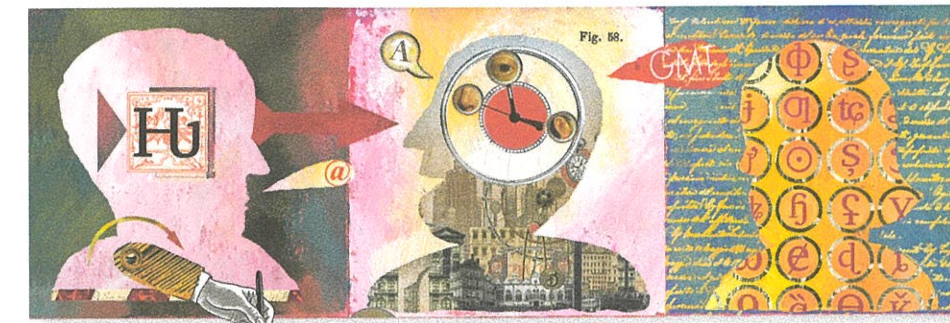
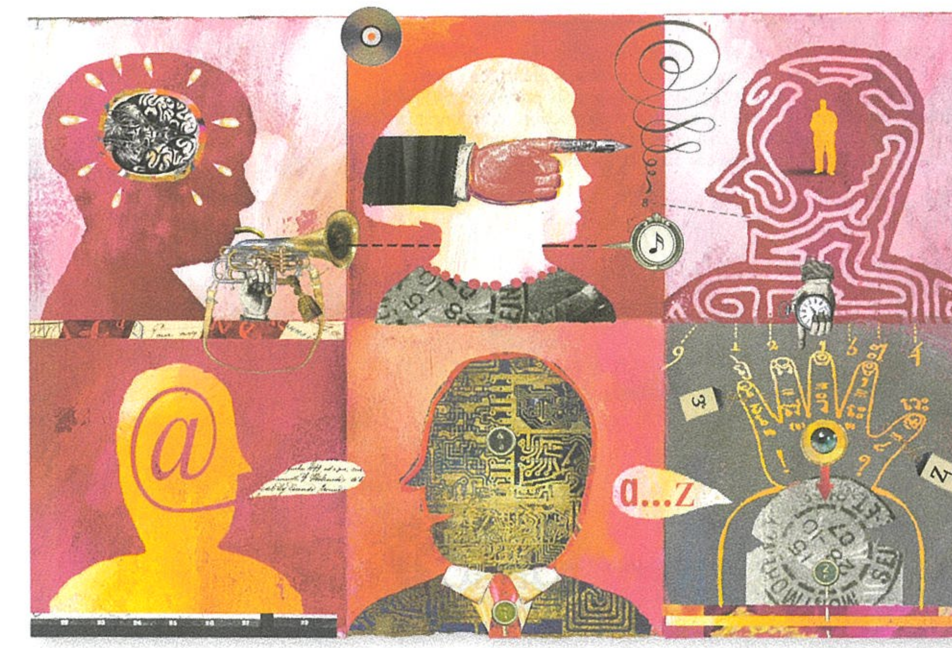


Language in Mind

An Introduction to Psycholinguistics



JULIE SEDIVY *University of Calgary*

SINAUER ASSOCIATES, INC. • Publishers • Sunderland, Massachusetts • USA

exercises in *Sesame Street* and other educational program like it. In fact, this sort of awareness of how syllables break apart into individual sounds is an important skill in reading (as seen in Box 3.2). Intrigued by the connection, researchers have started to look at whether ramping up musical skills through training can transfer over into reading performance. In one such study, a 6-month musical training program for children in elementary school led to improvements on measures of reading, whereas a 6-month program in painting did not (Moreno et al., 2009).

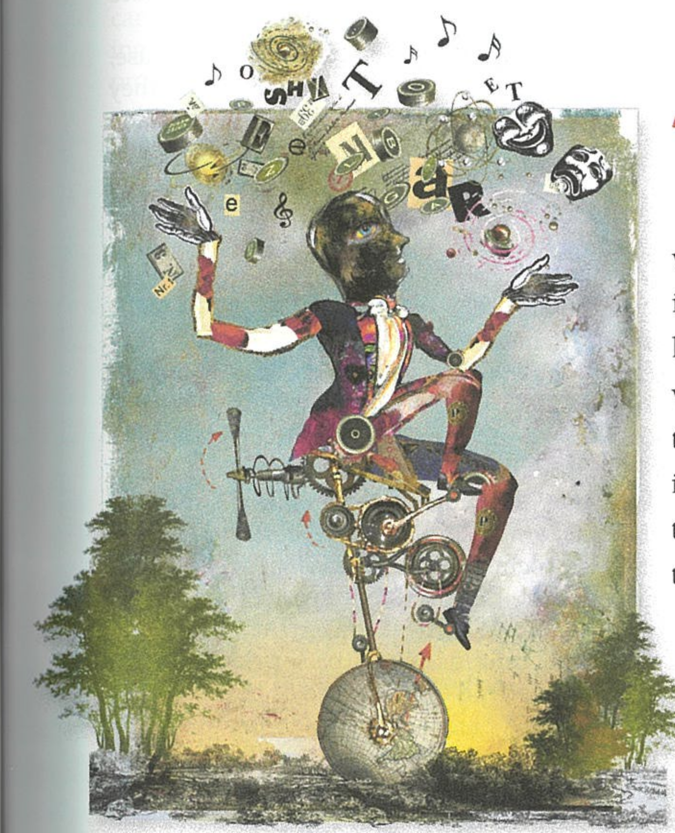
Finely tuned speech processing skills can also be seen in people with musical expertise and experience. Those who regularly take part in musical activity tend to perform better at distinguishing very fine aspects of speech sounds, and have an advantage when it comes to perceiving speech in noisy environments (e.g., Chobert et al., 2011; Parbery-Clark et al., 2009; Sadakata & Sekiyama, 2011).

Evidence is building for the idea that language and music make use of a number of similar resources and skills, even though they part ways in the brain for some of their important computations. It looks like they're especially closely linked when it comes to detailed processing of sound sequences, and structuring smaller units into sentences or melodies with predictable underlying structures. On the more practical side, this ongoing, intense research may show that the hours children log at the piano have broader benefits, and that music therapy might be appropriate for treating certain kinds of language problems. On the theoretical end, the relationship between language and music makes a fascinating case study of the possible ways in which two venerated human abilities might converge and diverge in the brain.

PROJECT

Music, like language, involves a complex *set* of activities and mental processes, and it demands multiple skills, not just one. Try to generate as detailed a list as you can of the various components that go into musical and linguistic activity, going beyond those discussed in this section. What makes someone good at music/language? Once you've generated your lists, identify which of the skills that are needed for music appear to have close analogues in language. Where do you think it would be most likely that you'd see crossover in cognitive processing? How would you go about finding out? Create a proposal for how you might gather evidence to support your idea of connections between music/language skills.

4 Learning Sound Patterns



To get a sense of what's really involved in learning language, cast your mind back to what it was like before you knew any words at all of your native tongue. Well, wait … since you obviously can't do that, the best you can do is to recall any experiences you may have had learning a *second* language at an age old enough to remember what the experience was like (or a third or fourth language, if you were lucky enough to learn more than one language as a tot). If these memories involve learning a language in a classroom setting, they turn out to be a useful point of departure for our purposes, especially to highlight the striking difference between how you learned language in the classroom, and how you learned it as a newborn initiated into your native language.

In a foreign language classroom, it's usual for the process to kick off with a teacher (or textbook) translating a list of vocabulary items from the new language into your native language. You then use a small but growing vocabulary to build up your knowledge of the language. You begin to insert words into prefabricated sentence frames, for example, and eventually you build sentences from scratch. This is simply not an approach that was available to you as an infant because then you had no words in *any* language that could be used as the basis of translation. Worse, you didn't even know what *words* were, or where words began or ended in the stream of speech you were listening to. You were basically swimming in a sea of sound, and there wasn't a whole lot anyone could do in the way of teaching that would have guided you through it.

If you were to have the unusual experience of learning a second language by simply showing up in a foreign country and plunging yourself into the language as best you could, without the benefit of language courses or tourist phrase books, that would be a bit closer to what you faced as an infant.

But still, you have advantages as an adult that you didn't have as an infant. You are much more sophisticated in your knowledge of the world, so you're not faced with learning how to describe the world using language while you're trying to figure out what the world is like. And your intellect allows you to be much more strategic in how you go about getting language samples from the speakers of that language; you can, for example, figure out ways to ask speakers about subtle distinctions—like whether there are different words for the concepts of *cat* and *kitten*, or how to interpret the difference in similar expressions, such as *screw up* and *screw off*. Not to mention that your motor skills allow you to pantomime or point to objects as a way to request native speakers to produce the correct words for you.

Deprived of many of the possible learning strategies that older people use, it might make sense that babies would postpone language learning until they develop in other areas that would help support this difficult task. And, given the fact that most babies don't start producing recognizable words until they're about a year old, and that they take quite a bit longer than that to string sentences together, it might seem that that's exactly what does happen. But in fact, babies begin learning their native language from the day they're born, or even earlier; it turns out that French babies tested within 4 days of birth could tell the difference between French and Russian, and sucked more enthusiastically on a pacifier when hearing French (Mehler et al., 1988). On the other hand, infants born into other linguistic households, such as Arabic, German, or Chinese, did not seem to be able to tell the difference between French and Russian speech, nor did French household babies seem to notice the difference between English and Italian. This study indicates that babies in utero can begin to learn something about their native language. Obviously, this can't be the result of recognizing actual words and their meanings, since in utero babies have no experience of the meanings that language communicates. Rather, it suggests that even through the walls of the womb and immersed in amniotic fluid, babies learn something about the patterns of sounds in the language they hear.

Humans are unlike honeybees and certain species of songbirds, which are genetically programmed for a specific type of bee dance or birdsong. The speech and accent of a child born of French parents but raised from infancy in the United States will usually be indistinguishable from that of a child born to U.S. parents and raised in that country. This linguistic flexibility reflects the fact that humans are powerful learning machines.

In this chapter, we'll look at what young children need to learn about the sounds of their language—and the sound system of any language is an intricate, delicately patterned thing. Not only does it have its own unique collection of sounds, but it has different "rules" for how these sounds can be combined into words. For example, even though English contains all the individual sounds of a word like *ptak*, it would never allow them to be strung in this order, though Czech speakers do so without batting an eye. And no word in Czech can ever end in a sound like "g," even though that consonant appears in abundance at the beginnings and in the middles of Czech words. English speakers, on the other hand, have no inhibitions about uttering a word like *dog*.

In fact, the sound pattern of a language is a complex code that infants manage to crack and mostly master within the first couple of years of life. A magnificent amount of learning happens within the first few months of birth. Long before they begin to produce words (or even really show that they understand their meanings), babies can:

1. Differentiate their native language from other languages.
2. Have a sense of how streams of sound are carved up into words.
3. Give special attention to distinctions in sounds that will be especially useful for signaling different meanings (for example, the distinction between "p" and "b" sounds; switch the "p" in *pat* to a "b" sound, and you get a word with a different meaning).
4. Figure out how sounds can be "legally" combined into words in their language.

Babies also develop many other nifty skills. Perhaps the only people who deserve as much admiration as these tiny, pre-verbal human beings are the scientists who study the whole process. Unlike foreign language teachers who can test students' mastery of a language via multiple choice exams and writing samples, language researchers have to rely on truly ingenious methods for probing the infant mind. In this chapter, you'll get a sense of the accomplishments of both groups: the very young who crack the sound code, and the scientists who study their feats.

4.1 Where Are the Words?

A stunningly *ineffective* way to learn language would be to simply memorize the meaning of every complete sentence you've ever heard, never bothering to break the sentence down into its component parts. In fact, if you took this route (even if you could actually memorize thousands upon thousands of complete sentences in the form of long strings of sound), you wouldn't really be learning language at all. No matter how many sentences you'd accumulated in your memory stash, you'd constantly find yourself in situations where you were required—but unable—to produce and understand sentences that you'd never encountered before. And without analyzing language in terms of its component, reusable parts, learning a sentence like *My mother's sister has diabetes* would be no help at all in understanding the very similar sentence *My father's dog has diabetes*. You'd treat the relationship in sound between these as purely coincidental, much as you would the relationship between the similar-sounding words *battle* and *cattle*; each one would have to be memorized completely independently of the other. As you saw in Chapter 2, the aspect of language that lets us combine meaningful units (like words) to produce larger meaningful units (like phrases or sentences) is one of the universal properties of human language, and one that gives it enormous expressive power.

Fundamentally, learning language involves figuring out which sounds clump together to form basic units, and learning how these units in turn can be combined with other units—which is why foreign language instruction for beginners puts so much emphasis on learning lists of words. One of the infant's earliest tasks, then, is to figure out which strings of sounds form these basic units—no trivial accomplishment. In talking to their babies, parents are not nearly as accommodating as Spanish or German textbooks, and they rarely speak to their children in single-word utterances (about 10 percent of the time). This means that babies are confronted with speech in which multiple words are sewn seamlessly together, and they have to figure out all on their own where the edges of words are. And unlike written language, where words are clearly separated by spaces (at least in most writing systems), spoken language doesn't present convenient breaks in sound to isolate words. To get an intuitive feel for what this task might feel like to a baby, see how you fare in Web Activity 4.1.

WEB ACTIVITY 4.1

Finding word boundaries In this activity, you'll hear speech in several different languages, and you'll be asked to guess where the word boundaries might be.

METHOD 4.1

The head-turn preference paradigm

The head-turn preference paradigm is an invaluable tool that's been used in hundreds of studies of infant cognition. It can be used with babies as young as about 4.5 months, and up to about 18 months. After this age, toddlers often become too fidgety to reliably sit still through this experimental task. Applied to speech perception, the technique is based on two simple principles:

1. Babies turn their heads to orient to sounds.
2. Babies spend more time orienting to sounds that they find interesting.

In a typical experiment, the baby sits on the lap of a parent or caregiver who is listening to music over a set of headphones; this prevents the adult from hearing the experimental stimuli and either purposely or inadvertently giving cues to the baby. A video camera is set up to face the child, recording the baby's responses, and an unseen observer monitors the experiment by watching the baby on video. (The observer can't hear the stimulus sounds and is usually not aware of which experimental condition the child has been assigned to, though the observer does control the sequence of events that occur.) A flashing light mounted next to the video camera and straight ahead in the child's view can be activated to draw the child's attention to a neutral point before any of the experimental stimuli are played. The stimuli of interest are then played on two speakers, mounted on the left and right walls (see Figure 4.1).

Each experiment usually consists of a *familiarization phase* and a *test phase*. There are three goals for the familiarization phase. The first is simply to have the infant become familiar with the sound stimuli. In some cases, if the purpose of the study is to find out whether babies will learn something about new sounds, the sounds played during the familiarization phase might consist of the new stimuli to be learned. The second goal is to train the baby to expect that sounds can come from the speaker on either the left or the right wall. The third goal of the familiarization phase is to tightly lock together the head-turn behavior to the infant's auditory attention. Babies tend to look in the direction of a sound that holds their attention anyway, but this connection can be strengthened by flashing a light

Continued on next page

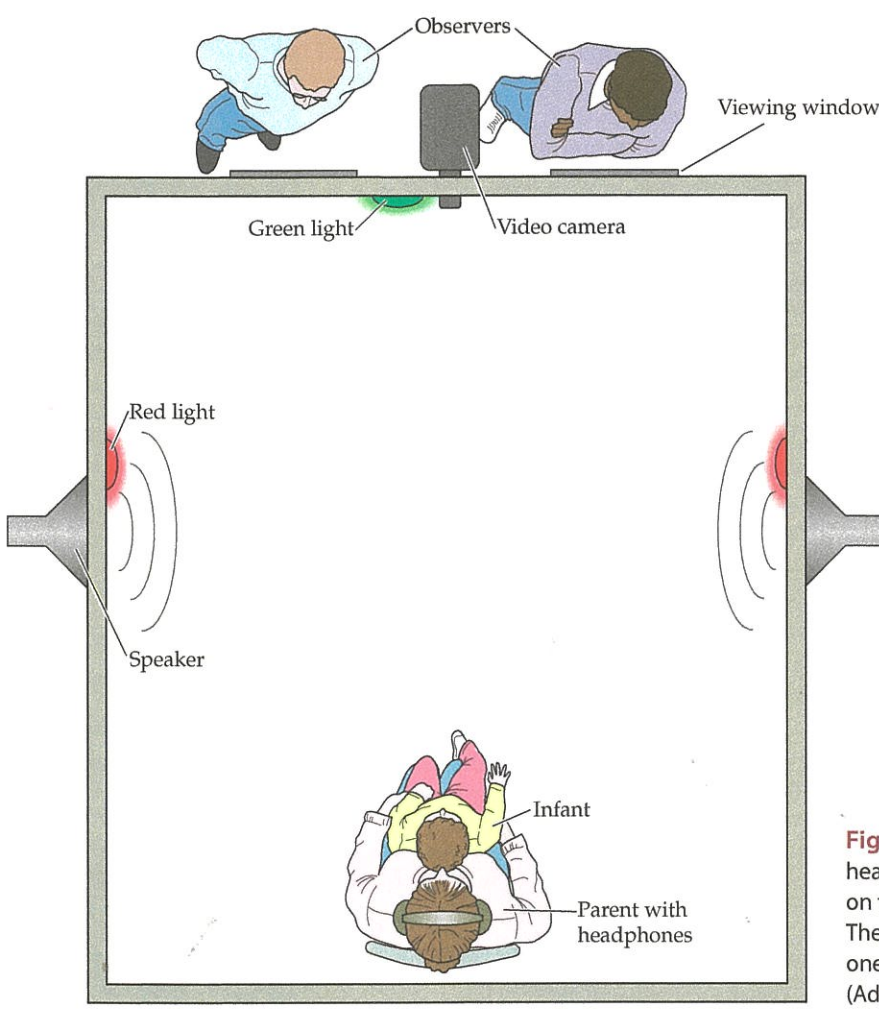


Figure 4.1 A testing booth set up for the head-turn preference paradigm. The baby sits on the caregiver's lap, facing the central panel. The observer looks through a small window or one-way mirror to note the baby's head turns. (Adapted from Nelson et al., 1995.)

METHOD 4.1 *(continued)*

in the location of the speaker before each sound, and by making sure that sounds during the familiarization phase are played only for as long as the baby looks in the direction of the sound. This signals to the child that if she wants to continue hearing a sound, she needs to be looking in its direction. After all these goals have been achieved, the baby is ready for the test phase.

During the test phase, the sounds of interest are played on either the left or the right speaker, and the baby's head-turn behavior is recorded by the video camera for later coding. Researchers then measure how long the baby spends looking in the direction of each sound, and these responses are averaged over stimulus type.

Sometimes researchers are interested in which sounds the infants prefer to listen to—do they prefer a female voice to a male voice, for example, or do they prefer to listen to sounds of their own language over an unknown language? Other times, the researchers are only interested in whether the infants discriminate between two categories of stimuli, and it doesn't matter which category is preferred. For instance, in a learning experiment, any distinction in looking times for familiar versus new stimuli should be an indication of learning, regardless of whether babies prefer to listen longer to the new or the familiar sounds. And in fact, it turns out that there isn't a clear preference for either new or familiar sounds—at times, babies show more interest in sounds that they recognize, and at other times, they show more interest in completely novel ones. The preferences seem to depend somewhat on the age of the child and just how often they've heard the familiar sounds (overly familiar sounds might cause a baby to become bored with them).

Probing infants' knowledge of words

Babies begin to produce their first words at about a year or so, but they start to identify word breaks at a much younger age than that. In fact, the whole process is under way by 6 or 7 months. Scientists can't exactly plop a transcript of speech in front of a baby, hand them a pencil, and ask them to mark down where the word breaks are. So how is it possible to peer into infants' minds and determine whether they are breaking sentences down into their component parts?

In studying the cognitive processes of infants, researchers have to content themselves with a fairly narrow range of infant behaviors as a way to measure hidden psychological mechanisms. So, when they decide to study infants of a particular age, they need to have a clear sense of what babies can do at that point in their development—and more specifically, which behaviors reflect meaningful cognitive activity. It turns out that a great deal of what we now know about infant cognition rests on one simple observation: when babies are faced with new sounds or images, they devote their attention to them differently than when they hear or see old, familiar sounds or images. And at the age of 6 or 7 months, one easy way to tell if a baby is paying attention to something is if she swivels her head in its direction to stare at it; the longer she keeps her gaze oriented in its direction, the longer she's paying attention to that stimulus. New sounds and sights tend to draw attention differently than familiar ones, and babies will usually orient to novel versus familiar stimuli for different lengths of time—sometimes they're more interested in something that's familiar to them (a sort of "Hey, I know what that is!" response), but sometimes they prefer the novelty of the new stimulus.

These simple observations about the habits of babies gave birth to a technique that psycholinguists now commonly use, called the **head-turn preference paradigm** (see Method 4.1). This technique compares how long babies keep their heads turned toward different stimuli, taking this as a measure of their attention. (If the target stimulus is a sound, it's usually coupled with a visual stimulus such as a light or a dancing puppet in order to best elicit the

head-turn preference paradigm An experimental framework in which infants' speech preference or learning is measured by the length of time they turn their heads in the direction of a sound.

familiarization phase A preparation period during which subjects are exposed to stimuli that will serve as the basis for the test phase to follow.

test phase The period in which subjects' responses to the critical experimental stimuli is tested following a familiarization phase.

head-turn response.) But what makes the method really powerful is that it can leverage the measure of looking time as a way to test whether or not the babies taking part in the study have *learned* a particular stimulus. For instance, let's say we give babies a new word to listen to during a **familiarization phase**. At some later time, during a **test phase**, we can see whether their looking times when hearing this word are different from those for a word they've never heard before. If babies spend either more or less time looking at the previously presented word than they do at a completely new word, this suggests that they've learned something about the first word and now treat it as "familiar." On the other hand, if they devote equal amounts of looking time to both words, it suggests that they haven't learned enough about the previously heard word to differentiate it from a completely novel word.

The head-turn preference paradigm has been used—for example, by Peter Jusczyk and Richard Aslin (1995)—to tackle the question of whether babies have learned where word breaks occur. Here's how: During the familiarization phase of that study, the baby participants heard a series of sentences that contained a target word, say, *bike*, in various different positions in the sentence:

His bike had big black wheels.

The girl rode her big bike.

Her bike could go very fast.

The bell on the bike was really loud.

The boy had a new red bike.

Your bike always stays in the garage.

During the test phase, the researchers measured how long the infants were interested in listening to repetitions of the target word *bike*, compared with a word (say, *dog*) that they hadn't heard during the familiarization phase. To first get the baby's attention before the test word was played, a flashing light appeared above the loudspeaker the word was to come from. Once the baby looked in this direction, the test word repetitions began to play. When the baby's interest flagged, causing him to look away from the loudspeaker, this was noted by a researcher, and the baby was scored for the amount of time spent looking in the direction of the loudspeaker. Jusczyk and Aslin found that overall, 7.5-month-old babies spent more time turning to the speaker when it played a familiar word (*bike*) than when it played an unfamiliar word (*dog*). This might not seem like a tremendous feat to you, but keep in mind that the babies must somehow have separated the unit *bike* from the other sounds in the sentences during the familiarization phase in order to be able to match that string of sounds with the repeated word during the test phase. Six-month-old babies didn't seem to have this ability yet.

This study shows that by the tender age of 7.5 months, babies seem to be equipped with some ability to separate or segment words from the speech stream—but it doesn't tell us how they manage to come by these skills, or what information they rely on to decide where the words are. Since Jusczyk and Aslin's initial study, dozens of published articles have explored the question of how babies pull this off. We'll investigate several ways that they might begin to crack the problem.

Familiar words break apart the speech stream

Here's one possibility. Remember that babies hear single-word utterances only about 10% of the time. That's not a lot, but it might be enough to use as a starting point for eventually breaking full sentences apart into individual words. It may be that babies can use those few words they do hear in isolation as a way to build up a small collection of known word units. These familiar words can then serve as anchoring points for breaking up the speech stream into more manageable chunks. For example, imagine hearing this procession of sounds in a fictional language:

bankiritubendudifin

Any guesses about what the word units are? The chances of getting it right are not very high. But suppose there are two words in this stream that you've heard repeatedly because they happen to be your name ("Kiri") and the name for your father ("Dudi"). You may have learned them because these are among the few words that are likely to be uttered as single words quite often, so they're especially easy to recognize. Perceptually, they'll leap out at you. If you've learned a foreign language, the experience of hearing sentences containing a few familiar words may be similar to the very early stages of learning a new language as a baby; you might have been able to easily pull out just one or two familiar words from an otherwise incomprehensible sentence. With this in mind, imagine hearing:

ban-**kiri**-tuben-**dudi**-fin

Now when you hear the names *kiri* and *dudi*, their familiarity allows you to pull them out of the speech stream—but it might also provide a way to identify other strings of sound as word units. It seems pretty likely that *ban* and *fin* are word units too, because they appear at the beginning and end of the utterance and are the only syllables that are left over after you've identified *kiri* and *dudi* as word units. So now, you can pull out four stand-alone units from the speech stream: *kiri, dudi, ban, fin*. You don't know what these last two mean, but once they're firmly enough fixed in your memory, they might in turn serve as clues for identifying other new words. So *tamfinatbankirisan* can now be pulled apart into:

tam-**fin**-at-**ban**-**kiri**-san

The residue from this new segmentation yields the probable units *tam, at*, and *san*, which can be applied in other sentences in which these units are combined with entirely new ones.

In principle, by starting with a very few highly familiar words and generating "hypotheses" about which adjacent clumps of sound correspond to units, an infant might begin to break down streams of continuous sound into smaller pieces. In fact, experimental evidence shows that babies are able to segment words that appear *next* to familiar names when they're as young as 6 months old, suggesting that this strategy might be especially useful in the very earliest stages of speech segmentation. (Remember that the Jusczyk and Aslin study showed that 6-month-olds did not yet show evidence of generally solid segmentation skills.) This was demonstrated in an interesting study led by Heather Bortfeld (2005). Using the head-turn preference paradigm, the researchers showed that even 6-month-olds could learn to segment words such as *bike* or *feet* out of sentences—but only if they appeared right next to their own names (in this example, the baby subject's name is *Maggie*) or the very familiar word *Mommy*:

Maggie's bike had big, black wheels.

The girl laughed at Mommy's feet.

That is, when babies heard these sentences during the familiarization phase, they later spent more time looking at loudspeakers that emitted the target word *bike* or *feet* than at speakers that played an entirely new word (*cup* or *dog*)—this shows they were treating *bike* and *feet* as familiar units. But when the sentences in the familiarization phase had the target words *bike* and *feet* right next to names that were *not* familiar to the child (for example, *The girl laughed at Tommy's feet*), the babies showed no greater interest in the target words than they did in the new words *cup* or *dog*. In other words, there's no evidence that the infants had managed to pull these words out of the stream of speech when they sat next to unknown words. At this age, then, it seems that babies can't yet segment words out of just any stream of speech, but that they *can* segment words that appear next to words that are already very familiar.

Discovering what words sound like

Relying on familiar words to bust their way into a stream of sound is just one of the tricks that babies have in their word-segmentation bag. Another trick has to do with developing some intuitions about which sounds or sequences of sounds are allowed at the beginnings and ends of words, and using these intuitions as a way to guess where word boundaries are likely to be.

To get a more concrete feel for how this might work, try pronouncing the following nonsense sentence in as "English-like" a way as you can, and take a stab at marking the word boundaries with a pencil (*hint:* there are six words in the sentence):

Banriptangbowpkesternladfloop.

If you compare your answer with those of your classmates, you might see some discrepancies, but you'll also find there are many similarities. For example, chances are, no one proposed a segmentation like this:

Ba-nri-ptangbow-pkester-nladfl-oop.

This has to do with what we think of as a "possible" word in English, in terms of the sequence of sounds it's allowed to contain. Just because your language includes a particular sound in its inventory doesn't mean that sound is allowed to pop up just anywhere.

Languages have patterns that correspond to what are considered "good" words as opposed to words that look like the linguistic equivalent of a patched-together Frankenstein creature. For example, suppose you're a marketing expert charged with creating a brand new word for a line of clothing, and you've decided to write a computer program to randomly generate words to kick-start the whole process. Looking at what the computer spat out, you could easily sort out the following list of words into those that are possible English-sounding words, and those that are not:

ptangb	sastashak
roffo	lululeming
spimton	ndela
skrs	srbridl

What counts as a well-behaved English word has little to do with what's actually pronounceable—you might think that it's impossible to pronounce sequences of consonants like *pt*, *nd*, and *dl*, but actually you do it all the time in words like *riptide*, *bandage*, *bed linen*—it's just that in English, these consonants have to straddle word boundaries or even just syllable boundaries. (Remember, there are no actual *pauses* between these consonants just by virtue of their belonging to different syllables or words.) You reject the alien words like *ptangb* or *ndela* in your computer-generated list, not because it takes acrobatic feats of

your mouth to pronounce them, but because you have ingrained word templates in your mind that you've implicitly learned, and these words don't match those mental templates. These templates differ from one language to another and are known as **phonotactic constraints**.

Using English phonotactic constraints to segment another language, though, could easily get you into trouble, especially if you try to import them into a language that's more lenient in allowing exotic consonant clusters inside its words. For instance, *skrs* and *ndela* are perfectly well-formed pronunciations of words in Czech and Swahili, respectively. So, trying to segment speech by relying on your English word templates would give you non-optimal results. You can see this if you try to use English templates to segment the following three-word stream of Swahili words:

nipemkatenzuri

You might be tempted to do either of the following:

nipem-katen-zuri

nip-emkat-enzuri

but the correct segmentation is:

nipe-mkate-nzuri

And while languages like Czech and Swahili are quite permissive when it comes to creating consonant clusters that would be banned by the rules of English, other languages have even tighter restrictions on clusters than English does. For instance, "sp" is not a legal word-initial cluster in Spanish (which is why speakers of that language often say "espanish"). (You can see some more examples of how languages apply different phonotactic constraints in Box 4.1.)

WEB ACTIVITY 4.2

Segmenting speech through phonotactic constraints In this activity, you'll hear sound files of nonsense words that conform to the phonotactic constraints of English, as well as clips from foreign languages that have different phonotactic constraints. You'll get a sense for how much easier it is to segment unknown speech when you can use the phonotactic templates you've already learned for your language, even when none of the words are familiar.

It turns out that by 9 months of age, babies have some knowledge of the templates for proper words in their language. Using the head-turn preference paradigm, researchers led by Peter Jusczyk (1993) have shown that American babies orient longer toward strings of sounds that are legal words in English (for example, *cubeb*, *dudgeon*) than they do to sequences that are legal Dutch words but illegal words in English (*zampljes*, *vlatke*). Dutch 9-month-olds show exactly the opposite pattern. This suggests that they're aware of what a "good" word of their language sounds like. And, just as neither you nor your classmates suggested that speech should be segmented in a way that allows bizarre words like *ptangb* or *nladf*, at 9 months of age, babies can use their phonotactic templates to segment units out of speech (Mattys & Jusczyk, 2001).

You might have noticed in Web Activity 4.2 that there was another bit of information that might have helped you segment words, in addition to clues about phonotactic constraints. In that exercise, English-like stress patterns were also present. In English, stress tends to alternate, so within a word, you usually get an unstressed syllable sitting next to a stressed one: *reTURN*, *BLACKmail*, *inVIgoRATE*. (In some other languages, such as French, syllables are more or less evenly stressed.) English words can follow either a **trochaic stress pattern**, in which the first syllable is stressed (as in *BLACKmail*), or an **iambic stress pattern**, in which the first syllable is unstressed (as in *reTURN*). But as it turns out, it's not an equal-opportunity distribution, and trochaic words far outnumber iambic words (on the order of 9 to 1 by some estimates). Chances are, you subconsciously made use of this knowledge in your segmentation answers in Web Activity 4.2. If babies have caught on to this pattern in

phonotactic constraints Language-specific constraints that determine how the sounds of a given language may be combined to form words or syllables.

trochaic stress pattern Syllable emphasis pattern in which the first syllable is stressed, as in *BLACKmail*.

iambic stress pattern Syllable emphasis pattern in which the first syllable is unstressed, as in *reTURN*.

BOX 4.1
Phonotactic constraints across languages

As languages go, English is reasonably loose in allowing a wide range of phonotactic templates. English allows consonants to gather together in sizable packs at the edges of words and syllables. For example, the single-syllable word *splints* has the structure CCCVCCC (where C = consonant and V = vowel). Many other languages are far more restrictive. To illustrate, the Hebrew, Hawaiian, and Indonesian languages allow only the following syllable structures:

Hebrew	Hawaiian	Indonesian
CV	V	V
CVC	CV	VC
CVCC		CV
		CVC

In addition to broadly specifying the consonant-vowel structure of syllables, languages have more stringent rules about which consonants can occur where. For example, in English, /rp/ is allowed at the end of the word, but not at the beginning; the reverse is true for the cluster /pr/.

Some constraints tend to recur across many languages, but others are highly arbitrary. The table gives a few examples of possible and impossible clusters that can occur at the beginnings of words and syllables in several languages.

To get a feel for how speech segmentation might be affected by the language-specific phonotactic constraints listed in the table, try listing all the possible ways to break down the following stream of sounds into word units that are legal, depending on whether your language is English, German, French, or Italian:

bakniskweriavrosbamanuesbivriknat

Allow words and syllables to start with

Language	/kn/	/skw/	/sb/	/vr/
English	no	yes	no	no
German	yes	no	no	no
French	no	no	no	yes
Italian	no	no	yes	no

the language, they might also be able to use this to make guesses about how words are segmented from speech. And indeed, they do: by 7.5 months, babies have no trouble slicing words with a trochaic pattern (like *DOCtor*) from the speech stream—but when they hear an iambic word like *guiTAR* embedded in running speech, they don't recognize it. In fact, if they've heard *guitar* followed by the word *is*, they behave as if they've segmented *TARis* as a word (Jusczyk et al., 1999).

But if you've been paying very close attention, you might have noticed a paradox: in order for babies to be able to use templates of permissible words to segment speech, they *already* have to have some notion of a word—or at the very least, they have to have some notion of a language unit that's made up of a stable collection of sounds that go together and can be separated from other collections of sounds. Otherwise, how can they possibly have learned that "ft" can occur as a sequence of sounds in the middle or at the end of a word unit, but not at its beginning? It's the same thing with stress patterns: How can babies rely on generalizations about the most likely stress patterns for words in their language unless they've already analyzed a bunch of words?

To get at generalizations like these, babies must already have segmented some word units, held them in memory, and "noticed" (unconsciously, of course) that they follow certain patterns. All this from speech that rarely has any words that stand alone without the added confusion of adjacent speech sounds.

Earlier we speculated that maybe isolated familiar words like the baby's name serve as the very first units; these first words then act as a wedge for

segmenting out other words, allowing the baby to build up an early collection of word units. Once this set gets large enough, the baby can learn some useful generalizations that can then accelerate the whole process of extracting additional new words from running speech. In principle, this is plausible, given that babies *can* use familiar words like names to figure out that neighboring bunches of sounds also form word units. But this puts quite a big burden on those very first few words. Presumably, the baby has managed to identify them because they've been pronounced as single-word utterances. But as it happens, parents are quite variable in how many words they produce in isolation—some produce quite a few, but others are more verbose and rarely provide their children with utterances of single words. If this were the crucial starting point for breaking into streams of speech, we might expect babies to show a lot more variability in their ability to segment speech than researchers typically find, with some lagging much further behind others in their segmentation abilities.

Luckily, youngsters aren't limited to using familiar, isolated words as a departure point for segmentation—they have other, more flexible and powerful tricks up their sleeves. Researchers have discovered that babies can segment streams of sounds from a completely unfamiliar language after as little as two *minutes* of exposure, without hearing a single word on its own, and without the benefit of any information about phonotactic constraints or stress patterns. How they manage this accomplishment is the topic of the next section.

4.2 Infant Statisticians

Tracking transitional probabilities: The information is out there

In a seminal study, Jenny Saffran and colleagues (1996) familiarized 8-month-old infants with unbroken 2-minute strings of flatly intoned, computer-generated speech. The stream of speech contained snippets such as:

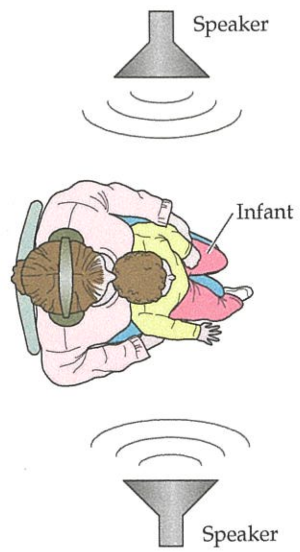
bidakupadotigolabubidaku

Notice that the sounds are sequenced so that they follow a repeating consonant-vowel structure. Because English allows any of the consonants in this string to appear either at the beginnings or the ends of syllables and words, nothing about the phonotactic constraints of English offers any clues at all about how the words are segmented, other than that the consonants need to be grouped with at least one vowel (since English has no words that consist of single consonants). For example, the word *bidakupa* could easily have any of the following segmentations, plus a few more:

bi-dak-upa	bid-aku-pa	bid-ak-u-pa
bi-da-kup-a	bid-ak-up-a	bidaku-pa
bid-akupa	bida-kupa	bi-dakup-a
bi-daku-pa	bid-akup-a	bidak-upa

If this seems like a lot, keep in mind that these are the segmentation possibilities of a speech snippet that involves just *four* syllables; imagine the challenges involved in segmenting a two-minute-long continuous stream of speech. This is precisely the task that Saffran and colleagues inflicted on the babies they studied.

In the Saffran et al. study, though, the stream of sound that the babies listened to during their familiarization phase was more than just a concatenation of consonant-vowel sequences. The stimuli were created in a way that repre-



Experimenters create an artificial "language" of four "words":
bidaku, golabu, padoti, dutaba

Familiarization phase

Infant hears each "word" repeated 45 times in random order, in an unbroken 2-minute synthesized speech stream:
bidaku-golabu-dutaba-golabu-padoti-bidaku-dutaba-padoti-golabu-dutaba-padoti-dutaba-golabu-bidaku-dutaba-bidaku-padoti-bidaku-padoti-golabu etc.

Test phase

Loudspeakers present infant either a "real" word:
bidaku, golabu
or a sequence of syllables with parts of two words:
dakugo, buduta

Results

Mean looking times for 8-month-olds	
"Real" words	6.77 seconds
Part-words	7.60 seconds

Figure 4.2 In this study, Saffran and colleagues prepared stimuli that amount to a miniature artificial language of four "words," each word consisting of three consonant-vowel syllables. Infants then heard an uninterrupted, 2-minute stream of random combinations of the four words. The researchers noted how much attention the babies paid to the four "words" from the familiarization phase and compared it with the attention the babies paid to three-syllable sequences that also occurred in the speech but that straddled "word" boundaries (part-words). (Adapted from Saffran et al., 1996.)

artificial language A "language" that is constructed to have certain specific properties for the purpose of testing an experimental hypothesis: strings of sounds correspond to "words," which may or may not have meaning, and whose combination may or may not be constrained by syntactic rules.

sents a miniature **artificial language**. That is, the string of sounds corresponded to concatenations of "word" units combining with each other. In this particular "language," each "word" consisted of three consonant-vowel syllables (see Figure 4.2). For example, *bidaku* in the above stream might form a word. The uninterrupted two-minute sound stream consisted of only four such "words" randomly combined to form a sequence of 180 "words" in total, which meant that each "word" appeared quite a few times during the sequence. (The fact that the words were randomly combined is obviously unrealistic when it comes to how real, natural languages work. In real languages, there's a whole layer of syntactic structure that constrains how words can be combined. However, for this study, the researchers were basically only interested in how infants might use very limited information from sound sequences to isolate words.)

Later, during the test phase, the researchers noted how much attention the babies paid to actual "words" they heard in the familiarization phase and compared it with the attention the babies paid to three-syllable sequences that also occurred in the speech but that straddled "word" boundaries—for example, *dakupa* (see Figure 4.2). The infants showed a distinction between "words" and "part-word" sequences. In this case, they were more riveted by the "part-words," listening to them longer than the "words"—possibly because they were already bored by the frequent repetition of the "word" units.

How did the 8-month-old infants Saffran et al. studied manage to do this? If the sound stimuli were stripped of all helpful features such as already-familiar words, stress patterns, and phonotactic cues, what information could the babies possibly have been using in order to pull the "words" out of the 2-minute flow of sound they'd heard? To see how you'd fare in such a task, try Web Activity 4.3.

The answer is that there's a wealth of information in the speech stream waiting to be mined, and it's there just by virtue of the fact that the stream is composed of word-like units that turn up multiple times. Saffran and her

WEB ACTIVITY 4.3

Segmenting "words": You be the baby! In this activity, you'll hear 2 minutes of stimuli from an artificial language very similar to that used by Saffran et al. (1996). You'll be asked to distinguish "words" from "non-words" to see if you, too, can manage to segment speech. (Ideally, you should attempt this exercise before you read any further.)

colleagues suggested that babies were acting like miniature statisticians analyzing the speech stream, and were keeping track of **transitional probabilities (TPs)** between syllables—this refers to the likelihood that one particular syllable will be followed by another specific syllable. Here's how such information would help to define likely word units: Think of any two syllables, say, a syllable like *ti* and a syllable like *bay*. Let's say you hear *ti* in a stream of normal English speech. What are the chances that the very next syllable you hear will be *bay*? It's not all that likely; you might hear it in a sequence like *drafty basement* or *pretty baby*, but *ti* could just as easily occur in sequences that are followed by different syllables, as in *T-bone steak, teasing Amy, teenage wasteland, Fawlty Towers*, and many, many others.

transitional probability (TP) The probability that a particular syllable will occur, given the previous occurrence of another particular syllable.

But notice that *ti* and a *bay* that follows it don't make up an English word. It turns out that when a word boundary sits between two syllables, the likelihood of predicting the second syllable on the basis of the first is vanishingly small. But the situation for predicting the second syllable based on the first looks very different when the two syllables occur together *within* a word. For example, take the sequence of syllables *pre* and *ti*, as in *pretty*. If you hear *pre*, as pronounced in this word, what are the chances that you'll hear *ti*? They're much higher now—in this case, you'd never hear the syllable *pre* at the end of a word, so that leaves only a handful of words that contain it, dramatically constraining the number of options for a following syllable. Generally, the transitional probabilities of syllable sequences are much higher for pairs of syllables that fall within a word than for syllables that belong to different words. This is simply because of the obvious fact that words are *units* in which sounds and syllables clump together to form a fairly indivisible whole. Since there's a finite number of words in the language that tend to get used over and over again, it stands to reason that the TPs of syllable sequences within a word will be much higher than the TPs of syllable pairs coming from different words.

How does all this help babies to segment speech? Well, if the little tykes can somehow manage to figure out that the likelihood of hearing *ti* after *pre* is quite high, whereas the likelihood of hearing *bay* after *ti* is low, they might be able to respond to this difference in transitional probabilities by "chunking" *pre* and *ti* together into a word-like unit, but avoid clumping *ti* and *bay* together.

Here's the math: The transitional probability can be quantified as $P(Y|X)$, that is, the probability that a syllable Y will occur given that the syllable X has just occurred. This is done by looking at a sample of a language and dividing the frequency of the syllable sequence XY by the frequency of the syllable X combined with any syllable:

$$TP = P(Y|X) = \frac{\text{frequency } (XY)}{\text{frequency } (X)}$$

In the study by Saffran and her colleagues, the only cues to word boundaries were the transitional probabilities between syllable pairs: within words, the transitional probability of syllable pairs (e.g., *bida*) was always 1.0, while the transitional probability for syllable pairs across word boundaries (e.g., *kupa*) was always 0.33.

That babies can extract such information might seem like a preposterous claim. It seems to be attributing a whole lot of sophistication to tiny babies. You might have even more trouble swallowing this claim if *you*, a reasonably intelligent adult, had trouble figuring out that transitional probabilities were the relevant source of information needed to segment the speech in Web Activity 4.3 (and you wouldn't have been alone in failing to come up with an explanation for how the speech might be segmented). How could infants possibly manage to home in on precisely these useful statistical patterns when you failed to see them, even after studying the speech sample and possibly thinking quite hard about it?

Though it might seem counterintuitive, there's a growing stack of evidence that a great deal of language learning "in the wild"—as opposed to in the classroom—actually does involve extracting patterns like these, and that babies and adults alike are very good at pulling statistical regularities out of speech samples, even though they may be lousy at actually manipulating math equations. As we'll see later, sensitivity to statistical information applies not just to segmenting words from an unfamiliar language, but also to learning how sounds make patterns in a language, how words can be combined with other words, or how to resolve ambiguities in the speech stream. The reason it doesn't feel intuitive is that all of this knowledge is *implicit* and can be hard to access at a conscious level. For example, you may have done reasonably well in identifying the "words" in the listening portion of Web Activity 4.3, even if you had trouble consciously identifying what information you were using in the analysis task. (Similarly, you may have had easy and quick intuitions about the phonotactic constraints of English but worked hard to articulate them.) That is, you may have had trouble identifying what it is you knew and how you learned it, even though you *did* seem to know it. It turns out that the vast majority of our knowledge of language has this character—throughout this book, you'll be seeing many more examples of a seeming disconnect between your explicit, conscious knowledge of language and your implicit, unconsciously learned linguistic prowess.

Being able to track transitional probabilities gives infants a powerful device for starting to make sense of a running river of speech sounds. It frees them up from the need to hear individual words in isolation in order to learn them, and it solves the problem of how they might build up enough of a stock of words to serve as the basis for more powerful generalizations about words—for example, that in English, words are more likely to have a trochaic stress pattern than an iambic one, or that the consonant cluster "ft" can't occur at the beginning of a word, though it can occur at its end. Ultimately, these generalizations, once in place, may turn out to be more robust and useful for word segmentation than transitional probabilities are. At times, such generalizations might even conflict with the information provided by transitional probabilities. Eventually, infants will need to learn how to attend to multiple levels of information, and to weight each one appropriately.

Is statistical learning a specialized human skill for language?

We've now seen some of the learning mechanisms that babies can use to pull word-like units out from the flow of speech they hear, including keeping track of various kinds of statistical regularities. Now let's step back and spend a bit of time thinking about how these learning mechanisms might connect with some of the bigger questions laid out in Chapters 2 and 3.

Much of Chapter 2 focused on the questions of whether language is unique to humans, and on whether certain skills have evolved purely because they make efficient language use possible. In that chapter, I emphasized that it was impossible to think of language as a monolithic thing; learning and using language involve an eclectic collection of skills and processes. Since we've now begun to isolate what some of those skills might look like, we can ask a much more precise question: Do non-humans have the ability to segment speech by keeping track of statistical regularities among sounds?

If you found yourself surprised and impressed at the capacity of babies to statistically analyze a stream of speech, you may find it all the more intriguing to learn that as a species, we're not alone in this ability. In a 2001 study, researchers led by Marc Hauser replicated the earlier studies by Saffran and colleagues; their subjects were not human infants, however, but cotton-top tamarins, a species of tiny monkey (**Figure 4.3A**). The monkeys heard about 20 minutes' worth of the same artificial language used by Saffran and her colleagues in their human experiments—four three-syllable "words" like *tupiro* and *bidaku*, strung together in random order and with no pauses between syllables or word units. Afterward, the monkeys went through a test phase, just as the human babies had, in which they heard a sequence of three syllables over a loudspeaker. As in the study with humans, these syllables could correspond to either a word from the artificial language (e.g., *bidaku*) or a part-word in which two syllables of an actual word were combined with a third syllable (for example, *tibida*). The researchers looked at how often the monkeys turned to face the speaker when they heard the test stimulus. They found that the monkeys distinguished between the two types of syllable sequences and oriented toward the speaker more often when they heard a part-word than when they'd heard a word (**Figure 4.3B**).

These results address the question of whether only humans are equipped to learn to chop up speech streams by paying attention to statistical regularities: apparently not. But the experiment doesn't fully get at the question of whether this statistical ability is exclusively enmeshed with our capacity for language. Cotton-top tamarins do have a fairly complex system of sequential calls, and maybe both their system of vocalizations and their statistical abilities reflect precursors of a fully linguistic system like ours. To better test for the connection or disconnection between statistical learning and language, it might make sense to study a species that shows no signs of having moved toward a human-like system of communication. Juan Toro and Josep Trobalón (2005) did just this in their study of speech segmentation in rats, using the same artificial language that previous researchers had used with human infants and cotton-top tamarins. They found that rats, too, were able to use statistical regularities in the speech to learn to differentiate between "words" and "non-words" of that language.

This suggests that picking up on statistical cues may be a very general cognitive skill—one that's not monopolized by species that have language or language-like communication systems, and one that might be useful in domains other than language. If that's so, then we should find that humans don't just pull this trick out of their hats for the purpose of learning language, but that we can also make use of it when confronted with very different kinds of stimuli. And indeed, this turns out to be true. Humans are capable of picking up on

(A) (B)

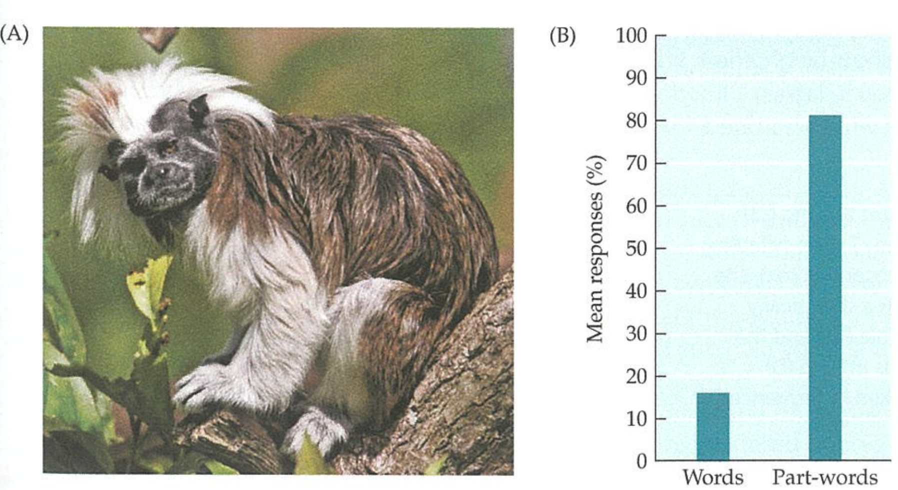


Figure 4.3 (A) An adult cotton-top tamarin (*Saguinus oedipus*), a species of Old World monkey. (B) The mean percentage of trials for which the tamarins oriented to the stimulus by turning to look at the speaker. "Words" in the familiarization phase were *tupiro*, *golabu*, *bidaku*, and *padoti*. During the test phase, the monkeys heard either the test words *tupiro* and *golabu* or the part-words *tibida* and *kupado*. (A © Danita Delimont/Alamy; B after Hauser et al., 2001.)