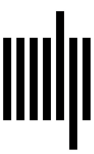
Evolvability

A Unifying Concept in Evolutionary Biology?

Edited by: Thomas F. Hansen, David Houle, Mihaela Pavličev, Christophe Pélabon

Citation:

Evolvability: A Unifying Concept in Evolutionary Biology?
Edited by: Thomas F. Hansen, David Houle, Mihaela Pavličev, Christophe Pélabon
DOI: 10.7551/mitpress/14126.001.0001
ISBN (electronic): 9780262374699
Publisher: The MIT Press
Published: 2023



The MIT Press

6 Measuring Evolvability

David Houle and Christophe Pélabon

The term evolvability is used in many different contexts in evolutionary biology, and this multiplicity has generated confusion. Here we examine the varied usages of evolvability from the perspective of conceptual measurement theory, the consideration of what attributes should be measured to quantify a concept. We argue that there is a shared conception of evolvability as a disposition, the propensity to evolve should the conditions promoting evolution occur. Even with that shared conception, identifying the properties that confer evolvability requires making explicit which entity's evolvability is of interest, the specific stimulus that would cause evolution, and the timescale over which evolution happens. We refer to these descriptors as *Of, Under,* and *Over,* respectively. Once *Of, Under,* and *Over* are clear, attributes of the properties that confer evolvability can be identified and then measured. Focusing on both the commonalities and the differences of evolvability concepts should help us develop the theoretical understanding that is a precursor to appropriate measurement.

6.1 Introduction

Evolvability is the disposition of a population to evolve (G. Wagner and Altenberg 1996; Love 2003; Hansen 2006; Brigandt et al., chapter 4).[1] The word is, however, used to refer to many somewhat different dispositions (Pigliucci 2008; Minelli 2017; Nuño de la Rosa 2017), and there is an even larger menagerie of attributes that are hypothesized to cause or shape evolvability. These attributes include mutation, pleiotropy, robustness, modularity, variability, and variation, among many others. On its face, this multiplicity of usages and attributes related to the concept suggests that evolvability is not a unifying concept.

The life cycle of concepts in science progresses from intuitive assertions through precise verbal models to increasingly rigorous mathematical models. In parallel, scientists identify relevant attributes in nature and learn to quantify them with increasing accuracy. With a mature understanding of evolvability, we could aspire to have the equivalent of a thermometer, call it an evometer, which would tell us how much disposition to evolve a population has. Such understanding implies that we would have a complementary theoretical appreciation for what that disposition actually is, and what it predicts about evolution.

1. References to chapter numbers in the text are to chapters in this volume.

From our current state of knowledge, the idea of an evometer seems plainly ridiculous. It is less ridiculous, however, if we remember that not so many centuries ago, there was no such thing as a thermometer either (McGee 1988; Chang 2004). Without a thermometer, an assertion that yesterday was hotter than today could not be verified without a component of storytelling, memory, or appeal to authority. And what does it mean to be hotter, anyway?

We offer the example of temperature and heat to make two points. First, having an evometer would imply knowing both what we mean by evolvability, and what role it plays in evolution. Chang (2004, 8) repeats the truism that "the scientific study of heat began with the invention of the thermometer," although physicists might substitute the term "energy" for "heat." Hand (2004, 2) quotes Lord Kelvin, who did so much to advance our understanding of the nature of energy on this point: "when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it . . . your knowledge of it is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts advanced to the stage of science."

Second, just because we cannot now conceive of an evometer to match each proposed concept of evolvability is no reason to discard that concept. If discussions of heat had been deemed too unscientific to be worthy of attention by Galileo and the others who rediscovered the thermoscope around 1600 (McGee 1988), we might have neither a well-developed theory of thermodynamics nor the ability to measure temperature today.

Many contributions to the literature on evolvability are at this pre-measurement level. For example, Hendrikse et al. (2007, 394) argue that evolvability, "the capacity of a developmental system to evolve" is the "central question" of evolutionary developmental biology. The authors focus on the ability of a developmental system to "generate" variation as the key to evolvability and separate this ability into bias in the direction of phenotypic effects and the amount of phenotypic variation generated. They further associate four other concepts with the study of this ability: integration, modularity, constraint, and canalization. This clearly argued essay has been influential and widely cited, and we agree that development shapes evolvability through features such as these. And yet this paper makes no reference to quantification whatsoever. We are only at the "beginning of knowledge" when it comes to some aspects of evolvability.

In this chapter, we consider evolvability and the attributes that may shape it from the perspective of measurement (Hand 2004). In particular, we focus on the principles that should guide our choice of attributes to be measured when measuring evolvability rather than on the details of how to quantify particular attributes. Houle et al. (2011) called this conceptual measurement theory.

We argue that the great complexities that immediately confront anyone wishing to measure the disposition to evolve can be surmounted by focusing on two important distinctions that are rarely made explicit. The first is that evolvability is caused by different attributes depending on the evolving entity, the evolutionary forces acting, and the timescale considered. The second is that many of the attributes that shape evolvability, including those featured in Hendrikse et al. (2007), have complex and nonmonotonic relationships with evolvability. Although we may be able to measure attributes such as modularity, they only measure evolvability under carefully delineated assumptions.

6.2 The Language of Dispositions

All conceptions of evolvability share a focus on the evolutionary future, that is, on predicting the possible state of a population in the future. This focus defines evolvability as a dispositional concept (Love 2003). We adopt much of the philosophical vocabulary concerning dispositions that Brigandt et al. (chapter 4) review. A disposition is the potential for some entity to manifest a particular change, when subjected to some set of stimulus conditions under appropriate background conditions. In the case of evolvability, the evolving entity must be a population of organisms or some set of populations. Most importantly for measurement, a disposition must be at least partly due to an intrinsic property of the bearers of the disposition. The bearer of a disposition may be distinct from the entity that manifests the change. Only populations can evolve, but the bearers of the properties hypothesized to generate evolvability can be either a population as a whole or a typical member of the population, such as an organism, a genome, or their components, a gene or an organ, which is termed a *type.* The individual bearers of the disposition are referred to as *tokens,* which may differ in their properties from the type.

This philosophical vocabulary only deals with qualitative outcomes in response to a discrete stimulus, such as the shattering of a fragile object when struck. The intrinsic properties that enable the disposition are also generally treated as discrete. An object is considered either fragile or not, even though fragile objects differ in their fragility. Bearers, however, can possess the relevant intrinsic property to different degrees. We refer to quantifiable intrinsic properties as *attributes.* Similarly, we seek to quantify the *stimulus strength* and the *manifestation strength.* The manifestation strength depends on both the amount of the attribute present and the stimulus strength.

To put this vocabulary into use, consider first a well-known dispositional property, solubility. The stimulus condition is placing a solid in a solvent, and the manifestation is solvation, passing into solution. What makes solubility a disposition is that a solid (e.g., sugar) will not pass into solution until exposed to a stimulus (e.g., water), whether that exposure happens today or next year. The attributes that determine solubility include the intermolecular interactions of the solid and solvent. These properties are responsible for differences in solubility, stimulus strength, and manifestation strength. For example, salt and sugar are both soluble in water. If we hold the stimulus strength constant at 1 liter of 20°C water, we can dissolve 2,000 g/l of sugar, but only 360 g/l of salt. Sugar is thus 5.5 times more soluble than salt in water.

With respect to evolvability, it is relatively straightforward to determine the manifestation strength (how much evolution has occurred). It is much more challenging to determine how much of that change is due to the disposition to evolve, and how much to stimulus strength (see Armbruster, chapter 15, and Jablonski, chapter 17, for further discussion in the context of macroevolution).

6.3 The Evolvability Multiverse

To apply the language of dispositions to evolvability, we need to recognize the diversity of usages of the concept (Pigliucci 2008; Nuño de la Rosa 2017). Nuño de La Rosa's

(2017) comprehensive review of the foundational literature suggests that the different conceptions of evolvability do share important common elements, such as the recognition that evolution consists of changes in the complement of genotypes in populations. Despite those commonalities, important differences preclude reducing evolvability to a single measurable disposition.

Nuño de La Rosa (2017) noted four "conceptual tensions" that characterize the differences among evolvability concepts: (1) Should research on evolvability focus on variability or variation? (2) Are we interested in all evolutionary changes, or just those of adaptive significance? (3) Do we want to focus on evolutionary innovations or novelties rather than other evolutionary changes? (4) Which organismal characteristics should evolvability be applied to and over what timescales? These questions partly echo the distinction that Pigliucci (2008) made between variation-, variability-, and innovation-based evolvability concepts. In Pigliucci's view, variation-based measures apply at short timescales, variability-based measures at intermediate timescales, and innovation-based ones at long timescales.

Part of the diversity in evolvability concepts arises because aspects of evolvability can be measured at four different biological levels: genetics, the genotype-phenotype (GP) map, variability, and variation (figure 6.1). The genetic level consists of processes that alter genotypes, including mutation, recombination, and segregation. The GP map level incorporates the tangled web of processes through which genotypes shape phenotypes. It includes development and physiology at cellular, tissue, organ, and whole-body scales. Variability is the disposition for a genetic change filtered through the GP map to generate phenotypic change. Variation refers to the differences in a particular population that result from the variability and the history of that population. This scheme explicitly distinguishes variation and variability, and it implicitly acknowledges the range of timescales by separating elements of evolvability that tend to evolve at a low rate (e.g., mutation, and the GP map) from variation, which can change rapidly (Pélabon et al., chapter 13). Some evolution is driven by stimuli that arise above the population level, such as extinction (Jablonski 2008). We did not try to represent those drivers in figure 6.1, which would require at least one additional level to represent species-level properties, such as species ranges (Grantham 2007; Villegas et al., chapter 3).

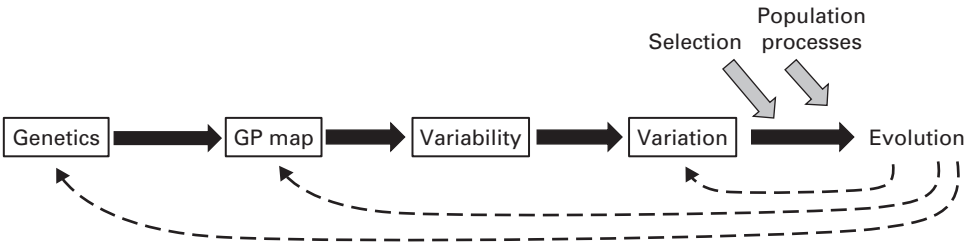


Figure 6.1
Flow of biological processes that shape evolution. Aspects of evolvability can be conceptualized at each of the levels shown in boxes: genetics, GP map, variability, and variation. Gray arrows represent processes that cause evolution through their effects on variation, notably natural selection and population processes, including drift and gene flow. The dashed lines represent the effects that evolution may have on variation, genetic processes (e.g., mutation and recombination), and the GP map.

To see how we can apply the language of dispositions to this scheme, let's consider mutation rate, the canonical evolvability attribute at the genetic level. The bearer of mutation rate is an individual organism, including especially its genome and internal physiological state. Mutation rate can be measured on the gametes produced by a single individual (see, e.g., Wang et al. 2012), although many estimates are made on the descendants of a set of individuals (e.g., Haag-Liautard et al. 2007). Such estimates of mutation rate refer to an individual as the representative of a type, specifically a typical genotype, itself a product of past population-level processes. Thus, the individual and its genome contain the relevant information necessary to study mutation rate. Although we might choose to measure mutation rate in only part of the genome, such as one gene, the actual rate is determined both by the properties of that gene and by the genomic and organismal milieu in which the gene resides. At the next level, the GP map specified by the entire genetic and developmental system of the individual determines whether a genomic variant has a phenotypic effect as well as the magnitude and direction of any effect. Mutation and the GP map together determine the variability of offspring potentially produced by each genotype. Ultimately, the mutation rate affects the rate of evolution of a population through its effect on the amount and nature of variation produced by the population, the evolving entity.

6.4 Of, Under, Over

Measurement is the process by which we assign numbers to the attributes of entities so that the mathematical relationships among the numbers capture the empirical relationships between the original entities (Krantz et al. 1971; Hand 2004). The theoretical context for measurement tells us what inferences we would like to make. In the case of evolvability, the entities are populations of organisms, or higher-level aggregations of populations, and the theoretical context is the hypothesis that intrinsic properties of individuals or populations give them the ability to evolve. Evolution is not a unitary process, and therefore the notion that "ability to evolve" is a single unitary disposition cannot be correct. The forces that cause evolution in one trait may differ from those affecting another trait; the attributes that enable evolution of novel organismal features are likely to be different from those that enable quantitative alterations of existing traits.

Therefore, to measure evolvability, we need to be more specific about the theoretical context in which the measurements will function. First, we must specify evolvability *Of* what class of organismal or population feature. Examples might be a trait, a DNA sequence, or a property of the GP map. Second, we must define the stimulus *Under* which we want to measure evolvability. Stimuli that create the opportunity for evolution include particular kinds of natural selection or genetic drift. Finally, we need to know *Over* what timescale we want to predict evolution. Attributes that predict short-term evolution might be irrelevant for long-term evolution. Specification of the *Of-Under-Over* conditions supplies the essential theoretical context for measurement.

Table 6.1 lists some of the many possible choices for *Of-Under-Over*. Note that properties in the *Of* column require further specification (i.e., which trait or GP map property, or fitness of what entity). Not all combinations of *Of-Under-Over* are sensible; for example, the evolution of novelty under strict stabilizing selection in all phenotypic dimensions is not likely.

Table 6.1
Examples of choices for *Of*, *Under*, and *Over*

Of	Under	Over
Allele	Directional selection	Cell cycle
Genome	Neutral process	One generation
Trait mean	Approach to an optimum	Hundreds of generations
Trait variance	Corridor selection	Macroevolutionary time†
Discrete state	Fluctuating selection	
Fitness	White noise motion of optimum	
GP map	Brownian motion of optimum	
Mutation rate	Mutation pressure	
Recombination		
Novel traits		
Speciation rate		
Species range		

† Time is frequently assessed in years when generation time is unknown, evolves, or when multiple organisms with different generation times are involved.

6.5 From Intrinsic Property to Attribute to Measurement: Two Examples

To illustrate the link between intrinsic properties, attributes, and measurements that are suggested by particular theoretical contexts, let's consider two cases where the measurement of evolvability is based on well-established theoretical models: short-term evolution of quantitative traits in response to directional selection, and the neutral evolution of discrete properties of genotypes.

6.5.1 Response of Quantitative Traits to Directional Selection

Additive genetic variation quantifies the evolvability *Of* the mean of a quantitative trait, z, in an outbred population *Under* directional selection *Over* one generation. The relevant theory is based on the Lande equation, $\Delta\bar{z} = V_A\beta$, where $\Delta\bar{z}$ is the change in the mean of trait z due to selection, V_A is the additive genetic variance in a trait z, and β is the selection gradient (Lande 1979). In this model, the manifestation strength is $\Delta\bar{z}$. The quantity V_A measures the evolvability of the population. Although it is often narrowly defined for random-mating diploid populations, we use V_A more broadly to encompass the inherited variation that causes offspring to resemble their parents, regardless of ploidy and mating system (Sztepanacz et al., chapter 12). It is estimated from the phenotypic resemblance of related and unrelated individuals and thus is an attribute of a population. The selection gradient β quantifies the stimulus strength, and it is the change in relative fitness w for a unit change in z. Mathematically, $\beta = COV(z, w)/V_z$, where $COV(z, w)$ is the covariance between z and relative fitness (w), and V_z is the phenotypic variance in z, the sum of V_A and the environmental and other causes of variance, which we refer to as V_R. Thus, β is not just a property of the population that evolves but the result of the interaction of the population with its environment. The derivation and significance of the Lande equation is developed in Hansen's chapter (chapter 5).

We often want to compare V_A or β across traits, populations, or species. When traits are measured in the same units, the values can be compared directly, bearing in mind that V_A

has units of the trait squared, while β has units of relative fitness/trait unit. When the traits to be compared are measured in different units, however, standardization is essential.

Since we often care about the proportional change in trait values, a natural way to standardize the Lande equation is by the trait mean, yielding response as a proportion of the trait's starting value

$$\frac{\Delta\bar{z}}{\bar{z}} = \frac{V_A}{\bar{z}^2}\bar{z}\beta = e_\mu\beta_\mu, \tag{6.1}$$

where e_μ is the mean-scaled evolvability, and β_μ is the mean-scaled selection gradient (Hansen et al. 2003b, 2011; Hereford et al. 2004. A key advantage of this standardization is that $\beta_\mu = 1$ when the trait is fitness, providing a natural marker for strong selection (Hereford et al. 2004). Mean-scaled evolvability, e_μ, offers a useful metric to compare changes in V_A when other components of the phenotypic variance are changing simultaneously, for example, when drift or environment affects both V_A and V_R (Hoffmann and Merilä 1999; Whitlock and Fowler 1999), or when the trait mean differs due to selection, environment, or inbreeding. Still, comparing evolvability using e_μ assumes that the variance can be meaningfully expressed as a proportion of the trait mean (Houle et al. 2011; Pélabon et al. 2020), which is not the case for variables on an absolute scale, such as probabilities, or for those on interval or ordinal scales. For many traits, particularly morphological traits (Gingerich 2000), variances are positively correlated with means, such that we expect that mean standardization will tend to homogenize e_μ of traits with different means. For other traits, such as clutch size in birds (Pélabon et al. 2020) or the traits described by G. Wagner (chapter 10), changing the trait mean will change e_μ and the proportional response to selection. For these traits, evolution of the trait mean causes evolution of evolvability (Hansen and Wagner, chapter 7).

Prior to Lande (1979), quantitative geneticists wrote the response equation as the "breeder's equation" $\Delta\bar{z} = V_A\beta = h^2S$, where $h^2 = V_A/V_p$ is the heritability, and $S = COV(z, w)$ is the selection differential (Falconer 1981). Because heritability is a dimensionless quantity, it may seem natural to standardize both sides of the equation by the phenotypic standard deviation, $\sqrt{V_z}$, yielding response in units of standard deviation,

$$\frac{\Delta\bar{z}}{\sqrt{V_z}} = h^2\frac{COV(z, w)}{\sqrt{V_z}} = h^2\sqrt{V_z}\beta = h^2\beta_\sigma, \tag{6.2}$$

where β_σ is the intensity of selection, symbolized i in the animal breeding literature. This standardization invites one to see h^2 as a measure of evolvability and β_σ as a measure of selection. However, because V_A is part of V_z, the factor that controls evolvability is part of the standardization (Houle 1992), making this variance standardization a "rubber ruler" (Houle et al. 2011). It also confounds evolvability and selection by multiplying β by a function of the evolvability. The pernicious impact of this standardization is perhaps most striking when there is a linear relationship between and z and w. In this case, by definition, an increase in V_A increases evolvability but leaves β unchanged. The Lande equation thus shows that $\Delta\bar{z}$ increases and that the reason for this is the increase in evolvability; the mean standardized equation (6.1) reflects this fact. However, in equation (6.2), when V_A increases, both h^2 and β_σ increase, obscuring the fact that selection has not changed. Furthermore, the rubber ruler effect increases the standard deviation used as a measuring stick. Comparison of response using equation (6.2) in populations with different levels of V_A will

therefore understate the effect of increased evolvability on response to selection and spuriously suggest that the strength of selection has increased.

6.5.2 Neutral Evolution

The neutral model predicts the evolution of organismal features when mutation and drift are the only evolutionary forces acting (Kimura 1983). Drift is the random sampling of alleles from one generation to the next due to stochasticity in reproduction or survival. Let's further restrict our attention to the rate at which discrete genotypic differences evolve between two populations. *Of* is thus any discrete neutral feature, such as a DNA or an amino acid sequence, and *Over* is whatever timescale we care to specify. The *Under* conditions are explicitly mutation, drift, and effective absence of selection.

Modeling this neutral evolutionary process requires two key parameters, the population size N, and the rate of mutation to new neutral variants, μ. When there are alternative genotypes at a given generation, the necessary outcome of the drift process is that eventually all the individual alleles in the population will descend from a copy of just one allele, which is called a fixation event. In a diploid population of N individuals and $2N$ alleles, there will be $2N\mu$ new neutral mutations in each generation and a probability of $1/(2N)$ that each mutation eventually becomes fixed. Thus, the rate of accumulation of differences between ancestor and descendant per generation, k, is

$$k = 2N\mu \times \frac{1}{2N} = \mu. \tag{6.3}$$

The simplicity of this outcome is a key to the usefulness of the neutral theory. This equation also shows that the neutral mutation rate, μ, is the attribute that generates evolvability under the neutral theory.

This seemingly straightforward conclusion hides two complications. First, how do we know which variants are neutral? The absence of selection that makes μ a dispositional parameter is difficult to measure directly, and variants with small fitness effects can be effectively neutral when the influence of drift at rate $1/(2N)$ is much greater than the effect of natural selection (Ohta 1992). This means that μ is entangled with N; as N decreases, the effective μ increases. Thus, even if we call μ a mutation rate, it is a function of a molecular mutation rate, the GP map that determines the phenotypic effect of the mutation, the shape of the fitness landscape, and N. The second complication is that the population size parameter, N, is not the census count of individuals in the population, but the effective population size, N_e, a complex function of the structure and history of the population that usually makes N_e much less than the census size (see e.g., Charlesworth and Charlesworth 2010, chapter 5.2).

We do have strong models for some aspects of sequence data, and in these cases, separating the innate mutation rate from assumptions about selection is relatively straightforward. For example, mutations to synonymous codons are likely to be effectively neutral in many species with population sizes that are not too large. Mutation rates at the molecular level are readily measured, and, at least to a first approximation (Hodgkinson and Eyre-Walker 2011), the mutation rate does not differ depending on whether the sequence is neutral. This makes it possible to parameterize models of robustness and evolvability for protein sequences (Ancel Meyers et al. 2005) and RNA secondary structure (A. Wagner 2008). However, attempts to measure μ for classes of traits that plausibly have neutral networks (Schuster

et al. 1994), such as gene regulatory networks, or morphological traits, are speculative, as we do not know when changes in those organismal features generate equivalent fitness.

The consistency of genomic evolutionary rates (Kumar 2005) suggests that μ usually evolves quite slowly, arguing that the model applies *Over* fairly long timescales, mostly because the number of generations from the lucky mutation event to when it takes over the population is of order $4N_e$ generations.

The stimulus for evolution by drift is population size, and its strength is $1/(2N_e)$. Taken literally, equation (6.3) suggests that the existence of neutral mutations ($\mu > 0$), is a sufficient cause for evolution to occur under the neutral model, similar to radioactive decay, where the disposition of an atom to decay is in itself the cause of its manifestation. This makes the typical genotype with its GP map the bearer of the mutation rate. This literal interpretation, however, ignores the fact that effective neutrality is also influenced by N_e. Taking this into account makes the population the bearer of the attribute μ. Somewhat uncomfortably, this result places N as both a stimulus and one of the background conditions that shapes μ. To place μ at the genotype level, we could assume a class of mutations that are effectively neutral unless the population size is very large. For example, vertebrates generally have effective population sizes of order 10,000 (chapter 4 in Lynch 2007), justifying a treatment of N_e as a background condition to μ for variants that will still evolve neutrally at all smaller values of N_e.

6.6 Measurement and Screening Off

We claimed that V_A and μ are the relevant dispositional attributes that dictate evolvability *Under* short-term directional selection and neutral evolution, respectively. Using the concept of screening-off (Salmon 1971), we can make the case that V_A and μ are not just relevant to evolvability *Under* directional selection and drift, but that they are also the best measures of evolvability under those scenarios.

Imagine a causal sequence where a distal cause (D) leads to a proximal cause (P) that in turn leads to a manifest change (M). P screens-off D when all you need to know to predict M is P. This is true when a prediction based on P alone is equivalent to prediction based on both P and D, and it is different from a prediction based on D alone. Screening-off is also helpful when trying to sort out correlates from causes. If we add a causally irrelevant feature C, which is correlated with P because it is also an outcome of cause D but has no direct effect on M, C is screened-off from prediction by both D and P. Screening-off has been used in evolutionary biology as a tool to identify which factors in a causal chain are "better" explanations for an outcome (Brandon 1982; Brandon et al. 1994). In particular, Brandon (1982) claimed that screening-off justifies Mayr's (1963, 184) intuition that "natural selection favors (or discriminates against) phenotypes, not genes or genotypes."

Applying the concept of screening-off to the evolvability *Of* a trait mean, *Under* directional selection, *Over* one generation, V_A is the only evolvability attribute that is appropriate in the Lande equation, because V_A screens-off the mutation rate, GP map and population history that cause V_A to have its actual value. We can build better understanding of evolvability by focusing on the distal attributes that affect V_A (Sober 1992), but we will not achieve better predictive power in the single generation considered in the Lande model. After the first generation, V_A itself can change due those very attributes that are screened off.

6.7 Evolvability Attributes

Several attributes have been hypothesized to affect population-level evolvability. We list the most important of these in table 6.2 and discuss them further in the following sections. These attributes are specific to the *Of, Under,* and *Over* conditions considered. For example, if *Of* is the transition from one base pair to another, we will measure evolvability as the probability that a mutation occurs per unit time. Alternatively, our *Of* might be the number of mutational changes in the entire genome *U*, which takes on any positive integer and is referable not just to the rate of mutation but also to the average number of base pairs involved in each mutational event. The *Under* column lists some simple evolutionary stimulus scenarios, including directional selection favoring only a change in the trait mean (D), evolution on a curved fitness landscape (C) that may favor a change in the mean, while simultaneously selecting on other aspects of variation, neutral evolution (N), or mutation pressure alone (M). The *Over* column denotes the relevant timescales from one generation to macroevolutionary trends. We do not include attributes that may apply above the population level.

6.7.1 Genetic

The first class of attributes concerns the processes of *mutation* and *recombination* on haplotypes inherited from parents to offspring. Sequencing allows the rates of these processes to be measured at the genotypic level, without any reference to phenotype, placing these processes at the far left of figure 6.1. Mutation and recombination are usually measured as the rate of discrete mutations or recombination events per generation per base

Table 6.2
Attributes hypothesized to affect population-level evolvability

Intrinsic property	Intrinsic level†	Bearer‡	*Of*		*Under*§	*Over*‖
			Discrete	Continuous		
Genomic mutation	G	I	Rate	Bp affected	Any	I, L
Recombination	G	I	Rate	Gene conversion tract size	Any	I, L
Mutational effect	GP	I	Robustness, probability	Canalization, effect size	N, M, D	I
Conditional effect	GP	I	Rate	Conditional effect size	C	I
Pleiotropy	GP	I	Conditional probability	Angle to trait	C	I
Plasticity	GP	I	Switch points	Reaction norm	D, C	S, I
Integration	GP	I	Covariance	Covariance	C	I
Modularity/autonomy	GP	I	Modularity	Modularity	C	I
Versatility	GP	I	Dimension	Dimension	Any	I, L
Mutational impact	Vy	I	Mutation number, bias	Mutational bias, variance	N, M, D	I
Conditional impact	Vy	I	Mutation number	Mutational variance	C	I
Genetic variation	Vn	P	Diversity H, π	Additive genetic variance	N, D	S
Conditional variation	Vn	P	Conditional rate	Conditional variance	C	S

† Level: G, genetic; GP, property of GP map; Vy, variability; Vn, variation.

‡ Bearer of the property: I, Individual organisms or their components, such as genotype, gene, or trait; P, population.

§ Under: D, directional selection; C, evolution on a curved fitness landscape; N, neutral; M, mutation.

‖ Over timescale: S, short (1 to 10s of generations); I, intermediate ($10^2 - 10^6$ generations); L, long ($>10^6$ generations).

pair. They also affect a continuously distributed number of base pairs. A single mutational event, for example, can alter a variable number of base pairs (Schrider et al. 2013), as does the recombinational process of gene conversion. On a genome-wide level, the total number of mutations or recombination events may be large integer values. These rates generally evolve slowly, suggesting that measurements are most relevant over intermediate to long timescales.

6.7.2 GP Map

We identify several evolvability attributes at the level of the GP map, because they determine phenotypic effects conditional on the existence of genetic (or environmental) variation. These are: mutational effects, conditional effects, pleiotropy, integration, modularity, versatility, and plasticity. One can measure them without regard to the rate of genetic changes. For example, one could engineer novel genetic variants to quantify these properties. This makes the typical genotype the bearer of the GP map properties.

Most fundamentally, a *mutational effect* is quantified by the probability that a mutation has a discrete effect, or as the average effect of a mutation on a continuous trait. Interactions between genotypes may cause mutational effects to differ either due to dominance of allelic variation at the site of the mutation or epistatic interactions of alleles at different sites in the genome. Epistatic interactions of mutational effects are what generates evolvability of the GP map (Hansen 2006, chapter 5).

Fixation of a mutation may alter genetic robustness or canalization, affecting variability in subsequent generations. The genetic robustness of a particular trait, such as an amino acid sequence or the 3-dimensional structure of a polymer can be measured as the probability of mutational effect on the phenotype. This makes robustness difficult to measure, as the absence of an effect can never be established experimentally. Measures of discrete effects are useful in models of mutation pressure, mutation load, neutral evolution, and developmental systems drift.

The term canalization was coined to refer to the process of evolving reduced effect size (Waddington 1942). It is now often used to refer to the relative effect sizes of a mutation in two different genotypes; the genotype in which a variant has a larger effect is less canalized than the one with a smaller effect (Flatt 2005). Continuous effect sizes can be measured directly for mutations with large effects, such as gene knockouts, or indirectly from the total mutational variance, coupled with an estimate of the number of mutations that potentially cause that variation.

Pleiotropy and *conditional effects* are complementary ways to characterize mutational effects on multiple traits. Conditional effect quantifies the probability or size of phenotypic effects, conditioned on the absence of some other effect(s), whereas pleiotropy is the tendency of variants to affect more than one trait. These two ways of looking at effects differ when our a priori definition of a trait is not the same as what selection sees. For example, if we measure long bone lengths, such as the femur and tibia, it is natural to characterize pleiotropy as the proportion of mutations that affect both bones, or as the angle between the multivariate direction of an effect and the trait axes. However, if selection favors an increase in both bones, it might be more natural to see the relevant trait as size, and a mutation that affects both traits as nonpleiotropic. Conditional effects are useful in the context of models of evolution on curved fitness landscapes, where, for example, fitness is maximized by

changing the relative lengths of the leg bones while holding leg length constant. Conditional effect sizes can sometimes be estimated for a given set of phenotypes, but identifying the full set of phenotypes pleiotropically affected by a given genetic variant is not currently feasible (Paaby and Rockman 2013).

Integration and *modularity* concern the collective pleiotropic and conditional effects of the genome-wide distribution of mutational effects. Integration is minimally defined as the degree of covariation among traits (e.g., Olson and Miller 1958; Cheverud 1996; Armbruster et al. 2014), and *modularity* is defined as the degree to which sets of integrated traits covary less with other sets of traits (Hendrikse et al. 2007; Klingenberg 2008). These definitions make no reference to adaptation and thus concern what Armbruster et al. (2014) term phenomenological integration and modularity, in contrast to the more restrictive sense of adaptive integration and modularity (see Pavličev et al., chapter 8). Many indices of integration have been proposed (Armbruster et al. 2014), and measures of modularity build on these measures of integration to identify clusters of integrated traits that are relatively independent of other such clusters (Zelditch and Goswami 2021). Much of the literature on integration and modularity focuses on morphology, but the concepts apply more generally to other suites of potentially correlated traits, such as gene expression or behavior.

Plasticity is the relationship between the phenotype and the environment in which an organism exists. It can be summarized as a function that relates the distribution of phenotypes to the environment, called a reaction norm. Although we presented mutational effects without explicit references to environment, these are more realistically approached as the study of the effects of genetic variation on reaction norms. In this context, genetic or environmental canalization is the evolution of those reaction norms, considering either the genetic or environmental background in which mutations occur. The importance of plasticity to the GP map and then for evolvability is evident (Sultan 2017) but extremely difficult to measure.

Vermeij (1973a,b) termed the number of dimensions in which the phenotype can vary *versatility* and further proposed that versatility enhances evolvability by providing more alternative phenotypes for selection to assay and increase the prospects for evolution of novel traits. For example, Vermeij observed that primitive mollusk shells were simple linear forms, while derived taxa evolved shells that coil in 2 and later 3 dimensions and speculated that a gain in dimensionality enabled the increased variety of forms.

6.7.3 Variability and Variation

Variability is the propensity of an individual to produce phenotypic variation due to genetic events filtered through the GP map, and is measured as the rate of increase in such variation. Variation measures the degree to which genomic or phenotypic properties differ among members of a population. Although closely related conceptually, a key distinction between variability and variation is that many evolutionary processes can alter variation on a short timescale (Pélabon et al., chapter 13), whereas properties at the mutational and GP map levels are likely to evolve more slowly, rendering estimates of variability relevant over longer timescales than variation. Attributes that we treated as part of the GP map, such as mutational and conditional effects, are quantified through their effect on variability and variation.

Mutational impact encompasses both the number and the effects of mutations. Discrete mutational impacts measure the number of mutations that have a specified effect. Measures

of continuous mutational impact characterize how mutation changes the distribution of phenotypes in the population. The effect on the trait mean results from mutational bias, while the effect on the variance is a function of the mutational variance, V_M, the increase in genetic variation from a single generation of mutation. Different definitions of V_M focus either on mutational effects that increase V_A (as described in section 6.5.1) or the effects of mutations once they are fixed (Lynch and Hill 1986).

Genetic variation can be measured for either discrete properties, as the probability of genetic differences between randomly chosen individuals (H or π), or for quantitative properties measured as variance. These variability and variation attributes are frequently used in models of mutation acting alone, of response to selection, and of change in a neutral model.

Measures of *conditional impact* quantify the effects of mutations on a focal trait or set of traits when effects on other traits are held constant (Hansen 2003; Hansen et al. 2003a; Hansen and Houle 2008). Conditional effects depend on the integration and modularity of the GP map but have a more direct relationship to evolvability than their GP map counterparts. If we are interested in the evolution of trait X, holding trait Y constant, we can directly ask: What is the evolvability of trait X if selection favors a change in X, while holding Y at its current value? Despite the conceptual clarity of this connection between conditional properties and an evolutionary model combining directional and stabilizing selection, it is often doubtful that all appropriate traits to condition on have been identified, particularly when pleiotropic interactions affect traits expressed at different ages or life stages.

The measurement of mutational variability is still challenging in most species, as mutations are individually rare, and their phenotypic effects are often small. Sequencing of parents and their descendants readily generates estimates of genomic mutational variability. At the phenotypic level, a typical design is a mutation-accumulation experiment, where mutations are allowed to build up in an initially homogeneous set of lines in the absence of natural selection (Houle and Kondrashov 2006). After a substantial number of generations, the cumulative change in mean and variance are measured. This can only be accomplished in model organisms that allow such designs. In contrast, genetic variation can readily be studied in any organism. Sequencing directly measures discrete genotypic variation. Quantitative genetic variation can be estimated by quantifying the relative similarity of related and unrelated individuals.

6.8 Measurement of Evolvability Attributes versus Measurements of Evolvability

In representational measurement, we assign numbers to the attributes of entities so that the mathematical relationships among the numbers capture empirical relationships in the real world (Krantz et al. 1971; Hand 2004). In the context of evolvability, representational measurements can allow quantitative predictions of the disposition to evolve when associated *Of* and *Under* conditions are met, as in our example of changes in trait mean under directional selection (see section 6.1). In the Lande model, when the additive genetic variance in a selected trait doubles, the population will evolve twice as fast.

We could hope that the attributes listed in table 6.2 have at least a monotonic relationship with evolvability. Close consideration, however, shows that this is often not the case. The assumptions under which many of these attributes measure evolvability are not general.

For some attributes, a change of assumptions can reverse the sign of their relationship to evolvability.

A familiar example is the relationship between recombination rate and evolution that figures prominently in the literature on the evolution of sex (Otto 2009). For each set of conditions under which recombination enhances the rate of adaptation, another set of conditions exists under which the converse is true. Under the assumptions of the Fisher-Muller model for the evolution of sex, mutations that enhance fitness act additively, such that the fitness of a genotype having two such mutations is always better than those with only one. In this case, sex and recombination enhance the rate of evolution by rapidly bringing together favored mutations in the same genotype. However, if sign epistasis is common, alleles favored in isolation will only spread in the presence of a subset of other alleles. In these cases, selection will create positive gametic disequilibrium between jointly favored alleles, but recombination will tend to separate them, reducing the rate of adaptation.

Modularity and integration provide another example (Houle and Rossoni 2022). For many authors, the phenomenological definition of integration and modularity is inadequate, and they instead define integration and modularity as the tendency of functionally related parts of an organism to covary, and for functional suites to be independent of each other (e.g., Olson and Miller 1958; Cheverud 1996; G. Wagner 1996; G. Wagner and Altenberg 1996; Pavličev et al., chapter 8). Under such a definition, modularity "is expected to improve evolvability by limiting the interference between the adaptation of different functions" (G. Wagner and Altenberg 1996, 967). This positive effect of modularity on the evolvability *Of* complex characters assumes that collections of integrated traits are *Under* selective regimes that differ from those affecting traits in different functional modules. Similarly, integration is expected to increase evolvability when the pattern of covariation among traits aligns with the orientation of the fitness landscape. However, if modularity and the directions of selection are not aligned, evolvability is actually reduced (Hansen 2003; Welch and Waxman 2003). The result of these conflicting considerations is that while everyone can agree that integration and modularity have an important role in determining evolvability, they have no monotonic relationship to evolvability (Armbruster et al. 2014; Houle and Rossoni 2022).

In most systems, we have relatively few data on the types of changes favored by natural selection in suites of traits for which integration and modularity are relevant. The major source of evidence for the adaptiveness of phenomenological integration and modularity is that evolution often proceeds by alteration of integrated modules, but this may be explained either by the assumption that modularity is adaptive or that phenomenological modularity constrains evolution in other possible directions (Houle and Rossoni 2022). The morphologically distinct and developmentally integrated liver of vertebrates provides a potential example of a mismatch between integration and selection. The hepatocytes of the liver are the site of a wide variety of biochemical and physiological functions, including synthesis, storage, detoxification, and digestion. It is, however, a challenge to unite the morphological integration of the liver with a hypothesis that relies on the coordinated evolution of these diverse functions. The complexity of hepatocyte function instead suggests that the liver is a locus for trade-offs among competing functions, rather than an organ individuated to optimize compatible functions.

The relationship of robustness to evolvability is similarly contingent on assumptions. On its face, robustness would seem to reduce evolvability due to the reduced probability that any mutation will change the phenotype. Andreas Wagner (2008) pointed out that this

is not necessarily so, because robust genotypes can reside in neutral networks of variants with no phenotypic effect that give access to high-fitness alternative phenotypes, enhancing evolvability. Conversely, Mayer and Hansen (2017) pointed to the possibility that robust genotypes are embedded in neutral networks that are not well connected, reducing evolvability, as suggested by the naïve expectation. These alternative assumptions about the nature of neutral networks can be tested only in exceptional circumstances (e.g., Zheng et al. 2020), leaving generalizations about the sign of the relationship between robustness and evolvability unverifiable in most cases.

6.9 Toward Better Measures of Evolvability

We have emphasized the importance of a specific theoretical context to the usefulness of measures of evolvability. Strong theory supports the relevance of measures of genomic mutation rates and of variability and genetic variance as evolvability attributes in many contexts. Other attributes, such as recombination and GP map properties, require more qualifying assumptions before their relationship to evolvability can be specified. The development of novel theoretical contexts may suggest more useful ways to connect these attributes to evolvability.

Predictive validity, the usefulness of the measure to predict some future outcome, represents the best method to understand the value of a measurement (Hand 2004), but is particularly difficult to assess for dispositional properties, for which we also need to know the stimulus strength to make such predictions. Even if we know the stimulus strength, the predictive ability can be compromised by inaccuracy of the measurements or shortcomings in the model used to relate measurements to predicted outcome. For example, the Lande equation assumes that the directional selection measured on the focal trait is the sole source of selection, and, in particular, that indirect responses of the focal trait to selection on other correlated traits are absent.

Predictive experiments have repeatedly been performed for short-term evolution by comparing realized evolutionary changes *Of* a character *Under* artificial selection *Over* a given number of generations with predictions obtained from the Lande equation. These experiments have provided somewhat inconsistent results, particularly so for indirect response to selection of genetically correlated traits, leading some authors to question the ability of additive variance and covariance to predict short-term evolution. These are readily, if unsatisfyingly, explained in principle by a wide variety of possible violations of the assumptions of the Lande model (Walsh and Lynch 2018, 504–506). In practice, limited predictive power mostly results from inaccuracies of measures of evolvability or from the effect of background conditions (e.g., small population size; see Pélabon et al. 2021). Accounting for these effects greatly improves the predictive validity of additive variance on short timescales, suggesting that the other possible violations of the Lande equation have little practical significance.

Over intermediate timescales, initial estimates of evolvability generally have diminished predictive power, particularly in small populations (Weber and Diggins 1990). This is expected, as we know that variation-based measures of evolvability may evolve on this timescale (Pélabon et al., chapter 13), for example, by loss of genetic variation.

When we are concerned with longer timescales, predictions are rarely possible (see Barrick et al. 2009 for a representative exception), and instead validation is only possible

by *retrodiction*, the relationship between the properties of extant populations and the prior rate of evolutionary change among related ancestral taxa. The prime example is the retrodiction of rates of molecular evolution by genomic mutation rate (Lynch et al. 2016). This relationship is so strong that rates of evolution have frequently been used to estimate mutation rates, following the neutral theory (Nachman and Crowell 2000). Somewhat surprisingly, retrodictions of evolutionary rates of quantitative traits among populations and species using variance and variability-based measures of evolvability suggest that these measures of evolvability have greater predictive validity than their performance at intermediate times scales would suggest (Voje et al., chapter 14). Although uncertainties characterizing measurements of the *Under* and *Over* conditions call for cautious interpretation, these results suggest that additive and mutational variance may represent fairly accurate measures of evolvability far beyond a handful of generations.

We believe that the main reason for the predictive power of short-term measures of evolvability lies in the well-defined model of evolution provided by the Lande equation, where attributes to measure evolvability and those to measure evolution are simultaneously defined (a similar argument could be made for mutation and molecular evolution). This approach contrasts with measurements of evolvability using GP map attributes that are not defined in the context of an evolutionary model. The validity of those evolvability measurements has not yet been demonstrated via predictions or retrodictions.

6.10 Evolvability Is a Unifying Concept

The ability to conceive and execute useful and predictive measurements is the hallmark of a maturing scientific field. The concept of evolvability is clearly measurable in some contexts, including the neutral theory and the Lande model. In other contexts, we may know what to measure but not how to measure it. For example, if we are interested in the ability to optimize one suite of traits while holding others constant, conditional evolvability is logically what we need to measure (Hansen et al. 2003a), but we rarely know what to condition on, as both pleiotropy and the actual shape of the fitness landscape are unknown. In still other contexts, like the ability to generate novel traits, we have only speculative notions of attributes that generate such a disposition.

For some observers, the diverse status of measurement for each of these variant concepts of evolvability may serve to enhance the claim that each conception of evolvability refers to a fundamentally different phenomenon (Pigliucci 2008; Brookfield 2009; Minelli 2017). We instead find unity in the growing recognition that all definitions of evolvability conceptualize the disposition of a population (or higher-level entity) to evolve (G. Wagner and Altenberg 1996; Love 2003; Hansen 2006; Brigandt et al., chapter 4). To apply that definition to the universe of entities with the potential to evolve, we need to specify the conditions *Of, Under,* and *Over*. The actual properties that constitute evolvability are different, depending on whether one is interested in the evolution of DNA sequences, proteins, or organism-level phenotypes. To some extent, this viewpoint echoes the positions taken in Villegas et al. (chapter 3) and Brigandt et al. (chapter 4), although those contributions note that, in addition to being applied to a wide variety of phenomena, the concept of evolvability plays a wide diversity of roles in scientific discourse.

This combination of unity of concept and diversity of application is already inherent in the typical definitions of evolution as, for example, "change over time in the characteristics of a population of living organisms" (Charlesworth and Charlesworth 2010, xxv). Few observers would maintain that we need to discard the word "evolution," depending on which characteristics, which timescale, or which forces shape those changes. The multifarious instantiations of evolvability are reminiscent of another disposition—energy, which quantifies the potential for matter to perform work. Energy is stored in many different forms, including elastic energy, chemical energy, potential energy, and the thermal energy that the humble thermometer quantifies. Each of these other forms of energy is quantified using a different measure appropriate to its nature. So it is with evolvability: one dispositional concept with many measures.

Acknowledgments

We thank Thomas Hansen, Günter Wagner, and Mihaela Pavličev for detailed and helpful comments on the manuscript. We thank the Center for Advanced Study (CAS) at the Norwegian Academy of Science and Letters for its generous funding of this workshop. The critical enthusiasm of the many participants was key to molding this contribution into its final form. DH thanks CP and Thomas Hansen for organizing and shepherding the evolvability project from its beginnings, and Alan Love for organizing the 2017 workshop titled "Generic and Genetic Explanations of Evolvability" at the Minnesota Center for Philosophy of Science, at which many of his ideas on this topic began their incubation. Preparation of this manuscript was supported by US National Science Foundation (www.nsf.gov) Division of Environmental Biology grant 1556774 to DH and by Norwegian Research Council grant no. 287214 to CP.

References

Ancel Meyers, L., F. D. Ancel, and M. Lachmann. 2005. Evolution of genetic potential. *PLoS Computational Biology* 1: e32.

Armbruster, W. S., C. Pélabon, G. H. Bolstad, and T. F. Hansen. 2014. Integrated phenotypes: Understanding trait covariation in plants and animals. *Philosophical Transactions of the Royal Society B* 369: 20130245.

Barrick, J. E., D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, et al. 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243–1274.

Brandon, R. 1982. The levels of selection. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1982: 315–323.

Brandon, R. N., J. Antonovics, R. Burian, S. Carson, G. Cooper, P. S. Davies, C. Horvath, et al. 1994. Sober on Brandon on screening-off and the levels of selection. *Philosophy of Science* 61: 475–486.

Brookfield, J. F. 2009. Evolution and evolvability: Celebrating Darwin 200. *Biology Letters* 5: 44–46.

Chang, H. 2004. *Inventing Temperature: Measurement and Scientific Progress.* Oxford: Oxford University Press.

Charlesworth, B., and D. Charlesworth. 2010. *Elements of Evolutionary Genetics.* Greenwood Village, CO: Roberts and Company.

Cheverud, J. M. 1996. Developmental integration and the evolution of pleiotropy. *American Zoologist* 36: 44–50.

Falconer, D. S. 1981. *Introduction to Quantitative Genetics.* London: Longman.

Flatt, T. 2005. The evolutionary genetics of canalization. *Quarterly Review of Biology* 80: 287–316.

Gingerich, P. D. 2000. Arithmetic or geometric normality of biological variation: An empirical test of theory. *Journal of Theoretical Biology* 204: 201–221.

Grantham, T. A. 2007. Is macroevolution more than successive rounds of microevolution? *Palaeontology* 50: 75–85.

Haag-Liautard, C., M. Dorris, X. Maside, S. Macaskill, D. L. Halligan, D. Houle, B. Charlesworth, et al. 2007. Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila. Nature* 445: 82–85.

Hand, D. J. 2004. *Measurement Theory and Practice: The World through Quantification.* London: Arnold.

Hansen, T. F. 2003. Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *Biosystems* 69: 83–94.

Hansen, T. F. 2006. The evolution of genetic architecture. *AREES* 37: 123–157.

Hansen, T. F., and D. Houle. 2008. Measuring and comparing evolvability and constraint in multivariate characters. *Journal of Evolutionary Biology* 21: 1201–1219.

Hansen, T. F., W. S. Armbruster, M. L. Carlson, and C. Pélabon. 2003a. Evolvability and genetic constraint in *Dalechampia* blossoms: Genetic correlations and conditional evolvability. *Journal of Experimental Zoology B* 296: 23–39.

Hansen, T. F., C. Pélabon, W. S. Armbruster, and M. L. Carlson. 2003b. Evolvability and genetic constraint in *Dalechampia* blossoms: Components of variance and measures of evolvability. *Journal of Evolutionary Biology* 16: 754–766.

Hansen, T. F., C. Pélabon, and D. Houle. 2011. Heritability is not evolvability. *Evolutionary Biology* 38: 258–277.

Hendrikse, J. L., T. E. Parsons, and B. Hallgrímsson. 2007. Evolvability as the proper focus of evolutionary developmental biology. *Evolution & Development* 9: 393–401.

Hereford, J., T. F. Hansen, and D. Houle. 2004. Comparing strengths of directional selection: How strong is strong? *Evolution* 58: 2133–2143.

Hodgkinson, A., and A. Eyre-Walker. 2011. Variation in the mutation rate across mammalian genomes. *Nature Reviews Genetics* 12: 756–766.

Hoffmann, A. A., and J. Merilä. 1999. Heritable variation and evolution under favourable and unfavourable conditions. *TREE* 14: 96–101.

Houle, D. 1992. Comparing evolvability and variability of quantitative traits. *Genetics* 130: 195–204.

Houle, D., and A. Kondrashov. 2006. Mutation. In *Evolutionary Genetics: Concepts and Case Studies,* edited by C. W. Fox, and J. B. Wolf, 32–48. Oxford: Oxford University Press.

Houle, D., C. Pélabon, G. P. Wagner, and T. F. Hansen. 2011. Measurement and meaning in biology. *Quarterly Review of Biology* 86: 3–34.

Houle D., and Rossoni D. M. 2022. Complexity, evolvability, and the process of adaptation. *AREES.* 53: 137–159.

Jablonski, D. 2008. Species selection: Theory and data. *AREES* 39: 501–524.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution.* Cambridge: Cambridge University Press.

Klingenberg, C. P. 2008. Morphological integration and developmental modularity. *AREES* 39: 115–132.

Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky. 1971. *Foundations of Measurement.* Vol. I: *Additive and Polynomial Representations.* New York: Academic Press.

Kumar, S. 2005. Molecular clocks: Four decades of evolution. *Nature Reviews Genetics* 6: 654–662.

Lande, R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain:body size allometry. *Evolution* 33: 402–416.

Love, A. C. 2003. Evolvability, dispositions, and intrinsicality. *Philosophy of Science* 70: 1015–1027.

Lynch, M. 2007. *The Origins of Genome Architecture.* Sunderland, MA: Sinauer.

Lynch, M., and W. G. Hill. 1986. Phenotypic evolution by neutral mutation. *Evolution* 40: 915–935.

Lynch, M., M. S. Ackerman, J.-F. Gout, H. Long, W. Sung, W. K. Thomas, and P. L. Foster. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nature Reviews Genetics* 17: 704.

Mayer, C., and T. F. Hansen. 2017. Evolvability and robustness: A paradox restored. *Journal of Theoretical Biology* 430: 78–85.

Mayr, E. 1963. *Animal Species and Evolution.* Cambridge, MA: Belknap Press.

McGee, T. D. 1988. *Principles and Methods of Temperature Measurement.* New York: John Wiley & Sons.

Minelli, A. 2017. Evolvability and its evolvability. In *Challenges to Evolutionary Theory: Development, Inheritance and Adaptation*, edited by P. Huneman, and D. M. Walsh, 211–238. New York: Oxford University Press.

Nachman, M. W., and S. L. Crowell. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297–304.

Nuño de la Rosa, L. 2017. Computing the extended synthesis: Mapping the dynamics and conceptual structure of the evolvability research front. *Journal of Experimental Zoology B* 328: 395–411.

Ohta, T. 1992. The nearly neutral theory of molecular evolution. *AREES* 23: 263–286.

Olson, E. D., and R. L. Miller. 1958. *Morphological Integration.* Chicago: University of Chicago Press.

Otto, S. P. 2009. The evolutionary enigma of sex. *American Naturalist* 174: S1–S14.

Paaby, A. B., and M. V. Rockman. 2013. The many faces of pleiotropy. *Trends in Genetics* 29: 66–73.

Pélabon, C., E. Albertsen, A. Le Rouzic, et al. 2021. Quantitative assessment of observed vs. predicted responses to selection. *Evolution* 75: 2217–2236.

Pélabon, C., C. H. Hilde, S. Einum, and M. Gamelon. 2020. On the use of the coefficient of variation to quantify and compare trait variation. *Evolution Letters* 4: 180–188.

Pigliucci, M. 2008. Opinion—Is evolvability evolvable? *Nature Review Genetics* 9: 75–82.

Salmon, W. C. 1971. *Statistical Explanation and Statistical Relevance.* Pittsburgh: University of Pittsburgh Press.

Schrider, Daniel R., D. Houle, M. Lynch, and M. W. Hahn. 2013. Rates and genomic consequences of spontaneous mutational events in *Drosophila melanogaster. Genetics* 194: 937–954.

Schuster, P., W. Fontana, P. F. Stadler, and I. L. Hofacker. 1994. From sequences to shapes and back: A case study in RNA secondary structures. *Proceedings of the Royal Society B* 255: 279–284.

Sober, E. 1992. Screening-off and the units of selection. *Philosophy of Science* 59: 142–152.

Sultan, S. E. 2017. Developmental plasticity: Re-conceiving the genotype. *Interface Focus* 7: 20170009.

Vermeij, G. J. 1973a. Adaptation, versatility, and evolution. *Systematic Zoology* 22: 466–477.

Vermeij, G. J. 1973b. Biological versatility and earth history. *PNAS* 70: 1936–1938.

Waddington, C. H. 1942. Growth and determination in the development of *Drosophila. Nature* 149: 264–265.

Wagner, A. 2008. Robustness and evolvability: A paradox resolved. *Proceedings of the Royal Society B* 275: 91–100.

Wagner, G. P. 1996. Homologues, natural kinds and the evolution of modularity. *American Zoologist* 36: 36–43.

Wagner, G. P., and L. Altenberg. 1996. Perspective: Complex adaptations and the evolution of evolvability. *Evolution* 50: 967–976.

Walsh, B., and M. Lynch. 2018. *Evolution and Selection of Quantitative Traits.* Oxford: Oxford University Press.

Wang, J., H. C. Fan, B. Behr, and Stephen R. Quake. 2012. Genome-wide single-cell analysis of recombination activity and *de novo* mutation rates in human sperm. *Cell* 150: 402–412.

Weber, K. E., and L. T. Diggins. 1990. Increased selection response in larger populations. II. Selection for ethanol vapor resistance in *Drosophila melanogaster* at two population sizes. *Genetics* 125: 585–597.

Welch, J. J., and D. Waxman. 2003. Modularity and the cost of complexity. *Evolution* 57: 1723–1734.

Whitlock, M. C., and K. Fowler. 1999. The changes in genetic and environmental variance with inbreeding in *Drosophila melanogaster. Genetics* 152: 345–353.

Zelditch, M. L., and A. Goswami. 2021. What does modularity mean? *Evolution & Development* 23: 377–403.

Zheng, J., N. Guo, and A. Wagner. 2020. Selection enhances protein evolvability by increasing mutational robustness and foldability. *Science* 370: eabb5962.

The MIT Press would like to thank the anonymous peer reviewers who provided comments on drafts of this book. The generous work of academic experts is essential for establishing the authority and quality of our publications. We acknowledge with gratitude the contributions of these otherwise uncredited readers.

This book was set in Times New Roman by Westchester Publishing Services.

Library of Congress Cataloging-in-Publication Data

Names: Hansen, Thomas F., editor.
Title: Evolvability : a unifying concept in evolutionary biology? / edited by Thomas F. Hansen, David Houle, Mihaela Pavličev, and Christophe Pélabon.
Description: Cambridge, Massachusetts : The MIT Press, [2023] | Series: Vienna series in theoretical biology | Includes bibliographical references and index.
Identifiers: LCCN 2022038288 (print) | LCCN 2022038289 (ebook) | ISBN 9780262545624 (paperback) | ISBN 9780262374705 (epub) | ISBN 9780262374699 (pdf)
Subjects: LCSH: Evolution (Biology)—Philosophy.
Classification: LCC QH360.5 .E99 2023 (print) | LCC QH360.5 (ebook) | DDC 576.801—dc23/eng/20220920
LC record available at https://lccn.loc.gov/2022038288
LC ebook record available at https://lccn.loc.gov/2022038289