# 1

# Why Now?

We are born crying, but those cries herald the first stirrings of language. German babies' cries mirror the melody of German speech; French babies mirror French speech—apparently acquired *in utero* (Mampe et al. 2009). Within the first year or so after birth, infants master the sound system of their language; then, after another few years have passed, they are engaging their caretakers in conversation. This remarkable, species-specific ability to acquire any human language—the "faculty of language"—has long raised important biological questions, including the following: What is the nature of language? How does it function? How has it evolved?

This collection of essays addresses the third question: the evolution of language. Despite claims to the contrary, in truth there has always been strong interest in the evolution of language since the very beginning of generative grammar in the mid-twentieth century. Generative grammar sought, for the first time, to provide explicit accounts of languages—grammars—that would explain what we will call the Basic Property of language: that a language is a finite computational system yielding an infinity of expressions, each of which has a definite interpretation in semantic-pragmatic and sensorimotor systems (informally, thought and sound). When this

problem was first addressed the task seemed overwhelming. Linguists scrambled to construct barely adequate grammars, and the results were so complex that it was clear at the time that they could not possibly be evolvable. For that reason, discussions about the evolution of language rarely reached publication, though there were some notable exceptions.

So what has changed? For starters, linguistic theory has matured. Complex linguistic rule systems are now a thing of the past; they have been replaced by much simpler, hence more evolutionarily plausible, approaches. Then too, certain key biological components associated with language, in particular the "input-output" system of vocal learning and production that constitutes part of the system we will call "externalization," have been clarified biologically and genetically, so much so that we can effectively use a "divide-and-conquer" strategy and place this sensorimotor aspect of externalization aside while we focus on language's more central properties.

While much must remain uncertain simply because we lack the required evidence, developments in linguistic theory over the past two decades have greatly clarified aspects of language's origin. In particular, we now have good reasons to believe that a key component of human language—the basic engine that drives language syntax—is far simpler than most would have thought just a few decades ago. This is a welcome result for both evolutionary biology and linguistics. Biologists well know that the more narrowly defined the "phenotype," literally the outward "form that shows," the better our biological grip on how that phenotype might have evolved—and equally, the narrower the gap between us and other species that lack language. With this better-defined phenotype in hand, we can begin to resolve the dilemma that plagued the Darwinian explanation of language evolution from the start.

In various places this has been called "Darwin's problem" or, more appropriately, "Wallace's problem"—after the codiscoverer of evolution by natural selection, Alfred Russel Wallace. Wallace was the first to call attention to the difficulties for any conventional Darwinian, adaptationist account of human language, since he could perceive no biological function that could not already be met by a species without language.[1]

Language does indeed pose a severe challenge for evolutionary explanation. On the one hand, Darwinian thinking typically calls for gradual descent from an ancestor via a sequence of slight modifications. On the other hand, since no other animal has language, it appears to be a biological leap, violating Linnaeus's and Darwin's principle, *natura non facit saltum*: "For natural selection can act only by taking advantage of slight successive variations; she can never take a leap, but must advance by the shortest and slowest steps" (Darwin 1859, 194). We firmly believe that this tension between Darwinian continuity and change can be resolved. That's one key goal of these essays.

What of Darwin? Never wavering from his strong principles of infinitesimal evolutionary change and continuity, in his *The Descent of Man* (1871) Darwin himself advanced a "Caruso" theory for the evolution of language: males who could sing better were sexually selected by females, and this, in turn, led to perfection of the vocal apparatus, like the peacock's tail. Better vocal competence went hand in hand with a general increase in brain size that led, in turn, to language—language used for internal mental thought:

> As the voice was used more and more, the vocal organs would have been strengthened and perfected through the principle of the inherited effects of use; and this would have reacted on the power of speech. But the relation between the continued use of language and

> the development of the brain has no doubt been far more important. The mental powers in some early progenitor of man must have been more highly developed than in any existing ape, before even the most imperfect form of speech could have come into use; but we may confidently believe that the continued use and advancement of this power would have reacted on the mind by enabling and encouraging it to carry on long trains of thought. A long and complex train of thought can no more be carried on without the aid of words, whether spoken or silent, than a long calculation without the use of figures or algebra. (Darwin 1871, 57)

Darwin's Caruso theory has recently undergone something of a revival. In fact, one of us (Berwick) advanced an updated version at the very first "Evolang" conference at Edinburgh in 1996, grounded on the modern linguistic theory of metrical structure.[2] Most recently perhaps, no one has done more to champion a version of Darwin's "musical protolanguage" theory than Fitch (2010). As he notes, Darwin's theory was in many ways remarkably prescient and modern. We share Darwin's view in the passage cited above that language is closely allied with thought, an "internal mental tool" in the words of the paleoneurologist Harry Jerison (1973, 55). We provide empirical linguistic support for this position in chapter 3.

Contrary to certain views, discussion of the evolution of language as "Darwin's problem" was not a taboo topic until its "revival" in the 1990s—like some quirky relative that had been squirreled away for thirty years in an upstairs attic. On the contrary, it was a subject of intense interest in Cambridge, Massachusetts, during the 1950s and 1960s and then throughout the 1970s. This deep interest is directly reflected in Eric Lenneberg's September 1966 preface to his *Biological Foundations of Language* (1967, viii), where he notes his debt "over the past 15 years" to a roll call of famous and familiar names: Roger Brown, Jerome Bruner, George Miller, Hans Teuber, Philip Liberman, Ernst Mayr, Charles Gross—and also Noam

Chomsky. In our view, Lenneberg's book remains highly pertinent today—in particular, his chapter 6, "Language in the Light of Evolution and Genetics," still stands as a model of nuanced evolutionary thinking, as does his even earlier work (Lenneberg 1964). In a certain sense, our essays update what Lenneberg had already written.

As far as we understand this history, it was Lenneberg who presciently proposed longitudinal collection of child-directed speech; discovered the spontaneous invention of sign language as a full human language (at the Watertown, Massachusetts, Perkins School for the Deaf); found that language acquisition still succeeded despite gross pathologies; presented the evidence for a critical period for language acquisition; noted dissociations between language syntax and other cognitive faculties; coined modern terminology such as the "language-ready brain;" used pedigree analysis of families with language impairment, echoing the *FOXP2* data to provide evidence that language has a genetic component; and noted that "there is no need to assume 'genes for language'" (Lenneberg 1967, 265). He also contrasted continuous versus discontinuous approaches to language's evolution, arguing for the discontinuous position—supported in part by key evidence such as the apparent uniformity of the language faculty: "The identical capacity for language among all races suggests that this phenomenon must have existed before racial diversification" (Lenneberg 1967, 266).

In truth, then, there has always been an abiding interest in the question of language and its evolution. To be sure, in the 1950s and 1960s not much more could be said about language evolution beyond what Lenneberg wrote. Typical generative grammars of the day consisted of many complex, ordered, transformational rules. A glance at appendix II of Chomsky's

*Syntactic Structures* (1957) with its twenty-six highly detailed rules for a fragment of English immediately reveals this intricacy. Nonetheless, interest in the evolution of language did not wane, and from time to time major conferences were held on the topic—for example, an international conference in 1975 at the New York Academy of Sciences (Harnad, Steklis, and Lancaster 1976). By that time, starting from the mid-1960s on, it was understood that while complex rule systems that varied radically from one language to the next might well meet the demands of adequate description for each particular language, they left children's easy language acquisition no matter what the language a total mystery. It was realized that some of this mystery could be dissolved by discovering constraints on the biological system for language acquisition—constraints on universal grammar, or UG, the theory of the genetic component of the language faculty.[3] In the 1975 New York Academy conference on the evolution of language, one of us (Chomsky) noted, just as at the start of this chapter, that there seemed to be constraints that restrict the language "phenotype," thereby narrowing the target of evolution. For example, linguistic rules are often restricted to particular domains, so that one can say *Who did Mary believe that Bill wanted her to see,* where *who* is interpreted as the object of *see,* but this is impossible when *who* is embedded with a Noun Phrase, as in, *Who did Mary believe the claim that John saw* (Chomsky 1976, 50). (See also chapter 4.) As that presentation concluded, "There is every reason to suppose that this mental organ, human language, develops in accordance with its genetically determined characteristics, with some minor modifications that give one language or another" (Chomsky 1976, 56). Questions like these arose at once as soon as efforts were made to construct a generative grammar for even a single language.

During the next ten years the pace of discoveries of this sort quickened, and a substantial array of systematic constraints on UG were accumulated that came to be known as the "Principles and Parameters framework" (P&P). In the P&P model, the detailed transformational rules of *Syntactic Structures*—for example, the "passive rule" that shifted Noun Phrases from Object to Subject positions in English, or the rule that moved words like *who* to the front of sentences in English questions—were combined into a single operation, "Move any phrase" ("Move alpha"), along with a set of constraints that winnowed out illicit movements, such as a more general form of the constraint described in the previous paragraph for *wh*-words like *who* or *what*. All this was parameterized via a finite array of allowable perturbations that captured differences from language to language—for instance, that Japanese is verb final, but English and French are verb initial. Linguistic theory took on some of the look of the Periodic Table, atoms combining into possible molecules, as noted in accounts like that of Mark Baker (2002).

By the 1990s, with the Principles and Parameters model accounting for a fair range of crosslinguistic variation, it became possible for the first time to step back and see whether one could boil down both the rules and the constraints into the smallest possible set that could be independently motivated, such as by principles of efficient or optimal computation. This pursuit of the *simplest* or most *minimal* system for human language has led to considerable simplification—a narrower language phenotype.

How can we characterize this narrower phenotype? The past sixty years of research into generative grammar has uncovered several basic, largely uncontroversial, principles about human language. Human language syntactic structure

has at least three key properties, all captured by minimalist system assumptions: (1) human language syntax is hierarchical, and is blind to considerations of linear order, with linear ordering constraints reserved for externalization; (2) the particular hierarchical structures associated with sentences affects their interpretation; and (3) there is no upper bound on the depth of relevant hierarchical structure. Note that if all this is true, then observation (1) implies that any adequate linguistic theory must have *some* way to construct arrays of hierarchically structured expressions, while ignoring linear order; while (2) implies that structure (in part) fixes interpretation at the level of "meaning." Finally, (3) implies that these expressions are potentially infinite. These then are the minimal properties any adequate syntactic theory must encompass and that's why they are part of the minimalist account.

To see that these properties do indeed hold in language, consider a simple example that we'll use later, in chapters 3 and 4: the contrast between *birds that fly instinctively swim* and *instinctively birds that fly swim*. The first example sentence is ambiguous. The adverb *instinctively* can modify either *fly* or *swim*—birds either fly instinctively, or else they swim instinctively. Now let's look at the second sentence. Placing *instinctively* at the front is a game-changer. With *instinctively birds that fly swim*, now *instinctively* can only modify *swim*. It cannot modify *fly*. This seems mysterious. After all, *instinctively* is closer to *fly* in terms of number of words than it is to *swim*; there are only two words between *instinctively* and *fly*, but three words between *instinctively* and *swim*. However, people don't associate *instinctively* with the closer word *fly*. Instead, they associate *instinctively* with the more distant word *swim*. The reason is that *instinctively* is actually closer to *swim* than it is to *fly* in terms of structural distance. *Swim* is

embedded only one level deep from *instinctively*, while *fly* is embedded one level deeper.than that. (Figure 4.1 in chapter 4 provides a picture.) Apparently, it is not linear distance that matters in human syntax, only structural distance.

Not only do hierarchical properties rule the roost in human syntax, they have no real upper bound, though of course processing difficulty may increase, as in an example such as *intuitively people know that instinctively birds that fly swim.* If one subscribes to the Church-Turing thesis along with the assumption that the brain is finite, then there is no way out: we *require* some notion of recursion to adequately describe such phenomena. So much is uncontroversial. Together, these three properties set out the *minimal* requirements for an adequate theory of human language syntax.

However, contemporary discussion of primate neuroscience has sometimes explicitly and strongly denied each one these three claims, arguing that only linear order-sensitive constraints are required, and, further, that there is no need to appeal to hierarchical constraints or a notion of recursion. This position has strong implications for both neurobiological language research and evolutionary modeling. But, it is incorrect.

For example, Bornkessel-Schlesewsky and colleagues (Bornkessel-Schlesewsky et al. 2015) argue for continuity between humans and other primates on this basis: "We do not subscribe to the notion...that a more elaborate and qualitatively distinct computational mechanism (i.e., discrete infinity produced by recursion) is required for human language. ... The ability to combine two elements A and B in an order-sensitive manner to yield the sequence AB forms the computational basis for the processing capacity...in human language (2015, 146).

They draw a potentially critical evolutionary conclusion: "there is compelling evidence to suggest that the computational architecture of the nonhuman primate...is qualitatively sufficient for performing the requisite computations (Bornkessel-Schlesewsky et al. 2015, 143). If true, this would have profound evolutionary consequences. Then "the basic computational biological prerequisites for human language, including sentence and discourse processing, are already present in nonhuman primates" (2015, 148).

But, as we have just seen, Bornkessel-Schlesewsky's claims are just plain wrong. Linear processing does not even come close to being adequate for human language. This means that the primate mechanisms identified by the Bornkessel-Shlesewsky et al. are *in principle insufficient* to account for what we typically find in human language. And if this is correct, it makes the nonhuman primate brain a poor candidate for modeling many aspects of human language.

Let's recap then what our minimalist analysis tells us. In the best case, there remains a single operation for building the hierarchical structure required for human language syntax, Merge. This operation takes any two syntactic elements and combines them into a new, larger hierarchically structured expression.

In its simplest terms, the Merge operation is just set formation. Given a syntactic object *X* (either a word-like atom or something that is itself a product of Merge) and another syntactic object *Y*, Merge forms a new, hierarchically structured object as the set {*X*, *Y*}; the new syntactic object is also assigned a label by some algorithm that satisfies the condition of minimal computation. For example, given *read* and *books*, Merge combines these into {*read, books*}, and the result is labeled via minimal search, which locates the features of the

"head" of the combination, in this case, the features of the verbal element *read*. This agrees with the traditional notion that the constituent structure for *read books* is a "verb phrase." This new syntactic expression can then enter into further computations, capturing what we called earlier the Basic Property of human language.

More about this approach may be found in the remaining chapters, but for the moment, it should be clear that narrowly focusing the phenotype in this way greatly eases the explanatory burden for evolutionary theory—we simply don't have as much to explain, reducing the Darwinian paradox. This recent refinement and narrowing of the human language phenotype is the first motivation behind this collection of essays.

Our second motivation is that our understanding of the biological basis for language has improved. We can now effectively use a "divide-and-conquer" strategy to carve the difficult evolutionary problem of "language" into the three parts as described by the Basic Property: (1) an internal computational system that builds hierarchically structured expressions with systematic interpretations at the interfaces with two other internal systems, namely (2) a sensorimotor system for externalization as production or parsing and (3) a conceptual system for inference, interpretation, planning, and the organization of action—what is informally called "thought." It is important to note that externalization includes much more than just vocal/motor learning and production, encompassing at least aspects of language such as word formation (morphology) and its relationship to language's sound systems (phonology and phonetics), readjustment in output to ease memory load during production, and prosody.

More importantly from our standpoint, though, in the case of language, apparently *any* sensory modality can be used for

input or output—sound, sign, or touch (thankfully, smell appears to be absent from this list). Note that the internal hierarchical structure itself carries no information about the left-to-right *order* of phrases, words, or other elements. For example, the verb-Object or Object-verb possibilities distinguishing Japanese from English and French are not even represented in the internal hierarchical structure. Rather, language's sequential temporal ordering is imposed by the demands of externalization. If the modality is auditory, this output is more familiarly called speech and includes vocal learning and production. But the output modality can also be visual and motor, as in signed languages.

Thanks in part to comparative and neurophysiological and genomic studies of songbirds, the biological basis for vocal learning is well on the way to being understood as an evolutionarily convergent system: identically but independently evolved in birds and us. It may well be that vocal learning—the ability to learn distinctive, ordered sounds—can be bootstrapped from perhaps 100–200 genes (Pfenning et al. 2014). Vocal learning in both songbirds and vocal-learning mammals apparently also comes with a distinctive neurobiology, projections from vocal cortex motor regions to brainstem vocal motor neurons, as shown in the top half of figure 1.1 These direct projections are conspicuously absent in nonvocal learners like the chicken or the macaque, as shown in the bottom half of figure 1.1.[4]

More recent findings by Comins and Gentner (2015) and by Engresser et al. (2015) suggest that this learning ability goes beyond just simple sequencing. Comins and Gentner report that starlings exhibit abstract category formation reminiscent of human sound systems, while Engresser and colleagues claim to have found one bird species, the chestnut-crowned babbler

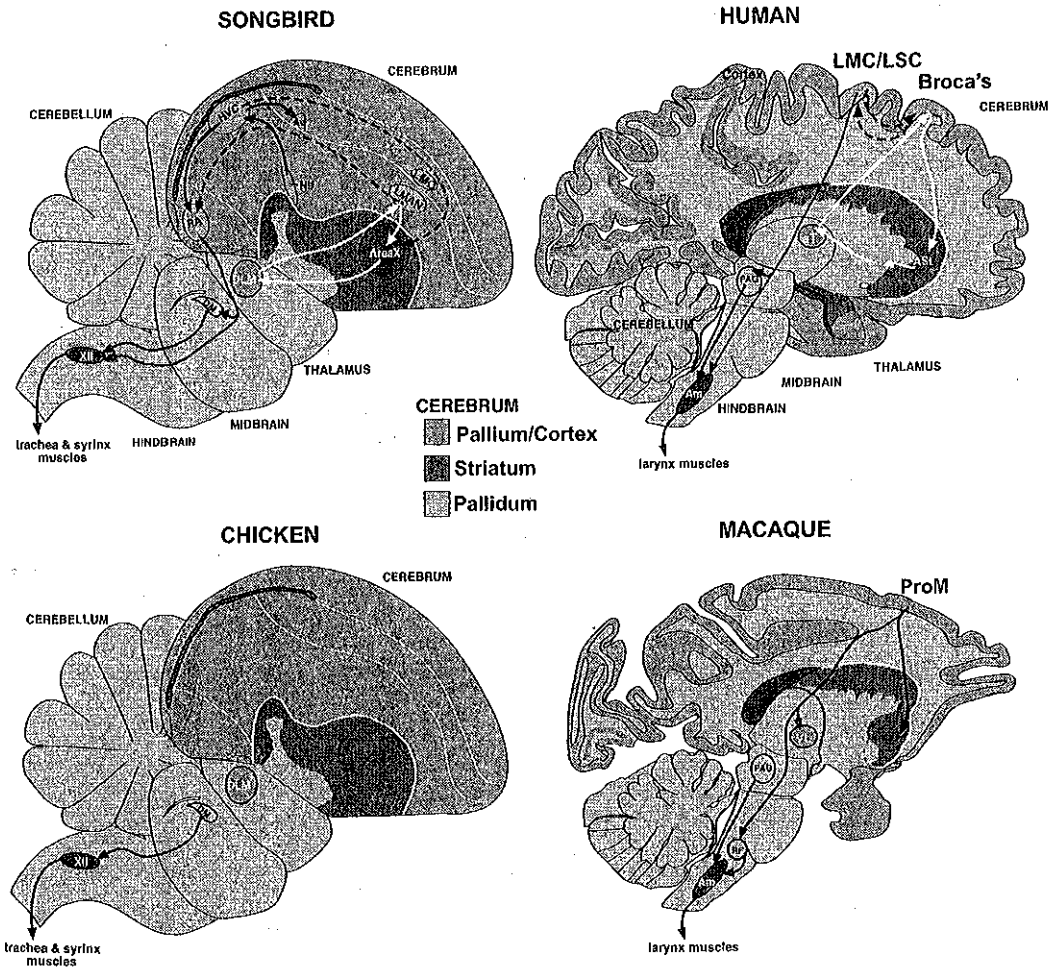


**Figure 1.1 (plate 1)**

Comparative brain relationships, connectivity, and cell types among vocal learners and nonvocal learners. Top panel: Only vocal learners (zebra finch male bird, human) have a direct projection from vocal motor cortex to brainstem vocal motor neurons, as marked by the red arrows. Abbreviations: (Finch) RA = robust nucleus of the arcopallium. (Human) LMC = laryngeal motor cortex in the precentral gyrus; LSC = laryngeal somatosensory cortex. Bottom panel: Nonvocal learners (chicken, macaque) lack this direct projection to the vocal motor neurons. Adapted from Pfenning et al. 2014. Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346: (6215), 1256846:1–10. With permission from AAAS.

(*Pomatostomus ruficeps*) with "phonemic contrasts." This species specific possibility was anticipated by Coen (2006). More recently still, Takahashi et al. (2015) have reported that baby marmosets "sharpen" their vocalizations in a manner that resembles human infant "tuning," a process that might be modeled in the way Coen envisioned. Berwick et al. (2011) have already demonstrated that the restricted linear sequencing in birdsong lends itself to acquisition from a computationally tractable number of positive examples. If all this is correct, it lets us set to one side this aspect of language's system for externalization and focus instead on the remaining central, human-specific aspects.

Finally, as just one bit of neurological evidence confirming our divide-and-conquer approach, there are even recent magnetoencephalographic (MEG) experimental results on dynamic cortical activity from David Poeppel's research group indicating that hierarchical entrainment to language structure is dissociated from linear entrainment to the word stream (Ding et al. 2015, in press). We have more to say about language and the brain in chapter 4.

Turning to our third motivation, it has seemed at least to us that Lenneberg's important insights regarding the biology and nature of language evolution were in danger of being lost. For example, he had a careful discussion of the pros and cons of evolutionary "continuity" approaches like Darwin's versus "discontinuity," his own choice. This seemed particularly poignant given recent advances in evolutionary thinking that have clarified these positions. Like any rich scientific field, modern evolutionary biology has moved on from Darwin's original view of evolution as adaptive change resulting from selection on individuals.

Darwin really *did* get some things wrong. Perhaps most familiar is what was repaired by the so-called Modern Synthesis—the mid-twentieth-century marriage of evolution by natural selection with Mendelism and particulate inheritance (genes), which remedied Darwin's lack of a good model of inheritance and eventually led to the modern genomic era in evolutionary analysis. Darwin had adopted the (incorrect) theory of inheritance of his day, "blending inheritance." On the blending account, if one breeds red flowers with white, all the offspring flower colors would fall somewhere in between: pink. Blending quickly wipes out the variation that natural selection feeds on—recall your childhood experience of taking a wet brush and letting it run up and down a palette of watercolors. The distinct color spectrum from purple to yellow turns a muddy brown. But if all offspring have the same muddy brown traits, there is nothing for natural selection to select. Nobody is above average, and nobody is below average; all are equal in natural selection's sieve. No variation—no natural selection, and the Darwinian machinery grinds to a halt in just a generation or two. What is needed is some way to preserve variation from generation to generation, even though red and white flower crosses sometimes turn out pink.

It was Mendel who discovered the answer: inheritance works via discrete particles—genes—though of course there was no way for him to know this at the time. In the first half of the twentieth century, it was left for the founders of the Modern Synthesis—Sewall Wright, Ronald A. Fisher, and J. B. S. Haldane—to show how to combine Mendel's particular inheritance with Darwin's evolution by natural selection in a systematic way, building mathematical models that explicitly demonstrated how the Darwinian machine could operate from

generation to generation to change the frequency of traits in populations.

But Darwin was also seriously wrong in his (generally tacit) assumption that biological populations are infinite, as well as his assumption that even in effectively infinite populations, evolution by natural selection is a purely deterministic process. *Every* cog in the evolutionary engine—fitness, migration, fertility, mating, development, and more—is subject to the slings and arrows of outrageous biological fortune. Quite often survival of the fittest boils down to survival of the luckiest—and this affects whether evolution might or might not be smoothly continuous in the way Darwin envisioned. To see this requires a more subtle mathematical analysis, and so far as we can make out, none of the recent books on the evolution of language seem to have grasped this in full. Darwin himself noted in his autobiography, "my power to follow a long and purely abstract train of thought is very limited; and therefore I could never have succeeded with metaphysics or mathematics" (Darwin 1887, 140).

In the remainder of this chapter, we unpack these last two motivations in reverse order, beginning with evolutionary theory and followed by a look at the divide-and-conquer approach along with evolution and genomics. We leave further details regarding the Minimalist Program and the Strong Minimalist Thesis for chapters 2 and 3.

## Evolutionary Theory's Evolution

To begin, what is so different about contemporary evolutionary theory and theories about the evolution of language? We might start with the historical setting around 1930, the heyday of the Modern Synthesis, as described just above. Most current

writers on language evolution seem to appreciate the history of Darwin's troubles with inheritance along with their resolution via the Modern Synthesis, and some even note some of the simple effects of finite population size on evolutionary change—for example, that sampling effects in small populations, sometimes called "genetic drift," might lead to the bad-luck-driven loss of advantageous traits (their frequency goes to 0 in the population) or the good-luck-driven complete fixation of nonadvantageous traits (their frequency goes to 1). It's not hard to see why. We can proceed as Sewall Wright and Ronald Fisher did: view a biological population as a finite collection of differently colored marbles in a jar, each marble an individual or a gene variant—say 80% white ones and 20% red. The population size is fixed—there is no selection, mutation, or migration to alter the color frequencies of the marbles in any other way. Now we simulate the generation of a small population of size 5. We do this by picking at random a marble from the jar, noting its color, and then putting it back in the jar until we have selected 5 marbles. The colors of the 5 selected marbles constitute the description of the new "offspring" generation. That counts as the first generation. Then we repeat the process, taking care that our second round of draws reflects any changes in frequency that might have occurred. So for example, we might wind up with 4 white marbles and 1 red one—this would match the frequency of white to red that we started with. But we might also wind up with, say, 3 white marbles and 2 red ones, 60% white and 40% red, in which case for the second generation we'd have a 2/5 chance of selecting a red marble. The game goes on, forever.

It's pretty clear that there's a real chance that we might not pick a red marble at all, and red would go extinct—once there

are no red marbles in the jar, there's no way for them to magically reappear (unless we assume that there's some way for white marbles to "mutate" into red ones). At the start, each time we draw from the jar, on average the red marble has a $1/5 = 20\%$ chance of being selected, just like any other "individual" in the population. Therefore, the probability that the red marble is *not* selected at any one draw is on average just one minus this probability, or $1-1/5 = 4/5$. The probability that the red marble would not be selected after *two* draws is just the product of not selecting it twice, $4/5 \times 4/5$ or $16/25$. And so on. On average, the probability in the first generation of not selecting the red five times is $(4/5)^5$ or about 0.328. So nearly a third of the time, the red marble might be "lost," and the frequency of red marbles would drop from 20% to 0. Similarly, if we picked the red marble 5 times in a row, the 80% frequency of the white marbles would drop to 0—this would happen on average $(1/5)^5 = 0.032\%$ of the time in the first generation, much less likely than the possibility of losing the red marble entirely. In this way the frequency of the mix of white and red marbles would "drift" between 0 and 1 from generation to generation in no particular direction—hence the terminology "genetic drift."

In fact it is not difficult to show that in this simple setting, given genetic drift any particular color will always wind up extinct or fixed. To picture this, it helps to think of "genetic drift" using another image, a "drunkard's walk." A drunk staggers away from their favorite bar, taking random steps at each tick of a clock in only one of just two directions: forward or backward. This is a random walk in one dimension. Where will the drunk go over time? Intuitively, since the drunkard begins to stagger just one step from the bar, it seems as though they ought to always wander back to their starting point. But

the intuition that random walks always fluctuate around their starting points is wrong. In fact, random walks always go somewhere—the expected distance from the starting point increases as the square root of time, which is the number of steps (Rice 2004, 76). If we recast the steps as trait or gene frequencies between 0 and 1, then on average half the time the drunk will reach 1—in which case the trait or gene has become fixed in the population and will stay at this point—and on average half the time the drunk will reach 0—in which case the trait has gone extinct and will also remain at 0. The leaders of the Modern Synthesis developed statistical models to demonstrate and predict these effects mathematically, at least in part.

However, as far as we have been able to determine, despite contemporary writers' embrace of the Modern Synthesis, *none* of the recent accounts of human language evolution seem to have completely grasped the shift from conventional Darwinism to its *fully* stochastic modern version—specifically, that there are stochastic effects not only due to sampling like directionless drift, but also due to *directed* stochastic variation in fitness, migration, and heritability—indeed, *all* the "forces" that affect individual or gene frequencies. Fitness is *not* some all-powerful "universal algorithmic acid" as some would have it. Contingency and chance play a large role. The space of possibilities is so vast that many, even most, "solutions" are unattainable to evolution by natural selection, despite the eons of time and billions of organisms at its disposal. Formal results along these lines have been recently established by Chatterjee et al. (2014), who prove that in general the time required for adaptation will be exponential in the length of the genomic sequence—which is to say, not enough time, even given geological eons. (The "parallel processing power" sometimes

attributed to evolution by natural selection because many organisms are in play turns out to be a chimera.)

Let's illustrate a stochastic effect with a real-world example. Steffanson and colleagues (Steffanson et al. 2005) discovered a particular large-scale disruption on human chromosome 17.[5] Icelandic women who carry this change on one chromosome have about 10% more offspring (0.0907) than Icelandic women who don't. Let's call these two groups C+ (for chromosomal change) and C (for no change). Per the usual Darwinian terminology, we say that the C+ women are 10% "*more fit*" than the C women or that the C+ women have a *selective advantage* of 0.10. In other words, for every child born to a C woman, a C+ woman has 1.1 children. (We use "scare quotes" around "fitness" for good reason.[6])

Now, from all we know about human reproduction, it's not hard to understand that in reality none of the C+ women could have actually had precisely 1.1 children more than a C woman. That would be particularly Solomon-esque. In reality, all the women the researchers tabulated (16, 959) bore either 0, 1, 2, 3, 4, or 5 or more children (2,657 women had 5 or more). So *on average* the C+ women had 10% more children than the C women—some of the "*more fit*" C+ women didn't have any children at all (in fact, a lot, 764 of them). And that is the nub of the point: any particular individual (or a gene) can be 10% "more fit" than the general population, yet still leave no offspring (or gene copies) behind. In fact, in our example, 764 "more fit" women had, in fact, *zero* fitness. Therefore, fitness is—must be—a random variable—it has an average and some variation about this average, which is to say a probability distribution. So fitness itself is stochastic—just like genetic drift (and migration, mutation, and the like). But unlike genetic drift, fitness or selective advantage has a definite direction—it doesn't wander around like the drunk.

All this can affect evolutionary outcomes—outcomes that as far as we can make out are not brought out in recent books on the evolution of language, yet would arise immediately in the case of any new genetic or individual innovation, precisely the kind of scenario likely to be in play when talking about language's emergence, when small groups and small breeding population size may have been the rule. Of course, whether or not models are sufficiently well specified to even reflect this level of detail.

Additionally, one might reply that fitness and Darwinian evolution are all about *population* averages and not individuals—what matters and what changes during evolution are the frequencies of fit versus less fit, not what happens to any particular woman. That's correct so far as it goes, but it does not apply when the number of individuals or gene copies is very small, and this happens to be precisely the situation of interest when considering the emergence of any genuine novel trait.

How so? If we pick a commonly used probability distribution to model situations like this, then a *single* individual (or gene) with a 10% fitness advantage has the (surprisingly large) probability of being lost in just one generation of more than one-third, about 0.33287.[7] And this is with a huge fitness advantage, perhaps 1 or 2 orders of magnitude larger than ordinarily measured in the field. Further, if a single individual or gene has *no* selective advantage at all—it is neutral, so it has a fitness of 1—then as one might expect its chance of being lost in one generation does indeed increase compared to its much more fit relative. However, the increase is slight: the chance of total loss rises to about 0.367 from 0.33, only 2%–3%. So contrary to what one might have initially thought—and contrary to what all the evolution-of-language books describe—this is *not* like the case of genetic drift, where

the smaller the population, the greater the chance of loss or gain. The size of the population does not play any role in the extinction-versus-survival probability across one generation when we are talking about small copy numbers of individuals or genes.

Why is this result important? Whenever a new gene variant or an individual with a new variant appears, then it will typically find itself alone in the world, or perhaps at most find itself one of four or five copies (if the new trait appeared in all the offspring of one particular individual due to a mutation). Population size will not govern the initial trajectory of this innovation—again contrary to the usual story one finds in the contemporary literature on the evolution of language. As Gillespie (2004, 92) puts it, "We judge [population size] to be irrelevant to the number of offspring produced by the lone [gene]. ... When the [gene] becomes more common and our interest turns from the number of copies to its *frequency*, its stochastic dynamics are more correctly said to be governed by genetic drift" [our emphasis]. In short, when new gene variants first appear, individuals with those traits must first climb out of a "stochastic gravity well" not governed by natural selection.

Once the number of such individuals (or gene copies) reaches a particular tipping point depending on fitness, then natural selection does take over the controls and the 10% more fit individuals ride the more familiar Darwinian roller coaster to the top, eventual total success, and fixation at frequency 1 in the population. (Why didn't the more fit Icelandic C+ women take over the entire country, or at least the Icelandic banks?)

And just what is that tipping point? If a new trait or gene variant has a selective advantage of 10% in order to be 99%

certain that this "new kid on the block" will not go extinct—that is, fix at frequency 1 rather than 0. This works out to be about 461 individuals. Importantly, this tipping point is also independent of population size. Gillespie (2004, 95) states the matter clearly: "In the initial generations, all that matters is the random number of offspring. ... There is no place for the $N$ [the population size] when modeling the fate of these individuals."

In short, to be a thoroughly modern evolutionary theorist, one really ought to move from a "gene's-eye view" to a "gambler's-eye view." (Readers interested in exploring this topic in greater depth are invited to consult Rice 2004, chaps. 8 and 9, or Rice, Papadapoulos, and Harting 2011.) What's the bottom line? We need to inject real-world biology and stochastic behavior into the evolutionary picture. This includes stochastic migration rates (Ellis Island yesterday and today); stochastic inheritance patterns (you don't look like your grandparents after all); interactions between genes (no single "gene for language"); and fitness fluctuating whenever frequency rises (overpopulation anyone?). If we do this, then the simple-minded view that adaptive evolution inexorably scales fitness peaks falls apart. It is difficult to simultaneously "satisfice" the effects of a thousand and one interacting genes, let alone tune them jointly to optimal fitness.

Some have claimed that these difficulties for natural selection can be fixed via the use of game theory applied in an evolutionary context—what are called "evolutionary stable strategies" (Maynard-Smith 1982), and, further that this has decisively "resolved" the problem associated with multidimensional fitness maximization (Fitch 2010, 54). This is not quite correct. There has been no such resolution, or at least, not yet. Game theory *does* have a very important place in modern

evolutionary thinking, because it is designed to consider what one individual should do given the actions or strategies of other individuals. As a result, it is particularly useful in the case of frequency-dependent selection, where fitness changes depending on how many other individuals are using the same strategy—for example, deciding to have offspring earlier in life rather than later. Such multidimensional frequency-dependent scenarios are typically extremely difficult to analyze in any other way. In fact, it seems to us that frequency-dependent effects might be exactly what would be expected in the case of human language evolution, with a dynamic interplay between individuals with/without language. We need Nowak's evolutionary dynamica models for language (2006).

We have not pursued the frequency-dependence/game-theoretic line of reasoning here because we are not certain whether the other assumptions it requires can be met. Game-theoretic evolutionary analysis is not the panacea it is sometimes made out to be, despite its widespread appearance at "Evolang" conferences. Game-theory analysis works best when population sizes are very large, at equilibrium, with no mutation, and when there is no sexual recombination—that is, precisely when we don't have to worry about stochastic effects, or when we want to know how populations moved toward equilibrium in the first place, and precisely contrary to some generally accepted assumptions that the human effective population size at that time was small and not at equilibrium. Finally, the game-theoretic approach has often been divorced from the insights we have gained from the study of population genetics and molecular evolution—and this happens to be a substantial part of what we have learned about evolution in the modern genomic era, and the vast bulk of the new data that has been and will be collected. To be sure,

there has been substantial recent progress in marrying classical Modern Synthesis population genetic models with game theoretic analysis by researchers such as Martin Nowak among others (Humplik, Hill, and Nowak 2014; McNamara 2013). Game theory remains an essential part of the modern evolutionary theorists' toolkit, but it has limitations, and these have yet to be fully worked out in the context of the rest of molecular evolution. (For further discussion, see Rice 2004, chap. 9; Rice, Papadapoulos, and Harting 2011). In short, Ecclesiastes 9:11 was right all along: "The race *is* not to the swift, nor the battle to the strong, neither yet bread to the wise, nor yet riches to men of understanding, nor yet favour to men of skill; but time and chance happeneth to them all."

If this conclusion is on the right track, it suggests that we need to take these stochastic effects into account when considering the evolution of language. Indeed, this element of chance seems to be implicated whenever one encounters the appearance of genuinely novel traits like the eye, as Gehring (2011) argues, and even as Darwin admitted—a bit grudgingly. We return to this point about the eye just below. More generally, we should understand that, as the evolutionary theorist H. Allen Orr has argued, "adaptation is not natural selection" (Orr 2005a, 119), so we need to be on alert whenever we find these two distinct notions casually run together.

This shift from deterministic Darwinism to its fully stochastic version is the result of a more sophisticated mathematical and biological understanding of evolution and stochastic processes developed since the publication of Darwin's *Origin* in 1859. Such progress is to be expected in any thriving scientific field—the evolution of evolutionary theory itself—but it seems as though many authors have not wavered from Darwin's original vision of evolution as solely adaptive selection on

individuals. We have known for some time now on both theoretical and empirical research that Darwin's and the Modern Synthesis views were not always accurate, and there is ample field evidence to back this up (Kimura 1983; Orr 1998, 2005a; Grant and Grant 2014; Thompson 2013)—all without the need to reject Darwinism wholesale; invoke viral transmission, large-scale horizontal gene flow, or miracle macromutations; or even incorporate legitimate insights from the field of evolution and development, or "evo-devo."

How then do organisms evolve? Is it evolution by creeps or evolution by jerks, as the famous exchange between Stephen J. Gould and his critics put it? (Turner 1984; Gould and Rose 2007). Both, of course. Sometimes adaptive evolutionary change is indeed very slow and plodding, operating over millions of years according to Darwin's classical vision. But sometimes evolutionary change, even large-scale behavioral changes, such as the food preferences of swallowtail butterflies (Thompson 2013, 65), can be relatively rapid, breathtakingly so. This speed has been confirmed in hundreds of different species across every major phylogenetic group, as noted recently in Thompson's magisterial survey (2013).

Here one must not muddy the waters simply by admitting, as some do, that Darwinian infinitesimal gradualism sometimes picks up its pace. We agree. But the crucial question is what's the pace regarding the evolutionary innovations at hand. Our view embraces both the long term possibilities—millions of years and hundreds of thousands of generations, as in the apparent evolution of a vocal learning toolkit antecedent to both avians and us—and the short term—a few thousands of years and hundreds or a thousand generations as in the case of relatively recent adaptations such as the Tibetan ability to thrive at high altitudes where there's less

oxygen; the ability to digest lactose past childhood in dairy farming cultures (Bersaglieri et al. 2004); or—our core belief—the innovative ability to assemble hierarchical syntactic structure.

Some of these traits skipped past the long haul of slow genetic change by following the biologist Lynn Margulis' advice: the quickest way to gain entire, innovative new genes is to eat them. The Tibetans evidently gained a snippet of regulatory DNA that is part our body's reaction to hypoxia by mating with our relatives, the Denisovans, so they gobbled up genes via introgression (Huerta-Sánchez et al. 2014). Apparently humans culled several important adaptive traits for surviving in Europe from the Neandertals and Denisovans, including skin pigment changes, immune system tweaks, and the like (Vernot and Akey 2014). To be sure, once eaten the genes had to prove their selective mettle—but this sort of genetic introgression can lift one out of the gravity well we mentioned earlier.

If there are any doubts that this kind of smuggling past the Darwinian entry gates is important, recall that it was Margulis who championed the theory, once decried but now confirmed, that organisms acquired the organelles called mitochondria that now power our cells by just such a free lunch, dining on another single cell via phagocytosis (Margulis 1970). This perhaps most ancient version of Manet's "luncheon on the grass" launched one of the eight "major transitions in evolutions," as identified by the evolutionary biologists John Maynard Smith and Eörs Szathmáry (1995). Maynard Smith and Szathmáry single out the important point that, of these eight transitions ranging from the origin of DNA to sexuality to the origin of language—six, including language, appear to have been unique evolutionary events confined to a single

lineage, with several transitions relatively rapid in the sense we've discussed above. Nothing here violates the most conventional Darwinism.

So there can indeed be abrupt genomic/phenotypic shifts, and what this does is "shift the starting point from where selection acts" as the biologist Nick Lane puts it (2015, 3112). Here Lane is commenting on the remarkable and apparently one-off and abrupt shift from simple cellular life, the prokaryotes—with circular DNA, no nucleus, no sex, and, essentially no death—to the gastronomy that led to complex life, the eukaryotes, including us—with linear DNA, mitochondria, a nucleus, complex organelles, and, ultimately beyond Woody Allen, sex, love, death, and language. As Lane remarks, "one must not confound genetic saltation with adaptation" (2015, 3113). From the perspective of geological time, these changes were swift.

All this underscores the role of chance, contingency, and biochemical-physical context in innovative evolutionary change—evolution by natural selection works blindly, with no "goal" of higher intelligence or language in mind. Some events happen only once and do not seem to be readily repeatable—the origin of cells with nuclei and mitochondria, and sex, and more. Other evolutionary biologists agree. Ernst Mayr, in a well-known debate with Carl Sagan, noted that our intelligence itself, and by implication language, probably also falls into the same category:

> Nothing demonstrates the improbability of the origin of high intelligence better than the millions of ... lineages that failed to achieve it. There have been billions, perhaps as many as 50 billion species since the origin of life. Only one of these achieved the kind of intelligence needed to establish a civilization. ... I can think of only two possible reasons for this rarity. One is that high intelligence is not at all favored by natural selection, contrary to what we would expect.

> In fact, all the other kinds of living organisms, millions of species, get along fine without high intelligence. The other possible reason for the rarity of intelligence is that it is extraordinarily difficult to acquire ... not surprisingly so because brains have extremely high energy requirements. ... a large brain, permitting high intelligence, developed in less than the last 6 percent of the life on the hominid line. It seems that it requires a complex combination of rare, favorable circumstances to produce high intelligence. (Mayr 1995)

Of course, given the results of Chatterjee et al. (2014), we now understand a bit more precisely the sense in which a trait might be "extraordinarily difficult to acquire": it might be computationally intractable to attain by natural selection.

Consider yet another example of rapid evolutionary change, one that may seem more concrete and secure because it's so recent and has been so thoroughly studied. One of the most thorough and long-running experimental observations of natural selection in the field is the forty-year study by P. R. Grant and B. R. Grant tracking the evolution of two species of Darwin's finches on the island Daphne Major in the Galápagos, *Geospiza fortis* and *G. scandens* (Grant and Grant 2014). This is evolutionary analysis as down to earth as one can get. What did the Grants discover? Evolutionary change was sometimes correlated with fitness differences, but equally sometimes it was not. As a result, fitness differences did not predict evolutionary outcomes. Selection varied from episodic to gradual. Singular events, like the appearance of a new finch species called "Big Bird" on Daphne Island, led to hybridization with existing finch species and spurts of evolutionary change prompted by external environmental events. All of these field observations bear witness to what one might actually expect in the case of human language evolution. As we noted above, intergroup hybridization from Denisovans and Neandertals has played a role in human adaptive human

evolution. While we do not mean to suggest that language arose this way—in fact, so far this seems to be specifically ruled out if we go by the evidence of genomic introgression—we do want to impress upon the reader that evolution can appeal to the hare just as well as the tortoise.

Why then is Darwinian evolution by natural selection generally assumed without question to be extremely gradual and slow? Darwin absorbed Lyell's influential three-volume *Principles of Geology* (Lyell 1830–1833) while on his *Beagle* voyage, along with its emphasis on "uniformitarianism"—forces in the present like those in the past, mountains slowly eroded to sand after eons. Darwin drank *Principles of Geology* neat. So do many origin of language theorists. Armed with Darwin and Lyell, they adopt a strong continuity assumption: like the eye and every other trait, language too *must* have evolved by "numerous, successive, slight modifications" (Darwin 1959, 189). But is this strictly so? Take "successive." On one reading, all "successive" means is that evolutionary events must follow one after the other in time. That's always true, so we can safely set aside this constraint.

That leaves "numerous" and "slight." Immediately after the publication of *Origin* "Darwin's bulldog" Huxley was openly critical of both, writing to Darwin on November 23 1859, "You have loaded yourself with an unnecessary difficulty in adopting 'Natura non facit saltum' so unreservedly" (Huxley 1859). Darwin himself could only push his gradual eye evolution story so far in *Origin*, certain only that natural selection would begin to act after a photoreceptor and a pigment cell had evolved to form a partially functional light-detecting prototype eye. He had no account of the actual origin of the pigment cell-photoreceptor pair, and nor should we have expected one.

Here modern molecular biology provides new insights. Darwin's prototype eye consisted of two parts: a light-sensitive cell (a "nerve") and a pigment cell to shadow the photoreceptor cell: "In the Articulata we can commence a series with an optic nerve merely coated with pigment" (Darwin 1859, 187). But Darwin could not find a way to reason further back in time before this point. In the end, Darwin resorted to the same option here that he set out for the origin of life itself—he relegated it to the realm of chance effects, beyond the explanatory purview of his theory: "How a nerve comes to be sensitive to light, hardly concerns us more than how life itself first originated; but I may remark that several facts make me suspect that any sensitive nerve may be rendered sensitive to light" (Darwin 1859, 187).

On reflection the same Darwinian dilemma arises with every true novelty. In the case of the eye's origin Gehring (2011) has provided a more subtle analysis. The eye is the product both of chance and necessity, just as Monod had anticipated (1970). Two components are required for the prototype eye, the photoreceptor cell and the pigment cell. The initial formation of the photoreceptor was a chance event; it did not occur by some laborious trial-and-error incremental search via selection: the capture of light-sensitive pigment molecules by cells, subsequently regulated by the *Pac-6* gene. What an observer would see from the outside is a very long period of geological time where life did not have photoreceptive cell pigment, and then the relatively rapid appearance of cells-plus-pigment—the pigment was either captured or it was not. All this occurred without the need for "numerous" and "slight modifications." To be sure, the molecule had to pass selection's sieve and has been fine-tuned since—but after the critical event. Similarly, the prototypical pigment cell arose

from the ubiquitous pigment melanin found in a single cell along with the now-captured photoreceptive pigment. At some point, this single cell then split into two, again a stochastic event, apparently under the control of a cell-differentiating regulatory gene. Here too, if viewed "from the outside" one would see a relatively long period of stasis, followed by the all-or-nothing split into two cells—the daughters were either produced or not. "We conclude from these considerations that the Darwinian eye prototype arose from a single cell by cellular differentiation, *Pax6* controlling the photoreceptor cell and *Mitf* the pigment cell" (Gehring 2011, 1058).

In short, the initial origin of Darwin's two-cell prototype eye does not seem to have followed the classical trial-and-error selectionist formula. Rather, there were two distinct, stochastic, and abrupt events responsible for this key innovation, the eye's "camera film." And since? While there have been many improvements and striking innovations to the eye's camera body, lens, and such and in just the way Darwin wrote, there has been far less tinkering with the film. It is not as if evolution ditched Kodak, then switched to Polaroid, and finally homed in on digital recording. The initial two key innovations were neither numerous nor slight.[8] On a timeline they stick out like two sore thumbs, two abrupt, large, and rapid changes in between nothing much happening at all—a pattern of stasis and innovation just like that in our own lineage, as we discuss just below.[9]

Nonetheless, a "Darwinian fundamentalist" might still insist on an ancestral chain requiring smooth, incremental continuity at all steps, and so a strong likelihood of finding contemporary species that share one or another of the traits that make up human language. In this framework, even the recent discovery that chimpanzees can cook food (Warneken

and Rosati 2015) literally adds fuel to the fire that our closest living relatives are also close to us in terms of language. However, as we saw earlier in this chapter in regard to the claims of Bornkessel-Schlesewsky et al. and Frank et al., and we'll see again in chapter 4, in fact chimpanzees are quite unlike us linguistically.

One might call this fundamentalist, uniformitarian picture the "micromutational view." The alternative often entertained in this conventional picture—largely as a caricatured strawman—has most often been its polar opposite, the so-called (and infamous) "hopeful monster" hypothesis proposed by Goldschmidt (1940). Goldschmidt posited giant-step genomic and morphological changes—perhaps even the appearance of a new species—after just one generation. Since "hopeful monsters" really do seem out of the question, many dismiss the possibility of any other sort of change *but* micromutation.

However, this is a false dichotomy. As we have already seen, there's good reason to believe that it's simply empirically false. Many evolutionary innovations—such as cell nuclei, linear DNA, and, we believe, echoing Lane (2015), language, fit uneasily on the micro vs. hopeful monster Procustean bed. From a theoretical point of view, the micromutational choice sits frozen in time at about the year 1930, near the culmination of the Modern Synthesis. In 1930, one of the three leaders of the Modern Synthesis, R. A. Fisher, published his *Genetical Theory of Natural Selection*, with a simple geometric mathematical model of adaptation, drawing a comparison to the focusing of a microscope (Fisher 1930, 40–41). The intuition is that if one is closing in on a pinpoint-focused image, then only very, very tiny changes will move us closer to a better focus. A large change in the focus wheel will in all likelihood move us far away from the desired spot. Intuitively plausible

and convincing, this single passage was enough to completely convince the next several generations of evolutionary biologists—that is, until recently.

Fisher used the results of his model to argue that all adaptive evolutionary change is micromutational—consisting of infinitesimally small changes whose phenotypic effects approach zero. As Orr (1998, 936) puts it, "This fact essentially guarantees that natural selection acts as the *sole source of creativity in evolution*. ... Because selection shapes adaptation from a supply of continuous, nearly fluid variation, mutation on its own provides little or no phenotypic form" (our emphasis).

In particular, Fisher's model suggests mutations with a vanishingly small phenotypic effect have a 50% chance of survival, while any larger mutations have an exponentially declining chance of survival. If we adopt Fisher's model, then by definition large-phenotypic-effect genes cannot play a role in adaptation. As Orr (1998, 936) notes:

> It would be hard to overestimate the historical significance of Fisher's model. His analysis single-handedly convinced most evolutionists that factors of large phenotypic effect play little or no role in adaptation (reviewed in Turner 1985; Orr and Coyne 1992). Indeed a review of the literature reveals that virtually every major figure from the modern synthesis cited the authority of Fisher's model as the sole support for micro-mutationism (see Orr and Coyne 1992; also see Dobzhansky 1937; Huxley 1963; Mayr 1963; Muller 1940; Wright 1948). J.B.S. Haldane appears to have been the sole exception.

And indeed, seemingly every work one turns to on the evolution of language embraces Fisher's position—and so along with it, the correspondingly completely dominant role for natural selection. Fitch's (2010, 47) remark is representative, following the "focus-the-microscope" metaphor: "The core argument against an adaptive role for major qualitative

changes is that the macromutations we observe in nature disrupt adaptive function rather than enhancing it. Organisms are fine-tuned systems, and individuals born with large random changes have a very small chance of ending up fitter to survive."

Tallerman (2014, 195), citing McMahon and McMahon (2012), indicates that both she and the two cited authors also adopt Fisher's gradualism: "McMahon and McMahon (2012, one linguist and one geneticist) note that 'biological evolution is typically slow and cumulative, not radical and sudden,' and, with regard to 'a macromutation causing an immediate and radical change' state that 'the latter is evolutionarily highly unlikely.'"

But Fisher was wrong. Experimental work in the 1980s on the genetics of adaptation demonstrated that individual genes could have surprisingly large effects on phenotypes. It is again worth quoting Orr in full:

> In the 1980s ... approaches were developed that finally allowed the collection of rigorous data on the genetics of adaptation—Quantitative trait locus (QTL) analysis. ... In QTL analysis, the genetic basis of phenotypic differences between populations or species can be analysed using a large suite of mapped molecular markers. In microbial evolution work, microbes are introduced into a new environment and their adaptation to this environment is allowed; genetic and molecular tools then allow the identification of some or all of the genetic changes that underlie this adaptation. The results of both approaches were surprising: evolution often involved genetic changes of relatively large effect and, at least in some cases, the total number of changes seemed to be modest ... [the results included] several classical studies, including those that analyse the evolution of reduced body armour or pelvic structure in lake stickleback, the loss of larval trichomes (fine "hairs") in *Drosophila* species, and the evolution of new morphologies in maize and the monkeyflower *Mimulus* species. Microbial studies further revealed that genetic changes occurring early in adaptation often have larger fitness effects than those that occur later, and that parallel adaptive evolution is surprisingly common (Orr 2005a, 120).

In fact, before Orr, Kimura (1983) noted a fundamental flaw in Fisher's model that follows from the stochastic nature of real biological evolution we discussed earlier: Fisher did not correctly take into account the likelihood of the stochastic loss of beneficial mutations. Kimura noted that changes with larger phenotypic effects are less likely to be lost. In Kimura's model, mutations of intermediate size ought to be more likely in adaptation. However, this model too has required some modification to capture the *series* of steps in any "adaptive walk" rather than any single step (Orr 1998). As Orr (2005a, 122) states, "Adaptation in Fisher's model therefore involves a few mutations of relatively large phenotypic effect and many of relatively small effect. ... adaptation is therefore characterized by a pattern of diminishing returns—larger-effect mutations are typically substituted early on and smaller-effect ones later." One can picture this evolutionary change as a bouncing ball, where the largest bounce comes first followed by successively smaller and smaller bounces—a sequence of diminishing returns. This finding has clear implications for any evolution-of-language scenario that insists on micromutational change at the first step. In short, rather than macromutational change being uncommon and unexpected, the reverse might hold at the first step, and sometimes does. Contemporary evolutionary theory, lab experiments, and field work all support this position—without the need to posit Goldschmidtian "hopeful monsters." There is in fact a secure middle ground. To be sure, what has actually happened in any particular situation remains an empirical question; as always, biology is more like case law, not Newtonian physics. The clues we have that we discuss just below and later on in chapter 4 point in the direction of relatively rapid change, sometime between the period when anatomically modern humans first appeared in Africa about

200,000 years ago, and their subsequent exodus out of Africa 60,000 years ago.

What is the lesson to learn from this modern take on Darwinism and evolutionary change? Essentially, you get what you pay for, and if you pay for it, you should understand what you have bought—the whole package with all its consequences. If you opt for Fisher's model, then you necessarily embrace micromutationism, and you have already ruled out by fiat everything except natural selection as the causal driver for the evolution of language. As we have seen, you also lose the ability to explain the origin of complex cells from simple-celled prokaryotes, the origin of eyes, and much else. On the other hand, if you don't buy Fisher's model, and move on to the more modern view, then you leave the door open for a richer set of possibilities.

Returning now to the human story, an examination of the paleoarcheological record for our lineage *Homo* supports the nongradualist picture, not the gradualist one: a recurring pattern of "disconnects between times of appearance (and disappearance) of new technologies and new species" (Tattersall 2008, 108). The basic point is easy to see. According to Tattersall, whenever a new, morphologically distinct, *Homo* species has appeared, there has been no simultaneous technological or cultural innovation. Rather, the technology/cultural innovations appear *long after* the appearance of each new *Homo* species—with that time measured in hundreds of thousands of years. In other words, as Tattersall (2008, 103) writes, "Technological innovations are *not* associated with the emergence of new kinds of hominid." For example, Mode 1 or Oldowan tools are first found about 2.5 million years before the present (BP). Quite recently, even older tools, dated at 3.3 million years (BP), have been found at Lomekwi in Kenya

(Harmand et al. 2015). These archaic tool types were then maintained for perhaps a million years until the innovation of the Mode 2 Acheulean hand axes. However, as Tattersall (2008, 104) notes, this technological innovation "significantly postdated the arrival on Earth of a new kind of hominid, often known nowadays as *Homo ergaster*." In a recent review, Svante Pääbo, the leading scientist behind the recovery of ancient DNA and the sequencing of the Neandertal and Denisovan genomes, echoes this sentiment: "Only some 2.6 million years ago did human ancestors start making stone tools that can be recognized as such when found by archeologists. But even then, the different tools produced did not change much for hundreds of thousands of years" (Pääbo 2014, 216).

Similarly, though brain size increased throughout the *Homo* lineage, with *Neandertal* cranial capacity becoming on average larger than modern humans, the behavioral and material record lags behind. It is not until the appearance of the first modern humans in Africa that we see the beginning of the rapid changes in both tools and the appearance of the first unambiguously symbolic artifacts, such as shell ornaments, pigment use, and particularly the geometric engravings found in Blombos Cave approximately 80,000 years ago (Henshilwood et al. 2002). Here too Pääbo agrees: he says that something must have set us apart from the Neandertals, to prompt the relentless spread of our species who had never crossed open water up and out of Africa and then on across the entire planet in just a few tens of thousands of years. What was it?

Along with Tattersall, Pääbo singles out the lack of figurative art and other trappings of modern symbolic behavior in Neandertals. That provides a strong clue (Pääbo 2014b). Evidently our ancestors moving out of Africa already had "it,"

and the "it," we suspect along with Tattersall, was language. Here Pääbo demurs. He suggests that what sets us apart is "our propensity for shared attention and the ability to learn complex things from others"—here taking language as one aspect of cultural learning, following the views of his colleague Michael Tomasello (Pääbo 2014b, 3757–3758). We feel that he is mistaken about language and how it is acquired. Pääbo seems to have returned to the "Boasian" anthropological view of the last century, as we describe in the next chapter.

In any case, the upshot of our ancestors' exodus out of Africa was that a particular *Homo* species—us—would eventually come to dominate the world, absorb whatever was good in the Neandertal and Denisova genomes, and leave the rest—perhaps a fanciful picture, but an all too familiar and unsettling one nonetheless, given what we know about the subsequent history of our species.

What we do not see is any kind of "gradualism" in new tool technologies or innovations like fire, shelters, or figurative art. While controlled use of fire appears approximately one million years ago, this is a full half-million years after the emergence of *Homo ergaster.* Tattersall points out that this typical pattern of stasis followed by innovative jumps is consistent with the notion of "exaptation"—that is, evolution by natural selection always co-opts existing traits for new uses; there cannot be any "foreknowledge" that a particular trait would be useful in the future. Innovations therefore arise independently of the functions that they will be eventually selected for. Acting like a sieve, natural selection can only differentially sift through what is presented to it. Any innovation must necessarily been created in some other way, as gold nuggets that pan out. The antecedent ingredients for language must in a sense already exist. But what were those ingredients?

## The Tripartite Model, Vocal Learning, and Genomics

Any account of the origin of language must come to grips with *what* has evolved. In our tripartite framework, that works out naturally as each of the three components we sketched earlier: (1) the combinatorial operator Merge along with word-like atomic elements, roughly the "CPU" of human language syntax; and the two interfaces, (2) the sensorimotor interface that is part of language's system for externalization, including vocal learning and production; and (3) the conceptual-intentional interface, for thought. Here we focus on (2), vocal learning and production, as mediated by the sensorimotor interface.

As mentioned at the beginning of the chapter, thanks to animal models like songbirds, researchers appear to be closing in on an understanding of vocal learning—apparently a genetically modular "input-output" sequential processing component. As Pfenning et al. (2014) suggest, this component might well be relatively uniform from one vocal learning species to the next because there may be only a few possible ways to build a vocal-learning system given evolutionary and biophysical constraints. This does not rule out the possibility of species-specific tuning, as in the case of human audition and speech, or gesture and visual perception.

This "input-output" picture matches up with the *FOXP2* story. Our view is that *FOXP2* is primarily a part of the system that builds component (2), the sensorimotor interface, involved in the externalization of narrow syntax—like the printer attached to a computer, rather than the computer's CPU. Chapter 3 discusses empirical linguistic evidence for this position. But there's other evidence as well. Recent work with transgenic mice raised with humanized Foxp2 suggests that

the human variant plays a role in "modifying cortico-basal ganglia circuits," boosting the ability to shift motor skills acquired declaratively to procedural memory, like learning to ride a bicycle (Schreiweis et al. 2014, 14253). This finding is quite compatible with the externalization view. This shift from declarative to (unconscious) motor skills seems to be exactly what human infants do when they learn how to perform the exquisite ballet dance of mouth, tongue, lips, vocal tract, or fingers that we call speech or gesture. Of course much remains unknown as the authors note, since "how these findings relate to the effect of [the] humanized version of Foxp2 in shaping the development of a human brain to enable traits such as language and speech acquisition is unknown" (Schreiweis et al. 2014, 14257).

To us at least, the Schreiweis experiment along with the findings of Pfenning and colleagues (Pfenning et al. 2014) strikingly confirms that the vocal learning and production aspect of language's externalization system is not human-specific. Roughly 600 million years of evolutionary time separate us from birds; nonetheless the specialized song and speech regions and genomic specialization of the vocal-learning songbird species (e.g., zebra finch, hummingbird) and those of the vocal-learning human species appear to be dramatically, and convergently, similar. In contrast, nonvocal avian learners (chickens, quail, doves) and nonvocal nonhuman primates (macaques) do not share these genomic specializations with vocal learners (either songbirds or humans).

Pfenning et al. sifted through thousands of genes and gene expression profiles in the brains of songbirds, parrots, hummingbirds, doves, quail, macaques, and humans, attempting to correlate distinctive gene expression levels (whether transcribed at a high level or a low level) against a sophisticated

hierarchical decomposition of known brain regions across the tested species. The aim was to discover whether subregions where certain genes were expressed more highly or not matched up to each other from one species to another in the case of vocal learners (songbirds, parrots, hummingbirds, humans) as opposed to nonvocal learners (doves, quail, macaques). The answer was yes: the same genomic transcriptional profiles could be aligned across all vocal learners, but not in vocal learners versus nonvocal learners. If we imagine the genes as some set of sound-tone controls in an amplifier, then they were all "tuned" in a parallel way in vocal-learning species—and the tuning was different as compared to nonvocal-learning species.

For example, both songbirds and humans have comparable down-regulation of the axon guidance gene *SLIT1* (a DNA target of FOXP2) in analogous brain regions, the so-called avian RA region ("robust nucleus of the arcopallium") and the human layrngeal motor cortex. As Pfenning et al. note, the protein product of *SLIT1* "works in conjunction with the *ROBO1* axon guidance receptor, and *ROBO1* mutations cause dyslexia and speech disorders in humans. ... *ROBO1* is one of five candidate genes with convergent amino acid substitutions in vocal-learning mammals" (2014, 2156846–10). The *SLIT1* gene evidently is part of a developmental network ensuring that songbird and human brains are properly "wired up."

Like *FOXP2*, many of the genes discovered by this approach up- or down-regulate DNA and its corresponding protein products. But we do not yet know how they are all causally woven together. Pfenning (personal communication) has planned out the next steps for tracking at least part of this down. It involves finding the DNA motifs that "regulate the

regulators." This is precisely the right approach, and it bears on what we have reviewed about evolution and evolutionary change. We have known since the pioneering analysis by King and Wilson (1975) that humans and chimpanzees are 99% identical at the macromolecular level—proteins involved in the working biochemistry of organisms—and that this identity is probably even stronger if we compared humans to our nonhuman ancestors. King and Wilson drew the obvious and important conclusion: the differences between humans and chimps must largely lie in regulatory elements. What this means is that changes in protein-coding genes might not be where the evolutionary action lies—perhaps especially in the evolution that made us human, since that has been a relatively recent event.

Over the past forty years, King and Wilson's important insight has been confirmed in spades, including both noncoding DNA as well as all the other components that regulate gene activity, from the chromatin scaffolding surrounding DNA, to the micro-RNA regulation of DNA during development, in particular brain development—part of the so-called evo-devo revolution (Somel, Liu, and Khaitovich 2011).

Here we will focus on just one factor in the gene regulatory system that controls DNA, so-called *enhancers*, and on why this kind of regulatory evolution has turned out to be so relevant. (We won't have space here to consider other genomic regions that appear to be relevant for evolutionary changes, for example, so-called *cis*-regulatory elements; see Wray, 2007.) An enhancer is a short stretch of DNA, about 1,500–2,000 DNA nucleotides (Adenine, Thymine, Cytosine, Guanine) long, that does not code for a functional protein like the *HBB* gene for the hemoglobin beta-globin protein chain, or the *FOXP2* gene for the FOXP2 protein. An enhancer does not code for any protein at all—so it's called *noncoding* DNA.

Its function? An enhancer lies some distance "upstream" or "downstream" from the start point of a protein-coding gene, perhaps up to a million DNA nucleotides away, and then "twists over" to contact that start point along with the other ingredients required to ignite DNA transcription—a promoter, RNA polymerase II, and any transcription factors (perhaps even FOXP2 itself). Once all components are in place, then (a bit fancifully) the promoter sparkplug fires and the DNA transcription engine starts running.

From the standpoint of evolution, enhancers are interesting for at least two reasons. First, they are more narrowly targeted than protein-coding DNA. Unlike protein-coding DNA that may (in fact usually does) play more than one role in an organism, employed in many different tissues and cells, an enhancer affects just one piece of DNA, and so is tuned to a single very particular context, in conjunction with promoters and transcription factors. Consequently, it is easier to mutate an enhancer without causing untoward nonlocal effects. An enhancer is *modular.* That's perfect for evolutionary experimentation—not so many worries about breaking a complicated machine by jamming a wrench in it. Second, an enhancer sits on just one of DNA's two strands (usually the same strand as the protein-coding DNA itself). This is unlike a protein-coding DNA gene, which might need to be on both DNA strands—in a so-called homozygous state, in order to surface as a phenotype—like the classic case of blue eyes. And this is a second evolutionary advantage: an organism doesn't have to wait for a change on both DNA strands. The bottom line is that evolutionary tinkering is in principle much easier with enhancers—there are over 100,000 of these in humans, all singling out specific gene contexts. It should come as no surprise that this is the first place that the avian researchers

will probe next to further our understanding of avian and human vocal learning. This line of thinking has recently been confirmed by the first functional confirmation of a human-chimpanzee DNA difference that promotes neuron cell division, as we describe below (Boyd et al. 2015).

Returning to the general picture, what are the evolutionary implications of these results for vocal learning? Pfenning et al. (2014, 1333) close their summary with this: "The finding that convergent neural circuits for vocal learning are accompanied by convergent molecular changes of multiple genes in species separated by millions of years from a common ancestor indicates that brain circuits for complex traits may have limited ways in which they could have evolved from that ancestor." In other words, the "toolkit" for building vocal learning might consist of a (conserved) package of perhaps 100–200 or so gene specializations no matter what the species that can be "booted up" quickly—and so evolved relatively rapidly. This fits into our general picture for the relatively rapid emergence of language, as well as with our methodology for distinguishing between the evolution of the input–output externalization system from the "central processor" of human language syntax.

What else can modern molecular biology tell us about the evolution of the human brain and language? We cannot do justice here to this rapidly expanding field, but instead single out a few key points along with well-known major roadblocks.

First, thanks to the recent work with ancient DNA, one can now figure out how many and what sort of genomic differences one might expect to find, and then see how this lines up with the known genomic differences between us and the sequenced Neandertal, Denisovan, and chimpanzee genomes.

As to the expected differences, the time since the split with our extinct *Homo* ancestors such as Neandertal is relatively recent—500,000 to 700,000 years ago—and modern humans appear in southern Africa about 200,000 years ago, so there are about 200,000 years of evolutionary time between these two events. We can use theoretical population genetics tools, including estimates of selective strength, population size, and DNA mutation rates, to calculate how many distinct, positively selected genomic regions one might expect to find that have been fixed in the human population—that is, with no variation in modern humans, so presumed to be functionally important—but that are different in nonhuman species. The so-called effective population size for humans 200.000 years ago has been estimated by several sources to be about 10,000—relatively small compared to many other mammals (Jobling et al. 2014). Selective strength—fitness, denoted *s*—is challenging to estimate in any situation, but one can use data from one of the strongest recent signals of selection in the population, that for the lactase persistence gene *LCT* (Tishkoff et al. 2007) to give an upper bound of 0.10. This is extremely high. Given all these parameters, one recent analysis estimates that there could have been 700 beneficial mutations, with only 14 of these surviving to fix in the human population, even given a strong selective advantage of $s = 0.01$ (Somel, Liu, and Khaitovich 2013). The low survival number is due to the "stochastic gravity well" effect described in the previous section, with the probability of loss approximately $(1–s/2)$, so 98% of 700 or 686 lost and 14 fixed.

This theoretical estimate turns out to be quite close to what has been found empirically. Whole-genome sequencing of Neandertals and Denisovans indicates that that there are 87 and 260 functional (amino-acid changing) genomic differences

respectively that are fixed in modern humans but not present in these two extinct species (Pääbo 2014a, supplementary table 1). As Pääbo writes, such differences are of special significance because at least from the genomic standpoint they highlight what makes us human. Focusing on the Neandertal-human differences, there are just 31,389 single DNA nucleotide differences (single-nucleotide polymorphisms, or SNPS) out of the approximately four billion possible; 125 DNA nucleotide insertions or deletions; 3117 regulatory region differences (using a particular definition of "regulatory"); and a mere 96 total amino acid differences, within 87 genes. (Some genes have more than a single amino acid difference.) What does this "difference list" tell us?

Many—most—of the 30,000-odd SNP differences presumably make no difference at all in natural selection's sieve—they are "neutral." Following Pääbo, let also put aside for a moment the 3,000 or so regulatory differences. We're left with just the 87 protein-coding differences between us and Neandertals—not many. For example, we apparently share the same FOXP2 protein with Neandertals, though there is some evidence of a regulatory region for *FOXP2* that is not fixed in the human population and whose variants are a bit different from Neandertal, as we discuss further in chapter 4.[10] Of the genes that do code for different proteins, some are almost surely unrelated to language and cognition. For example, at least three of the different genes are involved in the formation of skin, and this makes sense given the human loss of body hair and resulting changes in skin pigmentation.

Other genomic differences would seem more likely candidates for cognitive evolution. For example, Pääbo notes that there are three gene variants that we have but Neandertals don't—*CASC5*, *SPAG5*, and *KIF18A*. These are involved in

neural cell division in the so-called "proliferative zone" where stem cells divide to build the brain (Pääbo 2014a). However, at the time of this writing we don't know whether the proteins these genes code for actually lead to different developmental outcomes or phenotypes in us as opposed to Neandertals—bigger or different brains, or, more precisely, bigger brains in the right spots, since Neandertal cranial capacity was on average larger than ours, though perhaps more skewed to the rear, occipital part of the brain. And that's the main roadblock that has to be overcome: figuring out the road from genotype to phenotype.

We do know the answer to the functional question in the case of at least one regulatory genomic difference implicated in brain development—a difference between us and the other great apes, though, not with Neandertals (Boyd et al. 2015). There is a general increase in cranial capacity and brain size throughout the *Homo* lineage, from *Homo habilis* at about 2–2.8 million years ago, with a newly re-estimated cranial capacity of 727–846 $cm^3$, to *Homo erectus,* at about 850–1100 $cm^3$, and expanding from there. The *Homo* lineage differs here from the other great apes. What has driven brain expansion? If we look at enhancer regions in humans undergoing accelerated evolution, it turns out that many are located close to genes involved in building our brains (Prabhakar et al. 2006; Lindblad-Toh et al. 2011). Boyd and colleagues zeroed in on one of these enhancers that differ between us and chimpanzees, *HARE5*, and constructed transgenic mice with either the human or chimpanzee form of *HARE5*. Do the different mice exhibit different patterns of cortical growth? They do: the humanized mice had increased brain size about 12% compared to normal mice or those with the chimp-mouse *HARE5* form, apparently due to a boost in cell division rate

for neural progenitor cells. Just as described above, the *HARE5* enhancer works in tandem with the promoter region of a key gene involved in the pathway for neocortical development, *FZD8*. This research points to one path—albeit laborious—towards experimental confirmation of the phenotypic effects for all 87 genes in the Neandertal-human difference list. But we will need to know more. Even if we know that *HARE5* boosts brain growth, we will still need to know how this brain growth ties into the cognitive phenotype we call language.

What of the 3,000-odd regulatory differences? Somel and colleagues observe, "there is accumulating evidence that human brain development was fundamentally reshaped through several genetic events within the short time space between the human–Neandertal split and the emergence of modern humans" (Somel, Liu, and Khaitovich 2013, 119). They single out one particular difference between Neandertals and us: a stretch of regulatory DNA appearing upstream of a regulator of synaptic growth, *MEF2A* (myocyte enhancer factor 2). This they call a "potential transcriptional regulator of extended synaptic development in the human cerebral cortex"—one signal characteristic of human development, an extended period of childhood (Somel, Liu, and Khaitovich 2013, 119). That seems like a heavy explanatory burden for one small stretch of DNA to bear however.

Other novel genes and regulatory elements implicated in skull morphology and neural growth have accumulated en route from our last common ancestor with chimpanzees to the present day, again common to the *Homo* lineage. For example, the gene *SRGAP2* is known to play a role in human cortical development and neuron maturation. It has been duplicated three times on the lineage leading to us, with one duplication occurring just at about the time when the lineage *Homo*

appears, 2–3.5 million years ago (Jobling et al. 2014, 274). Such gene duplications are known to play important roles in evolutionary innovation since they allow one of the duplicates to "float free" and take on new functions (Ohno 1970). See note 9.

What's the bottom line? Perhaps the $64,000 question is whether Neandertals had language. The number of genomic differences between us and Neandertals and Denisovans is small enough that some authors answer yes. We remain skeptical. We don't understand the genomic or neural basis for the Basic Property. It is virtually impossible to say even whether anatomically modern humans 80,000 years ago had language. All we have to go by are the symbolic proxies for language behavior. Along with Tattersall (2010) we note that the material evidence for Neandertal symbolic behavior is exceptionally thin. In contrast, the anatomically modern humans in southern Africa around 80,000 years ago show clear signs of symbolic behavior—before their exodus to Europe. Chapter 4 revisits this question.

Our general problem is that we understand very little about how even the most basic computational operations might be carried out in neural "wetware." For example, as Randy Gallistel has repeatedly emphasized, the very first thing that any computer scientist would want to know about a computer is how it writes to memory and reads from memory—the essential operations of the Turing machine model and ultimately any computational device. Yet we do not really know how this most foundational element of computation is implemented in the brain (Gallistel and King 2009). For example, one of the common proposals for implementing hierarchical structure processing in language is as a kind of recurrent neural network with an exponential decay to emulate a "pushdown stack"

(Pulvermüller 2002). Unfortunately, simple bioenergetic calculations show that this is unlikely to be correct. As Gallistel observes, each action potential or "spike" requires the hydrolosis of $7 \times 10^8$ ATP molecules (the basic molecular "battery" storage for living cells). Assuming one operation per spike, Gallistel estimates that it would take on the order of $10^{14}$ spikes per second to achieve the required data processing power. Now, we do spend lots of time thinking and reading books like this to make our blood boil, but probably not that much. Similar issues plague any method based on neural spike trains, including dynamical state approaches, difficulties that seem to have been often ignored; (see Gallistel and King 2009 for details). Following the fashion of pinning names to key problems in the cognitive science of language, such as "Plato's problem," and "Darwin's problem," we call this "Gallistel's problem." Chapter 4 has more to say about Gallistel's problem in the context of computation and Merge.

Nearly fifty years ago, Marvin Minsky, in his 1967 book *Computation: Finite and Infinite Machines*, posed Gallistel's problem in virtually the same words, highlighting how little things have changed: "Unfortunately, there is still very little definite knowledge about, and not even any generally accepted theory of, how information is stored in nervous systems, i.e., how they *learn*. ... One form of theory would propose that short-term memory is 'dynamic'—stored in the form of pulses reverberating around closed chains of neurons. ... Recently, there have been a number of publications proposing that memory is stored, like genetic information, in the form of nucleic-acid chains, but I have not seen any of these theories worked out to include plausible read-in and read-out mechanisms" (Minsky 1967, 66). As far as we have been able to make out, Minsky's words still ring true, and Gallistel's

problem remains unsolved. Eörs Szathmáry is correct when he writes that "linguistics is at the stage at which genetics found itself immediately after Mendel. There are rules (of sentence production), but we do not yet know what mechanisms (neural networks) are responsible" (1996, 764).

Much as we would like to know what makes us human, and how language arose genetically, it is unsettling that scientists have yet to find any *unambiguous* evidence of natural selection's handiwork, a positive "selective sweep," occurring around the time *Homo sapiens* first emerged as a species. This may be an inevitable fact about our imperfect knowledge of our past demographic history as well as the relatively rarity of selective sweeps; evolution might simply be making use of variation already present in the population, as Coop and Przeworski (Jobling 2014, 204) argue.[11] In any case, as they go on to say, the genetic analysis of traits like language are "now a central challenge for human evolutionary genetics" (Jobling 2014, 204). We can only agree.