interaction model to test for PARs is problematic (and is something we ourselves have done [32, 64]). To demonstrate why, we next describe the empirical tests that follow from the formal definitions of the DC and PAR concepts that we provided in the Mathematical Definitions and Predictions section. We then show why a formal definition of PAR is not detectable using the interaction regression in (14), why it is mathematically incompatible with the definition of PARs, and why this incompatibility generates biased coefficient estimates. To evaluate the consequences, we use simulated data to demonstrate how often the use of interaction regressions will cause both Type I and Type II errors.

## Empirical specification for constructing theoretically-derived tests of the DC and PAR models

The definitions of DC and PARs are compatible with a general functional form:

$$y_1 = F(e_0, |\Delta e|), \tag{15}$$

where we use capital $F$ to indicate an empirical function involving observed data. We do not include adult environment in the function (thus the AEQ hypothesis) for the identification reasons given in the previous section.

Taking these theories to data requires a functional form or specification that we can estimate, for example,

$$y_1 = \beta + \beta_0 e_0 + \beta_d |e_1 - e_0| + u. \tag{16}$$

where $\beta_d$ is the marginal effect of the absolute value of the mismatch between developmental and adult environments.

[11] An alternative visualization could plot the absolute value of the mismatch between developmental and adult environments on the x-axis. However, this would not visually distinguish between adult environments being worse than or better than developmental environments, which the depiction in Figure 1b and c does.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

### Regression model derived from mathematical definitions

To obtain a functional form, a reasonable approach is to assume $F$ is differentiable over the relevant range and take a Taylor expansion, which yields a power series. The power series can have infinite terms, which cannot be estimated given finite data. Picking a finite limit on the number of powers in the series involves a trade-off between bias and variance. The fewer the number of powers, the greater the risk of omitted variable bias; the greater the number of terms, the lower the degrees of freedom and thus the greater the standard errors on estimated coefficients. Moreover, once the number of terms exceeds the sample size, it also sacrifices identification.

We choose a second-order expansion around $e_0 = e_1 = |\Delta e| = 0$ to limit the number of terms that must be estimated. Estimating a regression model with only first-order terms cannot capture both DC and PAR. Moreover, it would be difficult to relate to the interaction regression (14), since a first-order expansion does not include an interaction term. An interaction term requires a second order (or higher) expansion. A higher than second-order expansion is possible, but requires more data to be adequately powered relative to a model with with fewer terms. Sample sizes in the real world are often insufficient to accommodate this requirement. A second-order expansion implies the following regression model:

$$\begin{aligned} y_1 = \gamma + \gamma_0 e_0 + \gamma_d |\Delta e| + \gamma_{00} {e_0}^2 \\ + \gamma_{\mathrm{dd}} |\Delta e|^2 + \gamma_{0d} e_0 |\Delta e| + u, \end{aligned} \tag{17}$$

where the subscript 0 on $\gamma$ indicates the coefficient is on $x_0$, $d$ indicates it is on $|\Delta e|$, $0d$ indicates it is on $e_0|\Delta e|$, and so on. The error $u$ captures omitted, orthogonal factors that influence the observed outcomes in a given dataset. The first three terms are included in a first-order Taylor expansion; the last three terms, including the interaction term also in (14), are added by a second-order expansion.

The quadratic regression model in (17) is the primary regression model that we recommend and evaluate in the remainder of this paper. It balances capturing elements omitted from the interaction model (which we discuss in the next section) with power limitations of real world data sets. Little is lost by switching from an interaction to quadratic regression other than having to change one's tests for DC and PARs. Researchers with larger data sets may attempt to estimate higher order polynomials and apply the appropriate tests implied by the definitions of DC and PARs.

### Testing for DC and PARs with a quadratic regression

The quadratic regression (17) implies specific and different statistical tests for DC and PARs. The definition of DC in (1) implies the following test for DC:

$$\partial y_1 / \partial e_0 = \gamma_0 + 2\gamma_{00} e_0 + \gamma_{0d} |\Delta e| > 0. \tag{18}$$

If one cannot reject this hypothesis, then one cannot reject DC. Moreover, the prediction of the PAR hypothesis in (7) implies the following test for PARs:

$$\partial y_1 / \partial |\Delta e| = \gamma_d + 2\gamma_{\mathrm{dd}} |\Delta e| + \gamma_{0d} e_0 < 0 \tag{19}$$

If one cannot reject this hypothesis, one cannot reject PAR.

## Implications for the interaction regression model

Although the interaction regression in (14) is a popular functional form used to test for PARs, and also may be used to test for DC (via the coefficient on $e_0$), there are several problems with using that regression to test these hypotheses.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

First, if the interaction regression were a correct description of reality, PARs could not be true. To see why, we plug the data generating process $e_1 = e_0 + \Delta e$ into the interaction model (14) so we can take the derivative of the latter with respect to $\Delta e$. Using the chain rule, we can now take the total derivative of the interaction model with respect to $|\Delta e|$:

$$\begin{aligned}\frac{\mathrm{d}y_1}{\mathrm{d}|\Delta e|} &= \left[\frac{\partial y_1}{\partial(e_1 - e_0)}\right] \frac{\partial(e_1 - e_0)}{\partial|\Delta e|}\mathrm{d}|\Delta e| \\ &= \left[\beta_1 + \beta_{01}e_0\right] \frac{|\Delta e|}{(e_1 - e_0)}\mathrm{d}|\Delta e| < 0 \end{aligned} \tag{20}$$

for positive $\mathrm{d}|\Delta e|$. The prediction of the PAR hypothesis in (7) implies that PARs exists if this derivative is negative. Note that the derivative must be negative when $\Delta e$ is both positive and negative. But if $\Delta e > 0$, then $\Delta e/|\Delta e| > 0$, and if $\Delta e < 0$, then $\Delta e/|\Delta e| < 0$. So, for the inequality in (20) to hold, $\beta_1 + \beta_{01}e_0$ has to be negative for positive $\Delta e$ and positive for negative $\Delta e$. However, if we start from a given developmental environment $e_0$, this is impossible: $\beta_1 + \beta_{01}e_0$ is constant and does not flip signs. Importantly, PARs are not a theoretical impossibility with the quadratic regression because it includes a $|\Delta e|$ term; the derivative in (19) can theoretically be negative for the full range of $\Delta e$.

Strictly speaking, the interaction regression will only fail to find evidence of PARs if one tests PAR based on the prediction in (7). The interaction regression cannot generate results that pass that test. However, this impossibility likely also undermines the two-part test that supplements the interaction regression with a visualization of the interaction. That two part-test can be thought of as an effort to approximate testing the prediction in (7), but if directly testing that prediction with the interaction regression cannot find PARs, then the approximation that combines the interaction regression with a visualization is also unlikely to find it.

A second problem with the interaction regression is that the appropriate test for DC when using a linear regression model is not the correct test for DC in the interaction regression. A reasonable test for DC in a regression where adult outcome is the dependent variable is whether the coefficient on the predictor variable $e_0$ is positive. That is an appropriate test if one estimates a linear model along the lines of

$$y_1 = \beta_0 e_0 + \beta_1 e_1 + u \tag{21}$$

(assuming that a linear model is the correct specification). Recall that the definition of DC implies that the test for DC is that derivative $\partial y_1/\partial e_0$ is positive. In the linear model in (21), the derivative equals $\beta_0$. But if we are estimating an interaction regression as in (14), the correct derivative is instead $\beta_0 + \beta_{01}e_1$. Even if $\beta_0 > 0$, the derivative could be negative if $\beta_{01}e_1 > -\beta_0$.

Third, if the interaction regression is not a correct description of reality (e.g., if the quadratic regression is actually correct), then the coefficients estimated by the interaction regression suffer from omitted variable bias. Compared to the quadratic model, the interaction regression omits squared ($e_0^2$, $e_1^2$) and interaction ($e_0|\Delta e|$) terms. With ordinary least squares estimation, other coefficients to some extent absorb variance that should be attributed to omitted terms. As a result, the coefficient on the interaction is biased:

$$\begin{aligned}\mathrm{E}[\hat{\beta}_{01}] - \gamma_{01} = \gamma_{00}\delta_{00} + \gamma_{11}\delta_{11} + \gamma_{\mathrm{dd}}\delta_{\mathrm{dd}} \\ + \gamma_{0d}\delta_{0d} + \gamma_{1d}\delta_{1d}\end{aligned} \tag{22}$$

where $\delta_{\mathrm{xz}}$, for $x, z \in \{0, 1, d\}$, is the coefficient on $e_x e_y$ from a regression of $e_0 e_1$ on $e_0$, $e_1$, and

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

$e_x e_z$[12]. If one were to test whether PARs should be rejected by testing whether $\beta_{01} = 0$, the test may be inaccurate. The test could reject that $\beta_{01} = 0$ even though $\gamma_{01}$, the true measure of the interaction, is zero. Likewise, it could fail to reject $\beta_{01} = 0$ even though $\gamma_{01}$ is zero. When testing for DC, the coefficient on both the interaction and the $e_0$ term may be biased for similar reasons, leading to potentially incorrect results even if the test for DC (1) is employed. Even if the test obtains the right answer, it produces inaccurate conclusions about effect sizes.

## Simulations of alternative tests for DC and PARs

We conducted simulations to determine how often the interaction regression, as compared to the quadratic regression, correctly finds or rejects DC or PAR.

In order to ensure that our simulations reflects patterns taken from the real world, the covariate values we used for $(e_0, \Delta e)$ were drawn from data on observed early life and adult dominance ranks for wild female baboons. These baboons were the subjects of long-term, individual-based research by the Amboseli Baboon Research Project [1]. As is true for many primates, female dominance rank determines access to key resources and predicts adult fitness outcomes [37]. Females typically attain an adult dominance rank immediately below that of their mother, and dominance rank is usually relatively stable across a female's life. However, group fissions and revolutions in the hierarchy can lead some females who are born to low-ranking mothers to become high-ranking in adulthood, and vice versa [1]. Importantly, the observed distribution of developmental dominance rank ($e_0$, which we define as maternal dominance rank at the subject's birth), and the difference between subjects' developmental and adult dominance ranks ($\Delta e$), generate covariate values that are highly generalizable to many kinds of environmental quality measures in a wide variety of species. In the baboons, $e_0$ had a roughly even distribution spanning the range from 0 (lowest-rank or worst) to 1 (highest-rank or best). $\Delta e$ had an approximately normal distribution in the range $[-1, 1]$. More details about the baboon data can be found in the appendix.

### Simulation design

Each simulation posits a virtual reality[13] where, by assumption, a 3rd-order polynomial accurately describes the effect on adult outcomes of (a) an organism's developmental environment and (b) the mismatch between their developmental and adult environments[14]. Each 3rd-order polynomial can

[12]The interaction regression does not produce omitted variable bias if the two variables in the interaction term, here developmental and adult environment, are both truly binary variables. In that case, the squared terms are identical to and thus fully captured by the non-squared terms. However, the environment is virtually never binary. For example, an indicator for malnutrition is actually a caloric intake variable that is partitioned into high and low regions. If the operationalized version of an environmental variable is actually a partition of a continuous variable, the fact that a quadratic regression is identical to an interaction regression does not imply the interaction regression is adequate. The appropriate test for PAR employs a regression involving the continuous environmental variable without partitioning. Using the partitioned variable instead yields biased estimates of the coefficient from the regression using the continuous variable.

[13]An important limitation is that we were unable to achieve integration with an Oculus VR headset.

[14]Why a polynomial and why a third-order one? Reality permits a general function relating $y_1$ to $e_0$ and $|\Delta e|$. To make realistic virtual universes that are tractable for simulation, we took a Taylor series expansion of that general function around zero. We needed to select a finite order at which to stop the expansion. A 1st-order polynomial would cause the interaction term in an interaction regression to always be zero. A second-order term would always be estimated well by a quadratic and not an interaction, but by assumption. A 3rd- or higher-order polynomial allows both the interaction and quadratic to generate errors that can be compared, which is our goal. However, as polynomial order increases, exponentially more computation power is required to simulate reality given a fixed level of variation in $(e_0, |\Delta e|)$.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

have infinitely different coefficient values and thus describe infinitely different realities. We pared those realities down by, first, only simulating realities on the nodes of a 10-dimensional lattice where the value of each dimension takes 5 values ranging from -1 to 1. Second, we ruled out realities where at any value of $e_0 \in [0, 1]$ or $\Delta e \in [-1, 1]$ generated an outcome outside the range of $[0, 1]$. That is, we retained outcomes that could map onto binary outcomes (0/1), or to outcomes which are continuous but bounded[15]. This filtering produced 130,201 parameter combinations and thus "feasible" realities.

Each of these feasible realities was evaluated for PARs and DC by applying the tests in (7) and (1), respectively. With a 3rd-order polynomial, the derivatives in these tests depend on the value of $e_0$ and $\Delta e$. We evaluated the derivatives for each reality at 16 evenly spaced values of $(e_0, \Delta e)$[16]. A reality was labelled positive for PARs or DC if the relevant derivative test was satisfied on average across these values. Of the feasible realities, 2,755 were truly positive for PARs and 2,755 were truly positive for DC. These feasible and true positive realities are the benchmark against which tests based on each regression model are evaluated.

For each feasible reality, we generated a data set for estimating regressions. Each data set included 2000 observations generated by drawing values of $e_0$ uniformly from evenly spaced points between 0 and 1; values of $\Delta e$ from a normal distribution with mean $-0.03$ and standard deviation 0.21 (motivated by the mean and SD from the baboon data) but truncated at -1 and 1; error terms from a normal distribution with mean 0 and variance equal to that of the outcome at the mean value of $(e_0, \Delta e)$ in each reality. Continuous outcomes $y_1$ were generated from a third-order polynomial of the generated $(e_0, |\Delta e|)$ plus a generated error term.

On the data set from each of these feasible realities, we estimated an interaction regression and quadratic regression. From the regression results for each reality, we generated four test results for PARs in that reality:

1. **Visualization test for PARs using the interaction regression** (14). This test finds evidence for PARs if 1) $\beta_{01} \neq 0$, and 2) the visualization shows that the fit line depicting the adult environment/adult outcomes relationship for organisms from low-quality developmental environments (the dotted line in Figure 2b) *intersects from above* the same line for organisms from high-quality developmental environments (the solid line in Figure 2b).

2. **"Relaxed" version of the visualization test with the interaction regression** (14). In the presence of development constraints (DC), the fit line for adult environment/adult outcomes for organisms from low-quality developmental environments may be shifted downwards relative to organisms from medium- and high-quality environments (as depicted in Figure 1c). In this case, PARs might exist even if the lines for low- and high-quality developmental environments do not cross. This suggests a relaxed visualization test which finds evidence for PARs if 1) $\beta_{01} \neq 0$, and 2) the visualization shows that the fit line depicting the adult environment/adult outcomes relationship for organisms from low-quality developmental environments has *a lower slope* than the same line for organisms from high-quality developmental environments.

3. **Theoretically-motivated test** (7) **applied to the interaction regression** (14). This test finds evidence for PARs if the derivative of the interaction regression with respect to $|\Delta e|$ is negative (i.e., (20)). We implement this test notwithstanding the fact that the PAR hypothesis cannot be true if the interaction regression is a correct specification of reality.

[15] An affine function of a bounded, continuous variable can transform it into continuous variable with a range of [0,1]. Moreover, the test for PARs or DC can be adjusted to account for that transformation.

[16] We picked 4 evenly spaced values for $e_0$ and for $e_1$ and calculate $\Delta e$ from these.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

4. **Theoretically-motivated test** (7) **applied to the quadratic regression** (17). This test finds evidence for PARs if the derivative of the quadratic regression with respect to $|\Delta e|$ is negative (19).

We also generated three tests for DC in each reality:

1. **Naive test for DC with the interaction regression** (14). This test finds evidence for DC if $\beta_0$ in (14) is positive.
2. **Theoretically-motivated test** (1) **applied to the interaction regression** (14). This test finds evidence for DC if the derivative of the interaction regression with respect to $e_0$ is positive, i.e., $\beta_0 + \beta_{01} e_1$ is positive.
3. **Theoretically-motivated test** (1) **applied to the quadratic regression** (17). This test finds evidence for DC if the derivative of the quadratic regression with respect to $e_0$ is positive. This test is presented in (18).

The appendix contains additional details on the simulations.

## Simulation results

The top panel of Table 3 provides the sensitivity and specificity of each of the four methods of testing for PARs across all feasible realities (sensitivity is the percent of the 2,755 realities where PAR was true and where it was correctly detected; specificity is the percent of the remaining 127,504 realities where PAR was not true and where it was correctly not detected). The second column shows the performance of a coin-flip test, which we use as a benchmark of a data-uninformed test (i.e., a test with performance equal to chance).

The interaction regression has poor or imbalanced performance across a range of tests. For instance, the interaction regression with the visualization approach to testing for PARs has sensitivity that is only slightly better than a coin flip (58.84%). A relaxed visual test has higher sensitivity (79.24%) but lowered specificity, only marginally better than coin flip (56.64%); in other words, it incorrectly detected PAR 43.46% of the time. The interaction regression showed even worse performance with the theoretically-motivated test for PAR: sensitivity (9.76%) was much worse than a coin flip. This very poor sensitivity is to be expected: if the interaction regression is a correct description of reality, we demonstrated that one theoretically cannot find PAR. The visualization test using the interaction regression actually gets the right answer more often than the theoretically-motivated version precisely *because* the former is not actually testing the prediction generated by the PAR hypothesis (7). This allows it to perform at or marginally better than a coin flip, while a theoretically-motivated use of the interaction regression is specifically biased *against* finding PARs even when they are true.

The quadratic regression combined with a theoretically-motivated test performs best of all. Sensitivity (90.34%) is higher than any test using an interaction regression and specificity is roughly the same (71.61%) as the best tests under the interaction model. This test is not perfectly sensitive and specific because it too suffers from bias due to omitted variables: that is, the realities it approximates also have third-order terms.

The bottom panel of Table 3 provides the sensitivity and specificity of three methods of testing for DC across feasible realities. The interaction model performed poorly relative to the quadratic regression. The naive test using the interaction regression has good sensitivity, but much worse specificity than a coin-flip. Another way to say this is that the test often finds DC whether it is

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Table 3: Percent of simulated realities where PARs and DC were correctly detected (sensitivity) and correctly not detected (specificity) using different tests.

| Predictive Adaptive Response (PAR) | | | | | |
|---|---|---|---|---|---|
| Regression: | None | Interaction | Interaction | Interaction | Quadratic |
| Test: | Coin flip | $\beta_{01} \neq 0$ in (14) & strict visual test | $\beta_{01} \neq 0$ in (14) & relaxed visual test | Theory-motivated test (20) | Theory-motivated test (19) |
| Sensitivity | 51.54 | 58.84 | 79.24 | 9.76 | 90.74 |
| Specificity | 50.13 | 72.40 | 56.64 | 73.67 | 71.45 |
| **Developmental Constraints (DC)** | | | | | |
| Regression: | None | | Interaction | Interaction | Quadratic |
| Test: | Coin flip | | Naive test ($\beta_0 > 0$) | Theory-motivated test ($\beta_0 + \beta_{01}e_1 > 0$) | Theory-motivated test (18) |
| Sensitivity | 49.43 | | 93.21 | 100.00 | 100.00 |
| Specificity | 50.09 | | 10.11 | 56.76 | 66.57 |

true or not. The interaction regression combined with a theoretically-motivated test has perfect sensitivity, but specificity that is only marginally better than a coin flip. Switching to a quadratic regression and using a theoretically-motivated test for DC performs best of all. It too has perfect sensitivity, and it has somewhat better specificity (66.57%).

## Discussion

The existing literature on DC and PARs suffers from two problems. First, it lacks precise and consistent definitions of these phenomena. Second, perhaps as a result of this, at least one common existing test for DC and PARs has poor sensitivity and specificity, often little better than a coin flip.

Various assumptions are built into different conceptions of PARs. To ensure that studies are comparing apples to apples, these assumptions need to be made explicit rather than left implicit. Unraveling these assumptions has the benefit of making it clear where competing hypotheses are and are not generating differentiating predictions. Due to their overlapping definitions and predictions, DARs (developmental adaptive responses) and PARs, as well as external and internal PARs, are difficult or impossible to distinguish in the real world, at least using currently available methods. Focusing on the broader concept of environmental mismatch is likely to prove more fruitful than debating the specifics of what particular point in time (or what particular cues) organisms are adapting to.

Testing one of the key predictions generated by the PAR hypothesis requires researchers to detect the effects of environmental mismatches. Doing so with any accuracy is difficult with an interaction model and its complementary data visualization strategy. Indeed, the interaction model specification is theoretically incapable of testing the mismatch prediction of PAR. A test that applies the mismatch prediction to the quadratic regression does not suffer from a basic mathematical

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

incompatibility problem. Similar arguments justify the use of a reasonable, formal definition of DC along with a quadratic regression. Our simulations show that use of the theoretical predictions of PARs and DC with the quadratic regression dramatically improved sensitivity/specificity trade-offs (relative to any use of an interaction model) when testing for PARs and DC.

Currently, support for PARs in the literature is often weak and/or conflicting, especially in mammals [34]. However, the definitional and testing issues highlighted here raise the question of whether this is because PARs do not exist or because of flaws in the methods used to detect them. Our results show that statistical tests derived from mathematical definitions of the DC and PAR concepts, along with more flexible regression models, provide much clearer answers, and will improve our ability to compare results across studies.

Because there is already a literature that employs visualizations such as Figure 2 and the interaction regression in (14) to test for PARs, it is useful to know if there exist assumptions under which those approaches remain valid tests of the hypothesis. Sufficient conditions for the interaction regression to be valid are if (a) mismatches in only one direction reduce health/longevity/fitness, *and* (b) the relationship between outcome, developmental environment and the change in environment over a lifetime is quadratic, but with certain parameter restrictions that cause the quadratic model to be identical to the interaction model. For example, focusing on a positive change in environment, the second condition is that the following are true:

$$\begin{aligned} y = {} & \beta_0 e_0 + \beta_\Delta (e_1 - e_0) + \beta_{00} e_0^2 + \beta_{\Delta\Delta} (e_1 - e_0)^2 \\ & + \beta_{0\Delta} e_0 (e_1 - e_0) + e, \end{aligned} \tag{23}$$

and $\beta_{\Delta\Delta} = 0$, $\beta_{00} = -\beta_{0\Delta}$, and $\beta_{0\Delta} \neq 0$. Under these parametric restrictions, the quadratic regression above collapses to something like the interaction regression in (14). Of course, it is difficult to know ex ante whether the second condition is satisfied without first estimating a quadratic regression.

The concepts of DC and PARs are ubiquitous in the literature of many academic fields. Further, significant human and financial resources are being devoted to untangling their effects because of their important implications for public health and policy. Clear, careful definitions and appropriate statistical tests are absolutely essential for forward progress in this important research area. We recommend that researchers take the following steps when testing the DC and PAR hypotheses:

1. Rely on statistical tests that are derived from mathematical definitions, to avoid conflating different phenomena being evaluated in the same model.

2. Avoid using interaction models or first-order polynomial models to test for environmental mismatch effects; instead, use a quadratic or higher-order regression model.

3. Verify that data visualizations cleanly separate the concept(s) of interest. It is better to use multiple visualizations that each address a single phenomenon than a single visualization that potentially conflates different phenomena.

4. Keep in mind that the use of a plausible data generating process, the overlapping predictions made by DC and PAR hypotheses, and the non-mutually exclusive nature of these hypotheses mean that separately identifying the effects of early, adult, and mismatched environments is usually not possible.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Acknowledgements

We thank A. Lea, C. Weibel, B. Lerch, N. Grebe, and the members of the S.C.A., E.A.A., and J. Tung laboratories, whose insights greatly improved the manuscript. We gratefully acknowledge the support of the NIH, especially the National Institute on Aging, and the NSF for long-term support of the Amboseli Baboon Research Project (ABRP), which inspired our evaluation of these models. This work was supported through NIH Grant R01AG053330, R01AG053308, R01HD088558, and P01AG031719. In the past decade, ABRP acknowledges support from Awards/Grants IOS 1053461, IBN 9985910, IBN 0322613, IBN 0322781, BCS 0323553, BCS 0323596, DEB 0846286, DEB 0846532, IOS 0919200, R01AG034513-01, R21AG049936, and P01AG031719. We thank the Kenya Wildlife Service, Institute of Primate Research, National Museums of Kenya, National Council for Science and Technology, the Kajiado County Council, the members of the Amboseli–Longido pastoralist communities in Kenya, and the Enduimet Wildlife Management Area. We are particularly grateful for the work of the Amboseli Baboon Project long-term field team (R. S. Mututua, S. Sayialel, and J. K. Warutere), and V. Oudu and T. Wango for their assistance in Nairobi. We thank Karl Pinc for his database design and management expertise, as well as the Amboseli Baboon Research Project's database technicians, D. Onderdonk, C. Markham, T. Fenn, N. Learn, L. Maryott, P. Onyango, and J. Gordon. Malani acknowledges the support of the Barbara J. and B. Mark Fried Fund at the University of Chicago Law School.

## References

[1] Susan C. Alberts and Jeanne Altmann. The Amboseli Baboon Research Project: 40 Years of Continuity and Change, pages 261–287. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.

[2] SW Applebaum and Y Heifetz. Density-dependent physiological phase in insects. Annual review of entomology, 44(1):317–341, 1999.

[3] Kasey E Barton and Karina Boege. Future directions in the ontogeny of plant defence: understanding the evolutionary causes and consequences. Ecology Letters, 20(4):403–411, 2017.

[4] Patrick Bateson. Fetal experience and good adult designa. International Journal of Epidemiology, 30(5):928–934, 2001.

[5] Patrick Bateson, Peter Gluckman, and Mark Hanson. The biology of developmental plasticity and the predictive adaptive response hypothesis. The Journal of Physiology, 592(11):2357–2368, 2014.

[6] Andreas Berghänel, Michael Heistermann, Oliver Schülke, and Julia Ostner. Prenatal stress effects in a wild, long-lived primate: predictive adaptive responses in an unpredictable environment. Proceedings of the Royal Society B: Biological Sciences, 283(1839):20161304, 2016.

[7] Maarten Boersma, Piet Spaak, and Luc De Meester. Predator-mediated plasticity in morphology, life history, and behavior of daphnia: the uncoupling of responses. The American Naturalist, 152(2):237–248, 1998.

[8] Paul M. Brakefield, Julie Gates, Dave Keys, Fanja Kesbeke, Pieter J. Wijngaarden, Antónia Montelro, Vernon French, and Sean B. Carroll. Development, plasticity and evolution of butterfly eyespot patterns. Nature, 384(6606):236–242, 1996.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[9] Tim Burton and Neil B. Metcalfe. Can environmental conditions experienced in early life influence future generations? Proceedings of the Royal Society B: Biological Sciences, 281(1785):20140311, 2014.

[10] Gabriella Conti and James J Heckman. The economics of child well-being. Report 0898-2937, National Bureau of Economic Research, 2012.

[11] David Costantini, Pat Monaghan, and Neil B. Metcalfe. Prior hormetic priming is costly under environmental mismatch. Biology Letters, 10(2):20131010, 2014.

[12] A. Danese, A. Caspi, B. Williams, A. Ambler, K. Sugden, J. Mika, H. Werts, J. Freeman, C. M. Pariante, T. E. Moffitt, and L. Arseneault. Biological embedding of stress through inflammation processes in childhood. Molecular Psychiatry, 16(3):244–246, 2011.

[13] Panchanan Das. Time Series: Data Generating Process, pages 247–259. Springer Singapore, Singapore, 2019.

[14] Mathieu Douhard, Floriane Plard, Jean-Michel Gaillard, Gilles Capron, Daniel Delorme, François Klein, Patrick Duncan, Leif Egil Loe, and Christophe Bonenfant. Fitness consequences of environmental conditions at different life stages in a long-lived vertebrate. Proceedings of the Royal Society B: Biological Sciences, 281(1785):20140276, 2014.

[15] Martha M. C. Elwenspoek, Xenia Hengesch, Fleur A. D. Leenen, Anna Schritz, Krystel Sias, Violetta K. Schaan, Sophie B. Mériaux, Stephanie Schmitz, Fanny Bonnemberger, Hartmut Schächinger, Claus Vögele, Jonathan D. Turner, and Claude P. Muller. Proinflammatory t cell status associated with early life adversity. The Journal of Immunology, 199(12):4046–4055, 2017.

[16] Harrison J.F. Eyck, Katherine L. Buchanan, Ondi L. Crino, and Tim S. Jessop. Effects of developmental stress on animal phenotype and performance: a quantitative review. Biological Reviews, 94(3):1143–1160, 2019.

[17] Linda Marie Fedigan, Sarah D. Carnegie, and Katharine M. Jack. Predictors of reproductive success in female white-faced capuchins (cebus capucinus). American Journal of Physical Anthropology, 137(1):82–90, 2008.

[18] Kenneth F. Ferraro, Markus H. Schafer, and Lindsay R. Wilkinson. Childhood disadvantage and health problems in middle and later life:early imprints on physical health? American Sociological Review, 81(1):107–133, 2016.

[19] Willem E. Frankenhuis and Marco Del Giudice. When do adaptive developmental mechanisms yield maladaptive outcomes? Developmental Psychology, 48(3):628–642, 2012.

[20] P. D. Gluckman, M. A. Hanson, and T. Buklijas. A conceptual framework for the developmental origins of health and disease. Journal of Developmental Origins of Health and Disease, 1(1):6–18, 2010.

[21] Peter D. Gluckman and Mark A. Hanson. The developmental origins of the metabolic syndrome. Trends in Endocrinology & Metabolism, 15(4):183–187, 2004.

[22] Peter D. Gluckman, Mark A. Hanson, and Felicia M. Low. Evolutionary and developmental mismatches are consequences of adaptive developmental plasticity in humans and have implications for later disease risk. Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1770):20180109, 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[23] Peter D. Gluckman, Mark A. Hanson, and Hamish G. Spencer. Predictive adaptive responses and human evolution. Trends in Ecology & Evolution, 20(10):527–533, 2005.

[24] Peter D Gluckman, Mark A Hanson, Hamish G Spencer, and Patrick Bateson. Environmental influences during development and their later consequences for health and disease: implications for the interpretation of empirical studies. Proceedings of the Royal Society B: Biological Sciences, 272(1564):671–677, 2005.

[25] Alan Grafen. On the uses of data on lifetime reproductive success, pages 454–485. University of Chicago Press, Chicago, IL, 1988.

[26] Adam D. Hayward and Virpi Lummaa. Testing the evolutionary basis of the predictive adaptive response hypothesis in a preindustrial human population. Evolution, Medicine, and Public Health, 2013(1):106–117, 2013.

[27] Adam D. Hayward, Ian J. Rickard, and Virpi Lummaa. Influence of early-life nutrition on mortality and reproductive success during a subsequent famine in a preindustrial population. Proceedings of the National Academy of Sciences, 110(34):13886–13891, 2013.

[28] Mark D. Hayward and Bridget K. Gorman. The long arm of childhood: The influence of early-life social conditions on men's mortality. Demography, 41(1):87–107, 2004.

[29] Andrew Wooyoung Kim, Emma K. Adam, Sonny A. Bechayda, and Christopher W. Kuzawa. Early life stress and hpa axis function independently predict adult depressive symptoms in metropolitan cebu, philippines. American Journal of Physical Anthropology, 173(3):448–462, 2020.

[30] Munjong Kolss, Roshan K Vijendravarma, Geraldine Schwaller, and Tadeusz J Kawecki. Life-history consequences of adaptation to larval nutritional stress in drosophila. Evolution: International Journal of Organic Evolution, 63(9):2389–2401, 2009.

[31] Christopher W. Kuzawa and Elizabeth A. Quinn. Developmental origins of adult function and health: Evolutionary hypotheses. Annual Review of Anthropology, 38(1):131–147, 2009.

[32] Amanda J. Lea, Jeanne Altmann, Susan C. Alberts, and Jenny Tung. Developmental constraints in a wild primate. The American Naturalist, 185(6):809–821, 2015.

[33] Amanda J. Lea, Dino Martins, Joseph Kamau, Michael Gurven, and Julien F. Ayroles. Urbanization and market integration have strong, nonlinear effects on cardiometabolic health in the turkana. Science Advances, 6(43):eabb1430, 2020.

[34] Amanda J. Lea and Stacy Rosebaum. Understanding how early life effects evolve: progress, gaps, and future directions. Current Opinion in Behavioral Sciences, 36:29–35, 2020.

[35] Amanda J Lea, Jenny Tung, Elizabeth A Archie, and Susan C Alberts. Developmental plasticity: Bridging research in evolution and human health. Evolution, Medicine, and Public Health, 2017(1):162–175, 2018.

[36] THERESA M Lee and IRVING Zucker. Vole infant development is influenced perinatally by maternal photoperiodic history. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 255(5):R831–R838, 1988.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[37] Emily J. Levy, Laurence R. Gesquiere, Emily McLean, Mathias Franz, J. Kinyua Warutere, Serah N. Sayialel, Raphael S. Mututua, Tim L. Wango, Vivian K. Oudu, Jeanne Altmann, Elizabeth A. Archie, and Susan C. Alberts. Higher dominance rank is associated with lower glucocorticoids in wild female baboons: A rank metric comparison. Hormones and Behavior, 125:104826, 2020.

[38] Jan Lindström. Early development and fitness in birds and mammals. Trends in Ecology & Evolution, 14(9):343–348, 1999.

[39] L H Lumey and A D Stein. In utero exposure to famine and subsequent fertility: The dutch famine birth cohort study. American Journal of Public Health, 87(12):1962–1966, 1997.

[40] Kesson Magid, Robert T. Chatterton, Farid Uddin Ahamed, and Gillian R. Bentley. Childhood ecology influences salivary testosterone, pubertal age and stature of bangladeshi uk migrant men. Nature Ecology & Evolution, 2(7):1146–1154, 2018.

[41] Craig A. McEwen and Bruce S. McEwen. Social structure, adversity, toxic stress, and intergenerational poverty: An early childhood model. Annual Review of Sociology, 43(1):445–472, 2017.

[42] Kyeong Woon Min, Taehwan Jang, and Kwang Pum Lee. Thermal and nutritional environments during development exert different effects on adult reproductive success in drosophila melanogaster. Ecology and Evolution, 11(1):443–457, 2021.

[43] Pat Monaghan. Early growth conditions, phenotypic development and environmental change. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1497):1635–1645, 2008.

[44] Laura R. Nagy and Richard T. Holmes. Food limits annual fecundity of a migratory songbird: an experimental study. Ecology, 86(3):675–681, 2005.

[45] Charles A Nelson, Zulfiqar A Bhutta, Nadine Burke Harris, Andrea Danese, and Muthanna Samara. Adversity in childhood is linked to mental and physical health throughout life. BMJ, 371:m3048, 2020.

[46] Daniel Nettle and Melissa Bateson. Adaptive developmental plasticity: what is it, how can we recognize it and when can it evolve? Proceedings of the Royal Society B: Biological Sciences, 282(1812):20151005, 2015.

[47] Daniel Nettle and Melissa Bateson. Childhood and adult socioeconomic position interact to predict health in mid life in a cohort of british women. PeerJ, 5:e3528, 2017.

[48] Daniel Nettle, Willem E. Frankenhuis, and Ian J. Rickard. The evolution of predictive adaptive responses in human life history. Proceedings of the Royal Society B: Biological Sciences, 280(1766):20131343, 2013.

[49] M. Nores. The economics of early childhood interventions, book section 17, pages 229–238. Academic Press, 2nd edition, 2020.

[50] Robin Nusslock and Gregory E. Miller. Early-life adversity and physical and emotional health across the lifespan: A neuroimmune network hypothesis. Biological Psychiatry, 80(1):23–32, 2016.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[51] MP Pener and Yoram Yerushalmi. The physiology of locust phase polymorphism: an update. Journal of Insect Physiology, 44(5-6):365–377, 1998.

[52] Gabriel Pigeon, Leif Egil Loe, Richard Bischof, Christophe Bonenfant, Mads Forchhammer, R. Justin Irvine, Erik Ropstad, Audun Stien, Vebjørn Veiberg, and Steve Albon. Silver spoon effects are constrained under extreme adult environmental conditions. Ecology, 100(12):e02886, 2019.

[53] Nadine Provençal, Linda Booij, and Richard E. Tremblay. The developmental origins of chronic physical aggression: biological pathways triggered by early life adversity. The Journal of Experimental Biology, 218(1):123–133, 2015.

[54] Ian J. Rickard and Virpi Lummaa. The predictive adaptive response and metabolic syndrome: challenges for the hypothesis. Trends in Endocrinology & Metabolism, 18(3):94–99, 2007.

[55] Tessa Roseboom, Susanne de Rooij, and Rebecca Painter. The dutch famine and its long-term consequences for adult health. Early Human Development, 82(8):485–491, 2006.

[56] Stacy Rosenbaum, Shuxi Zeng, Fernando A. Campos, Laurence R. Gesquiere, Jeanne Altmann, Susan C. Alberts, Fan Li, and Elizabeth A. Archie. Social bonds do not mediate the relationship between early adversity and adult glucocorticoids in wild baboons. Proceedings of the National Academy of Sciences, 117(33):20052–20062, 2020.

[57] B. Savitsky, O. Manor, G. Lawrence, Y. Friedlander, D. S. Siscovick, and H. Hochner. Environmental mismatch and obesity in humans: The jerusalem perinatal family follow-up study. International Journal of Obesity, 45(7):1404–1417, 2021.

[58] Jack P. Shonkoff, W. Thomas Boyce, and Bruce S. McEwen. Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. JAMA, 301(21):2252–2259, 2009.

[59] Stephen J Simpson, Alan R McCaffery, and Bernd F Hägele. A behavioural analysis of phase change in the desert locust. Biological reviews, 74(4):461–480, 1999.

[60] J. Maynard Smith, R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, and L. Wolpert. Developmental constraints and evolution: A perspective from the mountain lake conference on development and evolution. The Quarterly Review of Biology, 60(3):265–287, 1985.

[61] Sara A. Stanner and John S. Yudkin. Fetal programming and the leningrad siege study. Twin Research, 4(05):287–292, 2012.

[62] Jenny Tung, Elizabeth A. Archie, Jeanne Altmann, and Susan C. Alberts. Cumulative early life adversity predicts longevity in wild baboons. Nature Communications, 7(1):11181, 2016.

[63] T. Uller, S. Nakagawa, and S. English. Weak evidence for anticipatory parental effects in plants and animals. Journal of Evolutionary Biology, 26(10):2161–2170, 2013.

[64] Chelsea J. Weibel, Jenny Tung, Susan C. Alberts, and Elizabeth A. Archie. Accelerated reproduction is not an adaptive response to early-life adversity in wild baboons. Proceedings of the National Academy of Sciences, 117(40):24909–24919, 2020.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[65] Jonathan C.K. Wells. Obesity as malnutrition: The role of capitalism in the obesity global epidemic. American Journal of Human Biology, 24(3):261–276, 2012.

[66] Mary Jane West-Eberhard. Phenotypic plasticity and the origins of diversity. Annual review of Ecology and Systematics, 20(1):249–278, 1989.

[67] Mary Jane West-Eberhard. Developmental Plasticity and Evolution. Oxford University Press, New York, NY, 2003.

[68] Kai P. Willführ and Mikko Myrskylä. Phenotype-environment mismatch due to epigenetic inheritance? programming the offspring's epigenome and the consequences of migration. American Journal of Human Biology, 25(3):318–328, 2013.

[69] Thomas C. Williams and Amanda J. Drake. Preterm birth in evolutionary context: a predictive adaptive response? Philosophical Transactions of the Royal Society B: Biological Sciences, 374(1770):20180121, 2019.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

# Appendix

## Additional notes on the conceptual problems with the interaction visualization

In the section "Issues with a popular visualization and testing strategy" in the main text, we discuss problems with separately plotting adult outcomes on adult environment for individuals who started in high-quality and low-quality developmental environments (Figure 2 in the main text). Other papers have noted the conceptual difficulty with testing for predictive adaptive responses (PARs) when researchers only have data on subjects that all experienced similar adult environments, as is the case in some canonical human studies [e.g., 39]. These papers note that it is impossible to know whether individuals who started in high-quality environments end up doing better because their developmental environment was high-quality, or because their developmental and adult environments match [32, 34, 54, 26]. While this is not strictly a criticism of plots like Figure 2 in the main text, it can be interpreted as an argument that Figure 2 cannot be plotted if there is no variation in adult environment.

However, if a visualization like Figure 2 can be plotted, i.e., subjects experienced variation in their adult environments, then one has to address whether Figure 2 and depictions like it are informative. As we discuss in the main text, plotting outcomes against adult environments is a cumbersome way to visualize PARs. It does not cleanly isolate the effects of starting point (i.e., developmental constraints (DC)), because it uses the variable of interest from yet a third hypothesis (the adult environmental quality hypothesis (AEQ)) as the x axis. Moreover, there is ambiguity about what pattern should be interpreted as evidence for PARs. The bisection from above of (a) the adult outcome-adult environment line for organisms from high-quality developmental environments by (b) the same line for organisms from low-quality developmental environments, such as illustrated in Figure 2, is a sufficient but not necessary condition for PARs. A sufficient condition is that the slope of the low-quality line is smaller than the slope of the high-quality line. Finally, if a plot like Figure 2 is generated with predictions from estimation of an interaction model (as opposed to plots of raw data), either the plot cannot find PARs or suffers from omitted variable bias and may not be informative. The reasons for this are discussed in detail in the section titled "Implications for the interaction regression model" in the main text.

## Testing for predictive adaptive responses using phenotypic adaptations and phenotype/environment interactions

As discussed in the main text in the section "Mathematical definitions and predictions," there are two strategies for testing PAR. One is to look at the impact of differences between early life and adult environments, i.e., test the mismatch hypothesis. We discussed this strategy in depth in the main text. The other strategy is to examine the phenotypic adaptations to developmental environments (sometimes referred to as developmental inputs [46]) and their impact. We explore this second strategy here.

Recall that the main text noted that the 4 steps in the causal chain for PAR can be expressed as 3 equations, and that if one of the equations (for step 1) is plugged into another (for steps 2 to 3), one obtains a simplified expression for PAR with just two equations:

$$p = f(e_0), \text{ where } \partial f/\partial e_0 < 0 \tag{24}$$

$$y_1 = g(p, e_1), \text{ where } \frac{\partial^2 y_1}{(\partial p)\partial e_1} < 0 \tag{25}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

where $p$ are phenotypic adaptions to low-quality environments. The first equation says that a low-quality developmental environment triggers phenotypic adaptations suitable for a low-quality adult environment (because the adult environment is expected to be the same as the developmental environment).

The second equation says that outcomes are better with the relevant adaptation if the adult environment is low-quality, than they would be if the adaptation had not occurred. As we explained in the main text, while our statement of PAR talks about external environment, one could substitute internal somatic state for environment[17] and the hypothesis would stand.

This simplified definition suggests two empirical tests for PAR that rely on observed phenotypic adaptations. One test uses the first equation (24) in the simplified definition, i.e., it checks if organisms from a low-quality developmental environment make a phenotypic adaptation to that environment. The most basic regression specification for this test is

$$p = \beta e_0 + v, \tag{26}$$

where the test for PAR is that $\beta < 0$. In the main text we recommend *against* specifications that employ first-order approximations for $y_1 = f(|\Delta e|)$ such as $y_1 = \gamma|\Delta e| + e$ due to the risk of omitted variable bias in estimates of $\gamma$. That is not a problem for (26) because there are no other possible (known) triggers for an adaptation that is correlated with developmental environment. [18]

A second empirical test for PAR that examines phenotypic adaptation focuses on the second equation (25) in the simplified expression of PAR. A second-order approximation yields a quadratic regression of the form

$$y_1 = \beta_p p + \beta_1 e_1 + \beta_{pp} p^2 + \beta_{11} e_1^2 + \beta_{p1} p e_1 + u. \tag{27}$$

The test for PAR is that

$$\frac{\partial^2 y_1}{(\partial p)\partial e_1} = \beta_{p1} < 0. \tag{28}$$

Given how simple this test is, one might be tempted to estimate an interaction model of the form:

$$y_1 = \beta_p p + \beta_1 e_1 + \beta_{p1} p e_1 + u. \tag{29}$$

However, this would yield a biased estimate for $\beta_{p1}$ because it omits squared terms $p^2$ and $e_1^2$, which are correlated with the interaction term $pe_1$.

[46] has suggested an empirical analysis that somewhat resembles the preceding interaction model. There are, however, two differences between the test proposed by [46] and estimating (29). First, the test proposed by [46] replaces the adult environment with a developmental input $i_0 = -e_0$ (which could be some feature of either the organism's somatic state or its environment):

$$y_1 = \beta_p p + (-\beta_0) i_0 + (-\beta_{p0}) p i_0 + u. \tag{30}$$

[17] The internal PAR hypothesis proposes that organisms use their external environment during development to predict their somatic state in adulthood [48, 46], but there is little reason to believe that internal state during development could not also be used to predict internal state in adulthood. Somatic state could theoretically be substituted for external environment either at $e_1$ only, or at $e_0$ and $e_1$.

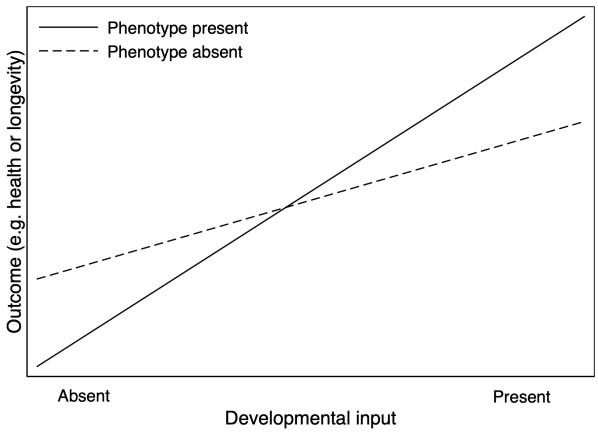
[18] It is true that a developmental adaptive response (DAR) could trigger adaptation. However, the problem there is that PAR and DAR are indistinguishable, not that the test for PAR is incorrect. Another way to put this is that the test for $\beta < 0$ is a test for both PAR and DAR, not that omitted variable bias means that this test is inaccurate. This is discussed in detail in the main text in the subsection entitled "Developmental adaptive response."

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Second, [46] recommends implementing this analysis via a regression and visualization similar to the visualization that [43] suggests for the environmental mismatch hypothesis. Figure 3 provides an example of the sort of visualization [46] recommends to complement the regression in (30). For convenience, we shall call this empirical strategy a "developmental input-phenotype interaction approach."

The approach suggested in [46] has several problems. First, the visualization that motivates the interaction regression above suffers from a problem similar to the one in Figure 2 in the main text (a visualization of a developmental environment/adult environment interaction). Both versions flatten a three (or higher) dimensional problem into two dimensions. In doing so, the figures omit an important variable and thus force what are typically un-articulated and/or untested assumptions. Figure 2 omits the developmental environment because it conflates the potential effects of the developmental environment with the potential effects of a change in the environment. Figure 3 groups organisms for whom the developmental input is a correct prediction of adult state with those for whom it is not. The prediction depicted in the figure is correct for the former set of organisms, but not for the latter. The picture can still be valid when both sets of organisms are examined together, but only with the additional assumption that the developmental input is *on average* a good prediction of adult state.

Figure 3: Visual depiction of an environment/phenotype interaction. This is conceptually similar to Figure 2 in the main text, but uses a developmental input on the x axis in place of adult environment. Figure taken from [46].




If evolution is responsible for the studied form of predictive response/developmental plasticity, it is certainly true that the prediction must be correct more often than not, unless there was a recent change in the environment and selective landscape. If the predictions are often incorrect, then this form of plasticity should not evolve. However, building this assumption into the figure and resulting interpretation seems problematic for two reasons. One is that a central question in the study of developmental plasticity is the degree to which it is an evolutionary phenomenon [67]. The other is that the conflation of organisms for whom the developmental input is a good prediction and those for whom it is not reduces the sensitivity of the suggested test. A more precise figure would condition on the set of organisms it is addressing, and allow for the fact that the crossing property is mitigated (or may disappear) for the organisms with "mismatched" developmental input and adult somatic state.

Second, an interaction regression that used adulthood instead of a developmental input might

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

still suffer from omitted variable bias. This means that the combination of a test for a significant interaction between adaptation and adult somatic state/environment ($\beta_{p0} \neq 1$) and the visualization will not yield an accurate test, because the estimate of $\beta_{p0}$ will be biased.

Third, the interaction regression in (30) suffers the same potential omitted variable bias problem as the interaction regression in Equation 1 in the main text. (30) above takes a strong stance on the functional form of the relationship between (a) the developmental input and phenotypic adaptation and (b) the adult outcome. A more modest assumption would be to start with a general functional form, e.g., $y = f(e_0, p^a)$ and then take a Taylor approximation. In this view, a quadratic regression of the form

$$y = \beta + \beta_0 i_0 + \beta_p p^a + \beta_{00}(i_0)^2 + \beta_{pp}(p^a)^2 + \beta_{0p} i_0 p^a + u, \tag{31}$$

is safer. If this functional form is correct, then the interaction regression suffers omitted variable bias due to misspecification. Even if the interaction regression is specified correctly, then a quadratic regression would yield unbiased estimates of $\beta_{0p}$, the trigger for a visualization. The test for whether the interaction regression was an acceptable specification is that estimates of $\beta_{00}$ and $\beta_{pp}$ are zero.

Fourth, because phenotypic adaptations are more likely be endogenous to an organism than adult environment is, the visualization and any regression that examines phenotypic adaptation rather than adult environment will be more likely to suffer selection bias (formally, a correlation between the explanatory variable–in this case, the phenotype–and the error term in the regression model). Consider an example in which the phenotype is how early an organism starts reproduction. Both the visualization and interaction regression suggest that a test for internal PARs is whether those who receive a negative input during development start reproduction earlier than those who receive a positive input. This test requires that there be some individuals who receive the negative input who do start reproducing early, and some who do not; without such variation one cannot estimate the interaction term. But one has to have a theory for why two organisms with the same input have different responses. Unless that response is unrelated to any observable variables, e.g., is random, the coefficient on the phenotypic adaptation and the interaction term will suffer from selection bias. If there is selection bias, then the test proposed in [48] may cause one to reject the internal PAR hypothesis even if it is true.

Fifth, if DC is also true, but has a differential effect on organisms that undertake a phenotypic adaptation and those that do not (which is certainly a biologically plausible scenario), then the developmental input-phenotype interaction approach will be inaccurate. DC changes the slope of the adult outcome-developmental input line. If it operates more on organisms that undertake a phenotypic adaptation, then it may suppress cross-over suggestive of a PAR even when there is a PAR. If DC operates more on organisms that do *not* undertake the adaptation, then it may generate cross-over suggestive of a PAR even when there is no PAR.

Finally, we want to address an important point about any test that relies on a specific phenotype, rather than the mismatch test strategy, regardless of whether the test is theoretically consistent with a PAR definition. A phenotypic test takes a narrow view of phenotypic adaptations, with the consequence of reducing the test's power to detect a PAR. Suppose that organisms who receive a negative developmental input can make several alternative adaptations. If the adaptations are substitutes for one another or mutually exclusive (due to, e.g., energetic or morphological constraints), those who adopt one adaptation will not adopt another and vice versa. If one only tests for one particular adaptation among those who receive the input, then one may reject PAR even when it is valid. In effect, if organisms that receive a developmental input can adopt a substitute phenotypic adaptation ($P'$) than the one examined in Figure 3, and that alternative adaptation is as effective as $P$, then fitness without $P$ will be the same as with $P$. This will also manifest in a non-significant

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

interaction. In short, any phenotypic test that hitches itself to a specific adaptation carries the implicit assumption that there are not other, alternative adaptations.

## Variants of PAR

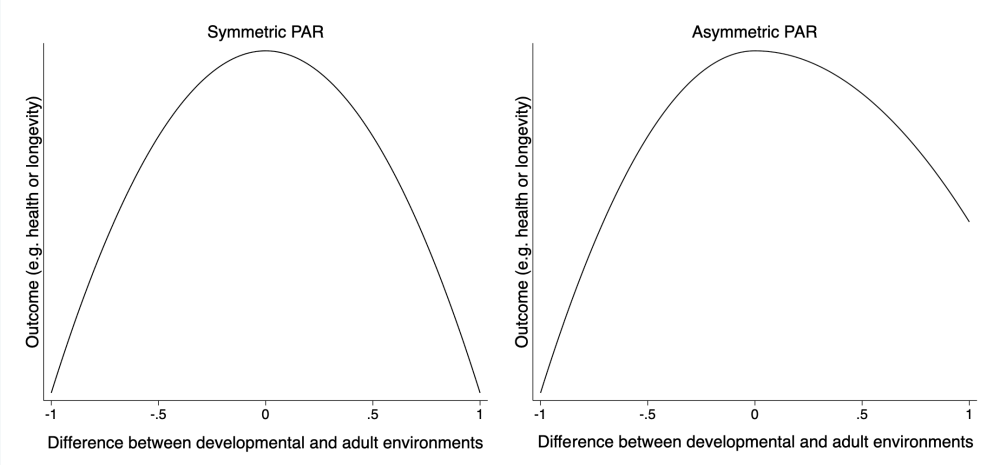
In the main text, we discussed the most prominent alternative variant of the PAR model: internal (as opposed to external) PAR. Here we discuss two other possible variants that one could define. The first variant would allow an organism to predict a future environment that differs from its developmental environment (contra steps 1 to 2 of the causal chain that defines PAR in the main text). For example, humans may predict that our future environment will be warmer; a lynx born during a season with abundant hares to eat might predict (perhaps not consciously) an eventual crash in the hare population. This is simply $E(e_1) \neq e_0$. More specific variants (i.e., that adult environment is going to be twice as good as developmental environment, or twice as bad) could easily be defined as well.

A second variant concerns whether overly optimistic and overly pessimistic predictions have differential effects [43]. We will define "symmetric mismatch" as a version of mismatch wherein positive and negative values of $\Delta e = e_1 - e_0$ have the same (negative) effect on outcomes:

$$\left.\frac{\partial y}{\partial \Delta e}\right|_{\Delta e>0,|\Delta e|=k} = \left.\frac{\partial y}{\partial \Delta e}\right|_{\Delta e<0,|\Delta e|=k} < 0. \tag{32}$$

A visualization of the prediction this version would generate can be found in the left panel of Figure 4. By contrast, we will define "asymmetric mismatch" as a version of mismatch wherein positive and negative mismatch each have deleterious effects on outcomes, but equal-magnitude positive and negative errors do not have the same magnitude of deleterious effects on outcomes. A visualization of this prediction can be found in the right panel of Figure 4. Note that these prediction errors could theoretically be asymmetric in the opposite direction; our choice to depict negative predictions as more deleterious than positive ones is arbitrary. The reason that these variants are important is because they affect whether the DC, PAR, and AEQ hypotheses can be empirically distinguished, a topic we discuss in the section below.

Figure 4: Visual depictions of the concept of symmetrical and asymmetrical predictive adaptive responses



bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Conditions under which it is possible to simultaneously test the developmental constraints (DC), predictive adaptive response (PAR), and adult environmental quality (AEQ) hypotheses

In "Conceptual issues in testing models" in the main text, we explain why it is generally impossible to test all three hypotheses (DC, PAR, and AEQ) simultaneously. This is because any data generating process cannot manipulate subjects' developmental environments, adult environments and the change in these environments independently; change in any one of these necessarily changes one of the others. Hence one cannot separately test for effects of any one of these environmental variables while holding the other environmental variables constant.

However, there is one exception (that we have identified) to this general principle. This exception relies on differences between what we refer to as "symmetric" and "asymmetric" PAR (see the section above, "Variants of PAR," for more details). In symmetric PAR, positive and negative differences between developmental and adult environments both have the same magnitude and direction of effects on adult outcomes (see the left panel of Figure 4); in asymmetric PAR, one direction of environmental change (i.e., either positive or negative) more strongly influences adult outcomes than the other (see, e.g., the right panel of Figure 4). If we assume that PARs are symmetric and that the AEQ hypothesis is false, the data should show that positive and negative changes in environment each have identical negative effects on adult outcomes, as described in (32) and the left panel of Figure 4. However, if the AEQ hypothesis is true, then—even in the presence of symmetric PARs—the data should show that positive changes in the adult compared to the developmental environment have less negative effects on adult outcomes than negative changes in environment do:

$$\left.\frac{\partial y}{\partial e_1}\right|_{\Delta e<0,|\Delta e|=k} < \left.\frac{\partial y}{\partial e_1}\right|_{\Delta e>0,|\Delta e|=k} < 0. \tag{33}$$

Thus, if we assume that symmetric PAR is true, then evidence of (33) is evidence for the AEQ hypothesis. However, if we are not a priori sure that PAR is symmetric, then evidence of (33) could be consistent with either (a) symmetric PAR plus AEQ or (b) asymmetric PAR where positive changes in the environment have less negative effects on adult outcomes than negative changes do ("left asymmetric PAR" as in right panel of Figure 4) and no AEQ. Indeed, one cannot even rule out the possibility of mild left asymmetric PAR that is made to appear *more* asymmetric by AEQ.

Building on this logic, it should also be clear that data ostensibly consistent with (33) (left panel of Figure 4) are also hard to interpret. If it is assumed that PARs are symmetric, then this can be interpreted as evidence against AEQ. But if it is assumed that PARs are right asymmetric, i.e., positive changes in environment are more harmful than negative changes in environment (e.g., the mirror image of the right panel of Figure 4), then this is evidence for the AEQ hypothesis. If one is not sure a priori whether PARs have symmetric effects or not, then one cannot rule in or out the AEQ hypothesis.

## Baboon data

In order to ensure that we selected covariate values that are representative of a real-world biological system, we summarized long-term data from the Amboseli Baboon Research Project and used the resulting $\Delta e$ values in our simulations (i.e., the value of the difference between developmental and adult environments). The subjects were wild female savannah baboons (primarily *Papio*

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

*cynocephalus* with some naturally occurring admixture from neighboring *Papio anubis* populations) living in the Amboseli ecosystem in southern Kenya. This baboon population has been studied on a near-daily basis since 1971 by the ABRP, which collects longitudinal demographic, ecological, life history, and behavioral data on individually known animals [1]. The data we summarized to obtain values used in the simulations spanned January 1980 to March 2021.

Like many other species including humans, baboons can experience a variety of developmental environments [32, 62, 56, 64]. We summarized a major source of variation in the developmental environment for baboons: dominance rank. Female baboons form strong, stable dominance hierarchies that determine access to important resources [37]. Female infants "inherit" their ranks from their mothers via a system of youngest ascendancy, where the infant becomes dominant over any older maternal sisters at birth. The functional consequence of the youngest ascendancy system is that female baboons can reasonably predict that they will hold a dominance rank similar to the one their mother held when they were born, when they themselves are adults. However, due to events such as group fissions and matriline overthrows [1], some animals occupy dominance ranks in adulthood that are not similar to the dominance rank their mother held when they were born. Since there is variation in both developmental environment (some baboons are born to low-ranking mothers, and some to high-ranking ones) and the how well the developmental and adult environments match (some females will hold nearly the same rank their whole lives, and some will experience change), rank would be an appropriate variable for testing the DC and PAR hypotheses.

We used relative (i.e., proportional) dominance ranks, which represent the proportion of adult female group members that the subject in question outranks [17, 37]. For example, a dominance rank of 0.9 means that the female outranks 90% of the adult females in her group. Each animal's rank was determined based on the outcomes of all observed, decided agonistic interactions between adult females. Trained observers, who could recognize individual animals via differences in morphological characteristics, recorded the identities of individuals participating in agonistic encounters and the outcome of each of these encounters. If one animal behaved submissively while the other was either aggressive or remained neutral, the interaction was recorded as "decided" with respect to the rank relationship between the two interacting animals. Any interactions without a clear outcome (e.g., where both animals displayed submissive signals) were excluded from dominance rank calculations.

For rank during development, each female subject was assigned the rank her mother held in the month the subject was born. In order to summarize the difference between developmental and adult rank, we needed time window(s) during adulthood in which to capture rank. We eliminated periods of rank instability (e.g., when groups were going through fissions, during which rank can be difficult to determine), then aggregated information about months in which females were cycling (i.e., months in which females were not either pregnant or nursing). A data set like this would be suitable for testing if developmental environments, or the difference between developmental and adult environments, predicted (for example) the chances that a female would successfully conceive, or the number of mating consortships she participated in. The difference between a female's mother's rank the month she was born and the rank she held during the cycling month(s) in question was the difference between her developmental and adult environment ($\delta e$). This resulted in a data set that contained information about 7,661 months for 281 individual female baboons. After weighting each baboon equally, the mean difference between developmental and adult rank was -0.03 (SD=0.21, range=-0.95-1). Data sets constructed across the course of pregnancies or lactation resulted in very similar values (e.g. for pregnancies mean =-0.02, SD=0.22, range=-0.92-1). We chose to use the summarized values obtained from cycling months since it was the largest of the data sets.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

## Further details on the simulations

Here we provide further details about our simulation strategy. Some of this information is repeated from the main text, but we provide it in both places to make it easier to follow. Our simulation strategy had five steps.

1. We prepared a set of possible "realities" where the PAR hypothesis was either true or false. We started with the 3rd-order polynomial for $y_1 = f(e_0, |\Delta e|)$, which generated an equation with 10 coefficients[19].

$$\begin{aligned} y_1 = \alpha + \alpha_0 e_0 + \alpha_d |\Delta e| + \alpha_{00} e_0^2 + \alpha_{d2} |\Delta e|^2 + \alpha_{0d} e_0 |\Delta e| \\ + \alpha_{03} e_0^3 + \alpha_{d3} |\Delta e|^3 + \alpha_{02d} e_0^2 |\Delta e| + \alpha_{0d2} e_0 |\Delta e|^2 \end{aligned} \tag{34}$$

   Then, we created a grid of points in a 10-dimensional space. Each dimension took 5 possible values (-1,-0.5,0,0.5,1), so the grid had $5^{10}$ points. This grid represents possible realities defined by different values of coefficients on up to 3 powers of $(e_0, |\Delta e|)$, i.e., each reality is a 10 x 1 vector $\theta \in \Theta = \{\alpha, \alpha_0, \alpha_d, \alpha_{02}, \alpha_{d2}, \alpha_{0d}, \alpha_{03}, \alpha_{d3}, \alpha_{02d}, \alpha_{0d2}\}$. We restricted the range of coefficients to $[-1, 1]$ because these are consistent with either binary outcomes or continuous, bounded outcomes, as well as environmental variables that range from 0 to 1. We chose 5 values in the [-1,1] range so that the grid was not so dense that there were too many possible realities to simulate.

2. We pruned realities that generated outcomes outside the range of 0 to 1 for any feasible value of $\mathbf{x} = (e_0, |\Delta e|)$, i.e., $e_0 \in [0, 1]$ and $|\Delta e| \in [0, 1]$. To implement this, we allowed $\mathbf{x} = (e_0, |\Delta e|)$ to each take 4 possible values (0, 1/3, 2/3, 1), and calculated the 16 possible outcomes that resulted from each coefficient vector $\theta$. If any outcomes were outside $[0, 1]$, we rejected that coefficient set. This left 130,201 "feasible" coefficient sets or realities ($\Theta^f \in \Theta$).

3. We identified those realities in which PARs and DC were true. We first started with the same 16 possible combinations of $(e_0, |\Delta e|)$ in the last step. For each feasible reality (i.e., coefficient vector surviving the previous step), we then calculated the derivative of (34) with respect to $e_0$ and then with respect to $|\Delta e|$ at each of the 16 values of $(e_0, |\Delta e|)$. If, for a given feasible reality $\theta$, the derivative with respect to $e_0$ was on average positive across those 16 values, DC was said to have been true in that reality. If, for a given feasible reality $\theta$, the derivative with respect to $|\Delta e|$ was on average negative across those 16 values, PAR was said to have been true in that reality.

4. We created a simulated data set for each pruned reality $\theta \in \Theta^f$. Our simulated data contained a set of 2000 predicted outcomes $\tilde{y}_i$ generated by the 3rd-order polynomial function:

$$\tilde{y}_i = y_i + v_i \tag{35}$$

   where $y_i$ is obtained from (34), the coefficients in that equation were given by the reality $\theta$, 100 independent values of the pair $\mathbf{x} = (e_0, e_1 - e_0))$ were drawn from each of 20 equally-spaced discrete points in the space $[0, 1] \times [-1, 1]$, and the error was drawn from a normal distribution with mean 0 and variance equal to $A(1-A)$, where $A = \alpha + \alpha_0 \times 0.5 + \alpha_{02} \times 0.25 + \alpha_{03} \times 0.125$, the value of outcomes at the mean of $\mathbf{x}$. This variance mimics the variance from a binomial distribution with a mean of $A$.

[19] We did not use a 4th-order polynomial because it requires 15 terms, increasing required memory and computational power dramatically.

bioRxiv preprint doi: https://doi.org/10.1101/2022.02.10.479998; this version posted February 11, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

5. We tested for PAR and DC in each reality. Specifically, we estimated an interaction regression and quadratic regression on the N=2000 simulated data set for each reality. From the regression results for each reality, we generated four test results for PARs and DC in that reality (described in Table 3 and the section "Simulations of alternative tests for PARs and DC" in the main text). We then compared these test results to the truth for each reality (determined in the prior step) to estimate the sensitivity and specificity of each test for PAR and DC. We also tested for PARs and DC in each reality using a coin flip, simulated with a Bernoulli random variable with mean 0.5. The comparison allowed us to calculate the fraction of realities where each of the four tests, along with the coin flip, generated results that differed from the ground truth (Table 3 in the main text).

The simulation code is available online at https://github.com/anup-malani/PAR/blob/d3d9588af6bb96c49f0550c94aa756e6a1261a9f/PAR_simulation_220117b.do.