

A Model for Optimal Reaction Norms: The Case of the Pregnant Garter Snake and Her Temperature-Sensitive Embryos

Stevan J. Arnold^{1,*} and Charles R. Peterson^{2,†}

1. Department of Zoology, Oregon State University, Corvallis, Oregon 97331;

2. Department of Biological Sciences, Idaho State University, Pocatello, Idaho 82309

Submitted June 19, 2001; Accepted February 6, 2002

ABSTRACT: We present a model to test Osgood's (1978) proposition that viviparous snakes have optimal reaction norms for temperature-sensitive meristic traits, such as scale counts. Our model predicts that traits that are subject to temperature effects during development will evolve a flat or U-shaped reaction norm (average scale count as a function of developmental temperature). We tested this prediction by maintaining 67 female garter snakes (*Thamnophis elegans*) at eight different constant temperatures (21°–33°C) during pregnancy and making a series of scale counts on their newborn offspring ($n = 491$). To insure that the experimental temperatures were ecologically relevant, we used automated radiotelemetry to record the body temperature of pregnant, free-ranging females. The resulting temperature data allowed us to test the prediction that the inflection points of reaction norms would correspond to the average temperature experienced by embryos in the field. In line with predictions of the Osgood model, reaction norms were flat or U-shaped. In the case of U-shaped reaction norms, the inflection point of the curves corresponded to the average temperature imposed on embryos by free-ranging females. In contrast to some past studies, none of the standard scale scores (ones commonly used in systematics) showed significant temperature effects in either sex. Reaction norms were flat. In contrast, incidences of various abnormalities showed U-shaped reaction norms. Temperature effects were more pronounced in males than in females. The results have implications for systematics and for the evolution of canalization and phenotypic plasticity.

Keywords: norms of reaction, reaction norms, canalization, environmental canalization, *Thamnophis elegans*, vertebral number, meristic traits, stabilizing selection, fluctuating asymmetry, scutellation, scallation, osteological abnormality.

* E-mail: arnolds@bcc.orst.edu.

† E-mail: petechar@isu.edu.

The ecological dimension is typically missing in studies of phenotypic plasticity and reaction norms. The omission happens because plasticity is usually studied in the laboratory. Two or more environmental treatments are imposed, and resulting phenotypic differences among treatments are labeled plasticity or described by functions called “reaction norms” (Scheiner 1993; Schlichting and Pigliucci 1998; Debat and David 2001). What is the ecological significance of the plasticity that has been measured? To answer that question, we need to know the statistical distribution of environments that is encountered in nature. If this environmental distribution is unknown, we might—in the worst case—choose a pair of treatments in the laboratory that is far outside the range of environments that the population experiences and to which it is adapted. It is difficult to say how many plasticity studies approach this worst case because many studies do not undertake crucial environmental measurements that would relate the laboratory treatments to natural conditions. Many plasticity studies may be exercises in ecological fantasy.

The phenotypic plasticity of scalation in viviparous snakes provides an opportunity both to assess the distribution of environments and to measure reaction norms over an ecologically relevant range of environments. During pregnancy, viviparous snakes such as garter snakes (*Thamnophis*) actively thermoregulate by behavioral means (Peterson et al. 1993). Furthermore, these snakes can be implanted with temperature-sensitive radiotransmitters so that the temperatures experienced by embryos can be monitored in free-ranging pregnant females in the field (Osgood 1970; Reinert and Cundall 1982; Peterson 1987; Peterson et al. 1993). The thermal environment of embryos in the field can then be compared with the thermal sensitivity of developmental processes by exposing pregnant females to different temperatures in the laboratory. Osgood (1978) conducted the first field and laboratory study of thermal sensitivity in viviparous snakes, focusing

on meristic (scale) counts in water snakes (*Nerodia*). The general picture emerging from Osgood's (1978) work was that scale counts were rather insensitive to temperature effects, a result that conflicted with earlier claims of large temperature effects (Fox 1948; Fox et al. 1961). Osgood (1978) proposed that the preferred body temperature of pregnant snakes was also an optimal temperature for the development of scalation. When this temperature is imposed on embryos in the laboratory, their scalation traits correspond to those of wild-caught snakes, but at higher or lower temperatures, their scalation is abnormal.

In this article, we test Osgood's proposition by making three improvements in his argument. First, we test the logical veracity of Osgood's proposal by building a model of optimal reaction norms and examining its consequences. Second, we pursue the ecological side of the question by using improved technology to monitor embryo temperatures in the field. Those improvements are based on Osgood's (1970) own pioneering work on miniature radiotransmitters. Third, to test our model, we make several improvements in laboratory protocols so that we can gain a detailed, ecologically relevant picture of reaction norms (i.e., temperature treatments focus on the range experienced in the field, the number of treatments is expanded, and family structure is taken into account in data analysis). In the laboratory work, we examine the effects of temperature on scale counts, especially ones that are intimately related to vertebral numbers. The temperature sensitivity of these meristic traits has been studied in many kinds of vertebrates because of their importance in systematics.

Temperature during development affects vertebral number and other meristic traits in all classes of vertebrates (Fox 1948; Yntema 1960; Lecyk 1965; Peabody and Brodie 1975; Jockusch 1997), but the phenomenon has been studied most in fish (Hubbs 1922; Tåning 1952; Barlow 1961; Lindsey 1966; Lindsey and Moodie 1967; Fowler 1970; Lindsey and Harrington 1972). The norm of reaction to temperature is commonly U-shaped in fish, with the lowest numbers of vertebrae at intermediate temperatures and higher numbers at both low and high temperatures (Fowler 1970). Linear norms of reaction also are common, with vertebral numbers decreasing as temperature increases (Lindsey and Harrington 1972). Temperature-sensitive periods have been demonstrated by changing incubation temperature at critical stages during development. Vertebral number in fish, for example, is sensitive to temperature at a variety of times during development. Surprisingly, the earliest sensitive periods occur at gastrulation, well before somite differentiation. Furthermore, the sensitive periods of different meristic traits (e.g., numbers of fin rays and vertebrae) do not coincide, and some traits are relatively unaffected by temperature, showing flat reaction norms

(Tåning 1952). An overarching framework for explaining the variety of shapes found in reaction norms is conspicuously lacking in all of these discussions. In recent years, however, a candidate framework has emerged in the field of quantitative genetics.

A series of population and quantitative genetic models have been proposed for the evolution of reaction norms (e.g., Via and Lande 1985; van Tienderen 1991) and canalization (Gavrilets and Hastings 1994; Wagner et al. 1997; Rice 1998). The most powerful of these models treat the reaction norm as an infinite-dimensional character so that the phenotype can be viewed as a continuous function of the environment (Kirkpatrick and Heckman 1989; Kirkpatrick et al. 1990; Gomulkiewicz and Kirkpatrick 1992; Kirkpatrick and Lofsvold 1992). These models have demonstrated that, so long as selection is frequency independent, the reaction norm will tend to evolve in a direction that maximizes average fitness in the population (Gomulkiewicz and Kirkpatrick 1992). These dynamic models, which can be used to predict evolutionary trajectories, also suggest specific ways for quantifying selection on and inheritance of reaction norms. The traits considered in this article are referred to as "nonlabile" by Gomulkiewicz and Kirkpatrick (1992). Such traits are fixed throughout life once development is completed. Models for the evolution of canalization have generally considered only linear reaction norms (Gavrilets and Hastings 1994; Wagner et al. 1997; but see Rice 1998). The goal of this article is the limited one of establishing which shape of reaction norm might be optimal when a nonlabile character is under stabilizing selection and close to its optimum; we will then determine whether those expectations are met in a natural population. To accomplish this goal, we use a model developed by Lande (1980) for the establishment of a zone of canalization when a trait under stabilizing selection is close to its adaptive optimum. In the "Discussion" section, we consider the fact that the environmental variable in our system (temperature of embryos) is, in large part, determined by maternal behavior.

Stabilizing selection on vertebral numbers has been revealed in both cross-sectional and longitudinal studies on snakes. Three studies have reported less variation in vertebral numbers in adult cohorts compared with juveniles, suggesting stabilizing selection (Dunn 1942; Inger 1942, 1943). In longitudinal studies, the highest growth rates occurred in snakes with intermediate numbers of vertebrae (Arnold 1988; Lindell et al. 1993). Correlational selection between numbers of body and tail vertebrae was documented in a study of crawling performance. The fastest snakes showed a characteristic positive relationship between the two vertebral numbers (Arnold and Bennett 1984). Selection acts against number of abnormal vertebrae in the field. The number of vertebrae with an extra

rib had a negative effect on growth rate, presumably because locomotion was impeded (Arnold and Bennett 1984; Kelley et al. 1997). These modes of selection on vertebral numbers have the potential to direct the evolution of reaction norms. Numerous studies have documented selection against asymmetries (Møller 1997), as in the bilateral scale counts that we will also consider.

A Model for Optimal Reaction Norms

A graphical overview of the model is provided in figure 1. Our focus is on a single, representative genotype. All of the phenotypic variation expressed by that genotype is, by definition, environmental. The model assumes that a reaction norm translates variation in some environmental variable into phenotypic variation that is expressed by the genotype. A fitness function then translates the genotype's phenotypic variation into fitness variation. Notice that

when the trait mean is in the vicinity of the peak of the fitness function (fig. 1), the net effect of trait variation will be to lower average fitness. Thus, the tendency of stabilizing selection is to reduce variation in the trait. That reduction in trait variation can be accomplished by changing the shape of the reaction norm and/or by reducing variation in the distribution of environments. A mathematical device, the Taylor series, is used to approximate variation in fitness as a function of the shapes of the fitness function and the reaction norm and the variation in imposed environments. The power of the Taylor series approach is that it makes its approximation using only the first and second derivatives of the reaction norm and fitness functions rather than the exact forms of those functions.

Turning to the details of the model, we assume that part of the environmental variance of a genotype for a trait, such as a scale count, is a function of two attributes: a

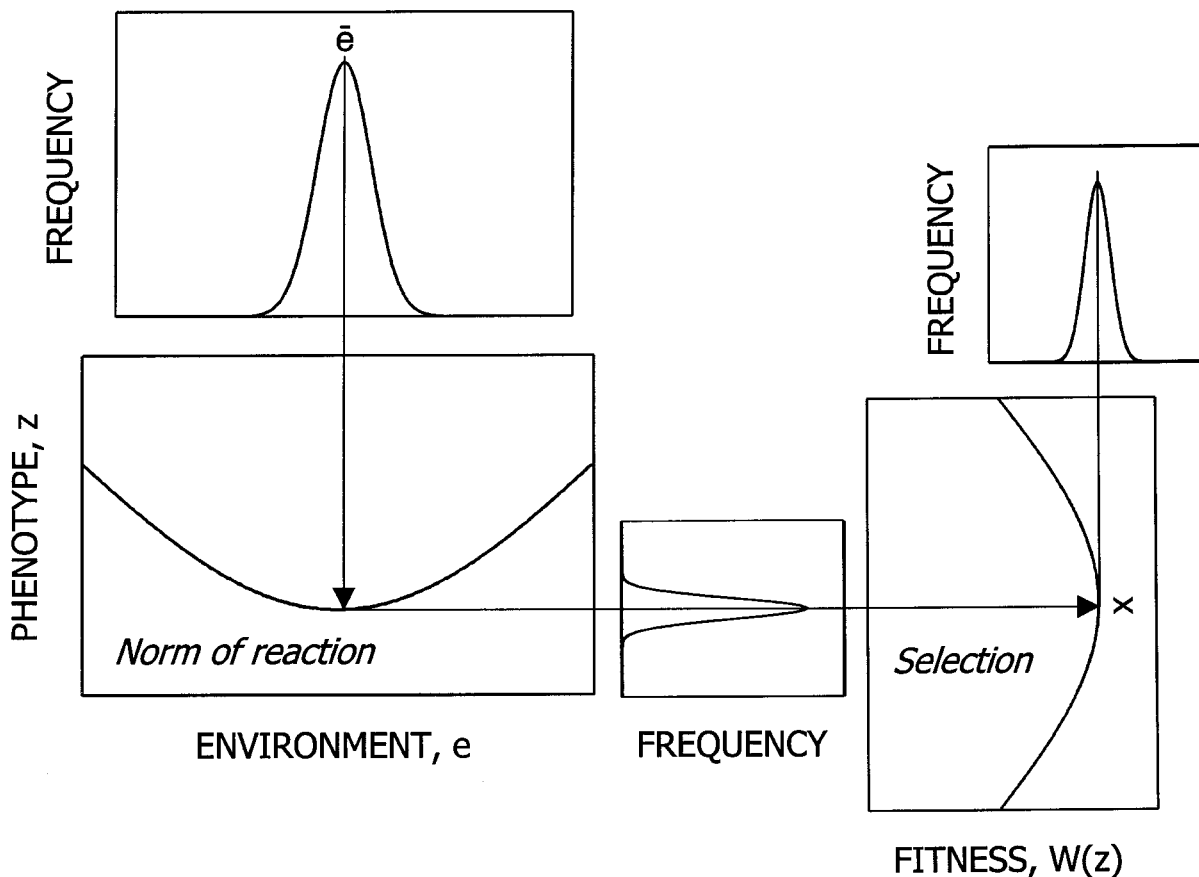


Figure 1: Model for an optimal reaction norm. The reaction norm (*lower left*) translates a distribution of environments (with mean \bar{e} and variance σ_e^2) into a distribution of phenotypes (with mean x and variance σ^2) for a particular genotype. The fitness function (*lower right*) translates that phenotypic distribution into a distribution of fitnesses. Note that this figure is diagrammatic. In an accurate portrayal, the two frequency distributions on the right would be asymmetrical.

distribution of environments encountered by the genotype and a reaction norm that translates those environments into phenotypes. Consequently, the value for environmental variance, σ^2 , is a function of the shape of the reaction norm as well as the distribution of encountered environments. Let $z(e)$ represent the reaction norm for our genotype. Suppose that our genotype experiences a normal distribution of environments with mean \bar{e} and variance σ_e^2 . We can approximate the expected value for the phenotype of our genotype by expanding $z(e)$ in a Taylor series about \bar{e} :

$$E[z(e)] \approx z(\bar{e}) + (e - \bar{e}) \left. \frac{dz(e)}{de} \right|_{e=\bar{e}} + \frac{(e - \bar{e})^2}{2} \left. \frac{d^2z(e)}{de^2} \right|_{e=\bar{e}}. \quad (1)$$

The first and second derivatives in this expression are constants, evaluated at the point $e = \bar{e}$; consequently, the environmental variance for our genotype is approximately

$$\sigma^2 = \left[\left(\frac{dz(e)}{de} \right)^2 - \bar{e}^2 \left(\frac{d^2z(e)}{de^2} \right)^2 \right] \sigma_e^2 + \frac{3}{4} \left(\frac{d^2z(e)}{de^2} \right)^2 \sigma_e^4 + \sigma_r^2, \quad (2)$$

where the first two terms represent the contribution of the reaction norm and σ_r^2 represents the residual environmental variance that remains if the reaction norm makes no contribution.

Suppose that the trait is under stabilizing selection so that the expected value of fitness for an individual with phenotypic value z , $W(z)$, is a curve that is concave downward. Our genotype has an average phenotypic value of x , situated in the vicinity of the optimum value of the trait (fig. 1). That genotype shows a distribution of phenotypic values caused by the differing environments encountered by different individuals. Let that distribution be symmetrical with variance σ^2 . We can obtain an approximation for the average fitness of the genotype by expanding the fitness function $W(x)$ in a Taylor series about x and then averaging that series over the environmental distribution. Following Lande (1980), we find that the average fitness of the genotype x is

$$\bar{W}(x) = W(x) + \frac{\sigma^2}{2} \left. \frac{d^2W(z)}{dz^2} \right|_{z=x} + \dots, \quad (3)$$

where higher-order terms are negligible if selection is weak

and $z = x$ denotes that the second derivative is to be evaluated at the point $z = x$. Because the selection function $W(z)$ is concave downward, the sign of $d^2W(z)/dz^2$ is negative in the vicinity of the optimum. Consequently, selection favors developmental processes (e.g., reaction norms) that result in the smallest possible value for environmental variance, σ^2 .

Thus, the optimal reaction norm is one that makes the minimum contribution to environmental variance in equation (2). Consequently, we can use equation (2) to evaluate various possible shapes of reaction norm. The best shape for the reaction norm is a flat line (fig. 2a), for then the environmental variance will take the minimum value of σ_r^2 . The next best shape is a curve that is concave upward or downward, with its inflection point at the environmental value corresponding to \bar{e} (fig. 2b). The minimum variance (σ_r^2) is achieved with this shape of reaction norm when the squared coefficient of variation for the environmental distribution equals four-thirds. The next best shape is a straight line with positive or negative slope (fig. 2c). Finally, the worst shape is a curved line (fig. 2d), with a first derivative at the point \bar{e} equal to the slope of the straight lines shown in figure 2c.

In our garter snake example, the environmental variance, σ_e^2 , is the variance in temperature produced by its thermoregulating mother and experienced by a developing

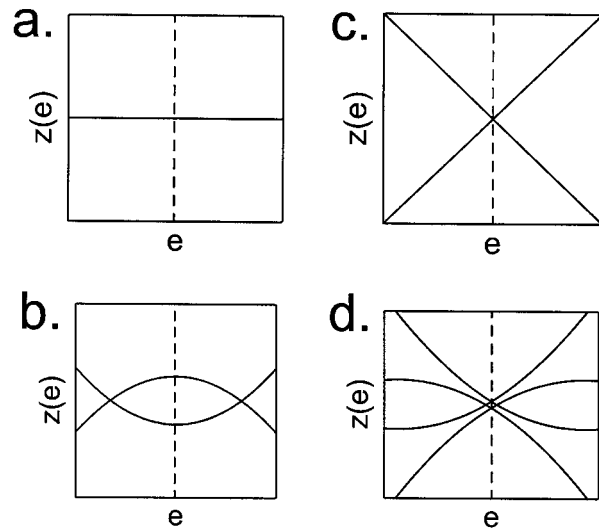


Figure 2: Various shapes for reaction norms. The vertical dashed line is the mean of the environmental distribution. *a*, Flat reaction norm. *b*, Reaction norms with nonzero second derivatives (*curvature*) at the environmental mean but with first derivatives equal to zero (no slope). *c*, Reaction norms with positive or negative slopes at the environmental mean but no curvature. *d*, Reaction norms with both slope and curvature at the environmental mean.

embryo. During embryonic development, a reaction norm translates this variation into environmental variation, σ^2 , for a particular scalation trait. The scalation trait for an individual is fixed in value by the time it is born. We assume that individuals experience stabilizing selection during their postnatal life.

Material and Methods

Maternal Thermoregulation

We studied maternal thermoregulation by monitoring the body temperatures of eight free-ranging pregnant female *Thamnophis elegans* that we surgically implanted with temperature-sensitive radiotransmitters. We captured pregnant females (50–75 g) early in pregnancy (April–May) at Papoose Meadow (elevation 1,646 m) near Eagle Lake (Lassen County, Calif.) and surgically implanted them with 4-g radiotransmitters (Telonics models CHP1 or CHP2) using methods described by Reinert and Cundall (1982). We used subcutaneous injections of brevit sodium to anesthetize the snakes. We released the females at the site of capture after a 5-d recovery period. We automatically recorded body temperatures with a Telonics custom-made computerized recording system with a fixed antenna every 5 min, 24 h/d, for periods ranging from 11 to 53 d (10,876 records).

Temperature Treatments

We captured 67 female *T. elegans* at eight localities in the vicinity of Eagle Lake (Lassen County, Calif.), shipped them by air to the laboratory, and maintained them in constant-temperature incubators until their litters were born. We immediately distributed the females among nine temperature treatments (21°, 24°, 26°, 27°, 28°, 29°, 30°, 32°, and 33°C). Details of maintenance procedures are described by Arnold et al. (1995). The range of temperature treatments reported here is greater than in Arnold et al. (1995) because at the extreme temperature treatments (21° and 33°C), there were few live births.

Trait Scores

We made scale counts on 491 neonates (224 females, 267 males) that were killed, fixed in 10% formalin, and preserved in 70% ethanol. We established sex by eversion of hemipenes in males. We made a standard series of scale counts on each specimen: ventrals (VENT), subcaudals (SUB), supralabials (SLAB), infralabials (ILAB), preoculars (PREOC), postoculars (POSTOC), and scale rows at mid-body (MID; Arnold and Phillips 1999). We used the sum of right and left sides in the case of bilateral characters.

We also recorded a series of abnormalities: number of half-ventral scales (HALF), number of undivided subcaudal scales (WHOLE), absolute difference between counts on the right and left sides for supralabials (SLAB ASYM), infralabials (ILAB ASYM), preoculars (PREOC ASYM), and postoculars (POSTOC ASYM). Half-ventral scales often reflect an osteological abnormality in which a vertebra has two ribs on one side and one rib on the other (King 1959; but see Clark and Callison 1967; Osgood 1970). We made the same scale counts on dams.

Statistical Methods

We based statistical analyses on the litter average for each scalation score. We analyzed the scores for male and female offspring separately. We made tests for temperature effects by fitting a general linear model (GLM) in SAS (SAS Institute 1989). We used several covariates: scalation score of the dam (DAM), Julian date of onset of constant temperature treatment (DATE), dam's locality (LOCALE), and average birth mass of offspring (MASS). We treated LOCALE as a categorical variable and the other covariates as continuous variables. We used DAM as a covariate to control for genetic differences among litters in scalation scores. The model included linear (TEMP) and quadratic terms (TEMP²) for temperature effects. We estimated the inflection point for reaction norms by setting the first derivative of the fitted quadratic regression equal to zero and solving for the temperature term. We approximated the standard errors of the inflection points using the delta method (Bulmer 1979).

Results

Maternal Thermoregulation

Embryos in the field experienced temperatures ranging from 12° to 36°C with a mean of 26.6°C (fig. 3). The distribution of temperatures showed a strong mode at 30°–31°C and was skewed to the left.

Standard Scale Scores

None of the standard scale scores showed significant temperature effects in either sex; DAM effects were significant ($P \leq .05$) for VENT, SUB, and ILAB in females and for VENT, SUB, and SLAB in males. Date of onset of temperature treatment was significant ($P = .04$) only for SLAB in females. Average birth weight (MASS) was significant ($P \leq .05$) only for SLAB and ILAB in males. The effect was positive in both cases.

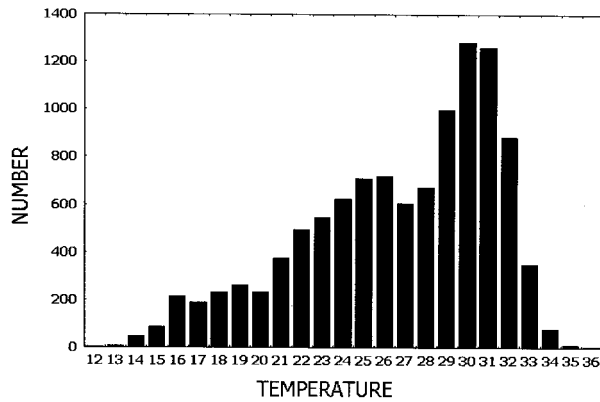


Figure 3: Distribution of body temperatures recorded in free-ranging pregnant garter snakes (*Thamnophis elegans*). The mean is 26.6°C.

Scale Abnormalities

In contrast to the standard scale scores, several scale abnormalities were affected by temperature. In females (table 1), only postocular asymmetries were significantly affected by temperature, but in males (table 2), four different types of abnormality were affected (half-ventrals and asymmetry of supralabials, infralabials, and postoculars). In all five cases, the linear temperature terms were negative and significant, while the quadratic temperature terms were positive and significant. These signs correspond to reaction norms that are U-shaped (fig. 4). The mean inflection point for these five reaction norms is 27.0°C ($\pm 0.7^\circ\text{C SE}$), with 95% confidence limits 26.1°–28.5°C.

Discussion

The Test for Optimal Reaction Norms

The *Thamnophis elegans* data support the optimal reaction norm model in two major ways. First, the shapes of the reaction norms were optimal in the sense that their shape minimized environmental variance in scalation traits and abnormalities. Reaction norms showed either the best shape (flat line), in the case of the standard counts or some measures of abnormality, or the next best shape (U), in the case of the other measures of abnormality. Second, when reaction norms were U-shaped, the inflection points were in the vicinity of the average temperature experienced by embryos. In particular, the 95% confidence limits of the inflection points for five U-shaped reaction norms overlapped the mean of the distribution of environmental temperatures (fig. 3). These two results offer strong support for the model and for Osgood's (1978) proposal.

Although the results correspond nicely with theoretical expectations, the experimental design could be improved

in several ways. First, embryos in nature do not experience constant temperatures. Although pregnant females behaviorally thermoregulate so that their body temperature is often in the range of 26°–31°C during the day, environmental cooling at night can cause the maternal temperature to drop to as low as 12°C (fig. 3; Peterson et al. 1993). The embryo experiences a complex temperature regime that cycles on a daily basis. Because reaction norms may have adapted to this cycling regime, temperature treatments should be cycling rather than constant. Second, embryos in our experiments experienced natural temperature regimes during the early portion of their development because we used wild-caught pregnant females with an unknown date of ovulation. A better but more difficult procedure would be to breed females in the laboratory and begin the onset of temperature treatment at or soon after ovulation.

Are females more buffered than males? In our experiment, only one kind of female abnormality was affected by developmental temperature (table 1), but in males, four kinds of abnormalities were affected (table 2). We cannot show that this sex difference is statistically significant in this one experiment; however, we have found this same sex effect in two additional experiments involving *Thamnophis sirtalis*, which we shall report at a later date. Although the gender difference in buffering appears not to be a statistical artifact, we have no explanation for it. One might suppose that the heterogametic sex would be less able to cope with environmental perturbations, but in snakes, females are the heterogametic sex (Baker et al. 1972).

Implications for Systematics

One of our most important results is that the standard scale counts used in snake systematics are remarkably buffered against temperature effects during development. Why did the earliest studies create an impression that these traits were thermally labile in snakes? Three design features contributed to this false impression. First, in the case of the studies by Fox and his colleagues (Fox 1948; Fox et al. 1961), a temperature treatment was used that was apparently outside the ecologically relevant range. The temperature was so low at the lowest treatment that most females failed to give birth to their litters. These females had to be sacrificed to obtain embryos for scale counting. The idea that the lowest temperature treatment was ecologically unrealistic is difficult to evaluate, however, because the investigators did not actually monitor temperatures in their two rooms. We know only the ranges of temperatures for the two treatments (18.3°–29.4°C vs. 23.8°–35°C), and these may represent only the daytime ranges. Second, Fox et al. (1961) and Osgood (1978) failed

Table 1: ANOVA in scalation abnormalities in female neonatal *Thamnophis elegans*

	df	HALF	WHOLE	SLAB ASYM	ILAB ASYM	PREOC ASYM	POSTOC ASYM
DAM	0–1	NS	NS	NS	NS	...	**
DATE	1	NS	NS	NS	NS	NS	NS
LOCALE	6–7	NS	NS	NS	NS	NS	**
MASS	1	NS	NS	NS	NS	NS	NS
TEMP	1	NS	NS	NS	NS	NS	*
TEMP ²	1	NS	NS	NS	NS	NS	*
Model	11–12	NS	NS	NS	NS	NS	***
R ²		.28	.32	.17	.10	.06	.53

Note: Degrees of freedom (df) vary from column to column because of missing values. Results are based on Type III sums of squares. Error: df = 41–53; total: df = 52–64.

* $P \leq .05$.

** $P < .01$.

*** $P < .001$.

to take family structure into account in their statistical analyses. Littermates were treated as independent data points in these analyses, giving a false sense of statistical power and Type II statistical errors. Third, relatively few temperature treatments were used, two in the case of Fox et al. and three in Osgood's case. When reaction norms are curvilinear, as in the case of some scalation abnormalities, a spurious linear relationship may appear if only a couple of temperature treatments are imposed, especially if one treatment lies outside the zone of canalization. This situation is especially likely in the Fox experiments. Thus, in past studies, an illusion of thermal lability was inadvertently created by imposing ecologically unrealistic temperature treatments, by inflating statistical power using littermates (rather than litters) as data points, and by using only two temperature treatments.

Our results suggest that geographic differences in meristic counts cannot be accounted for by temperature effects during snake development. We were unable to detect effects of developmental temperature on scale counts that are routinely used in systematics, despite numerous temperature treatments and substantial sample sizes in each treatment. This result, together with studies demonstrating heritable variation in these same counts (Beatson 1976; Arnold 1988; Dohm and Garland 1993; Arnold and Phillips 1999), suggests that geographic differences in meristic counts in snakes represent genetic differentiation caused by drift and/or selection rather than direct effects of the environment.

The Evolution of Reaction Norms

Our model for optimal reaction norms provides some insights into the evolution of reaction norms. Note that the model is for environmental canalization rather than adaptive phenotypic plasticity (Debat and David 2001). We are concerned with a trait that has a single optimum value

rather than multiple optima, depending on environment. When a trait is under stabilizing selection (single optimum), the primary effect of phenotypic plasticity is to reduce average fitness by producing environmental variance in the trait, σ^2 (eq. [3]). Selection will tend to eliminate plasticity entirely or mitigate its deleterious effects. Because of their contributions to environmental variance, selection will tend to favor two kinds of effects: reduction in the variance of environments, σ_e^2 , and the shape of the reaction norm in the vicinity of the environmental mean. In the garter snake case, thermoregulation by pregnant females can affect the thermal variance experienced by embryos (Peterson et al. 1993). Our model is telling us that selection should favor thermoregulatory behaviors that minimize the variance in temperatures experienced by embryos, σ_e^2 . More generally, selection should favor habitat selection and any other behaviors that minimize σ_e^2 .

Aside from variation in environments, the other major effect on average fitness comes from the shape of the reaction norm in the vicinity of the environmental mean, \bar{e} . If, for example, the reaction norm is U-shaped (as in fig. 4) and remains so on some short evolutionary time-scale, selection will favor evolution of the distribution of environments until \bar{e} corresponds to the inflection point of the reaction norm. This expectation is based on the fact that the first derivative of the reaction norm will initially be nonzero and then approach zero as \bar{e} approaches the inflection point. Presumably, an evolutionary history of this kind is responsible for the coincidence of thermal mean experienced by embryos (fig. 3) and the inflection points of reaction norms for scalation abnormalities in our garter snake example (fig. 4). This prediction could be further tested by determining whether the thermal average imposed on embryos by their mothers varies among garter snake species and co-varies with inflection points of thermal reaction norms. This prediction is of course

predicated on the assumption that thermoregulation is a heritable attribute, an assumption that gains some support from analysis of variation among litters (Arnold et al. 1995).

The joint evolution of thermoregulation and thermal reaction norms will be a complex process if the reaction norms of various traits differ in shape and position. In this study, scalation traits had either flat or U-shaped reaction norms, and in the later case, we were unable to demonstrate differences among traits in inflection points. Our sample of meristic traits is, however, only a small sample of the many behavioral, physiological, and morphological attributes that are potentially affected by temperature during development in snakes and lizards (Burger 1989; van Damme et al. 1992; Shine and Harlow 1993; Elphick and Shine 1998). To test for coincidence in the shape and position of reaction norms for this wider array of traits will require a departure from the experimental designs that have been used in the past. Most studies have imposed only two or three temperatures. Such a small number of temperature treatments is sufficient to give a test for temperature effects, but it is not sufficient to characterize the shape of the reaction norm. We advocate using at least five or six temperature treatments and centering the series at the average or modal temperature that is experienced in nature.

Another major lesson from this model is that the shape of the reaction norm itself should show a predictable evolutionary history. When a nonlabile trait is under stabilizing selection, selection will favor shape changes that progressively reduce and finally eliminate slope and curvature of the reaction norm in the vicinity of the environmental mean, \bar{e} (eq. [2]). Gomulkiewicz and Kirkpatrick (1992) have made the important point that quantitative inheritance may often constrain the possible shapes of reaction norms or at least the rates at which shape changes can be achieved. Although a flat reaction norm may be the best shape in the sense that it optimally buffers the

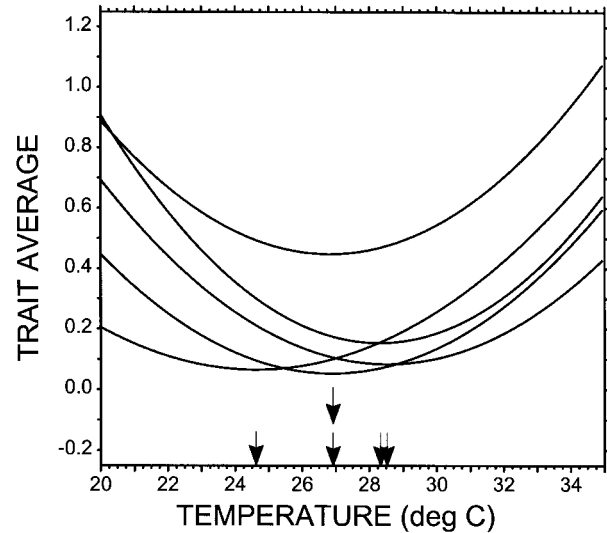


Figure 4: Norms of reaction to temperature during development for five kinds of scalation abnormalities in *Thamnophis elegans*. The curves represent quadratic regressions of litter averages for scalation abnormalities on temperature during development. Starting from the top on the far right side of the figure, the five curves are as follows: number of half ventrals in males (10 times average trait value), supralabial asymmetry in females, supralabial asymmetry in males, infralabial asymmetry in males, and postocular asymmetry in males. Arrowheads indicate the inflection points for the five curves.

organism from environmental effects, that shape may be difficult or even impossible to achieve depending on the values of additive genetic variation and covariation. For traits such as vertebral numbers that are likely to be under stabilizing selection, it remains an open question whether linear and curvilinear reaction norms represent transient evolutionary phenomena, departures from optima induced by genetic constraints, or a failure to focus on the ecologically relevant part of the reaction norm.

Table 2: ANOVA in scalation abnormalities in male neonatal *Thamnophis elegans*

	df	HALF	WHOLE	SLAB ASYM	ILAB ASYM	PREOC ASYM	POSTOC ASYM
DAM	1	NS	NS	NS	NS	*	NS
DATE	1	NS	NS	NS	NS	NS	NS
LOCALE	7	NS	NS	NS	NS	NS	**
MASS	1	*	NS	NS	NS	NS	NS
TEMP	1	*	NS	*	*	NS	*
TEMP ²	1	*	NS	*	*	NS	*
Model	12	NS	NS	NS	NS	NS	**
R ²		.28	.20	.20	.23	.17	.42

Note: Degrees of freedom (df) vary from column to column because of missing values. Results are based on Type III sums of squares. Error: df = 49–54; total: df = 52–66.

* $P \leq .05$.

** $P < .01$.

The Ecology of Reaction Norms

The significance of the shapes of reaction norms may be difficult to evaluate without field measurements of the environmental variable that is manipulated in a plasticity experiment (Via et al. 1995). In this experiment, the significance of the U-shaped reaction norms and their position was apparent because we had an estimate of the distribution of environments in nature. In much of the literature on reaction norms, such crucial environmental measurements are lacking, and the shapes of reaction norms remain mysterious. In the literature on the reactions of vertebral numbers in fish to temperature, for example, U-shaped reaction norms are relatively common (Fowler 1970), but the thermal ecology of embryos is seldom studied. This shape may find an explanation if the temperatures actually experienced by embryos could be determined. This failure to place reaction norms in an ecological context plagues much of the literature on phenotypic plasticity.

Curvilinear Maternal Effects and Reaction Norms

Our results underscore the need for theoretical models of nonlinear maternal effects and reaction norms. In the current generation of models, maternal characters (e.g., developmental temperature) have linear effects on offspring traits (e.g., scalation; Falconer 1965; Kirkpatrick and Lande 1989). Curvilinear reaction norms are common, however, for temperature effects on meristic traits in a variety of vertebrates (see the beginning of this article). Thus, we clearly need a model that includes the possibility of nonlinear maternal effects. Viewing our characters as the subjects of selection for canalization emphasizes the need for models with curvilinear reaction norms. One of the main contributions of this article is to point out that long-established stabilizing selection should affect the curvature of reaction norms as well as their average slopes.

Acknowledgments

We are grateful to J. Gladstone for her untiring efforts in snake husbandry and scale counting. We thank the many workers at Snake Camp for their help with the fieldwork. This study was supported by U.S. Public Health Service grant R01-GM35492 and National Science Foundation grants BSR 9119588 and DEB 9903934 (to S.J.A.).

Literature Cited

- Arnold, S. J. 1988. Quantitative genetics and selection in natural populations: microevolution of vertebral numbers in the garter snake *Thamnophis elegans*. Pages 619–638 in B. S. Weir, E. J. Eisen, M. M. Goodman, and G. Namkoong, eds. Proceedings of the Second International Conference on Quantitative Genetics. Sinauer, Sunderland, Mass.
- Arnold, S. J., and A. F. Bennett 1984. Behavioural variation in natural populations. V. Morphological correlates of locomotion in the garter snake (*Thamnophis radix*). *Biological Journal of the Linnean Society* 34:175–190.
- Arnold, S. J., and P. C. Phillips. 1999. Hierarchical comparison of genetic variance-covariance matrices. II. Coastal-inland divergence in the garter snake *Thamnophis elegans*. *Evolution* 53:1516–1527.
- Arnold, S. J., C. R. Peterson, and J. Gladstone. 1995. Behavioural variation in natural populations. VII. Maternal body temperature does not affect juvenile thermoregulation in a garter snake. *Animal Behaviour* 50:623–633.
- Baker, R. J., G. A. Mengden, and J. J. Bull. 1972. Karyotypic studies of thirty-eight species of North American snakes. *Copeia* 1972:257–265.
- Barlow, G. W. 1961. Causes and significance of morphological variation in fishes. *Systematic Zoology* 10:105–117.
- Beatson, R. R. 1976. Environmental and genetic correlates of disruptive coloration in the water snake, *Natrix s. sipedon*. *Evolution* 30:241–252.
- Bulmer, M. G. 1979. Principles of statistics. Dover, New York.
- Burger, J. 1989. Incubation temperature has long-term effects on behaviour of young pine snakes (*Pituophis melanoleucus*). *Behavioral Ecology and Sociobiology* 24:201–207.
- Clark, D. R., Jr., and G. L. Callison. 1967. Vertebral scute anomalies in a racer, *Coluber constrictor*. *Copeia* 1967:862–864.
- Debat, V., and P. David. 2001. Mapping phenotypes: canalization, plasticity and developmental stability. *Trends in Ecology & Evolution* 16:555–561.
- Dohm, M. R., and T. J. Garland, Jr. 1993. Quantitative genetics of scale counts in the garter snake *Thamnophis sirtalis*. *Copeia* 1993:987–1002.
- Dunn, E. R. 1942. Survival value of varietal characters in snakes. *American Naturalist* 76:104–109.
- Elphick, M. J., and R. Shine. 1998. Longterm effects of incubation temperature on the morphology and locomotor performance of hatchling lizards (*Bassiana duperreyi*, Scincidae). *Biological Journal of the Linnean Society* 63:429–447.
- Falconer, D. S. 1965. Maternal effects and selection response. *Proceedings of the Eleventh International Congress of Genetics* 3:763–774.
- Fowler, J. A. 1970. Control of vertebral number in teleosts: an embryological problem. *Quarterly Review of Biology* 45:148–167.
- Fox, W. 1948. Effect of temperature on development of

- scutellation in the garter snake, *Thamnophis elegans atratus*. *Copeia* 1948:252–262.
- Fox, W., C. Gordon, and M. H. Fox. 1961. Morphological effects of low temperature during embryonic development of the garter snake, *Thamnophis elegans*. *Zoologica* 46:57–71.
- Gavrilets, S., and A. Hastings. 1994. A quantitative-genetic model for selection on developmental noise. *Evolution* 48:1478–1486.
- Gomulkiewicz, R., and M. Kirkpatrick. 1992. Quantitative genetics and the evolution of reaction norms. *Evolution* 46:390–411.
- Hubbs, C. L. 1922. Variations in the number of vertebrae and other meristic characters of fishes correlated with temperature of water during development. *American Naturalist* 56:360–372.
- Inger, R. F. 1942. Differential selection on variant juvenile snakes. *American Naturalist* 76:527–528.
- . 1943. Further notes on differential selection of variant juvenile snakes. *American Naturalist* 77:87–90.
- Jockusch, E. L. 1997. Geographic variation and phenotypic plasticity of number of trunk vertebrae in slender salamanders, *Batrachoseps* (Caudata: Plethodontidae). *Evolution* 51:1966–1982.
- Kelley, K. C., S. J. Arnold, and J. Gladstone. 1997. The effects of substrate and vertebral number on locomotion in the garter snake *Thamnophis elegans*. *Functional Ecology* 11:189–198.
- King, W. 1959. Vertebral duplication, an osteological anomaly widespread in snakes. *Herpetologica* 15:87–88.
- Kirkpatrick, M., and N. Heckman. 1989. A quantitative genetic model for growth, shape, reaction norms, other infinite-dimensional characters. *Journal of Mathematical Biology* 27:429–450.
- Kirkpatrick, M., and R. Lande. 1989. The evolution of maternal characters. *Evolution* 43:485–503.
- Kirkpatrick, M., and D. Lofsvold. 1992. Measuring selection and constraint in the evolution of growth. *Evolution* 46:954–971.
- Kirkpatrick, M., D. Lofsvold, and M. Bulmer. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* 124:979–993.
- Lande, R. 1980. Genetic variation and phenotypic evolution during allopatric speciation. *American Naturalist* 116:463–479.
- Lecyk, M. 1965. The effect of hypothermia applied in the given stages of pregnancy on the number and form of vertebrae in the offspring of white mice. *Experientia* (Basel) 21:452–453.
- Lindell, L. E., A. Forsman, and J. Merila. 1993. Variation in number of ventral scales in snakes: effects of body size, growth rate and survival in the adder *Vipera berus*. *Journal of Zoology* (London) 230:101–115.
- Lindsey, C. C. 1966. Temperature-controlled meristic variation in the salamander *Ambystoma gracile*. *Nature* 209:1152–1153.
- Lindsey, C. C., and R. W. Harrington. 1972. Extreme vertebral variation induced by temperature in a homozygous clone of self-fertilizing cyprinodontid fish *Rivulus marmoratus*. *Canadian Journal of Zoology* 50:733–744.
- Lindsey, C. C., and G. E. E. Moodie. 1967. The effect of incubation temperature on vertebral count in the chicken. *Canadian Journal of Zoology* 43:99–104.
- Møller, A. P. 1997. Developmental stability and fitness: a review. *American Naturalist* 149:916–932.
- Osgood, D. W. 1970. Thermoregulation in water snakes studied by telemetry. *Copeia* 1970:568–571.
- . 1978. Effects of temperature on the development of meristic characters in *Natrix fasciata*. *Copeia* 1978:33–37.
- Peabody, R. B., and E. D. Brodie, Jr. 1975. Effect of temperature, salinity and photoperiod on the number of trunk vertebrae in *Ambystoma maculatum*. *Copeia* 1975:741–746.
- Peterson, C. R. 1987. Daily variation in the body temperatures of free-ranging garter snakes. *Ecology* 68:160–169.
- Peterson, C. R., A. R. Gibson, and M. E. Dorcas. 1993. Snake thermal ecology: the causes and consequences of body-temperature variation. Pages 241–314 in R. A. Seigel and J. T. Collins, eds. *Snakes, ecology and behavior*. McGraw-Hill, New York.
- Reinert, H. K., and D. Cundall. 1982. An improved surgical implantation method for radiotracking snakes. *Copeia* 1982:702–706.
- Rice, S. H. 1998. The evolution of canalization and the breaking of von Baer's laws: modeling the evolution of development with epistasis. *Evolution* 52:647–656.
- SAS Institute. 1989. *SAS/STAT user's guide*, version 6. SAS Institute, Cary, N.C.
- Scheiner, S. 1993. Genetics and evolution of phenotypic plasticity. *Annual Review of Ecology and Systematics* 24:35–68.
- Schlichting, C. D., and M. Pigliucci. 1998. Phenotypic evolution: a reaction norm perspective. Sinauer, Sunderland, Mass.
- Shine, R., and P. Harlow. 1993. Maternal thermoregulation influences offspring viability in a viviparous lizard. *Oecologia* (Berlin) 96:122–127.
- Tåning, A. V. 1952. Experimental study of meristic characters in fish. *Biological Reviews of the Cambridge Philosophical Society* 27:169–193.
- van Damme, R., D. Bauwens, F. Braña, and R. F. Verheyen. 1992. Incubation temperature differentially affects hatching time, survival, and hatching performance in the lizard *Podarcis muralis*. *Herpetologica* 48:220–228.

- van Tienderen, P. H. 1991. Evolution of generalists and specialists in spatially heterogeneous environments. *Evolution* 45:1317–1331.
- Via, S., and R. Lande. 1985. Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* 39:505–522.
- Via, S., R. Gomulkiewicz, G. de Jong, S. M. Scheiner, C. D. Schlichting, and P. H. Tienderen. 1995. Adaptive phenotypic plasticity: consensus and controversy. *Trends in Ecology & Evolution* 10:212–217.
- Wagner, G. P., G. Booth, and H. Bagheri-Chaichian. 1997. A population genetic theory of canalization. *Evolution* 51:329–347.
- Yntema, C. L. 1960. Effects of various temperatures on the embryonic development of *Chelydra serpentina*. *Anatomical Record* 136:305–306.

Associate Editor: Donald B. Miles