Darwin's theory is strong but also weak, and needs augmenting

One way brains are built is through natural processes like evolution and embryo development. The latter sort of brainbuilding occurs routinely in mothers' wombs everywhere. It fascinates me that some people calmly accept the assembly of a human over the course of months in a womb, while denying that a similar assembly might occur through millions of years of evolution. One might think that the assembly of a human over millions of years would be *less* shocking than the same thing accomplished in mere months! Both forms of brainbuilding are miraculous to me, and that inspires me to study them.

It also fascinates me that many scientists and inveterate skeptics think that Charles Darwin's theory of evolution adequately accounts for human origins. It does not. Darwin's theory is only an excellent first step in a long journey of understanding. To prepare for later steps I now give a brief account of the strengths and weaknesses of Darwin's theory of evolution.

Charles Darwin's theory of evolution is like a theory of lost things. Its strength is that, like a theory of lost balls, it accurately identifies where organisms come from—they descend from common ancestors, as a ball comes from a hand—and its weakness is that at no point in the history can it tell what any organism's descendants will become, a fate as hidden as any ball's in the bush. This is a weakness because an important measure a scientific theory's power is how well it predicts the future. The theory of celestial mechanics predicts eclipses centuries in advance. Darwin's theory can do nothing of the kind.

It may be hard to see the connection between a theory of evolution and a lost ball, so now I offer a little more detail.

Darwin's Theory of Evolution is simple enough to outline in a single sentence: Organism traits are

1) heritable,
2) variable, and
3) differently affect reproductive success,

so that long-term change in organism traits is inevitable.

The functioning of Darwin's three-part engine also provides that

1) organism traits are broadly adapted to their circumstances,
2) all organisms descend by degrees from other dissimilar organisms, and
3) any two organisms share a common ancestor.

The last two inferences offended people who would rather have sprung up quickly in full human glory than slowly from something like a monkey: In 1925CE schoolteacher John

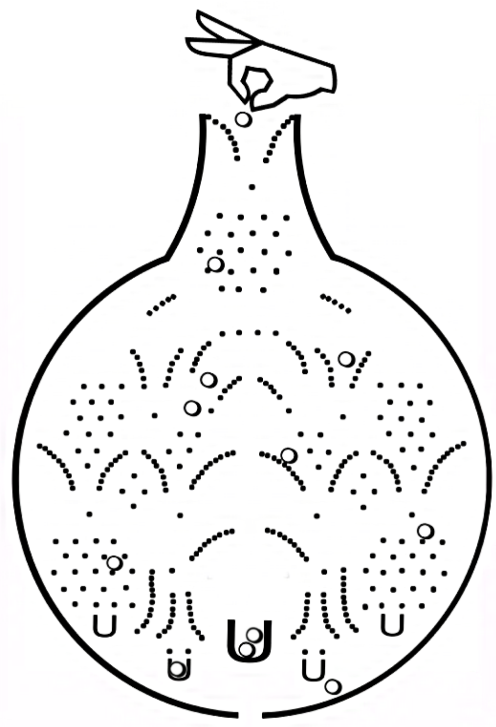
Scopes was convicted of advancing Darwin's theory in a Tennessee classroom, and the court case became popularly known as the *Monkey Trial*.[1]

As noted, the logic of Charles Darwin’s Theory of Evolution provides that one can trace the history of any two organisms to a common ancestor. In addition, scientifically literate observers generally find the fossil and DNA evidence for it overwhelming. In this way Darwin’s theory is good at *retro*dicting history. This is its great strength. However, also as noted, it is poor at *pre*dicting history, and this is its main weakness. Note that retrodicting common ancestry does entail a weak form of prediction: One predicts, for example, that future evidence will confirm the shared ancestry of humans and apes. However, this form of prediction, bearing as it does on the human past, remains essentially retrodictive.

Predicting the course life will take from a given point is a more profound problem than confirming the course it did take in the past. In this light there are features of life on Earth—features reasonably called *important*—for which Darwin’s theory does not adequately account. In particular **Darwin’s theory fails to predict the rise of human-like brains** or to say much about what brain structure must be. A further metaphor illustrates the coupled retrodictive strength and predictive weakness of Darwin's theory.

Pachinko is a vertical pinball game invented in Japan, and Darwin's theory may be compared to a theory of Pachinko. In Pachinko steel balls about the size of chickpeas are dropped from an upper location and then descend through a forest of metal pins nailed in various patterns to a board. The balls tumble down, colliding with the pins and bouncing and rolling to left and right as they fall. Balls dropped at the same top location follow very different routes through the pins. Through one route or another the balls make their way toward the bottom, and occasionally a ball rolls into a special hole or cup, earning a prize jackpot. The pattern of pins varies from machine to machine, but the basic principle of operation remains the same. The simple Pachinko game pictured below is unusual for its pomegranate shape and for the fact that the balls are dropped manually from a single location at the top. The pins, seen end-on, appear in the picture as black dots. Five U-shaped prize cups and a dozen descending balls are also shown.

[1] L. Sprague de Camp, *The Great Monkey Trial* Doubleday (1968)



Pachinko

Darwin's theory may be compared to a Theory of Pachinko that says, "All balls at the bottom of a Pachinko game started at the top and their varied positions when arriving at the bottom are due to the many small influences from collisions on the way down." This theory, like Darwin's, can retrodict, that is, detail the past: It says that any ball at the bottom started in one place at the top. In the Pachinko metaphor, people who object to a common descent with apes are like the two Pachinko balls in the large **U** cup, whispering that they cannot possibly share an origin with the ball in a smaller cup.

As a theory the Theory of Pachinko is strong, reasonable, and testable. It's obvious from inspection that the balls at the bottom start at the top and that each ball's course through the maze of pins is strongly influenced by collisions with them. However, to a Pachinko player what's important is not where a ball starts but where it will go—it is the precise location of a ball at the bottom that is most important, for that is what determines whether there is a jackpot.

Like Darwin's theory the Pachinko theory predicts mildly: It predicts that balls will move toward the bottom by way of many small influences. However, beyond that the Pachinko Theory's predictions are vague and weak: By its sights it is essentially uncertain where exactly a ball dropped at the top will hit the bottom.

The Theory of Pachinko can be made more predictive. One can predict, for example, that a ball falling deeply into one of the funnel-shaped pin arrangements will drop through its bottom. This is a local prediction affecting only small regions of the Pachinko layout. It's also possible to make global predictions of a statistical nature. For example, as a crude first approximation one might predict that whatever fraction of total distance across the game's bottom is occupied by cups, that is the fraction of balls that will end up in cups. With a little physics and statistics this estimate can easily be improved. Doubtless there is somewhere a technical paper in Japanese with a title like *From Layout to Payout: Estimating a priori odds in Pachinko.*

Note that the forgoing predictive extensions of the Theory of Pachinko are not part of the original Theory of Pachinko: They are *additions* to the theory and in a sense form a new theory of their own. The additions also depend strongly on the precise pin layout of a particular Pachinko game and are in this sense ad hoc additions to the original Pachinko theory—i.e. they are special-purpose additions tailored to the job of predicting ball behavior in particular cases.

Darwin's Theory of Evolution is, like the Theory of Pachinko, strong, reasonable, and testable: To deny the three-part engine of Darwinian evolution is to deny that any organism traits are heritable, variable, or differently affect reproductive success. Denying even one of the three provisions is absurd, and so *some* form of species evolution must occur. As Thomas Henry Huxley put it, "…Mr. Darwin does not so much prove that natural selection does occur as that it must occur…"[2] Furthermore, as mentioned, fossil and DNA evidence confirm the retrodictive expectations of Darwin's theory. However, it is not retrodiction but prediction that interests us here. For example, we would like to know the direction, if any exists, that a world with apes but no humans is ordained to go.

Many casual observers imagine that Darwin's theory in its original form predicts evolution. They think, for example, that when a moth population becomes darker overall in a newly soot-polluted world, or when finch beaks change shape over generations as seed prevalence changes, or when bacteria become resistant to antibiotics—that these events are somehow foreordained by the Darwinian machine—as if they were guided by pins nailed like funnels to a board. Such casual observers may be heard to intone, "These developments are what we expect from Darwinian evolution."

Expect means predict, and in each of the forgoing scenarios prediction depends on ad hoc assumptions—assumptions for example that moths, finches and bacteria had the genetic wherewithal to produce the observed changes, and that the needed changes carried no deadly risk to the changing organisms themselves. The assumptions are ad hoc because there is no incontrovertible necessity to them but rather they exist specifically to rationalize a hypothetical mechanism in a special case. The assumptions are like pitons installed on a cliff to make it climbable.

It seems *reasonable* that finch beaks will change to suit food supply—until it turns out that the requisite beak changes are beyond the reach of genetic variation, or that beak changes affect flight, mating, defense ability, parasite load, and so on, with fatal results. It seems reasonable that moths will develop dark wings in a soot-stained world, until it turns out that dark coloration predisposes moths to fatal moth melanoma. The

[2] Huxley, Thomas H., *The Origin of Species* London: The Westminster Review, April (1860)

assumptions that these complications will *not* arise are ad hoc, and most predictive readings of Darwin's theory are riddled with unacknowledged assumption leakage of this kind. Without the assumptions prediction remains vague and general at best.

Predicting evolution by means of ad hoc assumptions is like predicting the future course of a highway from a straight piece of roadway. If only reasonable—and no radical and unpredictable—changes were permitted then today all living things arguably would resemble primeval organisms. In any case, whatever the assumptions allowing prediction, they are not part of Darwin's original theory, but are extensions of it, and in the first century and a half after the publication of Darwin's theory the extensions were almost always unacknowledged and ad hoc.

To review, one can say with confidence that a Pachinko ball found at the bottom of a game started at the top and bounced it's way down, but the theory has little to say about where precisely it ends up. Similarly, as far as Darwin's theory is concerned absent further ad hoc assumptions a monkey might become a marmot or rutabaga: The emergence of human brains remains essentially unexplained by Darwin's theory and is by its sight as good as miraculous. If a monkey gives rise to a marmot Darwin happily retrodicts a series of intermediate forms, and also variants that perish along the way, but that's as far as his theory reaches. Only through the addition of assumptions can Darwin's theory be made more predictive. The question then becomes whether one can make such assumptions both explicit and incontrovertible rather than ad hoc.

In an alternative metaphor, Darwin's theory may be compared to a theory that all of the Sistine Chapel paintings are exhaustively composed of brush strokes and of atoms. Indeed all of the images can be traced to a sequence of brush strokes and an agglomeration of atoms, but that theory overlooks what arguably is most interesting and important about the paintings.

As I write it is common to hear scientists dismiss Darwinian evolution as having no predictable direction. However, it's odd that people dismiss the idea of humans arising in a lawlike fashion over the course of millions of years from something not recognizably human—when, as earlier noted, a similarly radical transition occurs routinely and in mere months in wombs worldwide. If the transition from an aliquot of water and chemicals—a fertilized egg cell—to a brainy human could not occur in a lawlike fashion our species would have perished long ago. It seems a little odd then that the self-organization of a brain in a womb is almost a paradigm of mechanistic determination when, as I write, the same self-organization occurring over millions of years is commonly regarded as blind. The highly predictable miracle of gestation and human birth makes a highly predictive augmentation of Darwin's theory at least plausible.

We ask whether we can augment Darwin's theory so that the augmented theory joins the company of more rigorous predictive theories. Can we say with confidence not only *this did happen* but also with some rigor and detail *this must have happened* and *this will* *happen*? The shining goal here is the human brain or its parallel elsewhere in the universe. Not all assumptions are ad hoc. Global statistical augmentations of Darwin's theory are conceivable, for example, and some of these may apply with such necessity that they cannot properly be called ad hoc. Thus there may be unavoidable constraints that together with Darwin *do* predict something like a human brain. It is precisely such augmentations of Darwin that we seek here.

It is important to understand that an augmentation of Darwin's theory is not necessarily an extension of it. Newtonian celestial mechanics can be thought of as an augmentation of atomic theory without being an extension of it: The equations describing planetary motion do not formally depend on the existence of atoms. In the same way an augmentation of Darwin's theory need not strictly depend on it.

The reasoning going forward [in later chapters or essays] suggests that a more rigorously predictive theory of evolution is indeed possible, one that begins to reveal inevitable features of humanlike brains and that yields formally testable retrodictions and predictions alike. We begin the quest by tempering our ambition with the limits of prediction. There are two great limits to prediction, the first practical and the second philosophical.

The Asteroid Problem

The trouble with Darwin's theory is that it does not predict well, but there is trouble with prediction itself. In an old story a little Dutch boy wanders one day along one of the dikes that holds back the sea from his native land. By the sheerest chance he passes a small but dangerous leak in the dike. Seeing the threat to his lowland nation the boy plunges his finger into the hole, stopping the flow. He remains stalwartly in place until adults find him, gratefully relieve him, and effect the necessary repairs to the dike. The boy's heroism is then sung throughout the land, for if he had not found the leak and performed his digital act the dike would have been breached and the nation drowned.[3]

The story of the Little Dutch Boy is of course fanciful, but it illustrates a phenomenon common in nature: Important events can depend on events as unpredictable as a coin toss or an able boy's chancing upon a hole in a dike. In the story it is only the sheerest luck that takes the boy to the leak when it is small enough to plug, with the fate of the nation in the balance. Examples of small and unpredictable events with momentous outcomes are common in sports: A ball takes an odd hop granting a team the winning score; a player barely makes a tackle, blocking a game-changing open-field run, and so on. It is true that the stronger team usually wins, but it does not always, and in any case what is most interesting is the actual play of the game.

The history of life on Earth is filled with events of consequence precipitated by events as small and unpredictable as any in sport. The whim of a meteor's path erases all big animals on Earth in a virtual instant. Continents collide and drift apart. Volcanoes bring

[3] "The Little Hero of Haarlem". *Sharpe's London Journal of Entertainment and Instruction* **12**:8–9 (1850)

global winters and new lands. A landslide buries an ecosystem or a hurricane wipes it clean. A tree falls and a tiny species is no more. At every scale the important depends on the unpredictable, and varied forms of life perish or thrive accordingly. There are averages to be sure but important features always escape these blurry averages.

To predict the course of evolution in robust detail is to predict the hops of balls, the vagaries of winds and seas, and the private life of meteors. To boast too loudly of a theory's power, then, is to invite humility at the hands of chance. Unless one magically prevents the unpredictable, or anything depending on the unpredictable, one's theories will be imperfect prophets. **No theory of evolution can be very strongly predictive.** Even where asteroids, seas and winds have all been tamed, important but unpredictable events will arise and be indistinguishable from miracles.

Doubt

The trouble with prediction runs deeper than the whims of chance. The world of the Little Dutch Boy is a world of chance but it is also a world of regularity: In it perforated dikes leak and patient waiting is rewarded by aid. What if the regularities of the universe suddenly ended? We can easily imagine a universe in which regularities change suddenly in an instant but we have no way of knowing in advance whether ours is such a universe. That something has always been so does not mean it must always be so. It doesn't even mean it *probably* must be so. In the face of this frightening fact people commonly comfort themselves with a laugh and a protest—"Oh that's just silly nonsense!"—and then quickly turn their attention elsewhere.

Against the backdrop of meteor strikes and doubt we nevertheless predict. We predict the motion of planets, the rising of the sun, and that wind will lift a wing. The strongest scientific theories enable clear and testable predictions, and there are many such powerful theories. Because of our practical ability to predict we trust our lives to wings and go to bed comfortable in the knowledge that a dawn awaits.

Evolution Must Have* Some *Direction

Skeptics sometimes seize on the asteroid problem as evidence that evolution can have no direction. However, one might as well protest that asteroid impacts prove that gravity has no direction because the impacts send so much rock flying up to the heavens. The question then is whether something analogous to gravity urges life's evolution in a particular direction, even if the evolution can be knocked off course by blows, or is forced to incorporate a blow's effects.

A simple *reductio ad absurdum* argument shows that evolution must have *some* direction—that it does not head off haphazardly toward any pole at any moment. Suppose that evolution actually had no systematic direction. It would follow that in any moment a monkey might give birth to a rutabaga, or become one. All seedlings might start growing toward darkness rather than light, and send roots toward the driest possible ground. Foals might universally gallop into walls and cliffs. If evolution were completely unconstrained life would end in a few generations at best. The very existence of life tells that evolution *is* constrained—that there are directions evolution does not go, and thus that there are general directions evolution does go.

The quest for prediction leads to this: On the one hand we cannot erase the effects of chance or prove that regularities will persist forever; and on the other hand there are regularities that enable predictions confirmed every day. With ambitions tempered by chance and doubt we nevertheless attempt a predictive theory of evolution, evolution that we see must have at least some direction. We cannot hope for a perfectly predictive theory but we can attempt something like a weather forecast—a theory that, assuming the universe preserves its regular ways, yields probability without certainty.