So the overall picture is that there is a partly specialized Universal Musical Grammar, which, like language, is a species-specific adaptation. An adaptation for what?—That's not so clear. It's easy to see what evolutionary advantage is conferred by having language, but it's hard to imagine how music does any good for our survival as a species. I consider this a real puzzle: Why should there be such a thing as music among our abilities?

In fact, music presents a double puzzle for evolution. First, there is the more general puzzle of why human nature should at all include such a thing as aesthetic pleasure, which we can experience in response both to natural phenomena such as landscapes and bird song and to human creations such as literature, art, music, dance, and food. Aesthetic pleasure doesn't enhance survival of the species, it only makes us feel good. But we undeniably have aesthetic experiences, and they are a primary reason behind our sense that it diminishes us to view people as mere machines.

The second and more specific puzzle is why there should be such an aesthetic object as *music*. All our other sources of aesthetic pleasure make use of materials already present in experience: the visual arts make use of visual perception, literature (including poetry and drama) makes use of language and our understanding of social situations, dance makes use of our preexisting sense of movement, and artistic preparation of food makes use of the preexisting sense of taste. Initially we might think that music just uses preexisting materials from auditory perception. But in fact, as I pointed out a moment ago, it's more specialized than that. Not any old collection of sounds can be made into music (except perhaps music of the John Cage sort). So why should the somewhat specialized principles of Universal Musical Grammar be present in our genetic heritage, when their only apparent use is for organizing aesthetic objects?

Having no coherent answer to these puzzles, I'll go on to the Argument for the Construction of Experience. This one is easy. We speak of music as happy or sad, dignified or poignant or sensuous. But where are these feelings? Music can't have feelings, it's nothing but sound waves! Rather, it's clear that the feelings are emotional responses evoked *in us* by the sound waves. That is, the emotional content that we experience in music must be constructed by our minds—it's not out there in the physical signal that strikes our ears. (Just as Sam's experience of the word "tree" in Chapter 12 is a product of his perceptual processes.)

Some of this emotional content may be fairly immediate and primitive. Fast loud music is often exciting and soft slow music is

often soothing, probably because of rather simple effects on the autonomic nervous system. (Note well, though, that even these "simple" effects are actually not so simple: it still takes substantial brain processing to convert speed and volume into autonomic effects.) On the other hand, in order to experience the music of Haydn as witty, we need a rather sophisticated grasp of the conventional melodic, harmonic, and rhythmic patterns in Haydn's style, so that we can appreciate how he uses these patterns to make jokes.

But in either of these cases, we experience *in the music* some character of excitement or wit that in fact comes from *within us*, that is actively but unconsciously constructed by our brains, using principles of auditory perception and musical grammar. That's the point of the Argument for the Construction of Experience.

We see, then, that language is not splendidly isolated among human mental capacities. All its basic characteristics are mirrored in our ability to understand music.

Vision

Next I'd like to show that vision, too, is governed by a mental grammar. What could this mean? Don't we just see things out there? I'll try to make clear what the problem is, drawing on a long and rich tradition of research on visual perception.

Remember in Chapter 2, when we had Harry looking at a tree, and we represented what he saw by drawing a little tree inside a cloud in his head? Here it is again.

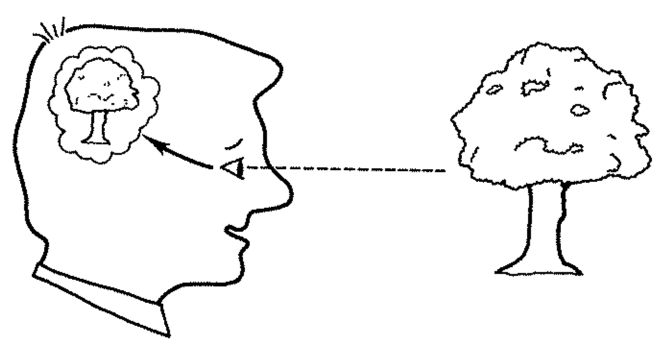

Figure 13.2 *The way we think about Harry seeing a tree*

This is the way we normally think about seeing: we think of producing inside our heads an image of the thing being seen. But let's think about it more closely. If we cut open our brains, we don't find any little images in there, any more than we find consonants and vowels.

And if there were little images in there, what good would they do? Who would look at them? There's no little person inside the head who sees the images. If there were, who would look at the images inside *that* person's head?

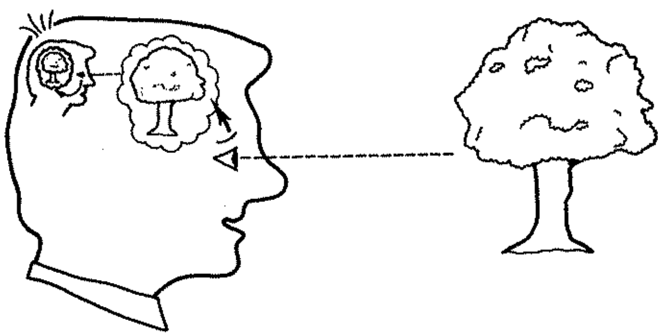

Figure 13.3 *Who looks at the image of a tree inside Harry's head?*

We might be tempted to answer that the brain *interprets* the images in the head. However, the brain can't interpret the images by *seeing* them—it doesn't have visual organs.

We have to think about the process somewhat differently. When light strikes the eyes, the lenses focus it to produce images on the retinas. But nobody looks at those images either. Rather, the retinas convert the light into patterns of neural impulses, and it's nothing but neural impulses from there on out.

Some areas of the brain reproduce the spatial arrangement of the neural impulses coming from the retinas; these areas are often said to have "retinotopic maps." But again, these maps aren't images as such; they're not at all like pictures in the head. They're merely more patterns of neural impulses, and they have to be processed further to arrive at what I'll call *visual understanding*.

To see what I mean by visual understanding, let's start with a very simple example, Figure 13.4.

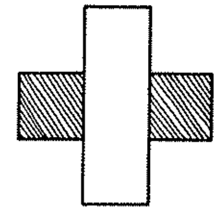

Figure 13.4 *Does the shaded rectangle go behind the white rectangle?*

This looks like a white vertical rectangle in front of a shaded horizontal rectangle. But think carefully about the physical basis for this perception. There is indeed a white vertical rectangle, but there's no single shaded horizontal rectangle behind it—only two shaded rectangles on opposite sides of it. Nevertheless, our brains construct an experienced configuration in which the two shaded rectangles are unified behind the white rectangle—in a place where the eye can provide no evidence.

Moreover, the brain constructs the *simplest possible* way of unifying the two shaded rectangles. It connects them with a part that is bounded by straight lines and homogeneously shaded, as in Figure 13.5:

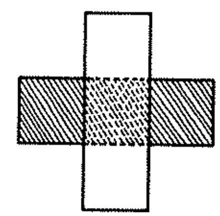

Figure 13.5 *The way we see Figure 13.4*

Wouldn't it be a little surprising if I removed the white rectangle and revealed one of the things in Figure 13.6?

Of course, *actually* there's nothing at all behind the white rectangle except the page. It's only a drawing. Yet our brains construct the "unseen" part completely reliably and with no effort—we don't *guess* that the shaded rectangle goes behind the white one, we *see* it, as if directly.

Figure 13.6 *Two ways we don't see Figure 13.4*

Of course I see it! I'm used to seeing things behind other things, and that's what they look like.

Right. But seeing a fence behind a tree trunk doesn't look exactly like Figure 13.4, and seeing a car behind a flagpole doesn't look exactly like Figure 13.4, and seeing a belt behind a belt-loop doesn't look exactly like Figure 13.4. All of these are instances of a common *pattern* of one object partially occluding the view of another object. The pattern occurs in indefinitely many different ways in daily experience.

In order for us to use this pattern to understand Figure 13.4, it has to be there somewhere in our brains. And it has to be different from a visual image (or a retinotopic map), because it is more general than any image can be—it applies to vast numbers of different situations. It is a tool with which we *comprehend* visual situations. In addition, it contains a part that is not present in any of the visual situations it helps us comprehend, namely the "virtual contours" indicated by the dotted lines in Figure 13.5.

Let's look at some more examples of patterns.

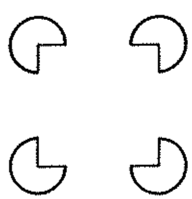

Figure 13.7 *Is there a square in the middle?*

Everybody sees Figure 13.7 as four circles, with a sort of "invisible square" in front that occludes part of each one. Of course, there's no real invisible square there, just blank paper, but, still, that's the way it looks. How is it that we see Figure 13.7 this way? It isn't as though we're are "used to" seeing invisible squares! And again, we can't see Figure 13.7 just any way we want to—it's invariably interpreted like the left-hand configuration in Figure 13.8, not like the other possibilities.

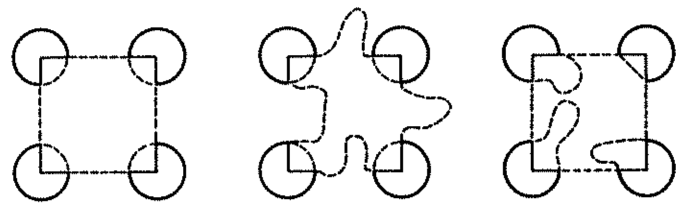

Figure 13.8 *The way we see Figure 13.7, plus two ways we don't see it*

What appears to be going on—very roughly—is that the brain unconsciously finds the simplest way to construct an interpretation of the visual field out of known patterns. In Figure 13.7, it's apparently simpler to see the "Pac-Man" shapes as occluded circles, and the occlusions adding up to a square—even an invisible square—than to see four isolated "Pac-Men." But to do this, the brain must be making choices among possible interpretations, rating them against known patterns.

Another example. It's possible to see Figure 13.9 in two different ways: (1) as a black square with a white disk in front of it; or (2) as a black square with a circular hole in it, through which can be seen the white background behind the square.

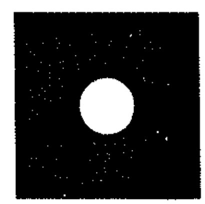

Figure 13.9 *A black square with a disk on it, or a black square with a white hole in the middle*

Now notice that even though there are are two ways of seeing Figure 13.9, it produces only one pattern on your retinas. This again shows that seeing involves more that just recording the patterns of light: it requires interpreting the patterns in principled ways.

In this case the principle is, very roughly, that a contour (or boundary) in the visual field has to separate the inside of a region from the outside background in a consistent way. (For convenience, I'll call this the Contour Principle.) For instance, in Figure 13.9 the square contour is understood as separating the inside of the square region from the outside background. But the circular contour can be understood in two ways. It can be understood as separating the inside of a disk from the outside background, in which case the disk's background is the square region. Or it can be understood as separating the inside of the square region from the outside background, in which case the round region is seen as a hole in the square, as part of the background that also surrounds the square. In other words, the Contour Principle allows two consistent analyses of the visual field.

This same principle accounts for the strangeness of some of the "impossible figures" that appeared in Chapter 4. Figure 13.10 is one of them, the "snakes."

Figure 13.10 *"Snakes": an impossible figure*

Intuitively, the oddness here has to do with insides and outsides: somehow the inside of the top snake merges indistinguishably into the outside of the bottom snake, and vice versa, violating the Contour Principle. In other words, a very general principle accounts for both the ambiguity of Figure 13.9 and the anomaly of Figure 13.10.

So what? What do these silly illusions have to do with the real world?

Here's what: as far as we can tell, the processes responsible for these unusual perceptual experiences are exactly the same as those responsible for perception of the ordinary world. It's not as though we turn on a special set of processes when we're doing a psychological experiment. So if we unconsciously apply principles such as

the Contour Principle in the sorts of situations I've just illustrated, we must be doing so in ordinary seeing as well.

It should be clear that I'm going through these examples in order to construct a visual analogue of the Argument for Mental Grammar. Our being able to comprehend an unlimited number of visual situations depends on our having in our brains a set of unconscious patterns and principles that can analyze a visual image and create an interpretation. We can call this set of patterns and principles a visual grammar—though it will probably bear little substantive resemblance to a linguistic grammar. (I don't care much whether we actually use the term "grammar," which some readers may find strained. The point is that the principles in the brain abstract away from particular visual images and help to organize what we see.)

Skipping the Argument for Innate Knowledge for a moment, it is not hard to see how the Argument for the Construction of Experience applies to visual understanding. The "invisible contours" in Figures 13.4 and 13.7, the visual ambiguity in Figure 13.9, and the strangeness of Figure 13.10 are certainly part of our visual experience. But there is no sense in which they are actually in the physical world—any more than words are physically present in an acoustic signal. Yet we can't help experiencing these figures the way we do: unconscious processes in the visual system make use of stored visual patterns and principles to construct an optimal understanding of the optical signal presented to the eyes. Every time we see something behind something else, and every time we see something standing out against its background, we are making use of principles of visual understanding illustrated here. And many more such principles, which I don't have space to mention here, have been studied in great detail by experimental psychologists.

Now let's go back to the Argument for Innate Knowledge. How do we acquire the patterns and principles of visual understanding? As usual, the right balance must be found between learning from the environment and using innate resources that make that learning possible.

Now environmental input certainly plays an important role in learning to identify particular kinds of objects. Someone who has never seen a car won't be immediately able to tell a Fiat from a Volkswagen, and someone who has never looked under the hood of a car won't be able to tell a generator from a carburetor. To draw on a parallel with language, this is sort of like learning a "visual vocabulary." We must carry around in our heads representations of thousands of objects whose appearance we are familiar with.

But recall that, in the case of language, learning a vocabulary is not sufficient for learning the language. Many principles of mental grammar are abstract; they determine how the vocabulary can be constructed from smaller parts such as distinctive features, and how items of the vocabulary can be combined into larger utterances. In the case of vision, the parallel questions are: What are the abstract pieces available for constructing items in the visual vocabulary? And how can individual items be combined in order to understand the full configuration of objects in the visual field (including, for instance, objects behind others)?

The evidence that has accumulated suggests that an overwhelming proportion of the basic principles of visual perception (for instance, the Contour Principle) are innate. For one thing, the visual system is scattered over numerous highly specialized areas or modules in the brain. It is unlikely that such differentiation in the brain is the result of learning, especially since (1) it is quite consistent across individuals and (2) the specialized areas are paralleled in other species such as monkeys. Rather, this points to a genetic predisposition for the brain to develop in a certain way.

Evidence has also arisen from infant studies. Infants' visual acuity isn't very good. And they don't have very much visual vocabulary. *But*—to the extent that they can be tested, infants as young as a few hours show command of some of the principles of visual perception. For example, suppose we show an infant Figure 13.9, and then gradually change it in one of the two ways shown in Figure 13.11.

The infant won't react very strongly to the change in Figure 13.11a: it just looks as though the circle moves to the left a little bit. But there will be a stronger reaction to the change in Figure 13.11b, which violates the integrity of the shapes: half of the square and half of the circle have moved together. (The strength of reaction is measured in terms of how long the infant looks at the display and how hard the infant sucks on a pressure-sensitive nipple.) That is, as early as we can measure it, infants are already making use of principles of visual understanding.

To be sure, as I've mentioned a couple of times before, it's necessary to have visual input in order for the visual system to develop properly. If nothing else, the neurons of the visual system require stimulation in order to acquire and tune their connections. But such stimulation is more like nourishment than like teaching. The baby comes to see the world in the way it does because of the brain's genetic disposition to develop in a certain way in response to visual stimulation.

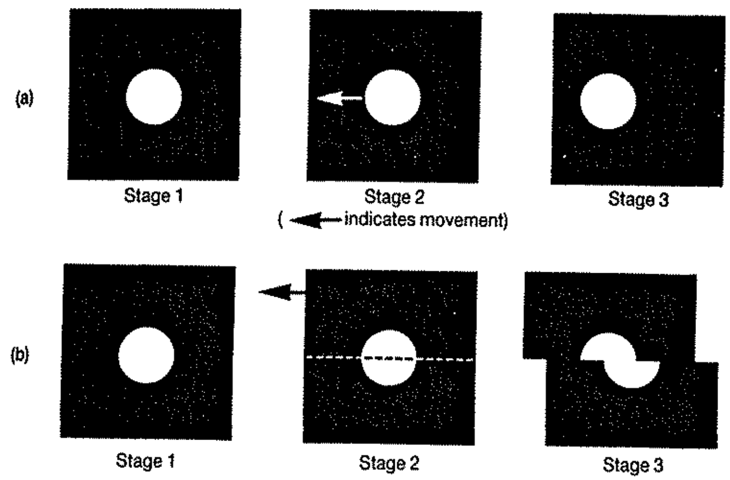

Figure 13.11 *Infants find the change in part a. unremarkable, but are surprised by the change in part b.*

To sum up, we now have musical and visual analogies to the three Fundamental Arguments for the nature of language.

The Argument for Mental Grammar: In each of these domains, our ability to make sense of novel stimuli is supported by a set of abstract patterns that are specialized for that domain.

The Argument for Innate Knowledge: In each of these domains, we learn the patterns we do in part because our brains are genetically programmed with substantial aspects of these patterns in advance. Learning is not "soaking up" of patterns, but rather active tuning and elaboration of the innate specialized mental "proto-patterns." That is, a great deal of nature lies behind learning through nurture.

The Argument for the Construction of Experience: Our experience and understanding of stimuli in each domain is actively constructed by our minds, making essential use of the abstract mental patterns specific to that domain.

Making visual understanding conform to ideal forms

I want to conclude this chapter by illustrating a phenomenon that is absolutely central in talking about visual understanding. I bring it up because it connects to our discussion in Chapter 12 about the possible extent of variation in interpretation, and because it will play an important role in the next two chapters.

Consider the configuration in Figure 13.12. What is it?

Figure 13.12 *An A or an H?*

Seen in the context of Figure 13.13a, it's the one that's not an A; in the context of Figure 13.13b, it's the one that's not an H.

Figure 13.13 *a. It's not an A b. It's not an H*

But in the context of Figure 13.14, it's a funny H in the first word and a funny A in the second word.

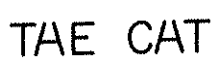

Figure 13.14 *It's either an A or an H, given the proper context*

So, depending on the context, it can be an A or not an A, an H or not an H. How can this be?

For a very rough approximation—which is all we'll need here—it seems that elements of the visual vocabulary, such as letters of the alphabet, are stored in the brain in terms of sets of criteria. For example, the criteria for a capital A are two slanted sides that meet at the top, plus a horizontal crossbar that joins the two sides in the vicinity of their midpoints. A configuration that meets these criteria is seen as an "ideal" or "stereotypical" A.

However, these criteria are not entirely rigid. Figures 13.15a and 13.15b push the crossbar up and down. Figure 13.15c has a crossbar, but it's not horizontal. Figure 13.15d has slanted sides, but they don't meet at the top (it's the same as Figure 13.12). Figure 13.15e doesn't have slanted sides; the sides meet at the top by virtue of another crossbar. All of these are acceptable as capital A's, especially if we put them between a C and a T.

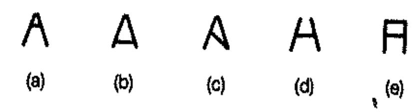

Figure 13.15 *Configurations that deviate minimally from an ideal A are acceptable A's, especially given context*

Still, we can't go too far. Look at the configurations in Figure 13.16, which fail two or more of the criteria. Figures 13.16a and 13.16b have nonslanted sides and the crossbar in the wrong place; Figure 13.16c has sides that don't meet and a nonhorizontal crossbar. These are unlikely to be seen as capital A's, even if they are put between a C and a T. (Try it!) Figure 13.16d has nonslanted sides that don't meet, so it falls under an alternative set of ideal criteria, namely those for an H. As a result, it definitely can't be seen as an A.

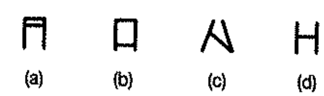

Figure 13.16 *Configurations that deviate by two criteria from an ideal A are not acceptable as A's, even with context*

These little experiments show us that our unconscious criteria for elements of the visual vocabulary tolerate a certain amount of deviation, but not too much. The way deviations are tolerated in these examples is altogether typical:

1. A configuration that meets all the criteria for some category is seen as an "ideal" or stereotypical example of the category, independent of context. For instance, an ideal capital A will still look like a capital A if put someplace where it makes no sense, say between a T and an E—or even between a cat and a mouse!

2. A configuration that violates too many criteria is seen as not belonging to the category, no matter in what context. For instance, no amount of context can make the configurations in Figure 13.16 look like capital A's.

   Corollary: A configuration that not only violates several criteria for a category but also satisfies the ideal criteria for another category (for example, Figure 13.16d) is seen especially readily as not belonging to the first category.

3. A configuration that satisfies most but not all of the criteria for a category can be seen as a marginal or deviant member of the category. Such a judgment, unlike the ones above, *is* sensitive to context. For instance, in Figure 13.13a, where the context involves choosing the configuration that is not an A, Figure 13.12 is judged not to be an A. But in the second word of Figure 13.14, where the context demands making sense of a written word, Figure 13.12 *is* judged to be an A.

   Corollary: A configuration that "falls in the cracks" between the criteria of two categories can be seen as belonging to either one, depending on the context, as seen in Figure 13.14.

In other words, context can affect the way we see things—but not all the time. Context plays a role just when the things we're looking at *almost* satisfy the "ideal" criteria of some category; in such a situation, different contexts can either coerce the configuration into the category or exclude it from the category.

There is a final wrinkle in this story. We don't tolerate marginal members of a category without some associated cognitive stress. Suppose I ask, "Are you perfectly happy with Figure 13.14 as it is—would you like to change anything?" (Maybe we also have to suppose that we haven't yet been through all this discussion.) You would probably prefer to make the deviant configuration more like an ideal H in the first word and more like an ideal A in the second. (It's this impulse that makes people want to straighten a picture

hanging crookedly on the wall.) That is, ideal configurations in some sense make us happier than marginal ones.

It stands to reason that our perceptual systems have to be equipped to deal with marginal members of a category. The world out there isn't populated exclusively by things that conform completely to our ideals. Things are always going to be falling in the cracks, and if we couldn't deal with them, we'd be in trouble.

On the other hand, the specific way we respond to a marginal configuration is not so self-evident: we are more sensitive to context, and we often experience a certain degree of uneasiness and/or a desire to "improve" the configuration toward the ideal. Unless we work through an analysis of the sort we've just done, these aspects of the response are largely unconscious.

I think it is an important fact about human nature that we have this particular complex way of fitting the things we perceive into categories (and it is a fact that has eluded most philosophers and psychologists). It's not clear that we had to be built this way, but so it is. In the context of innocuous examples like Figure 13.12, it doesn't seem very important, but as we'll see, its effects in our lives are pervasive.