It is the judicious excess over minimum requisite support. It is why a good bridge does not crumble when subjected to stress beyond what reasonably could have been foreseen. It is fallback and failsafe. It is why we address our mail to city and state in so many words, despite the zip code. One indistinct digit in the zip code would spoil everything. . . . A kingdom, legend tells us, was lost for want of a horseshoe nail. Redundancy is our safeguard against such instability.

Thanks to the redundancy of language, yxx cxx xndxrstxd whx xt nxm xrttxng xvsn xf x rxplcx xll thx xvwxlsw xwh th xnx “x” (t gts ltrl hrdr f y dnt vn kn whr th vwl s r). In the comprehension of speech, the redundancy conferred by phonological rules can compensate for some of the ambiguity in the sound wave. For example, a listener can know that “thisrip” must be *this rip* and not the *rip* because the English consonant cluster *sr* is illegal.

So why is it that a nation that can put a man on the moon cannot build a computer that can take dictation? According to what I have explained so far, each phoneme should have a telltale acoustic signature: a set of resonances for vowels, a noise band for fricatives, a silence-burst-transition sequence for stops. The sequences of phonemes are massaged in predictable ways by ordered phonological rules, whose effects could presumably be undone by applying them in reverse.

The reason that speech recognition is so hard is that there’s many a slip twixt brain and lip. No two people’s voices are alike, either in the shape of the vocal tract that sculpts the sounds, or in the person’s precise habits of articulation. Phonemes also sound very different depending on how much they are stressed and how quickly they are spoken; in rapid speech, many are swallowed outright.

But the main reason an electric stenographer is not just around the corner has to do with a general phenomenon in muscle control called coarticulation. Put a saucer in front of you and a coffee cup a foot or so away from it on one side. Now quickly touch the saucer and pick up the cup. You probably touched the saucer at the edge nearest the cup, not dead center. Your fingers probably assumed the handle-grasping posture while your hand was making its way to the cup, well before it arrived. This graceful smoothing and overlapping of gestures is ubiquitous in motor control. It reduces the forces necessary to move body parts around and lessens the wear and tear on the joints. The tongue and throat are no different. When we want to articulate a phoneme, our tongue cannot assume the target posture instantaneously; it is a heavy slab of meat that takes time to heft into place. So while we are moving it, our brains are anticipating the next posture in planning the trajectory, just like the cup-and-saucer maneuver. Among the range of positions in the mouth that can define a phoneme, we place the tongue in the one that offers the shortest path to the target for the next phoneme. If the current phoneme does not specify where a speech organ should be, we anticipate where the next phoneme wants it to be and put it there in advance. Most of us are completely unaware of these adjustments until they are called to our attention. Say Cape Cod. Until now you probably never noticed that your tongue body is in different positions for the two sounds. In *horseshoe*, the first *s* becomes a *sh*; in NPR, the *n* becomes an *m*; in *month* and *width*, the *n* and *d* are articulated at the teeth, not the usual gum ridge.

Because sound waves are minutely sensitive to the shapes of the cavities they pass through, this coarticulation wreaks havoc with the speech sound. Each phoneme’s sound signature is colored by the phonemes that come before and after, sometimes to the point of having nothing in common with its sound signature in the company of a different set of phonemes. That is why you cannot cut up a tape of the sound *cat* and hope to find a beginning piece that contains the *k* alone. As you make earlier and earlier cuts, the piece may go from sounding like *ka* to sounding like a chirp or whistle. This shuffling of phonemes in the speech stream could, in principle, be a boon to an optimally designed speech recognizer. Consonant and vowels are being signaled simultaneously, greatly increasing the rate of phonemes per second, as I noted at the beginning of this chapter, and there are
many redundant sound cues to a given phoneme. But this advantage can be enjoyed only by a high-tech speech recognizer, one that has some kind of knowledge of how vocal tracts blend sounds.

The human brain, of course, is a high-tech speech recognizer, but no one knows how it succeeds. For this reason psychologists who study speech perception and engineers who build speech recognition machines keep a close eye on each other's work. Speech recognition may be so hard that there are only a few ways it could be solved in principle. If so, the way the brain does it may offer hints as to the best way to build a machine to do it, and how a successful machine does it may suggest hypotheses about how the brain does it.

Early in the history of speech research, it became clear that human listeners might somehow take advantage of their expectations of the kinds of things a speaker is likely to say. This could narrow down the alternatives left open by the acoustic analysis of the speech signal. We have already noted that the rules of phonology provide one sort of redundancy that can be exploited, but people might go even farther. The psychologist George Miller played tapes of sentences in background noise and asked people to repeat back exactly what they heard. Some of the sentences followed the rules of English syntax and made sense.

Furry wildcats fight furious battles.
Respectable jewelers give accurate appraisals.
Lighted cigarettes create smoky fumes.
Gallant gentlemen save distressed damsels.
Soapy detergents dissolve greasy stains.

Others were created by scrambling the words within phrases to create colorless-green-ideas sentences, grammatical but nonsensical:

Furry jewelers create distressed stains.
Respectable cigarettes save greasy battles.
Lighted gentlemen dissolve furious appraisals.
Gallant detergents fight accurate fumes.
Soapy wildcats give smoky damselflies.

A third kind was created by scrambling the phrase structure but keeping related words together, as in

Furry fight furious wildcat battles.
Jewelers respectable appraisals accurate give.

Finally, some sentences were utter word salad, like

Furry create distressed jewelers stains.
Cigarettes respectable battles greasy save.

People did best with the grammatical sensible sentences, worse with the grammatical nonsense and the ungrammatical sense, and worst of all with the ungrammatical nonsense. A few years later the psychologist Richard Warren taped sentences like The state governors met with their respective legislatures convening in the capital city, excised the first s from legislatures, and spliced in a cough. Listeners could not tell that any sound was missing.

If one thinks of the sound wave as sitting at the bottom of a hierarchy from sounds to phonemes to words to phrases to the meanings of sentences to general knowledge, these demonstrations seem to imply that human speech perception works from the top down rather than just from the bottom up. Maybe we are constantly guessing what a speaker will say next, using every scrap of conscious and unconscious knowledge at our disposal, from how coarticulation distorts sounds, to the rules of English phonology, to the rules of English syntax, to stereotypes about who tends to do what to whom in the world, to hunches about what our conversational partner has in mind at that very moment. If the expectations are accurate enough, the acoustic analysis can be fairly crude; what the sound wave lacks, the context can fill in. For example, if you are listening to a discussion about the destruction of ecological habitats, you might be on the lookout for words pertaining to threatened animals and plants, and then when you hear speech sounds whose phonemes you cannot pick out like “ceces,” you would perceive it correctly as species—unless you are Emily Litella, the hearing-impaired editorialist on *Saturday Night Live* who argued passionately against the campaign to protect endan-
The Language Instinct

The top-down theory of speech perception exerts a powerful emotional tug on some people. It confirms the relativist philosophy that we hear what we expect to hear, that our knowledge determines our perception, and ultimately that we are not in direct contact with any objective reality. In a sense, perception that is strongly driven from the top down would be a barely controlled hallucination, and that is the problem. A perceiver forced to rely on its expectations is at a severe disadvantage in a world that is unpredictable even under the best of circumstances. There is a reason to believe that human speech perception is, in fact, driven quite strongly by acoustics. If you have an indulgent friend, you can try the following experiment. Pick ten words at random out of a dictionary, phone up the friend, and say the words clearly. Chances are the friend will reproduce them perfectly, relying only on the information in the sound wave and knowledge of English vocabulary and phonology. The friend could not have been using any higher-level expectations about phrase structure, context, or story line because a list of words blurted out of the blue has none. Though we may call upon high-level conceptual knowledge in noisy or degraded circumstances (and even here it is not clear whether the knowledge alters perception or just allows us to guess intelligently after the fact), our brains seem designed to squeeze every last drop of phonetic information out of the sound wave itself. Our sixth sense may perceive speech as language, not as sound, but it is a sense, something that connects us to the world, and not just a form of suggestibility.

Another demonstration that speech perception is not the same thing as fleshing out expectations comes from an illusion that the columnist Jon Carroll has called the mondegreen, after his mis-hearing of the folk ballad “The Bonnie Earl O’Moray”:

```
Oh, ye hielands and ye lowlands,
Oh, where hae ye been?
They have slain the Earl of Moray,
And laid him on the green.
```

He had always thought that the lines were “They have slain the Earl of Moray, And Lady Mondegreen.” Mondegreens are fairly common (they are an extreme version of the Pullet Surprises and Pencil Vaneas mentioned earlier); here are some examples:

A girl with colitis goes by. [A girl with kaleidoscope eyes. From the Beatles song “Lucy in the Sky with Diamonds.”]
Our father wishart in heaven; Harold be thy name . . . Lead us not into Penn Station.
Our father which art in Heaven; hallowed by thy name . . . Lead us not into temptation. From the Lord’s Prayer.
He is trampling out the vintage where the grapes are wrapped and stored. [ . . . grapes of wrath are stored. From “The Battle Hymn of the Republic.”]
Gladly the cross-eyed bear. [Gladly the cross I’d bear.]
I’ll never be your pizza burnin’. [ . . . your beast or burden. From the Rolling Stones song.]
It’s a happy enchilada, and you think you’re gonna drown. (It’s a half an inch of water . . . From the John Prine song “That’s the Way the World Goes ’Round.”]

The interesting thing about mondegreens is that the mishearings are generally less plausible than the intended lyrics. In no way do they bear out any sane listener’s general expectations of what a speaker is likely to say or mean. (In one case a student stubbornly misheard the Shocking Blue hit song “I’m Your Venus” as “I’m Your Penis” and wondered how it was allowed on the radio.) The mondegreens do conform to English phonology, English syntax (sometimes), and English vocabulary (though not always, as in the word mondegreen itself). Apparently, listeners lock in to some set of words that fit the sound and that hang together more or less as English
words and phrases, but plausibility and general expectations are not running the show.

The history of artificial speech recognizers offers a similar moral. In the 1970s a team of artificial intelligence researchers at Carnegie-Mellon University, headed by Raj Reddy, designed a computer program called HEARSAY that interpreted spoken commands to move chess pieces. Influenced by the top-down theory of speech perception, they designed the program as a "community" of "expert" subprograms cooperating to give the most likely interpretation of the signal. There were subprograms that specialized in acoustic analysis, in phonology, in the dictionary, in syntax, in rules for the legal moves of chess, even in chess strategy as applied to the game in progress. According to one story, a general from the defense agency that was funding the research came up for a demonstration. As the scientists sweated he was seated in front of a chessboard and a microphone hooked up to the computer. The general cleared his throat. The program printed "Pawn to King 4."

The recent program DragonDictate, mentioned earlier in the chapter, places the burden more on good acoustic, phonological, and lexical analyses, and that seems to be responsible for its greater success. The program has a dictionary of words and their sequences of phonemes. To help anticipate the effects of phonological rules and coarticulation, the program is told what every English phoneme sounds like in the context of every possible preceding phoneme and every possible following phoneme. For each word, these phonemes-in-context are arranged into a little chain, with a probability attached to each transition from one sound unit to the next. This chain serves as a crude model of the speaker, and when a real speaker uses the system, the probabilities in the chain are adjusted to capture that person's manner of speaking. The entire word, too, has a probability attached to it, which depends on its frequency in the language and on the speaker's habits. In some versions of the program, the probability value for a word is adjusted depending on which word precedes it. This is the only top-down information that the program uses. All this knowledge allows the program to calculate which word is most likely to have come out of the mouth of the speaker given the input sound. Even then, DragonDictate relies more on expectancies than an able-eared human does. In the demonstration I saw, the program had to be coaxed into recognizing word and worm, even when they were pronounced as clear as a bell, because it kept playing the odds and guessing higher-frequency were instead.

Now that you know how individual speech units are produced, how they are represented in the mental dictionary, and how they are rearranged and smeared before they emerge from the mouth, you have reached the prize at the bottom of this chapter: why English spelling is not as deranged as it first appears.

The complaint about English spelling, of course, is that it pretends to capture the sounds of words but does not. There is a long tradition of doggerel making this point, of which this stanza is a typical example:

Beware of heard, a dreadful word
That looks like beard and sounds like bird,
And dead: it's said like bed, not bead—
For goodness' sake don't call it "deed"!
Watch out for meat and great and threat
(They rhyme with suite and straight and debt).

George Bernard Shaw led a vigorous campaign to reform the English alphabet, a system so illogical, he said, that it could spell fish as "ghoti"—gh as in tough, o as in woman, ti as in nation. ("Mnemonocode" for minute and "mnopspteich" for mistake are other examples.) In his will Shaw bequeathed a cash prize to be awarded to the designer of a replacement alphabet for English, in which each sound in the spoken language would be recognizable by a single symbol: He wrote:

To realize the annual difference in favour of a forty-two letter phonetic alphabet . . . you must multiply the number of minutes in the year, the number of people in the world who are
continuously writing English words, casting types, manufacturing printing and writing machines, by which time the total figure will have become so astronomical that you will realize that the cost of spelling even one sound has cost us centuries of unnecessary labour. A new British 42 letter alphabet would pay for itself a million times over not only in hours but in moments. When this is grasped, all the useless twaddle about enough and cough and laugh and simplified spelling will be dropped, and the economists and statisticians will be set to work to gather in the orthographic Golconda.

My defense of English spelling will be halffhearted. For although language is an instinct, written language is not. Writing was invented a small number of times in history, and alphabetic writing, where one character corresponds to one sound, seems to have been invented only once. Most societies have lacked written language, and those that have it inherited it or borrowed it from one of the inventors. Children must be taught to read and write in laborious lessons, and knowledge of spelling involves no daring leaps from the training examples like the leaps we saw in Simon, Mayela, and the Jabba and *mice-eater* experiments in Chapters 3 and 5. And people do not uniformly succeed. Illiteracy, the result of insufficient teaching, is the rule in much of the world, and dyslexia, a presumed congenital difficulty in learning to read even with sufficient teaching, is a severe problem even in industrial societies, found in five to ten percent of the population.

But though writing is an artificial contraption connecting vision and language, it must tap into the language system at well-demarcated points, and that gives it a modicum of logic. In all known writing systems, the symbols designate only three kinds of linguistic structure: the morpheme, the syllable, and the phoneme. Mesopotamian cuneiform, Egyptian hieroglyphs, Chinese logograms, and Japanese kanji encode morphemes. Cherokee, Ancient Cypriot, and Japanese kana are syllable-based. All modern phonemic alphabets appear to be descended from a system invented by the Canaanites around 1700 B.C. No writing system has symbols for actual sound units that can be identified on an oscilloscope or spectrogram, such as a phoneme as it is pronounced in a particular context or a syllable chopped in half.

Why has no writing system ever met Shaw's ideal of one symbol per sound? As Shaw himself said elsewhere, "There are two tragedies in life. One is not to get your heart's desire. The other is to get it." Just think back to the workings of phonology and coarticulation. A true Shavian alphabet would mandate different vowels in *write* and *ride*, different consonants in *write* and *writing*, and different spellings for the past-tense suffix in *slapped*, *sobbed*, and *sorted*. Cape Cod would lose its visual alliteration. A *horse* would be spelled differently from its *horseshoe*, and National Public Radio would have the enigmatic abbreviation MPR. We would need brand-new letters for the *n* in *month* and the *d* in *width*. I would spell *often* differently from *orphan*, but my neighbors here in the Hub would not, and their spelling of *career* would be my spelling of *Korea* and vice versa.

Obviously, alphabets do not and should not correspond to sounds; at best they correspond to the phonemes specified in the mental dictionary. The actual sounds are different in different contexts, so true phonetic spelling would only obscure their underlying identity. The surface sounds are predictable by phonological rules, though, so there is no need to clutter up the page with symbols for the actual sounds; the reader needs only the abstract blueprint for a word and can flesh out the sound if needed. Indeed, for about eighty-four percent of English words, spelling is completely predictable from regular rules. Moreover, since dialects separated by time and space often differ most in the phonological rules that convert mental dictionary entries into pronunciations, a spelling corresponding to the underlying entries, not the sounds, can be widely shared. The words with truly weird spellings (like *of*, *people*, *women*, *have*, *said*, *do*, *done*, and *give*) generally are the commonest ones in the language, so there is ample opportunity for everyone to memorize them.

Even the less predictable aspects of spelling bespeak hidden linguistic regularities. Consider the following pairs of words where the same letters get different pronunciations:
Once again the similar spellings, despite differences in pronunciation, are there for a reason: they are identifying two words as being based on the same root morpheme. This shows that English spelling is not completely phonemic; sometimes letters encode phonemes, but sometimes a sequence of letters is specific to a morpheme. And a morphemic writing system is more useful than you might think. The goal of reading, after all, is to understand the text, not to pronounce it. A morphemic spelling can help a reader distinguishing homophones, like meet and mete. It can also tip off a reader that one word contains another (and not just a phonologically identical impostor). For example, spelling tells us that overcome contains come, so we know that its past tense must be overcame, whereas succumb just contains the sound "kum," not the morpheme come, so its past tense is not succame but succumbed. Similarly, when something recedes, one has a recession, but when someone re-seeds a lawn, we have a re-seeding.

In some ways, a morphemic writing system has served the Chinese well, despite the inherent disadvantage that readers are at a loss when they face a new or rare word. Mutually unintelligible dialects can share texts (even if their speakers pronounce the words very differently), and many documents that are thousands of years old are readable by modern speakers. Mark Twain alluded to such inertia in our own Roman writing system when he wrote, "They spell it Vinci and pronounce it Vinchy; foreigners always spell better than they pronounce."