VOL. 182, NO. 3 THE AMERICAN NATURALIST SEPTEMBER 2013

When Unreliable Cues Are Good Enough

Matina C. Donaldson-Matasci,[1,*] **Carl T. Bergstrom,**[2] **and Michael Lachmann**[3]

1. Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721; 2. Department of Biology, University of Washington, Seattle, Washington 98195; and Santa Fe Institute, Santa Fe, New Mexico 87501; 3. Department of Evolutionary Genetics, Max Planck Institute for Evolutionary Anthropology, Leipzig 04103, Germany

Submitted June 20, 2012; Accepted March 15, 2013; Electronically published July 12, 2013

ABSTRACT: In many species, nongenetic phenotypic variation helps mitigate risk associated with an uncertain environment. In some cases, developmental cues can be used to match phenotype to environment—a strategy known as predictive plasticity. When environmental conditions are entirely unpredictable, generating random phenotypic diversity may improve the long-term success of a lineage—a strategy known as diversified bet hedging. When partially reliable information is available, a well-adapted developmental strategy may strike a balance between the two strategies. We use information theory to analyze a model of development in an uncertain environment, where cue reliability is affected by variation both within and between generations. We show that within-generation variation in cues decreases the reliability of cues without affecting their fitness value. This transpires because the optimal balance of predictive plasticity and diversified bet hedging is unchanged. However, within-generation variation in cues does change the developmental mechanisms used to create that balance: developmental sensitivity to such cues not only helps match phenotype to environment but also creates phenotypic diversity that may be useful for hedging bets against environmental change. Understanding the adaptive role of developmental sensitivity thus depends on a proper assessment of both the predictive power and the structure of variation in environmental cues.

Keywords: bet hedging, developmental plasticity, variable environment, delayed germination, information, environmental cue.

Introduction

For species living in variable environments, often there is no single phenotype that is well suited for all conditions. Such species may evolve strategies that produce a variety of phenotypes, each performing well under different conditions. We can classify these strategies into two types. First, predictive plasticity is the development of phenotypes that match the environmental conditions in which they are found. For example, acorn barnacles develop either a bent or a conical shell shape, depending on the presence of predatory snails (Lively 1986). Second, a diversified bet-hedging strategy produces a variety of phenotypes that are not matched to the environmental conditions as a way of managing the risk associated with widespread events like drought and floods (Seger and Brockmann 1987). Desert annual plants, for example, must germinate, grow, and reproduce all within 1 year—but extreme year-to-year variation in precipitation means that many years are unsuitable. Such plants often use a strategy of delayed germination: in any given year, only a fraction of the seeds produced by one individual will germinate, reducing the risk of losing all offspring to drought (Cohen 1966).

What ecological factors affect the adaptive evolution of predictive plasticity and/or diversified bet hedging? The first important factor is the level at which environmental variation occurs: within and/or between generations. Theory shows that diversified bet hedging is useful only when environmental conditions vary from one generation to the next. This is because when conditions vary only within generations the mean fitness is the same in every generation. The strategy that is most likely to prevail is to produce the phenotype that does best on average. On the other hand, when conditions vary between generations mean fitness varies between generations as well. Which strategy prevails depends on the sequence of environmental conditions, but the most likely winner is the strategy that maximizes the geometric mean fitness across environments (Dempster 1955). This strategy may involve producing a mixture of several phenotypes. The second important factor is how well the selective environment can be predicted using cues available during development: without informative developmental cues, predictive plasticity would be impossible. Putting these two factors together, when the environment varies from one generation to the next and partially informative cues help predict that environment, we may see the evolution of integrated strategies involving both diversified bet hedging and predictive plasticity (Cohen 1967; Haccou and Iwasa 1995).

In all this research, however, a third potentially important factor has been almost completely overlooked: the level at which variation in developmental cues occurs. Pre-

---

* Corresponding author; e-mail: matina@email.arizona.edu.

Am. Nat. 2013. Vol. 182, pp. 313–327. © 2013 by The University of Chicago.
0003-0147/2013/18203-53926$15.00. All rights reserved.
DOI: 10.1086/671161

vious theoretical models of evolution in a fluctuating environment have explored the consequences of unreliability in shared cues, with no variation between individuals within a generation (but see Donaldson-Matasci 2008; Rivoire and Leibler 2011). However, like the selective environment itself, the developmental cues used to predict that environment may vary within as well as between generations. For example, some multivoltine butterflies show seasonal polyphenism: the ones that pupate in cooler parts of the year show greater wing melanism and thus stay warm. In some cases wing melanism seems to be controlled by day length during the pupal period, while in others it may be controlled by temperature during the pupal period or even some combination of the two (Hoffmann 1978). Day length is a cue that is shared by all individuals in the same generation, and in predictable environments it is a good predictor of the thermal regime that adults will experience. However, when it is wrong—as in an unseasonably cool summer—the consequences will affect the entire generation of butterflies. Pupal temperature, instead, is a cue that may vary from individual to individual, depending on the microhabitat of the cocoon. It might be misleading for two reasons: either the individual happens to be in a particularly warm or cool spot or there is a sudden change in temperature that occurs between the pupal period and the adult stage. The former source of unreliability affects just one individual, while the latter affects the entire generation; in general, we can expect a cue to have a mixture of both.

In addition, previous work has generally focused on the adaptive function of phenotypic diversity (e.g., diversified bet hedging or predictive plasticity) with relatively little discussion of the developmental mechanisms used to generate that diversity. For predictive plasticity, only one mechanism will do: the developmental process must be sensitive to environmental cues that can help predict the selective environment. For diversified bet hedging, however, phenotypic diversity may be generated in a number of ways; in Cooper and Kaplan's (1982) terminology, there are different "coins" that can be used for "adaptive coin-flipping." For example, stochastic gene expression may drive the developmental trajectory along one of several alternative pathways (McAdams and Arkin 1997). Alternatively, the developmental pathway may be sensitive to subtle differences in microclimate that are completely independent of the shared selective environment and that themselves have no direct bearing on fitness (e.g., Simons and Johnston 2006). In general, we will classify the mechanisms for generating phenotypic diversity into two groups: developmental stochasticity, which is due to intrinsic developmental noise, and developmental sensitivity, which is shaped by extrinsic environmental conditions.

In this article, we use an information theoretic approach to show that the adaptive evolution of developmental plasticity depends critically on how unreliability in developmental cues arises: from variation among individual cues within and/or between generations. Both kinds of variation make individuals less certain about what environmental conditions they will experience. However, only variation in cues between generations affects the environmental uncertainty that is shared by all individuals within a generation. It is this shared uncertainty that determines the optimal amount of diversified bet hedging and thus the long-term fitness that may be achieved. How diversified bet hedging is actually realized—via developmental stochasticity and/or sensitivity—depends instead on the amount of variation in developmental cues within generations.

An Example: Drought Prediction in Desert Annuals

Cohen's classic model of bet hedging is based on the biology of desert annual plants (Cohen 1966). Because they live in such a harsh environment and have only one chance to reproduce, these plants are particularly vulnerable to the risk of drought. Many desert annual plants avoid this problem by diversified bet hedging: even under highly controlled, ideal laboratory conditions, only a fraction of seeds will germinate, while the rest remain dormant (Freas and Kemp 1983; Philippi 1993). In good years, those seeds that do germinate are likely to reproduce successfully. In drought years, even though some seeds germinate and die, the ones that remain dormant retain the chance to perpetuate the lineage in a future year. Cohen showed that when there is no chance of survival in bad years, the optimal fraction of seeds germinating each year is equal to the fraction of good years. Thus, if severe drought occurs in about 20% of years, we would expect approximately 80% of seeds to germinate each season.

When environmental cues can help predict future environmental conditions, it may be possible to do even better (Cohen 1967). For example, early and plentiful rains during the germination season may indicate that a good growing season is likely, while late or sparse rains may indicate the opposite (Philippi 1993). Cohen showed that the optimal fraction of seeds germinating in response to a cue is equal to the conditional probability of a good year given that cue. Say that severe drought occurs in about 40% of years overall, but in years with a good germination season there is just a 20% chance of drought, while with a poor germination season there is a 60% chance of drought. In this situation we would expect about 80% of seeds to germinate in response to a good germination season but only about 40% of seeds to germinate in response to a poor germination season. Such a strategy combines elements of predictive plasticity, because seeds change their probability of germination in response to cues, and diversified bet hedging, because not all seeds will germinate even under the best conditions.

The above calculations assume that all individuals receive exactly the same cue, that is, all seeds receive exactly the same amount of rain during the germination season. However, due to the patchy nature of desert rainfall, different seeds may receive slightly different amounts of rain. Say that in years with a good germination season (i.e., years where the average seed receives early and plentiful germination rain) 20% of seeds receive poor germination rain, while in years with a poor germination season 40% of seeds nonetheless receive good germination rain. If every seed that received good germination rain were to germinate while those that received poor germination rain did not, this would produce just the same patterns of phenotypic variation as the developmental strategy described above: 80% germinate in a good germination season, while 40% germinate in a poor germination season. Critically, however, now the within-generation variation in phenotype is caused not by developmental stochasticity but by within-generation variation in the environmental cue that is received. This is particularly surprising if we think about forecasting from an individual seed's perspective. Across years, of all seeds that receive good germination rains, we can expect only two-thirds of them to escape drought (see app. A for calculations). Nonetheless, all of them germinate. On the other hand, of all seeds that receive poor germination rains, we would expect half of them to escape drought—yet none of them germinate.

This simple (if contrived) example illustrates three important points. First, variation in the cues that individuals receive means that individuals obtain less information about the shared environmental conditions they will experience. Second, strong sensitivity to such cues can create within-generation phenotypic diversity, spreading the risk among individuals and obviating the need for developmental stochasticity. Third, the phenotypic distribution created by the optimal strategy in response to cues with individual variation may in some cases be the same as for the optimal strategy in response to cues without individual variation. Together, these observations suggest that within-generation variation in cues is not necessarily disadvantageous, even though it reduces individuals' abilities to predict fitness-relevant aspects of the environment. In the following section, we develop a model of developmental plasticity in an uncertain environment. The model is designed to highlight the role of variation in developmental cues within and between generations and will allow us to explore these points more thoroughly.

A Model of Developmental Plasticity in Response to Unreliable Cues

We model a species with discrete, nonoverlapping generations inhabiting a variable environment. Population-level risk, in which the environmental state varies from one generation to the next, affects every individual in the population. In each generation, the state of the selective environment e is drawn from a distribution $\Pr(e)$ over a discrete set of possible environmental states. An individual's phenotype is determined by its developmental trajectory, which is shaped by its genotype, the conditions it experiences during development, and developmental noise. Each individual observes a developmental cue c and develops a phenotype x, which remains fixed over its lifetime. For simplicity, we assume that each individual produces one of several discrete phenotypes. We model this by assuming that the individual's genotype encodes a developmental strategy g, which specifies the chance of developing each phenotype in response to each developmental cue. The developmental strategy may thus involve developmental stochasticity, developmental sensitivity to the cue, or some combination of the two.

The developmental cue c received by an individual conveys only partial information about the environmental state e. If developmental cues within a generation are strongly correlated, a misleading cue might cause many individuals to develop the wrong phenotype and die without being able to reproduce (shared uncertainty). On the other hand, if developmental cues are weakly correlated, even if some individuals receive misleading cues others will not, so some individuals will always develop the correct phenotype (individual uncertainty). We will separate these two types of uncertainty by identifying a global predictor q—some feature of the shared environment that imperfectly predicts the selective environment but that itself does not directly affect fitness. The developmental cue received by each individual in turn imperfectly reflects the global predictor but is otherwise independent of the developmental cues others receive. Shared uncertainty is the result of a mismatch between the selective environment and the global predictor, while individual uncertainty is the result of a mismatch between the global predictor and the individual's developmental cue. In the example of desert annual plants, the global predictor would be the overall quality of the germination season, as determined by the amount and timing of germination rain received on average. The selective environmental state would be the overall quality of the growing season, and the individual developmental cue would be the amount of rain an individual seed receives during the germination season.

Practically, it may not always be easy to identify an environmental characteristic that acts as a global predictor—a feature shared by a whole generation that is correlated with the environmental state and upon which all individual developmental cues are based. In appendix B, we argue that under fairly general conditions it is nonetheless possible to identify one mathematically if the distribution of developmental cues within and between generations is known. In fact, the global predictor may be thought of as indicating what proportion of all individuals receive each possible cue. For example, if there are only two different cue types, then q could be a real number between 0 and 1 and indicate the probability of receiving the first cue type. However, if there are really discrete types of environments, as we assume (e.g., "good" vs. "bad" years), and developmental cues are reflecting that structure, we might expect to see discrete types of generations when characterized in terms of the proportions of individuals receiving each type of cue (e.g., either 75% of individuals receive a cue suggesting a good year or only 25% do). Even if there are not discrete types of generations in this sense, we argue in appendix B that for the purposes of finding the optimal strategy we can cluster different types of generations together into k different groups (if there are k different possible developmental cues) and treat them as if they were discrete. For this reason, we will assume throughout the main text that the global predictor is drawn from a discrete set of possibilities of the same size as the set of possible developmental cues.

Mathematically, the whole process can be described as follows (see fig. 1). In each generation, an environmental state e is chosen from the distribution $\Pr(e)$. Then, a global predictor q is chosen from the conditional distribution $\Pr(q|e)$. These two variables are shared by the entire population. Next, for each individual in the population an individual cue c is chosen from the conditional distribution $\Pr(c|q)$; the cue for each individual is chosen independently from the cues for all other individuals. Finally, for each individual in the population a phenotype x is chosen from the conditional distribution $g(x|c)$, which characterizes the developmental strategy g.

The fitness of each individual is determined by its phenotype x and the shared environmental state e according to the fitness function $f(x, e)$. In the main body of this article, we will assume for simplicity that each phenotype x survives in just one environment, e_x; that is, $f(x, e) > 0$ if and only if $e = e_x$. In appendix C, we extend the results to more general fitness functions using the framework introduced in Donaldson-Matasci et al. (2008). We also assume that any differences between individuals in the developmental cues they receive do not reflect fitness-relevant differences in their environment. That is, we assume that fitness is directly influenced only by e, not by c. In "Discussion," we will consider the implications of this assumption.

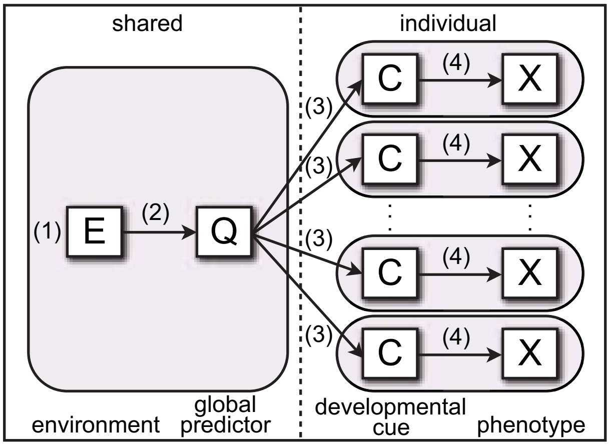

Figure 1: Stochastic model of developmental plasticity in response to individually variable cues: (1) the environmental state e is drawn from the distribution $\Pr(e)$; (2) the global predictor q is drawn from the conditional distribution $\Pr(q|e)$, depending on the environmental state e; (3) for each individual, the developmental cue c is drawn (independently of all other individual cues) from the conditional distribution $\Pr(c|q)$, depending on the global predictor q; and (4) for each individual, the phenotype x is drawn from the conditional distribution $g(x|c)$, determined by its developmental strategy g and its developmental cue c.

In a temporally varying environment, the genotype that is most likely to fix over the long term is the one whose expected long-term growth rate is greatest (as long as there is no interaction between phenotypes, e.g., frequency-dependent selection). A good estimator of the long-term growth rate of a genotype is the mean log fitness, or, equivalently, the log of the geometric mean fitness (Dempster 1955; Cohen 1966; Cooper and Kaplan 1982). We write the expected long-term growth rate as the mean (over generations) of the log of the mean fitness (within generations):

$$\begin{aligned} r(g) &= \sum_{e,q} \Pr(e)\Pr(q|e)\log_2 \bar{f}(g|e,q) \\ &= \sum_{e,q} \Pr(e)\Pr(q|e) \qquad (1) \\ &\quad \times \log_2\left(\sum_c \Pr(c|q)\sum_x g(x|c)f(x,e)\right). \end{aligned}$$

(Note that while any base could be used for the logarithm, we will use base 2 because that is the standard in information theory.) The mean within-generation fitness of a genotype is determined by the distribution of the phenotypes it creates, which depends not only on its developmental strategy g but also on the distribution of developmental cues, as given by $\Pr(c|q)$. The resulting growth rate is linear in $\Pr(q|e)$ but not in $\Pr(c|q)$ (see eq. [1]).

This means that the two sources of unreliability—variation in cues between individuals and year-to-year variation in cue distributions—have fundamentally different effects on the optimal strategy and on the maximum attainable growth rate.

An Information-Theoretic Perspective

The field of information theory was developed by Claude Shannon as a way to quantify the efficiency of information transfer over unreliable channels (Shannon 1948). We will use it to quantify the relationship between cue and environment (environmental uncertainty) and the relationship between cue and phenotype (developmental strategy).

The fundamental measure in information theory is entropy, which describes the amount of uncertainty about a, the outcome of some random process A:

$$H(A) = -\sum_{a} \Pr(a) \log_2 (\Pr(a)). \tag{2}$$

Entropy depends on the probabilities of different outcomes; it is highest when there are many different possibilities and each of them is equally likely. One "bit" of entropy corresponds to two different possible outcomes that are equally likely, while n bits of entropy correspond to 2^n equally likely outcomes. If we cannot observe the outcome of A directly but can instead observe the outcome of a linked random process B, our uncertainty about a may be reduced. To measure the uncertainty that remains once the outcome of B has been observed, we can calculate the conditional entropy:

$$H(A|B) = -\sum_{b} \Pr(b) \sum_{a} \Pr(a|b) \log_2 (\Pr(a|b)). \tag{3}$$

The mutual information between the two random processes is the difference between the entropy and the conditional entropy:

$$\begin{aligned} I(A; B) &= H(A) - H(A|B) \\ &= \sum_{b} \Pr(b) \sum_{a} \Pr(a|b) \log_2 \left(\frac{\Pr(a|b)}{\Pr(a)} \right). \end{aligned} \tag{4}$$

Mutual information is a measure of how much the uncertainty about the outcome of one random process is reduced by observing the outcome of another process. It is maximal when knowing one outcome allows the other to be predicted perfectly; it is 0 when knowing one outcome does not change our prediction about the other at all.

We will first use the relationship between uncertainty and information to quantify the relationship between the selective environment, e, and the developmental cue, c:

$$H(E) = H(E|C) + I(E; C). \tag{5}$$

This says that the total uncertainty about the environmental state can be divided into the uncertainty that remains once an individual's developmental cue has been observed and the information the cue contains about the environmental state.

Similarly, we can quantify the relationship between the selective environment, e, and the global predictor, q:

$$H(E) = H(E|Q) + I(E; Q). \tag{6}$$

This says that the total uncertainty about the environmental state can be divided into the uncertainty that would remain if the state of the global predictor could be directly observed and the information the global predictor contains.

We can also use information theory to describe the developmental strategy. The entropy $H(X)$ represents our uncertainty about which phenotype will be produced—a measure of phenotypic diversity similar to the widespread use of Shannon entropy in ecology as a measure of species diversity (e.g., Pielou 1969). We quantify the phenotypic diversity according to our uncertainty about which phenotype an individual develops, x, and how that relates to the developmental cue it receives, c:

$$H(X) = H(X|C) + I(X; C). \tag{7}$$

This says that the total uncertainty about the phenotype can be divided into the uncertainty remaining once the developmental cue is known and the information the developmental cue gives about which phenotype will develop. The former, $H(X|C)$, is a measure of the amount of developmental stochasticity, while the latter, $I(X; C)$, is a measure of the amount of developmental sensitivity to the cue.

Similarly, we can also describe the consequences of the developmental strategy for phenotypic diversity produced within and between generations. To do this, we quantify the relationship between the phenotype, x, and the global predictor, q:

$$H(X) = H(X|Q) + I(X; Q). \tag{8}$$

This says that the total uncertainty about the phenotype can be divided into the uncertainty remaining once the global predictor is known and the information the global predictor gives about which phenotype will develop. We will argue that the former, $H(X|Q)$, is a measure of the amount of diversified bet hedging, while the latter, $I(X; Q)$, is a measure of the amount of predictive plasticity.

Model Results

To solve for the optimal developmental strategy, we must find the conditional probability distribution $g(x|c)$ that maximizes the long-term growth rate given in equation

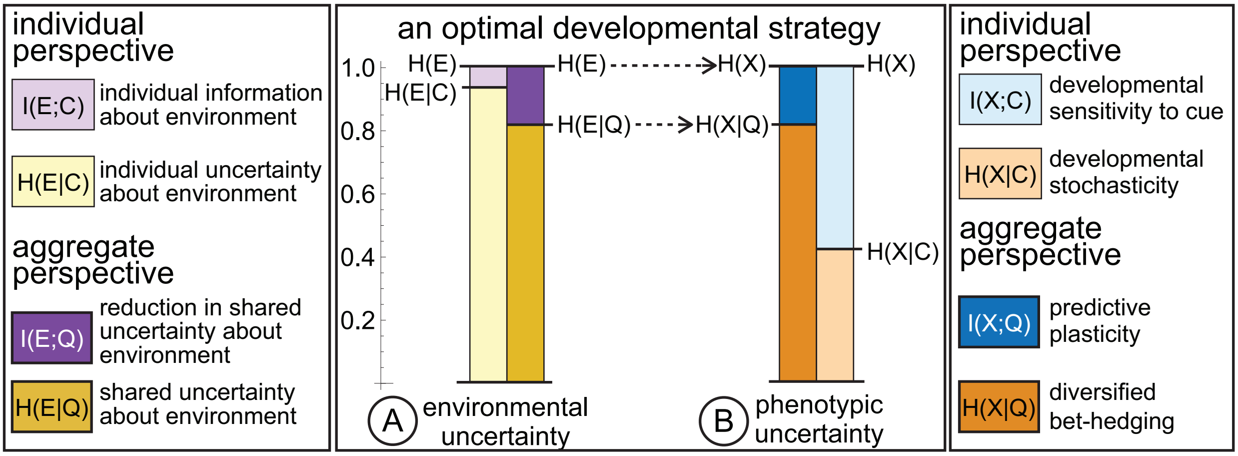

Figure 2: Relationship between environmental uncertainty and information (*A*) and phenotypic diversity produced by the optimal strategy (*B*), as viewed from the individual perspective (in response to the developmental cue) and from the aggregate perspective (in response to the global predictor). In the text, we show that taking the aggregate perspective allows us to see the connection between environmental uncertainty and adaptive phenotypic diversity: $H(E) = H(X)$ and $H(E|Q) = H(X|Q)$. Parameter values: $\Pr(e_1) = \Pr(e_2) = 0.5$, $\Pr(q_1|e_1) = \Pr(q_2|e_2) = 0.75$, and $\Pr(c_1|q_1) = \Pr(c_2|q_2) = 0.8$.

(1). The solution is best understood by applying the principle of proportional betting: when several outcomes are possible and only betting on the correct outcome yields any payoff, the optimal scheme for long-term investment is to place money on each outcome according to its probability of occurring—regardless of the associated payoff (Kelly 1956). For example, in Cohen's classic model of diapause in desert annual plants, the lineage that maximizes its long-term growth rate is the one whose conditional probability of germination in response to a cue equals the conditional probability of a good year (Cohen 1967). This simple, intuitive result depends on two special assumptions: first, that there is no chance of reproduction in a bad year; and second, that all individuals receive the same cue. Our focus in this article is relaxation of the second assumption; in appendix C, we show that our results still hold when the first assumption is relaxed as well.

When developmental cues vary from individual to individual, the principle of proportional betting does not directly apply. However, we can still solve for the optimal response to the global predictor as if individuals could observe it directly and then ask whether that response could still be achieved by a developmental strategy that was sensitive only to individual developmental cues. According to the principle of proportional betting, the optimal strategy in response to the global predictor, $\hat{g}(x|q)$, should match the probability of each phenotype to the conditional probability that its environment occurs:

$$\hat{g}(x|q) = \Pr(e_x|q) = \frac{\Pr(q|e_x)\Pr(e_x)}{\Pr(q)} \quad \text{for all } x, q. \tag{9}$$

When developmental cues vary from one individual to the next, the response to the global predictor depends both on within-generation variation in the cues and on the developmental strategy in response to those cues. We call the resulting aggregate response to the global predictor the effective strategy:

$$\bar{g}(x|q) = \sum_c g(x|c)\Pr(c|q). \tag{10}$$

A developmental strategy $g(x|c)$ that can achieve the optimal effective strategy is the optimal developmental strategy. Combining equations (9) and (10), we get the conditions for the optimal developmental strategy, $\hat{g}(x|c)$:

$$\sum_c \hat{g}(x|c)\Pr(c|q) = \frac{\Pr(q|e_x)\Pr(e_x)}{\Pr(q)} \quad \text{for all } x, q. \tag{11}$$

Note that this criterion may not always be achievable, given the requirement that $\hat{g}$ represent a feasible developmental strategy.

Recasting equation (9) in the information-theoretic framework established in the section titled "An Information-Theoretic Perspective," we see that, if the optimal effective strategy can be achieved, $H(X|Q) = H(E|Q)$. This also means that $H(X) = H(E)$ and that

$$I(X;Q) = I(E;Q). \tag{12}$$

That is, the phenotypic diversity that is produced within and between generations as an indirect response to the global predictor mirrors the uncertainty and information about the environment given the global predictor (see fig. 2). This means that, by looking at both environmental

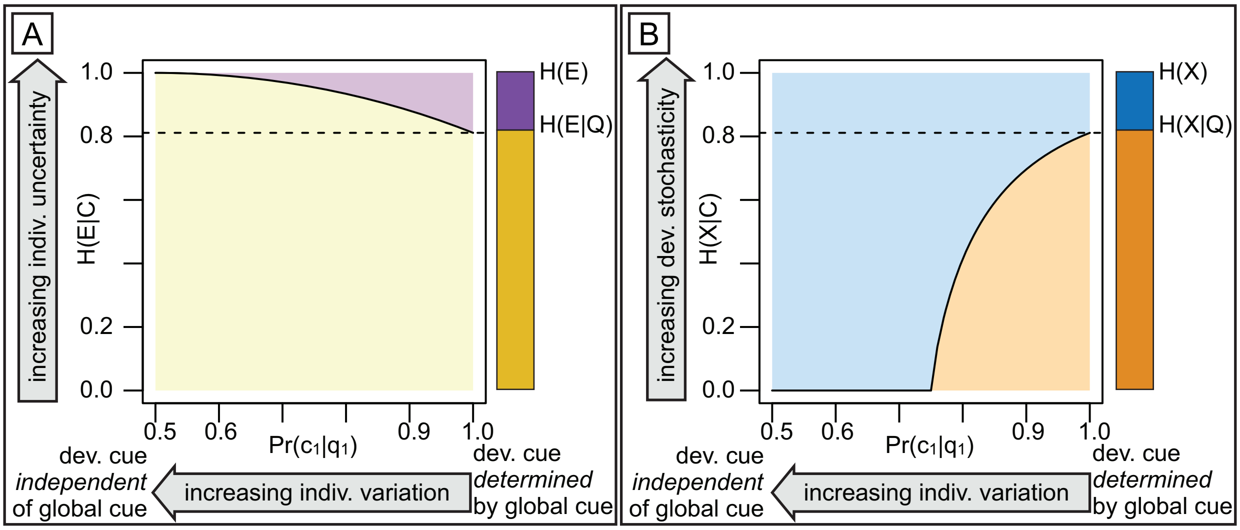

Figure 3: Individual variation in developmental cues increases uncertainty about the environmental state: $H(E|C) > H(E|Q)$, except when the developmental cue is completely determined by the global predictor (*A*). However, it also decreases the optimal amount of developmental stochasticity: $H(X|C) < H(X|Q)$, again except when the developmental cue perfectly reflects the global predictor (*B*). Note that if there is too much within-generation variation in the cue, the optimal developmental strategy is to use no stochasticity at all (as shown here, for $\Pr(c_1|q_1) = \Pr(c_2|q_2) < 0.75$). Both panels: $\Pr(e_1) = \Pr(e_2) = 0.5$ and $\Pr(q_1|e_1) = \Pr(q_2|e_2) = 0.8$; the *X*-axis shows the value of $\Pr(c_1|q_1) = \Pr(c_2|q_2)$, as they vary together.

uncertainty and phenotypic diversity from the aggregate perspective (how they relate to the global predictor Q), we can see the connection between the environmental characteristics and the optimal developmental strategy.

How can we interpret this aggregate perspective on the developmental strategy? As we saw in equation (1), it is the aggregate response to the global predictor that determines the long-term growth rate and is therefore the metric by which adaptive fit can be determined. For this reason, we argue that it is appropriate to measure diversified bet hedging and predictive plasticity at the level of the effective strategy. Thus, we use $H(X|Q)$, the uncertainty about the phenotype once the global predictor is known, to measure the amount of diversified bet hedging and $I(X; Q)$, the amount of information that the global predictor provides about the phenotype, to measure the amount of predictive plasticity.

Here we have shown that the optimal amount of predictive plasticity is exactly equal to the amount of information the global predictor provides about the environment (see eq. [12]). However, this strong equivalence depends on the assumption that each phenotype survives in just one environment. In appendix C, we show that even when that assumption is relaxed, it is nonetheless the case that the relationship between the environmental state and the global predictor determines the optimal balance of diversified bet hedging and predictive plasticity.

Individuals Receive Less Information about the Environment

How much information does an individual gain about the environment from a developmental cue? How does individual variation in developmental cues affect the amount of information they contain? To answer this, we need to compare the amount of information in the developmental cue, $I(E; C)$, to the amount of information in the global predictor, $I(E; Q)$. For the example shown in figure 2, the global predictor provides 0.189 bits of information about the environment ($\Pr(e_1|q_1) = \Pr(e_2|q_2) = 0.75$ and $\Pr(e_1|q_2) = \Pr(e_2|q_1) = 0.25$), while the developmental cue provides only 0.066 bits of information ($\Pr(e_1|c_1) = \Pr(e_2|c_2) = 0.65$ and $\Pr(e_1|c_2) = \Pr(e_2|c_1) = 0.35$). In figure 3*A*, we show an example where the developmental cue varies from being completely independent of the global predictor (i.e., having maximal within-generation variation) to being completely dependent on the global predictor (i.e., having no within-generation variation). Except when there is no variation within generations at all, the developmental cue always leaves an individual with more environmental uncertainty than the global predictor would.

In fact, it is generally true that the global predictor carries more information about the environment than the developmental cue does:

$$I(E;Q) \geq I(E;C). \tag{13}$$

This is because the developmental cue depends on the environment only via the global predictor; that is, if we knew the global predictor, then knowing the developmental cue would not give any additional information about the environment (see fig. 1). Under these conditions, we can apply a fundamental theorem of information theory known as the data processing inequality (see, e.g., Cover and Thomas 1991). This theorem tells us that only when the developmental cue is exactly equal to the global predictor—that is, there is no variation between individuals in the cues received—do they provide the same amount of information about the environment. Otherwise, the developmental cue must provide less information.

The Optimal Strategy Is More Sensitive to Developmental Cues

We have seen that variation between individuals in developmental cues reduces the amount of information the cues transmit, making it harder to predict environmental conditions. One might expect individuals to pay more attention to more reliable cues and less attention to less reliable cues. To examine this, we need to see how individual variation in cues affects the optimal balance of developmental sensitivity and stochasticity in response to those cues. For the example shown in figure 2, the optimal strategy produces 1 bit of phenotypic diversity (i.e., an equal mixture of the two phenotypes) via a combination of 58.6% sensitivity and 41.4% stochasticity ($\Pr(x_1|c_1) = \Pr(x_2|c_2) = 0.917$ and $\Pr(x_1|c_2) = \Pr(x_2|c_1) = 0.083$; $H(X|C) = 0.414$ and $I(X;C) = 0.586$). In contrast, if there were no variation in the environmental cues (i.e., individuals could directly observe the global predictor), the optimal strategy would be to produce the same total amount of phenotypic diversity but via only 18.9% sensitivity and 81.1% stochasticity. This is just the opposite of our expectation: the optimal strategy is in fact to be less stochastic and more sensitive to a variable and thus less reliable cue. For the example shown in figure 3*B*, we see that this holds true for any amount of individual variation between developmental cues.

To show this more generally, we need to compare the optimal amount of developmental sensitivity, $I(X;C)$, to the amount that would be optimal if there were no variation between developmental cues. We have already shown that the optimal amount of predictive plasticity, $I(X;Q) = I(E;Q)$, is fixed regardless of the amount of variation in developmental cues (see eq. [12]). To see the relationship between $I(X;C)$ and $I(X;Q)$, we can again use the data processing inequality. The phenotype depends on the global predictor only indirectly, via the developmental cue (see fig. 1). This means that the global predictor generally provides less information about the phenotype than the developmental cue does:

$$I(X;Q) \leq I(X;C). \tag{14}$$

These quantities are equal only when there is no individual variation between cues, so the developmental cue is a perfect reflection of the global predictor and they both provide the same amount of information about the phenotype. This means that as long as there is some individual variation in developmental cues, the optimal amount of developmental sensitivity exceeds the optimal amount of predictive plasticity.

The Fitness Value of Information Stays the Same

To measure the value of using a developmental cue, we can compare the fitness for the optimal developmental strategy in response to that cue to the fitness for the optimal strategy without it. This indicates the cue's potential to increase fitness—in our case, the long-term growth rate of a lineage. We have previously shown that, for cues that are shared by all individuals in a generation, the amount of information the cue provides about the environment is an upper limit on the fitness value it confers (Donaldson-Matasci et al. 2010).

What happens if we add variation in the developmental cues within generations—how does the value of a cue relate to the amount of information it carries about the environment? In general, adding within-generation variation actually has no effect on its value at all, as long as the amount of variation is not too large. This is because whenever the conditions for equation (11) are met, the optimal effective strategy in response to the global predictor—proportional betting—can actually be achieved. This means that even though individuals do not observe the global predictor directly, they manage to do just as well as if they could. In such cases, the value of the cue does not depend on how much information individuals actually have, $I(E;C)$; it depends on how much information they would have if they could directly observe the global predictor, $I(E;Q)$.

What happens if within-generation variation in developmental cues is so high that proportional betting is impossible? We know that proportional betting is optimal if it can be achieved, so any other strategy must be strictly worse. The information provided by the global predictor thus places an upper limit on how much the growth rate can be improved—and as long as within-generation variation is not too high, that limit can in fact be achieved.

It may seem surprising that the value of using a cue is determined by how much information the global predictor provides about the environment, $I(E;Q)$, even though that is more information than individuals actually have. The reason is this: the value of using a cue lies in its ability to reduce the amount of diversified bet hedging that the optimal strategy requires (Donaldson-Matasci et al. 2010). Diversified bet hedging is driven by uncertainty about the environment that is shared by all individuals in the population. The global predictor, because it is shared by all, reduces shared uncertainty by an amount $I(E; Q)$ (see fig. 2*A*). Adding variation in the cue that is not shared while increasing the uncertainty an individual has about the environment does not affect the shared uncertainty—so it does not affect the optimal amount of diversified bet hedging. However, it does affect the means by which diversified bet hedging is achieved. When variation in cues occurs within generations, a combination of developmental stochasticity and sensitivity to cues is used to generate the right amount of phenotypic diversity (see fig. 2*B*).

Discussion

The importance of the distinction between environmental variation within and between generations is well recognized where fitness functions are concerned (e.g., Seger and Brockmann 1987; Frank and Slatkin 1990; Moran 1992; Robson 1996). It stands to reason that, for developmentally plastic organisms, distinguishing within- and between-generation variation in developmental cues might be just as important. A number of authors have discussed the connection between information theory and fitness in a fluctuating environment, showing that the fitness value of a shared environmental cue is in many cases equivalent to the amount by which it reduces uncertainty about shared environmental conditions (Cohen 1967; Haccou and Iwasa 1995; Bergstrom and Lachmann 2004; Donaldson-Matasci et al. 2010). In this article, we extend that connection to include developmental cues that are not shared (see also Rivoire and Leibler 2011). When all individuals in the same generation receive the same misleading cue, the uncertainty created by any error in the cue is shared uncertainty. It is this kind of uncertainty that determines how a developmental strategy should invest in different phenotypes within and between generations. On the other hand, if some individuals receive a misleading developmental cue while others do not, this creates uncertainty about the environment at the individual level—and diversified bet hedging does not help with this type of uncertainty. The value of a developmental cue lies therefore not in how much it reduces an individual's uncertainty about the environment it will encounter but rather in how much it reduces shared uncertainty about the environment. This reduction in shared uncertainty determines how much the optimal strategy can reduce the amount of diversified bet hedging, which in turn determines the increase in long-term growth rate (or geometric mean fitness) that can be achieved by using the cue.

In our model, we have discussed developmental cues that provide information only about shared environmental conditions, not about individual variation in the selective environment. In the example of butterfly wing melanism, the individual variation in pupal temperature reflects differences in cocoon placement—something that will not affect the temperature regime experienced by the adult butterfly. However, in other cases differences between individuals in developmental cues may reflect real differences in the selective environment that those individuals experience. For example, tadpoles developing in temporary pools may metamorphose late, developing into large, high-fitness adults, or metamorphose early, to evade the risk that the pool dries up before metamorphosis is complete. Some species living in highly variable environments may adjust their development rate in response to cues indicating pool drying, such as increased salinity or reduced swimming volume (Denver et al. 1998; Richter-Boix et al. 2006). These cues provide an individual not only with information about the warm, dry weather conditions affecting all other tadpoles in the same generation but also with information that its pool in particular is drying up. Both of these would suggest that accelerating metamorphosis would be a good strategy, but the former is information about shared environmental variation, while the latter is information about individual environmental variation. In general, it is only uncertainty about shared environmental conditions that drives the adaptive evolution of diversified bet-hedging strategies. We have shown that developmental cues that vary within generations are more valuable than would be expected from the amount of information they provide about the shared environment. If cues could provide information about an individual's own environmental conditions as well, they would become more valuable still.

Numerous empirical tests have now suggested that diversified bet hedging could play an adaptive role in a variety of life-history traits, such as delayed germination of seeds (Philippi 1993; Evans et al. 2007; Venable 2007; Petru and Tielboerger 2008; Simons 2009), timing of metamorphosis in anurans (Lane and Mahony 2002; Morey and Reznick 2004; Richter-Boix et al. 2006), arthropod diapause (Saiah and Perrin 1990; Danforth 1999; Philippi et al. 2001; Menu and Desouhant 2002), the trade-off between egg size and number (Koops et al. 2003), and bacterial persistence (Balaban et al. 2004; Kussell et al. 2005; Acar et al. 2008). Empirical studies that look at both predictive plasticity and diversified bet hedging as potential adaptations to environmental uncertainty are still quite rare, but a few examples suggest strategies that combine the two (e.g., Danforth 1999; Richter-Boix et al. 2006;

Sadeh et al. 2009; Khatchikian et al. 2010). Most of these studies focus on identifying features of the shared environment that could be acting as predictive cues. For example, germination in desert annual plants is known to be highly sensitive to the amount of rain that falls within a restricted germination period; the amount of germination rain is thought to be an environmental cue that could help predict the total amount of rainfall for the year (Freas and Kemp 1983; Philippi 1993; Khatchikian et al. 2010). Some studies have also used experimental manipulations to try to pinpoint which features might directly cause the developmental response. For example, in desert annual plants, plasticity in germination fraction has been demonstrated in response to variation in day length and temperature conditions (Adondakis and Venable 2004). In amphicarpic plants (which produce two types of offspring with different dispersal abilities), both nutrient availability and local plant density affect the fraction of each type produced (Sadeh et al. 2009). Although some anuran tadpoles are known to respond to pool drying by accelerating metamorphosis, the proximate cues used to recognize that the pool is drying are as yet unclear (Denver et al. 1998; Richter-Boix et al. 2006).

Our theoretical results have two important implications for these studies. First, we show that regardless of the developmental mechanisms involved, environmental variation and information can select for certain levels of diversified bet hedging and predictive plasticity. This means that an adaptive fit can be assessed just by looking at the effective response to shared cues like the weather, even if there is variation in the way individuals experience those cues. Second, we argue that in order to understand the adaptive role of the proximate developmental mechanisms at play, it is necessary to account for natural variation in developmental cues both within and between generations. A strong developmental response to an informative cue that varies within generations may look like a clear-cut case of predictive plasticity even though the result is substantial within-generation phenotypic variation. A closer look at the patterns of variation in both cue and environment could reveal that the phenotypic diversity is advantageous in a variable environment—and that the cue therefore also acts as an adaptive coin flip, used by the developmental strategy to hedge bets against environmental variation between generations.

Acknowledgments

This material is based on work supported in part by a National Science Foundation (NSF) Graduate Research Fellowship to M.C.D.-M., the Max Planck Society, and National Institutes of Health training grant 1K12GM000708. C.T.B. was supported in part by NSF grant ATB-1038590. J. Felsenstein, R. Huey, and B. Kerr provided very helpful feedback on an early version of the manuscript. N. Bertschinger and E. Olbricht contributed useful discussion. A. Robson suggested the use of de Finetti's theorem in appendix B.

APPENDIX A

Calculations for an Example: Drought Prediction in Desert Annuals

In this appendix, we show how to calculate the chance that an individual experiences a good growing season on the basis of the germination rain it has received. We use the numerical values given in the section titled "An Example: Drought Prediction in Desert Annuals." Let $\Pr(e_g) = 3/5$ be the probability of a good growing season and $\Pr(e_p) = 2/5$ be the probability of a poor growing season. Let $\Pr(e_g|q_g) = 4/5$ be the conditional probability of a good growing season given a good germination season and $\Pr(e_g|q_p) = 2/5$ be the conditional probability of good growing season given a poor germination season. From these constraints, we can calculate the marginal probabilities of good and poor germination seasons as $\Pr(q_g) = \Pr(q_p) = 1/2$, because this satisfies the equations $\Pr(e_g|q_g)\Pr(q_g) + \Pr(e_g|q_p)\Pr(q_p) = \Pr(e_g)$ and $\Pr(q_g) + \Pr(q_p) = 1$. Finally, let $\Pr(c_g|q_g) = 4/5$ be the conditional probability that an individual receives good germination rain during a good germination season, and let $\Pr(c_g|q_p) = 2/5$ be the conditional probability that an individual receives good germination rain during a poor germination season. We can calculate the marginal probability that an individual receives good or poor germination rain by averaging over different types of germinations seasons: $\Pr(c_g) = \Pr(c_g|q_g)\Pr(q_g) + \Pr(c_g|q_p)\Pr(q_p) = 3/5$, and, similarly, $\Pr(c_p) = \Pr(c_p|q_g)\Pr(q_g) + \Pr(c_p|q_p)\Pr(q_p) = 2/5$.

We would like to calculate the conditional probability of a good growing season for an individual that has received good germination rain, $\Pr(e_g|c_g)$. To do this, we need to distinguish two cases, one in which a good germination season has occurred and one in which a poor germination season has occurred:

$$\begin{aligned}\Pr(e_g|c_g) &= \Pr(e_g, q_g|c_g) + \Pr(e_g, q_p|c_g)\\ &= \Pr(e_g|q_g)\Pr(q_g|c_g) + \Pr(e_g|q_p)\Pr(q_p|c_g).\end{aligned} \tag{A1}$$

The conditional probabilities of each type of germination season given that good germination rain was received can be calculated using Bayes's rule:

$$\begin{aligned}\Pr(q_g|c_g) &= \frac{\Pr(c_g|q_g)\Pr(q_g)}{\Pr(c_g)} = 2/3 \\ \Pr(q_p|c_g) &= \frac{\Pr(c_g|q_p)\Pr(q_p)}{\Pr(c_g)} = 1/3.\end{aligned} \tag{A2}$$

Plugging these numerical values into equation (A1), we find that

$$\Pr(e_g|c_g) = 4/5 \times 2/3 + 2/5 \times 1/3 = 2/3. \tag{A3}$$

Similar calculations show that if an individual receives poor germination rain, the conditional probability of a good growing season is $\Pr(e_g|c_p) = 1/2$.

APPENDIX B

The Global Predictor Describes the Distribution of Individual Cues

We model the developmental cues that individuals receive according to a two-step stochastic process. First, a shared global predictor q is generated conditional on the shared selective environment e; second, individual cues c are generated conditional on the global predictor. Two assumptions are implicit here: first that a global predictor exists, and second that individual developmental cues are independent of one another given the value of that global predictor. These may seem like rather restrictive assumptions, particularly for those cases in which it is not clear which environmental variable might correspond to that global predictor. Here we argue that under a wide variety of correlational structures among cues received by different individuals we can nonetheless mathematically identify a shared feature of a particular generation that can be treated like a global predictor. That shared feature turns out to be the probability distribution of individual cues, that is, what proportion of all individuals receive each possible cue.

Say that in each generation we can observe the environmental state e as well as the developmental cue c_i for each individual i. In each generation, we will keep track of the fate of every individual in a single lineage by giving each individual an index i from 1 to n. We write a sequence representing one developmental cue for each individual in the lineage as $c = c_1, \dots, c_n$. Similarly, we use a sequence $x = x_1, \dots, x_n$ to represent the phenotypes x_i adopted by all individuals in the lineage. The overall strategy for the lineage, $g(x|c)$, represents the probability that the sequence of phenotypes x is produced given the sequence of developmental cues c. The total reproductive output of the lineage in one generation, $f(x, e)$, depends on the phenotype of each individual and on the environmental state.

Very generally, we can describe the relationship between the environmental state and the individual cues using the joint distribution $\Pr(e, c)$. These two things together can be used to define the type of a generation in terms of its effect on the per capita fitness of a genotype. The expected long-term growth rate for the strategy g is, therefore,

$$\begin{aligned} r(g) &= \sum_{e,\vec{c}} \Pr(e, \vec{c}) \log_2 \bar{f}(g|e, \vec{c}) \\ &= \sum_{e,\vec{c}} \Pr(e, \vec{c}) \log_2 \frac{1}{n} \sum_{\vec{x}} g(\vec{x}\,|\,\vec{c}) f(\vec{x}, e). \end{aligned} \tag{B1}$$

To find a simpler form, we would like to distinguish between generations where the per capita fitness is different but lump together those that are the same. We will argue that, under quite reasonable assumptions, it is not important to keep track of exactly which individual received which developmental cue.

The first assumption is that each individual's fitness depends only on its own phenotype, not on the phenotypes of other individuals. The second assumption is that each individual's phenotype depends only on its own developmental cue, not on the cues received by others. Given these two assumptions, the per capita fitness of a strategy within a particular generation can be rewritten

$$\bar{f}(g|e, \vec{c}) = \frac{1}{n} \sum_{i=1}^{n} \sum_{x} g(x_i = x|c_i) f(x, e). \tag{B2}$$

Note that this quantity does not depend on which individual received which cue but rather what proportion of all individuals received each different type of cue. It is therefore sufficient to distinguish between generations that have different distributions of developmental cues. We use the parameter θ to describe a distribution of individual developmental cues, yielding a simpler form for the per capita fitness of a strategy within a generation:

$$\bar{f}(g|e, \theta) = \sum_{c} \Pr(c|\theta) \sum_{x} g(x|c) f(x, e). \tag{B3}$$

A comparison to equation (1) in the main text shows that θ plays the same role as the global predictor q.

A well-known theorem of de Finetti proves a related result for exchangeable variables (see, e.g., Feller 1966). A finite sequence of random variables $C_1, \dots, C_n$ is exchangeable if every permutation of these variables has the same probability distribution, and an infinite sequence is exchangeable if every finite subsequence is exchangeable. De Finetti's theorem states that any infinite, exchangeable sequence of binary random variables can be understood as a draw from some probability distribution of a single random parameter Θ, followed by a sequence of independent Bernoulli trials weighted according to the parameter θ. When generalized to discrete-valued (rather than only binary) random variables, the parameter θ describes the probability distribution of individual cues (Hewitt and Savage 1955). We argue that even if the developmental cues received by different individuals in the same generation are not exchangeable variables, they nonetheless act as if exchangeable, as far as per capita within-generation fitness is concerned. The only caveat comes when we consider competition between lineages. If the developmental cues are more strongly correlated within a lineage than they are between two competing lineages, this simple breakdown into global predictor and individual developmental cues may not be warranted.

The next question is, what kind of distribution is the global predictor drawn from? If there are k different possible values of the developmental cue, any distribution of such cues lies on a unit $(k-1)$-simplex. Since the global predictor q represents one such distribution, the most general assumption is to say that it is drawn from a continuous distribution over that simplex. If the number of individuals n is consistent from one generation to the next, we could instead identify q not with the probability distribution of cues but with a partition of n individuals into k groups, each group receiving one of the possible developmental cues. For example, if there were 100 individuals in a generation and two possible developmental cues indicating two possible environmental states, q could take on 101 different possible values. To find the optimal developmental strategy—that is, what proportion of each phenotype to produce in response to each of the two different cues—we would need to solve a system of 101 different equations (eq. [11], one for each q), with only two free parameters, one for each developmental cue. This system is highly overdetermined. It would thus be impossible to actually achieve the optimal proportional betting solution for all possible types of generations—instead we must do the best we can and choose the strategy that does best overall by optimizing the growth rate given in equation (1).

We will now show that the same solution can be found by redefining the global predictor as coming from a discrete set with as many different possible values as there are different developmental cues. As we have previously shown (Donaldson-Matasci et al. 2010), when the optimal proportional betting strategy cannot be achieved, the resulting loss in growth rate depends on the difference between the optimal strategy and the effective strategy:

$$\begin{aligned} r(g) &= \sum_{e,q} \Pr(e,q)\log_2 \bar{g}(x_e|q) \\ &= \sum_{e,q} \Pr(e,q)\log_2 \Pr(e|q) \\ &\quad - \sum_{e,q} \Pr(e,q)\log_2 \frac{\Pr(e|q)}{\bar{g}(x_e|q)} \qquad \text{(B4)} \\ &= H(E|Q) - D_{\mathrm{KL}}(\Pr(e|q)\|\,\bar{g}(x_e|q)). \end{aligned}$$

The second term is a measure of the difference between probability distributions $\Pr(e|q)$ and $g(x_e|q)$, known as the Kullback-Leibler (KL) divergence (e.g., Cover and Thomas 1991). We must find the strategy that minimizes that difference. First, we can cluster the different types of generations into k different groups, q'_1 through q'_k. For each group, we can calculate the marginal probability of each environment as $\Pr(e|q') = \sum_{q\in q'}\Pr(e|q)\Pr(q|q')$. This defines the best effective strategy for the group of generations q', $\hat{g}(x_e|q') = \Pr(e|q')$, because it minimizes the average KL divergence for that group. The problem, then, is to ensure that the clustering of q's is done in a way that minimizes the overall KL divergence (a classic centroid-based clustering problem). Assuming that this has been done, the best achievable growth rate can be written as

$$\begin{aligned} \hat{r}(g) &= \sum_{e,q'} \Pr(e,q')\log_2 \bar{g}(x_e|q') \\ &= \sum_{e,q'} \Pr(e,q')\log_2 \Pr(e|q') \\ &\quad - \sum_{e,q'} \Pr(e,q')\sum_{q\in q'}\Pr(q|q')\log_2 \frac{\Pr(e|q)}{\Pr(e|q')} \qquad \text{(B5)} \\ &= H(E|Q') - D_{\mathrm{KL}}(\Pr(e|q)\|\,\Pr(e|q')). \end{aligned}$$

The first term is exactly what the optimal growth rate would be if there were really only k different types of generations, and the second term (which is always positive) represents the loss due to the fact that not all generations in each group q' are the same. This shows that if we cluster generations together in the right way, we can find the optimal strategy and an upper bound on the growth rate so that even if q seems to be drawn from a continuous distribution we can transform the problem and treat it as if it were discrete. In the main text, however, we will simply assume that q is drawn from a discrete distribution.

APPENDIX C

Results Hold When Phenotypes Survive in Multiple Environments

The principle of proportional betting—that the optimal strategy is to produce phenotypes in proportion to the probability of each environmental state—holds only when each phenotype can survive in a single environmental state. In the main body of this article, we have made this assumption for clarity of exposition. However, some of our results (eqq. [9], [11], and [12]) depend explicitly on this assumption. In this appendix, we explain how our results generalize when phenotypes can survive in multiple environments.

In a previous article, we showed how to generalize the principle of proportional betting under general fitness functions (Donaldson-Matasci et al. 2008). We define the level of reproductive investment, or specialization, that each phenotype puts into each possible environment and show that the optimal strategy mixes phenotypes so that its reproductive investment in each environmental state matches the probability of that state. To do this, we define a set of completely specialized phenotypes y, each of which survives in only one environmental state e. We rewrite the fitness profile of each phenotype x across all environments as a mixture of these completely specialized phenotypes as follows:

$$\begin{aligned} f(x, e) &= \sum_y s(y|x)f(y, e) \qquad \text{for all } x, e \\ &= s(y_e|x)f(y_e, e), \end{aligned} \tag{C1}$$

where $\sum_y s(y|x) = 1$ for all x and $f(y, e) > 0$ if and only if $y = y_e$, that is, y_e is the one phenotype that survives in environment e. The function $s(y|x)$ is a measure of the relative investment the phenotype x puts into the environmental state associated with each completely specialized phenotype y. We show in Donaldson-Matasci et al. (2008) that a set of these completely specialized phenotypes can be defined uniquely for a quite general fitness function.

We can now rewrite the mean fitness of a genotype g within a generation (see eqq. [1] and [10]) as follows:

$$\begin{aligned} \bar{f}(g|e, q) &= \sum_x \bar{g}(x|q)f(x, e) \\ &= \sum_x \bar{g}(x|q)\sum_y s(y|x)f(y, e) \\ &= \sum_y g'(y|q)f(y, e), \end{aligned} \tag{C2}$$

where $g'(y|q) = \sum_x \bar{g}(x|q)s(y|x)$. This shows that the effective strategy $\bar{g}(x|q)$, which describes a mixture of phenotypes x produced in response to each possible global predictor q, is equivalent to an effective strategy $g'(y|q)$, which uses the completely specialized phenotypes y. We know that the optimal mixture for g' is proportional betting (see eq. [9]). This gives the conditions we need to solve for the optimal developmental strategy $\hat{g}(x|c)$:

$$\begin{aligned} \Pr(e|q) &= \hat{g}'(y_e|q) \\ &= \sum_x \hat{g}(x|q)s(y_e|x) \\ &= \sum_c \Pr(c|q)\sum_x \hat{g}(x|c)s(y_e|x). \end{aligned} \tag{C3}$$

We now proceed to show that three results from the main body of this article still hold with generalized fitness functions.

In the section titled "The Optimal Strategy Is More Sensitive to Developmental Cues," we claim that the optimal developmental strategy shows greater developmental sensitivity when developmental cues vary within a generation. This proof hinges on the fact that the optimal relationship between the global predictor and the phenotype is unchanged by adding within-generation variation in developmental cues—that is, $I(X; Q)$ is fixed. The same is true here: from equation (C3) we see that the optimal effective strategy $\hat{g}(x|q)$ depends only on the relationship between the environmental state and the global predictor, via $\Pr(e|q)$, and the fitness matrix, via $s(y_e|x)$, but not on the relationship between the global predictor and the individual cues. The rest of the proof follows without modification.

In the section titled "The Fitness Value of Information Stays the Same," we claim that the fitness value of using a cue is not limited by the amount of information an individual receives from that cue but rather by the amount of information the shared global predictor would provide—even though individuals do not observe it. This is because when proportional betting can be achieved both with and without the cue the difference in growth rate is the mutual information that the global predictor provides about the environment. Otherwise, the mutual information is an upper limit on the value. For general fitness functions, the optimal strategy is no longer to do proportional betting, but it is nonetheless equivalent to a strategy that does proportional betting with the completely specialized phenotypes y (see eq. [C3]). As long as the optimal strategy (with and without the cue) can actually be achieved even with the set of phenotypes x, the maximal growth rates will be unchanged—and the upper limit on the value of the cue, $I(E; Q)$, can actually be achieved.

What happens when the constraints of the fitness function make it impossible to achieve the optimal strategy? In Donaldson-Matasci et al. (2010), we showed that the value of a shared cue is nonetheless limited by the information it provides about the environment: $r(\hat{g}(x|q)) - r(\hat{g}(x)) \leq I(E; Q)$. How does this change for cues that are not shared but that may vary between individuals within a generation? The maximal growth rate without the cue, $r(\hat{g}(x))$, is of course unchanged. The maximal growth rate with an individually variable cue, $r(\hat{g}(x|c))$, can only decrease compared with the growth rate of the optimal strategy responding directly to the global predictor, $r(\hat{g}(x|q))$. This is because the effective strategies produced by a developmental strategy responding to an individually variable cue are a subset of the strategies that could be used in direct response to the global predictor. Therefore, the value of using a cue that varies within generations is limited by the information carried in the global predictor even for general fitness functions.

In the main text, we argue that it is shared uncertainty that drives the optimal amount of diversified bet hedging. In fact, when each phenotype survives in just one environment, we can make the stronger statement that the optimal amount of diversified bet hedging and predictive plasticity exactly mirrors the amount of shared uncertainty and information about the environment: $H(X|Q) = H(E|Q)$ and $I(X; Q) = I(E; Q)$ (see fig. 2). For more general fitness functions, however, we can say only that $H(Y|Q) = H(E|Q)$ and that $I(Y; Q) = I(E; Q)$ (see eq. [C3]). We can interpret this as saying that the optimal strategy diversifies its investment into different environments according to the shared environmental uncertainty. The important point is that adding within-generation variation in developmental cues does not change the optimal amount of bet hedging because it does not alter the optimal effective strategy (see eq. [C3]).

Literature Cited

Acar, M., J. T. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. Nature Genetics 40:471–475.

Adondakis, S., and D. Venable. 2004. Dormancy and germination in a guild of Sonoran Desert annuals. Ecology 85:2582–2590.

Balaban, N., J. Merrin, R. Chait, L. Kowalik, and S. Leibler. 2004. Bacterial persistence as a phenotypic switch. Science 305:1622–1625.

Bergstrom, C. T., and M. Lachmann. 2004. Shannon information and biological fitness. Pages 50–54 *in* IEEE Information Theory Workshop 2004. IEEE, New York.

Cohen, D. 1966. Optimizing reproduction in a randomly varying environment. Journal of Theoretical Biology 12:119–129.

———. 1967. Optimizing reproduction in a randomly varying environment when a correlation may exist between the conditions at the time a choice has to be made and the subsequent outcome. Journal of Theoretical Biology 16:1–14.

Cooper, W. S., and R. H. Kaplan. 1982. Adaptive "coin-flipping": a decision-theoretic examination of natural selection for random individual variation. Journal of Theoretical Biology 94:135–151.

Cover, T. M., and J. A. Thomas. 1991. Elements of information theory. Wiley Series in Telecommunications. Wiley, New York.

Danforth, B. N. 1999. Emergence dynamics and bet hedging in a desert bee, *Perdita portalis*. Proceedings of the Royal Society B: Biological Sciences 266:1985–1994.

Dempster, E. R. 1955. Maintenance of genetic heterogeneity. Cold Spring Harbor Symposia on Quantitative Biology 20:25–32.

Denver, R. J., N. Mirhadi, and M. Phillips. 1998. Adaptive plasticity in amphibian metamorphosis: response of *Scaphiopus hammonii* tadpoles to habitat desiccation. Ecology 79:1859–1872.

Donaldson-Matasci, M. C. 2008. Adaptation in a changing environment: phenotypic diversity in response to environmental uncertainty and information. PhD diss. University of Washington, Seattle.

Donaldson-Matasci, M. C., C. T. Bergstrom, and M. Lachmann. 2010. The fitness value of information. Oikos 119:219–230.

Donaldson-Matasci, M. C., M. Lachmann, and C. T. Bergstrom. 2008. Phenotypic diversity as an adaptation to environmental uncertainty. Evolutionary Ecology Research 10:493–515.

Evans, M. E. K., R. Ferrière, M. J. Kane, and D. L. Venable. 2007. Bet hedging via seed banking in desert evening primroses (*Oenothera*, Onagraceae): demographic evidence from natural populations. American Naturalist 169:184–194.

Feller, W. 1966. An introduction to probability theory and its applications. Vol. 2. Wiley, New York.

Frank, S. A., and M. Slatkin. 1990. Evolution in a variable environment. American Naturalist 136:244–260.

Freas, K. E., and P. R. Kemp. 1983. Some relationships between environmental reliability and seed dormancy in desert annual plants. Journal of Ecology 71:211–217.

Haccou, P., and Y. Iwasa. 1995. Optimal mixed strategies in stochastic environments. Theoretical Population Biology 47:212–243.

Hewitt, E., and L. J. Savage. 1955. Symmetric measures on Cartesian products. Transactions of the American Mathematical Society 80: 470–501.

Hoffmann, R. J. 1978. Environmental uncertainty and evolution of physiological adaptation in *Colias* butterflies. American Naturalist 112:999–1015.

Kelly, J. L. 1956. A new interpretation of information rate. Bell System Technical Journal 35:917–926.

Khatchikian, C. E., J. J. Dennehy, C. J. Vitek, and T. P. Livdahl. 2010. Environmental effects on bet hedging in *Aedes* mosquito egg hatch. Evolutionary Ecology 24:1159–1169.

Koops, M. A., J. A. Hutchings, and B. K. Adams. 2003. Environmental predictability and the cost of imperfect information: influences on offspring size variability. Evolutionary Ecology Research 5:29–42.

Kussell, E., R. Kishony, N. Balaban, and S. Leibler. 2005. Bacterial persistence: a model of survival in changing environments. Genetics 169:1807–1814.

Lane, S. J., and M. J. Mahony. 2002. Larval anurans with synchronous and asynchronous development periods: contrasting responses to water reduction and predator presence. Journal of Animal Ecology 71:780–792.

Lively, C. M. 1986. Predator-induced shell dimorphism in the acorn barnacle *Chthamalus anisopoma*. Evolution 40:232–242.

McAdams, H. H., and A. Arkin. 1997. Stochastic mechanisms in gene expression. Proceedings of the National Academy of Sciences of the USA 94:814–819.

Menu, F., and E. Desouhant. 2002. Bet-hedging for variability in life cycle duration: bigger and later-emerging chestnut weevils have increased probability of a prolonged diapause. Oecologia (Berlin) 132:167–174.

Moran, N. A. 1992. The evolutionary maintenance of alternative phenotypes. American Naturalist 139:971–989.

Morey, S. R., and D. N. Reznick. 2004. The relationship between habitat permanence and larval development in California spadefoot toads: field and laboratory comparisons of developmental plasticity. Oikos 104:172–190.

Petru, M., and K. Tielboerger. 2008. Germination behaviour of annual plants under changing climatic conditions: separating local and regional environmental effects. Oecologia (Berlin) 155:717–728.

Philippi, T. 1993. Bet-hedging germination of desert annuals: variation among populations and maternal effects in *Lepidium lasiocarpum*. American Naturalist 142:488–507.

Philippi, T., M. Simovich, E. Bauder, and J. Moorad. 2001. Habitat ephemerality and hatching fractions of a diapausing anostracan (Crustacea: Branchiopoda). Israel Journal of Zoology 47:387–395.

Pielou, E. 1969. An introduction to mathematical ecology. Wiley, New York.

Richter-Boix, A., G. A. Llorente, and A. Montori. 2006. A comparative analysis of the adaptive developmental plasticity hypothesis in six Mediterranean anuran species along a pond permanency gradient. Evolutionary Ecology Research 8:1139–1154.

Rivoire, O., and S. Leibler. 2011. The value of information for populations in varying environments. Journal of Statistical Physics 142: 1124–1166.

Robson, A. J. 1996. A biological basis for expected utility. Journal of Economic Theory 68:397–424.

Sadeh, A., H. Guterman, M. Gersani, and O. Ovadia. 2009. Plastic bet-hedging in an amphicarpic annual: an integrated strategy under variable conditions. Evolutionary Ecology 23:373–388.

Saiah, H., and N. Perrin. 1990. Autumnal vs. spring hatching in the fairy shrimp *Siphonophanes grubii* (Dybowski) (Crustacea, Anostraca): diversified bet-hedging strategy? Functional Ecology 4:769–775.

Seger, J., and H. J. Brockmann. 1987. What is bet-hedging? Pages 182–211 *in* P. Harvey and L. Partridge, eds. Oxford surveys in evolutionary biology. Vol. 4. Oxford University Press, Oxford.

Shannon, C. E. 1948. A mathematical theory of communication. Bell System Technical Journal 27:379–423, 623–656.

Simons, A. M. 2009. Fluctuating natural selection accounts for the evolution of diversification bet hedging. Proceedings of the Royal Society B: Biological Sciences 276:1987–1992.

Simons, A. M., and M. O. Johnston. 2006. Environmental and genetic sources of diversification in the timing of seed germination: implications for the evolution of bet hedging. Evolution 60:2280–2292.

Venable, D. L. 2007. Bet hedging in a guild of desert annuals. Ecology 88:1086–1090.

Associate Editor: Yannis Michalakis
Editor: Mark A. McPeek

Spring wildflowers (Mexican gold poppy and Coulter's lupine) blooming in a good year in the Sonoran desert. Photo by M. C. Donaldson-Matasci.