

WHY IS NATURAL SELECTION SO HARD TO UNDERSTAND?

In 1959, at a symposium celebrating the 100th anniversary of the publication of Charles Darwin's *On the Origin of Species*, the geneticist and Nobel Laureate H.J. Muller presented a lecture with the provocative title, "One Hundred Years Without Darwin Are Enough" in which he lamented many people, even dedicated Darwinians, misunderstand the theory of natural selection. Sixty-five years later, Muller's lament is still pertinent. If my teaching experiences serve as a guide, natural selection is still poorly understood today.

Earlier in my career as a philosopher of science, I taught a continuing education class for Rhode Island high school biology teachers on how to teach evolutionary theory. Creationism runs rampant in the US; biology teachers sought methods on how to handle students' queries and worries about how evolution impinges on their religious beliefs. The first topic of discussion was how best to articulate the theory of evolution by natural selection. The teachers reported that among high school students, the theory is known by the slogan, "the survival of the fittest," which is not even a declarative sentence, let alone a theory. The inability to articulate Darwin's theory of natural selection beyond slogans is a problem that is not particular to our Rhode Island biology students: it is, as Muller suggested in 1959, a persuasive and long-standing problem. What can educators do to remedy misunderstandings about natural selection, both to the creationist-minded and ones who are receptive to scientific biology?

Darwin's *On the Origin of Species* is very long and complex. Darwin's monumental task was to convince hard-headed creationists that species are related by common descent and that natural selection can make sense of a complex organic world full of variation. How does one make sense of a theory covering so much ground from individual adaptations to whole ecological systems?

If you think Darwin's book is hard to understand, open a modern evolutionary textbook. It is full of math. This is because modern evolutionary textbooks have to consider what Darwin didn't know then: that inheritance involves genes and that reproductive schedules and population size (among other factors) matter for the direction of evolutionary change. Today, the theory of natural selection is embedded in a more encompassing theory of evolutionary change. Although Darwin didn't express it in 1859, math is the scientific language for understanding the dynamics of evolving populations, especially those whose individuals vary in their features.

No wonder people, even committed Darwinians, don't get it. The theory of evolution by natural selection is hard because the framework for understanding evolving systems is hard. To succeed in raising the public's understanding of evolutionary theory, our Rhode Island high school teachers are faced with preparing students to comprehend the probability and statistical ideas before they can begin to teach the theory of natural selection.

“Not so,” says the late great evolutionary theorist and historian S. J. Gould. I encountered Gould’s response in an old VHS tape I found in the local library from a series called “Into the Classroom Video.” (There is no information on the box about where the lecture was held or what year it was given. From the camera angles, you can tell it was recorded in a large college auditorium.) Gould has a ready answer to our pedagogical problem contrary to mine: despite appearances, the theory, in its “barebones, is simplicity itself.” Gould advocates that we first teach the barebones version of the theory, presumably leaving the more difficult ideas for later.

According to Gould, the theory of natural selection is based upon three facts “that no one can deny in a simple, almost syllogistic inference.”

First, all organisms produce more offspring than can possibly survive.

Second, all organisms vary within a species.

Third, at least some of that variation is inherited.

From these three facts, it follows that “on average, those organisms that fortuitously are better adapted to changing local environments will tend to survive better and produce more offspring, and the average of the population will shift in their direction.” To cite a “not way-off” example, there is: “population of elephants in Russia. The ice is advancing. There is variation in the amount of hair on elephants. Elephants that are a little more hairy on average will do better—the hairiest might still fall down a crevasse and die—but as a statistical statement, on average, they will do better in surviving and reproduction. A hundred generations down the line, you will have a woolly mammoth”. Gould concludes that the theory of natural selection is based upon a “principle of adaptation to changing local conditions. It isn’t hard.”

According to Gould, the correct answer to Muller’s timely challenge—why was it “and still is” misunderstood—has nothing to do with the complexity of the theory’s barebones mechanics but rather society’s reception to the philosophically radical ideas that Darwin’s theory engendered. The implications for the existence of God and the placement of humans in the order of nature stand against many of Western culture’s traditional social hopes and psychological biases.

Surely, Gould is right; the profound implications of Darwin’s theory must cause some confusion among some members of the public. Perhaps expounding on the barebones version of natural selection would suffice for those high school biology students struggling with creationism. However, Muller’s remarks about the lack of understanding of Darwinism were also aimed at committed Darwinians, including many of our high school students, who are prepared to accept that Darwin’s theory upends some of our traditional beliefs about the place of humans in relation to other animals. The problem that Muller’s challenge posed is that the wider public appears to lack a clear understanding of how the theory of natural selection is embedded into a larger theory of evolution. Let us look more carefully at Gould’s account to see why it doesn’t help convey this post-20th-century advance to understanding nature.

The limits of Gould's barebones account arise when we put meat on the bones. Philosophers would spot the problem right away: there is no such thing as an "almost syllogistic inference." Syllogisms are, like pregnancies, non-monotonic. There is no in-between state; syllogisms are either valid or not. If the structure of natural selection theory were indeed in the form of a deductively valid syllogism, then, from the three facts about reproductive rates, variation, and inheritance, it would follow (as a matter of deductive logic) that evolution over the course of generations will change the character of the population, making the most adapted feature the most prevalent one. But that's not true. Evolutionary processes do not abide by deductive logic. As Gould acknowledges, even the hairiest elephant might still "fall down a crevasse and die." If the population of elephants is small enough, losing the most adapted variant could have (with some probability) changed the direction of evolution entirely.

The logical point is that Gould's barebones account is not a deductive syllogism. It is an inductive inference assessed by a degree of probability. To compensate for this, Gould added qualifications to his description: "*On average, as a statistical phenomenon*, not every time, *on average* those organisms that are fortuitously adapted to changing local environments will tend to survive better and produce more offspring and the average of the population will shift in their direction." With all these references to probability, "on average," "statistical phenomenon," "tend to," "the average of the population," there is a lot of biology and math to unpack if we want to assess the probability that evolution by natural selection occurs from the conditions of variation, overpopulation, and principle of inheritance.

The problem with Gould's barebones account of natural selection is not just a matter of logic; its biology is also potentially misleading. We know today that in natural populations, factors besides variation, inheritance, and over-reproduction determine the direction of evolutionary change. This is why modern evolutionary textbook treatments of evolution are so complicated. Take Gould's "not far off" example of the evolution of woolly mammoth. Paleontological and mitogenomic studies suggest various contingent factors played a role in the evolution of woolly mammoths in N. America: distinctly timed migrations and population dispersals, geographical partitions, and "hybridization" (i.e., sex) between species. How much of a factor depends in part on the size of interbreeding populations; if small, then the effect of these additional factors would have been as significant or even greater than those cited in the barebones account. (See Chang, Dan, Michael Knapp, Jacob Enk, Sebastian Lippold, Martin Kircher, Adrian Lister, Ross DE MacPhee et al. "The evolutionary and phylogeographic history of woolly mammoths: a comprehensive mitogenomic analysis." *Scientific reports* 7 (2017): 44585).

Again, Gould knew all about the probabilistic basis of natural selection and the existence of contingent factors that determine the direction of evolutionary change. In the lecture, he presented the barebones mechanics to demonstrate how easy it is to understand natural selection as a first approximation. This is a common pedagogical

practice across the sciences: First, teach the simpler theory, which is only approximate but does not involve complex ideas. The problem is Gould did not articulate what natural selection entails beyond the barebones. For guidance, let us turn to a parallel discussion in teaching basic physics.

As the great physicist and educator Richard Feynman explained, “simple” laws of physics are, at best, approximations to the truth, so they can’t be explained using deductive structures. Physics is not like Euclidean geometry, in which we deduce results from axioms. Even when physicists know enough to correct the laws, their re-statement is challenging to understand; it requires applications of advanced mathematics and, hence, a considerable amount of preparatory training.

For example, take the Newtonian “constant-mass” law whereby mass is independent of speed. This can be inexpensively demonstrated in a physics classroom with a toy spinning top and a simple kitchen scale. Yet, like the barebones account of natural selection, the constant mass law is only approximately true; it breaks down at velocities approaching light speed, where mass increases with velocity. The approximate law requires modification to reflect the truth: As Feynman puts it, “If an object moves with a speed of less than one hundred miles a second, the mass is constant to within one part in a million.” The modification is closer to the truth but more challenging to teach. What should an educator do? Teach the simple approximate theory, leaving the complex and difficult-to-understand correctives for a future course, or begin with the correct but unfamiliar theory?

There are good reasons to begin with the simple approximate theory. Demonstrating the correct law is impractical. A physics instructor would require sophisticated and expensive equipment to demonstrate how the constant-mass law breaks down in special conditions, with speeds approaching light. The simpler constant-mass law applies widely to most objects that humans manipulate and at speeds much slower than that of light: after all, it is said that Newton’s laws are sufficient for sending rockets to space. So, there is practical value in teaching the false constant-mass law, especially considering that some students will become engineers and rocket scientists. We live in a mostly Newtonian world where mass is independent of speed, just like the simpler law says.

However, the pragmatic approach sacrifices truth. While easier to teach, the simpler approximations represent a false view of the nature of the universe. The constant-mass law is a vestige of late 17th-century Newtonian physics, which includes beliefs about absolute space and time. The lesson learned from the historical experiments on the constant mass law performed at near-light speeds was that time and space are relative, not absolute, as Einstein championed.

Here’s where truth comes in. Even though in modern experiments, mass changed just a little bit, the shift from a Newtonian to an Einsteinian framework effectively altered the entire view of the universe. As Feynman put it, “philosophically we are completely wrong with the approximate law. Our entire picture of the world has to be altered even though

the mass changes only by a little bit". Feynman concludes with one of my favorite remarks from scientists about philosophy: "this is a very peculiar thing about the philosophy, or the ideas, behind the laws. Even a very small effect sometimes requires profound changes in our ideas." The point is, by taking the practical route and teaching the simpler but false Newtonian law, teachers miss out on an important conceptual consequence from 20th-century work on objects approaching the speeds of light, the whole understanding of the universe had to be overhauled. Teaching science is not merely a technical exercise in replacing a true law with a false one; it is an exciting, mind-expanding, and profound endeavor, introducing concepts—even complex ones—that are universal to the human experience.

Nevertheless, Feynman gives good pedagogical advice. Start with simple approximate law, good enough for practical purposes (and engineers), and leave teaching the more exciting, intellectually enriching, and correct law for philosophers and theoretically minded physicists, those willing first to learn some difficult conceptual ideas. Is this how we should approach the teaching of evolutionary theory? Well, it depends.

On the one hand, there are relevant parallels between physics and evolutionary theory. Darwin's theory of natural selection was a vestige of the middle 18th-century Newtonian view of nature. He aimed to demonstrate that evolution was the product of natural laws, like those that keep planets spinning in their orbits. In the early 20th century, theoretical biologists, armed with Mendel's statistical treatment of inheritance, realized that evolutionary theory is not like Newtonian physics; a proper understanding of the evolving populations requires first some understanding of variability, how it emerges, how it is maintained, and how it changes over generational time. So, teaching the corrected evolutionary theory requires first teaching some probability and evolutionary theory, the framework for modern evolutionary theorizing. This means that, like in the case of physics, a proper course in evolutionary theory requires some mathematical prerequisites.

Following Feynman's advice, our biology teachers could begin teaching the incomplete barebones account of natural selection, as Gould prescribes, and leave the task of teaching the difficult but correct statistical versions of evolutionary theory to senior or AP biology classes.

The problem with this advice, however, is that there are relevant disanalogies between physics and evolution that make Feynman's pedagogical advice sub-optimal. I will discuss two. The first concerns the applicability of laws in normal conditions. The simple approximate laws in physics apply to a large range of normal conditions on Earth, but the same is not true for evolutionary biology (in a sense the opposite is true, as I will explain). The second disanalogy concerns the role variation plays in theory. For physical forces, variations from what is expected are often ignored or explained away. But, in evolutionary theory, variation is essential, both as the phenomenon to be explained and

the explanation itself. I'll explain both disanalogies in turn, the discussion of each is part of the peculiar things about nature that philosophers of science like to think about.

Recall Feynman's exposition of the constant-mass law. For ordinary speeds, the simpler approximation is true enough. It breaks down only in rare cases where speed approaches that of light. Hence, there are good practical reasons to teach the approximate law first, as it applies to a wide range of conditions. In evolution, the opposite is true. The barebones account of natural selection is approximately applicable in unrealistic situations, for example, when the population size approaches infinite or mating is purely random. Natural selection's effect on evolution is analogous to a crooked gambler playing with imperceptibly loaded dice. The gambler's sneaky advantage will be evident only after numerous games, practically undetectable when the number of games is small and limited. Many evolving populations are relatively small in number, and sexual creatures are biased in their mating choices. Just like the case of the crooked gambler, the slight reproductive advantage of better-adapted variants could get swamped by other effects simply due to sampling in small and non-random trials of life. In those cases, the best adapted and the crooked gambler will lose out. For a real-world example, recall that the evolution origin story of the woolly mammoth involved factors whose genomic effects would be most pronounced in limited population sizes, migration, geographic isolation, and selective mating.

So, there is a practical difference between teaching physical laws and evolutionary theory: approximate laws in physics tend to be good enough to describe many realistic cases and break down only in extreme cases; the opposite is true for evolutionary theory. Natural selection has a relatively smaller effect on populations of smaller sizes and is more pronounced when the population approaches infinity.

The second disanalogy between physics and evolution concerns the differences between how systems change over time. Physical systems often involve transformational changes, while evolutionary change is variational. In a transformational process, the ensemble changes over time because each individual obeys the same law-like history over their lifetime: my fellow Generation Xers are turning grey-haired and developing wrinkles because each of us is affected by the same forces of age; the cosmos is evolving because all stars go through the same thermonuclear and gravitational sequences. Consequently, to explain transformational change involves discovering the deterministic laws that each individual in the population undergoes.

In contrast, the evolutionary theory of organic change is a variational model of change. Ensembles of individuals change, not because each is undergoing the same transformational processes in development. The opposite is true; individuals vary, and the trait composition of subsequent variation includes some of those variants. Bacteria that cause infection are becoming more resistant to pharmaceutical antibiotics not because each bacterium is acquiring greater resistance but because pre-existing resistant individuals live and reproduce and (recalling Gould's correctives) '*on average*—

as a statistical phenomenon—are better represented in subsequent generations. As the great evolutionist and philosopher Richard Lewontin described, “Evolution is a theory of proportional representation of different variants, which are themselves unchanging in their properties.”

And therein lies a difference between physical and evolutionary changes. In the transformational world, we can pretty much ignore the variation between objects and look at what they have in common. Drop a handful of stones from atop the leaning tower of Pisa and, sure, the stones’ downward trajectory won’t all be identical according to Newtonian laws of gravity. However variation in transformational theories like Newton’s is treated as extraneous (sometimes called “error”). Newton’s laws tell us what would happen to each stone had there not been a variety of external forces acting differently for each stone.

For evolutionary theory variation is not extraneous as it is in transformational theories, rather it is essential, in both the description of the phenomenon to be explained and the way natural selection works. Proto-mammoths varied in the amount of hair they had before the advance of the ice age, and this had consequences for the survivability and reproductive success of the aboriginal lineages. Variation is a requirement for natural selection to operate, without it, evolution grinds to a halt. Consequently, the appropriate language of variational theories like evolutionary is not, as Gould, with his barebones account depicts it, deductive logic, simple fact, and universal laws; rather, the appropriate language of variational change is statistics and probability.

Perhaps the two disanalogies between physics and evolutionary theories are insufficient to change the pedagogical strategy of first teaching natural selection in its barebones first and leaving the complex and philosophically interesting ideas about stochasticity and variation for more advanced biology courses. Yet, it puts a bit of urgency on teaching the more difficult but truer versions of natural selection, because (unlike the physics case) students who are taught only the barebones account of natural selection are left with a theory with limited application and a wrong-headed view of the role variation plays in evolutionary processes.

In the second part of the essay, I will focus on the task of teaching more complex ideas that should be covered in an advanced biology class. Specifically, I will outline the different ways in which variational thinking is involved in evolutionary theory, starting from the simplest concepts and progressing to the most complex and philosophically intriguing ones.

How do probabilities enter into evolutionary theory? We already mentioned one way when we considered population size and whether mating is random or directed. The barebones account is approximately applicable under unrealistic assumptions when populations are infinite and mating is purely random. An advantageous variant is akin to a crooked gambler playing with weighted dice. Depending on how great the advantage—how imbalanced the dice are—the crooked gambler stands to win a lot of money in the

long run of many games. But, in the short run, the crooked gambler will appear to win and lose games roughly as often as the fair-minded gamblers. And that's when things get tricky. Contemporary evolution textbooks are often divided into two parts, "deterministic" (with unrealistic infinite/random assumptions) and "stochastic" models of evolutionary change. As one evolutionary textbook puts it, "Some of the most elaborate mathematics in evolutionary theory has been devised to study" (Rice, p. 73) stochastic models. Why so difficult? With the barebones account, the direction of evolutionary change follows a single trajectory, towards the variants best suited for local environmental conditions. But, with smaller populations and non-random mating, the direction of evolution could go a variety of ways and in different directions at different points in time. To see why, think about our crooked gambler again. If she plays an infinite number of games, the direction of her win/loss will tend towards wins at a rate to be determined by the propensity of her loaded dice. But, with a finite number of games and limited funds, there's a chance—even with her loaded dice—that her funds go bust before she accumulates her fortune. An investigator tracking evolutionary change or gambling wins then has to consider all the possible ways the system could change.

Another way probabilities enter into evolutionary theory concerns the nature of variation and inheritance. To see the problem, let's go back to Gould's description of the biological facts about inheritance that form the 'almost syllogism' of the barebones account of natural selection: at least some individual variation within a species is inherited. "Folk wisdom," says Gould; but he immediately follows with a qualifier: "Sure Darwin didn't know the mechanism—the world didn't until the Mendelian rediscovery of 1900—but you don't need to know the mechanism, you just need to know that there is a principle of inheritance, and that's folk wisdom. We know that tall parents tend to have tall kids, black parents, black children, white parents, white children, we know there is a principle of inheritance even if we don't know how it works."

Yet, the mechanism matters a lot. To see why, consider that variation is the fuel for natural selection. The inheritance system must produce enough variation for natural selection to operate continuously and keep up with the demand. So, for the barebones version of natural selection to maintain plausibility, we need more than folk wisdom; we need to understand how nature produces and maintains the superabundance of variation we see all around us. As Gould alludes, the world didn't know how until the 1900s when theoretical biologists rediscovered Mendel's account of the inheritance of hybrids.

However, even Mendel's account of inheritance didn't fully solve the problem. To solve it 20th century evolutionists relied on math, not biology, in particular, applications of probability theory. The key, as R.A. Fisher said in 1930, was to take a page from the theory of gases, and attend to the statistical concept of variance. The account is fascinating, full of philosophical wonders, including how math can be incorporated to provide insights into biological processes. Students who are not exposed to it, those left hanging with the impractical barebones account, will not only miss out on understanding

a critical scientific/mathematical insight, they will harbor a misunderstanding about how natural selection really works.

Under the idealizing assumptions that we associated (above) with the deterministic conception of natural selection, infinitely large populations mating at random, along with the idea, taken from Mendel, that inheritance is particulate in nature, a population generates a statistical distribution generated by the central limit theorem, whose graphical representation is a bell-shaped frequency curve. Importantly, under the particulate theory of inheritance, the variance (the shape of the curve) is “conserved...for an indefinite period”. That is to say, the nature of genetic variation on the particulate theory and accompanying assumptions is the maintenance of enough variation that there is no need to postulate a distinct biological mechanism producing an inordinate amount of mutation.

On the face of it, this sounds like voodoo—a piece of statistical mathematics serving as the basis for a biological problem of maintaining genetic variation. But, it isn't. We can demonstrate how it works by utilizing an inexpensive mechanical toy (virtual ones are all over the internet). The toy will simulate statistical and probabilistic ideas without having to avert to any mathematical formulae.

The toy, a “quincunx”, resembles a pachinko machine. Inside the glass face, at the top, is a funnel from which to drop numerous pellets. At the bottom are several segregated compartments. Upon passing the narrow neck of the funnel, the pellets cascade across an array of 17 rows of offset spikes. The pattern of spikes is where the name “quincunx” comes from; it resembles the mesh pattern like those of fishermen's nets. At the bottom of the machine is a row of compartments separated by a thin wall. This is the final resting place for the dropped pellets.



The quincunx was conceived and constructed in the 1870s by Francis Galton, Darwin's cousin (by marriage). He employed it to demonstrate to a public audience the rather surprising fact that the variation of many organic characteristics is not chaotic as one might suspect by casual observation, but it conforms to a pattern that is seen when

we look at the data in a certain way.

To illustrate, Galton displayed a statistical table originally produced and published decades earlier, featuring the heights of American, French, and Belgian conscripts.

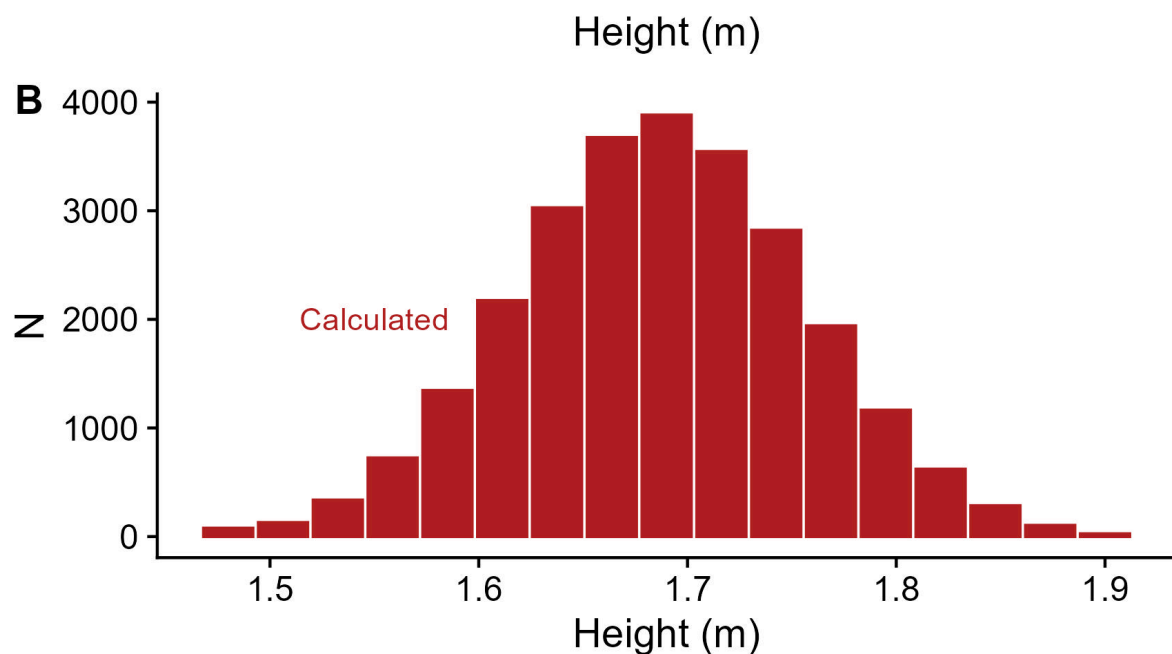
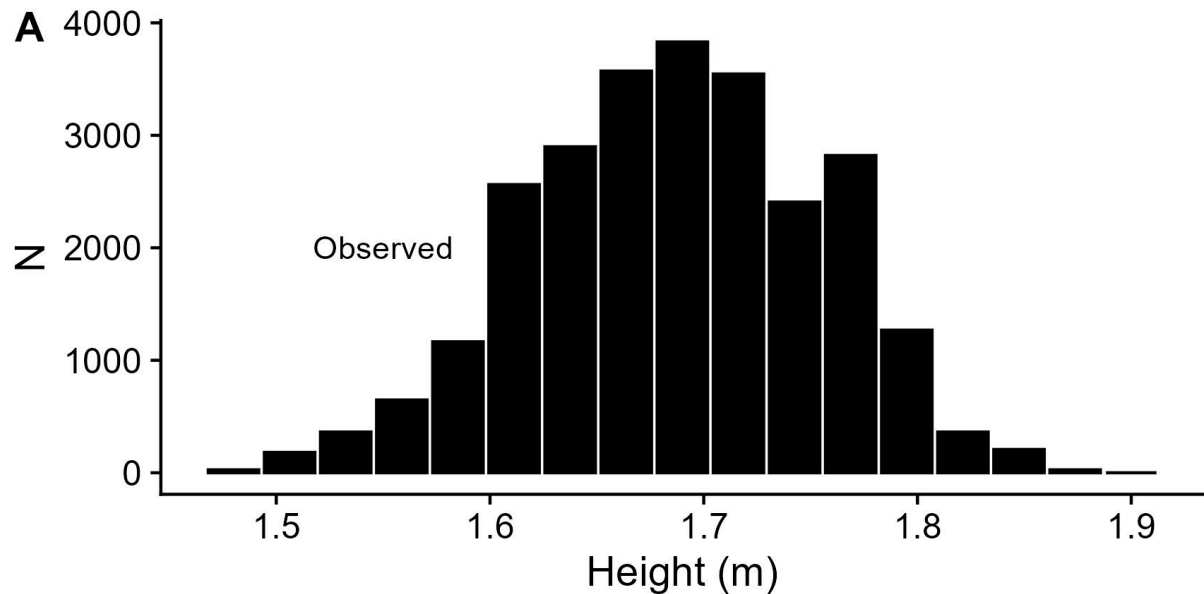
Scale of Heights.	American Soldiers (25,878 Observations).		France (Hargenvilliers).		Belgium, Quetelet. 20 years' Observations.	
	Observed.	Calculated.	Observed.	Calculated.	Observed.	Calculated.
mètres.						
1.90	1	3				
1.90	7	5	..	..	1	1
.87	14	13	..	1	2	3
.84	25	28	25	3	7	7
.81	45	52		16	14	14
.79	99	84	32	32	34	28
.76	112	117	55	55	48	53
.73	138	142	88	87	102	107
.70	148	150	114	118	138	136
.68	137	137	144	140	129	150
.65	93	109	140	145	162	150
.62	109	75	116	132	106	136
.60	49	45	..	105	110	107
.57	14	24		73		53
.54	8	11		44		28
.51	1	4		24		14
.48	..	1	286	11	147	7
.45	..	..		4		3
.42	..	..		2		1
.39	..	..		1		
.36	..	..				
	1000	1000	1000	1000	1000	1000

The data is arrayed in ascending order from tallest (in meters). In each series, there are two columns entitled "observed" and "calculated". The numbers therein represent the frequency of individuals that correspond to the unit measured. "Calculated" refers to frequencies of an abstract population that would conform to statistical law in the long run.

The "striking" fact that Galton wants us to see from the charts is that there is close conformity between "observed" and "calculated". The data reveal that the frequency distribution for heights for Americans, French, and Belgians are different: Americans are taller, on average, than the French, and the French are taller than the Belgians. But, in all instances, the chart reveals a robust pattern, a bell-shaped distribution curve. You can see the outlines of the curve reproduced by the digits displayed in each row. The height

of the curve is represented by the calculated number of individuals of each height, the statistical “mean”: the 150 American soldiers who are 1.7 meters tall, 145 French at 1.62 meters, and 150 Belgian at 1.65 or 1.62 meters. At either side of the center mean, the numbers slope downwards until they reach the rare height extremes, the shortest and tallest of the populations which are represented in each instance by a single person.

The following figure isolates the bell-shaped distribution curve from the data on American soldiers (thanks Rob Walker!).



At the time of his lecture, Galton echoed what a lot of scientists thought, that the

“curve of deviation” was a universal feature of natural variation. Today we know this is an exaggeration, but still, it is surprisingly ubiquitous.

The mystery Galton was posing was this: why would the processes of biological development maintain a particular statistical pattern? Galton employed his quincunx toy to shed some light on this interesting phenomenon.

As each pellet drops from the funnel at the top of the quincunx, it immediately encounters a pin, forcing its trajectory to change either to the right or left of the pin, with equal probability. Once the pellet clears the first pin, it immediately encounters the pin affixed in the next row down. That’s the purpose of the quincunx pattern, in each row, the pellet will always encounter another pin, forcing the alteration in trajectory. Once the pellet has dropped through all the rows, it rests at the bottom of the device, separated by the compartments. We could calculate the chances of any pellet landing in any resting compartment by resorting to a series of coin tosses. But, the key to the quincunx demonstration is to pay attention to the shape of piles of pellets they have all completed their tumbles down the array of pins. What you see is roughly the same bell-shaped curve that you see in the heights of soldiers.

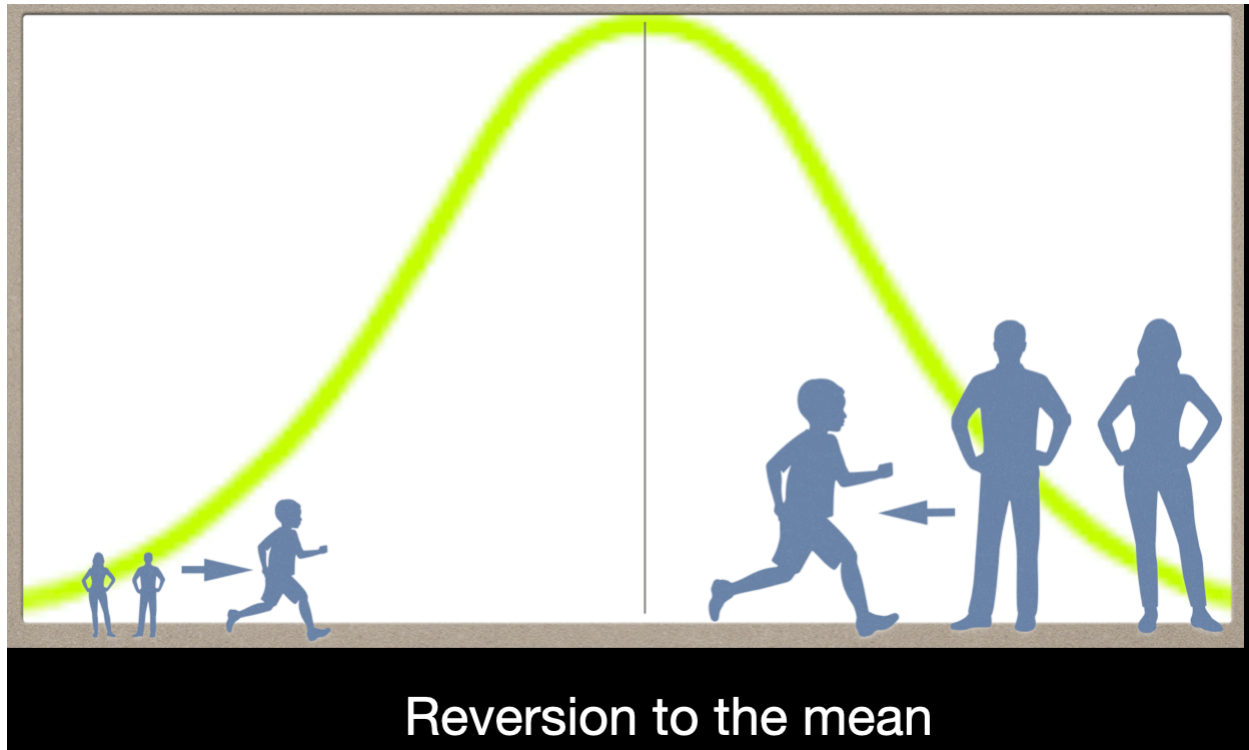
How can these two entirely different physical processes, the aggregate effect of pellets in the quincunx and the developmental processes that determine human heights, produce the same curve? This is an interesting question, inviting philosophical speculation. We might infer that both processes at base share these features: they are the product of random sampling processes where the discrete trials are numerous (the more the better), and the outcomes are probabilistically independent of one another. As you learn from most statistic textbooks, you can substitute coin tosses for the quincunx (I don’t know why you would, though—the quincunx produces the effect more vividly). Toss a fair coin 100 times and note the proportion of heads and tails. Now do this again and again, each time noting the heads/tails ratios. The resulting graphical display of your results will increasingly look like a bell-shaped curve.

To be sure, Galton, like his cousin Darwin, did not know that inheritance was particulate (let alone genetic). It took the 20th-century theoretical biologists to put these two ideas together and show that, like the pellets in a quincunx and coin tossing, the graphical representation of the aggregate effect of Mendelian particulate inheritance on the development of characters distributes in its familiar bell shape.

Now that we’ve established the (abstract) conditions that would produce character variation in a single generation we can proceed to our original question about how enough variation is maintained over the course of generations for natural selection to continuously operate. Once again, Galton, armed with a modified quincunx, provided the conceptual answer (albeit without the Mendelian insight about particulate inheritance). And, once again, the demonstration had an empirical and a mathematical component.

Recall Gould’s discussion of the essence of inheritance, tall parents tend to produce tall children, short parents, short children. True enough. But, what Gould doesn’t tell us is

that inheritance seems to have a quantitative limit, for instance, the children of tall parents tend not to be taller than themselves, rather they tend to “revert” to the mean height of the population; children are, as it were, meaner than their parents. Same is true for other characteristics.



Darwin was well aware of this phenomenon; he had heard about it from domestic breeders who would complain that despite the selection of a profitable or prized feature, sometimes the breeding stock would revert to aboriginal characteristics. Darwin understood this to be a significant obstacle to the power of natural selection. But, here we're interested in reversion (later called “regression”) as an important feature of variation. As Galton put the mysterious phenomenon, “the processes of heredity are found to be so wonderfully balanced, and their equilibrium to be so stable, that they concur in maintaining a perfect statistical resemblance so long as the external conditions remain unaltered...the processes of heredity must work harmoniously with the law of deviation, and be themselves in some sense conformable to it.” How do we explain the reversion effect?

The original quincunx was insufficient to demonstrate it because it required a second stage to represent heredity or the passing of character traits from one generation (the upper stage) to another (the lower stage). The picture below is a drawing from Galton's autobiographer, Karl Pearson, based on a letter from Galton to George Darwin (Charles's son). Separating the two stages are trap doors at the bottom of the verticle compartments holding the pellets at the bottom of the upper stage. When the trap doors are open the pellets would run through a second field of pins also arrayed in

the quincunx pattern.

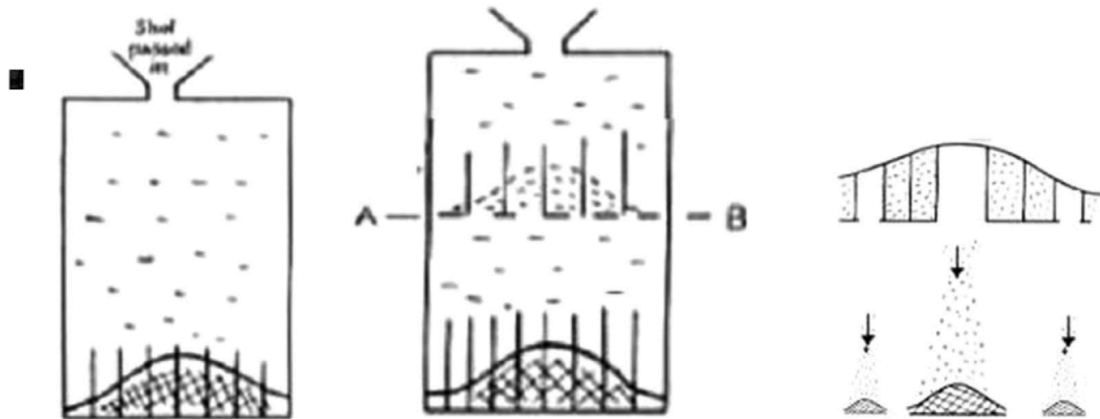
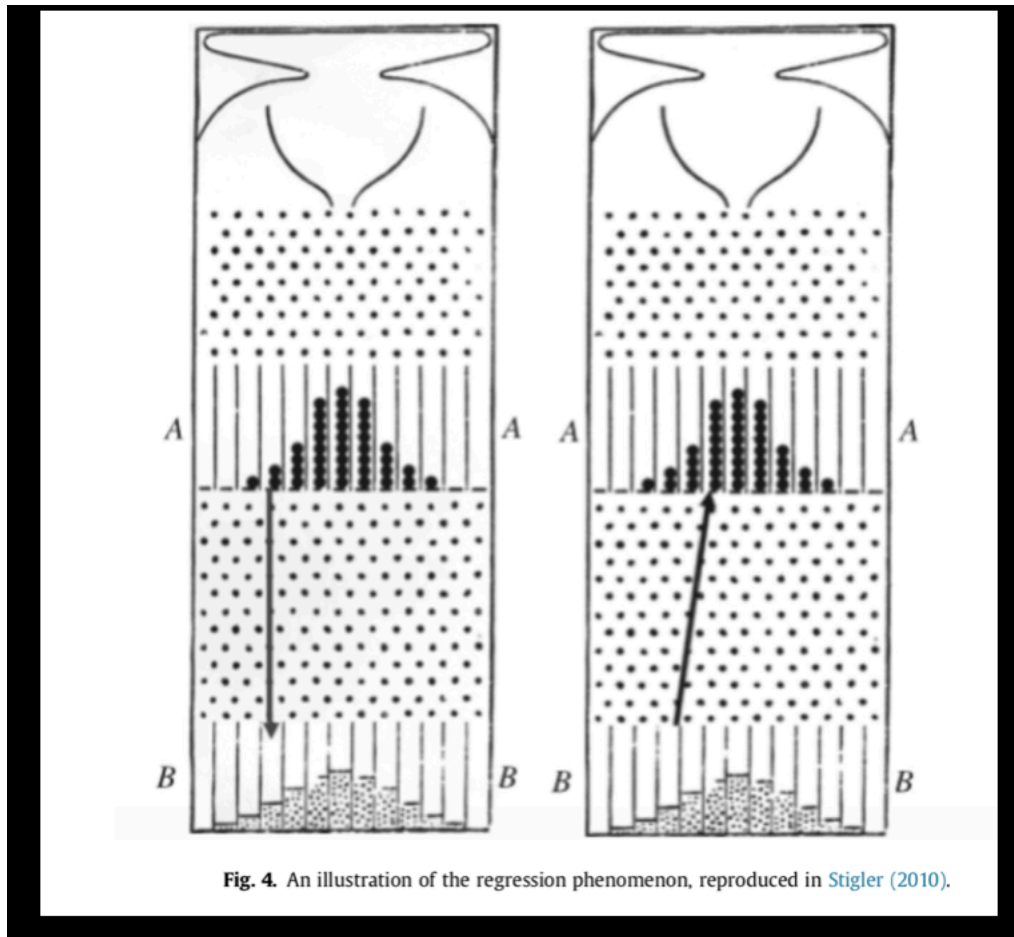


Fig. 3. A drawing from Pearson's biography based on a letter from Galton to George Darwin, reproduced in Stigler (1986, p. 278).

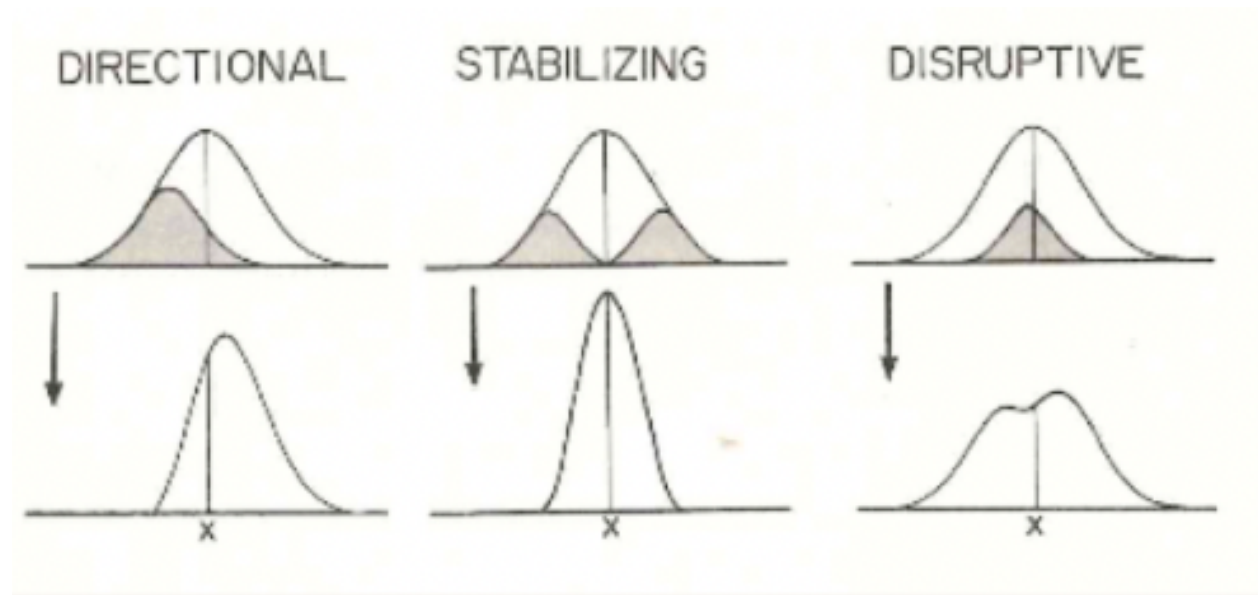
The second and third illustrations from the left demonstrate the reversion effect: the normal distribution at the top stage is reproduced in the bottom stage. Why? As the third illustration shows, the overall pile of pellets is composed of numerous subsets of pellet piles originating in the first stage, the little piles at the extreme of the first stage beget (on average) little piles at the extreme of the second stage. To see this from another point of view, consider the next figure below. Start with the illustration at the left. It shows where a pellet originating at the right side of the first stage is likely, on average, to land in the second stage; straight down. The rightmost illustration, however, illustrates the reversion effect. The key is to do a bit of mental jujitsu—ask of a pellet at the bottom of the second stage where is it likely to have come from? The answer is going to be, from a bigger rather than a smaller pile of pellets at the first stage. Why? Simply because there are more pellets hence more chances that a pellet from a bigger pile will shift to the extreme in the second stage.



So, now we have the answer to the question that dogged the early versions of Darwin's theory about the maintenance of character traits over the course of generations. Turns out that there is no need to search for a separate mechanism that produces a superabundance of variation at every generation. As historian Ian Hacking put it, Galton noticed that reversion to the mean (the preservation of variation over the course of generations) is "a mathematical consequence of the Normal curve." Assuming that the external conditions remain the same throughout the process, if a population of characters is distributing according to the statistical law of deviation then it can be deduced that the second generation will feature a distribution curve of the same mean and dispersion. I say "deduced" to indicate that the quincunx demonstration is simulating a mathematical result well known in statistical mathematics.

Galton's statistical insight is preserved in modern versions of evolutionary theory. Natural selection is a process that determines the frequency distribution of heritable traits in a population over the course of generational time. Referring to the illustration below (taken from Endler's book "Natural Selection in the wild"): In "directional" selection, circumstances favor features to the right side of the variation curve (shade is disfavored) thereby modifying the distribution of characters rightwards. In "stabilizing" selection, the extreme characters are at a selective disadvantage thereby redistributing the next generation of characters towards the center of the curve. "Disruptive" selection

favors the extreme ends skewing the distribution of characters in the next generation towards the extremes.



Muller challenged us to articulate why natural selection is hard to understand, and in this essay, I have presented us with an answer: because statistical thinking is hard to understand, or at least it requires knowledge of some mathematical ideas that high school biology students might find strange and conceptually difficult. What should we teach first? While Gould's barebones account is very easy to understand, it is misleading because it ignores the variational character of evolutionary theory. If we are seeking a real understanding of nature, then it is worthwhile to tend to the statistical ideas; in this essay, I attempted to showcase how one might want to do it with the use of simple and inexpensive toy models and without reference to any mathematical equations.