In the previous chapter I introduced the idea of nonconceptual representation as a form of perceptually derived representation that does not involve concepts and therefore does not enter into beliefs and thoughts (and probably also memories, since they typically involve inferential reconstructions). Nonconceptual representations have been widely discussed in philosophical circles for a number of reasons. One reason concerns the basic problem that we have encountered in several places earlier in this book: the need for a way to get information from its distal causes through proximal effects (e.g., the retinal image) to perceptual beliefs, the latter being conceptual by definition. The interface is thought to involve a type of information-bearing state whose content is more concrete and detailed than is the content of beliefs, but which nonetheless qualifies as being a form of representation because it carries information about some state of affairs in the world. This form of representation does not represent the visual scene in terms of conceptual categories but is more iconic and uninterpreted.

The primary evidence cited in favor of such a form of representation is the disparity between appearances and beliefs (the hallmark of perceptual illusions) or between the panoramic, uniformly fine-grained nature of our conscious visual experiences and the relatively abstract, categorical, and variable-grained nature of our thoughts, beliefs, and recollections. Moreover, the mapping between the proximal stimulus and how the scene appears to us is not fixed, which itself needs to be explained. For example, when we look at an ambiguous figure such as a Necker cube (shown in the left panel of figure 4.1), something changes over time as we watch, and that something is not the physical stimulus, nor is it generally thought to be our beliefs about what we see. It is what we usually refer to as the appearance of
the figure or how we consciously experience it. Since the content of our experience seems to be distinct both from the proximal stimulus (the optical projection of the figure on our retina) and from what we believe about the figure, this suggests that we need another vehicle of representation for that type of content.

When we examine a visual scene, the content of our experience is very different from what we know to be the information that enters the visual system. The evidence is clear that the incoming information is highly incomplete and has a narrow scope (it is literally a moving peephole no more than about two degrees of visual angle) compared with how we experience it (this point will be discussed later in connection with the special case of the experience of space). But the experience also seems intuitively different from what we might plausibly capture in terms of the vocabulary of concepts we are likely to have. The argument from the richness of experience compared with the relative poverty of our conceptual resources depends both on how we characterize experience and on what we think are the conceptual resources of the mind. But even without considering the fine points of what are reasonable bounds on our conceptual apparatus, it seems clear that we are unlikely to have as many distinct concepts for, say, colors, as there are colors that we can discriminate. Certainly if we consider the number of color terms in known languages we find that the number is actually very small (languages have no more than about 11 monolexemic words; see Berlin and Kay 1969). Yet we can distinguish well over a million different colors (Halsey and Chapanis 1951), so it is unlikely that we encode each of these as a separate concept or code. (Of course there are many more concepts than there are words. But since each word corresponds to some concept, it seems reasonable that each concept is at least a potential word in some language. A million different colors is far more than the total human vocabulary, so it seems unreasonable to suppose that every discriminable color could have a corresponding concept.)

Notwithstanding such plausible arguments for nonconceptual representations, there are several questions that need to be considered and several tacit assumptions that need to be exposed before the hypothesis that there is a nonconceptual representation of the sort generally accepted in philosophy can be taken as established. The most contentious of these is the assumption that the content of this nonconceptual representation is the same as the content of conscious experience. This view assumes that the content of conscious experience corresponds to a level of representation in an information-processing or functional analysis of the cognitive system. In other words, it assumes that the content of conscious experience
is a natural kind for purposes of psychological explanation. I take up this and other issues in the next few sections.

4.2 The Role of Conscious Experience in the Study of Perception and Cognition

Cognitive science, and particularly vision science, has had a deeply ambivalent relation with the phenomenon of conscious experience. On one hand, the way things appear or what they look like has always been an important, if not the primary, source of data, at least for vision science. When one thing looks bigger in one condition than in another or looks to be moving faster under one condition than another, or when colors appear different under one lighting condition than another, these are considered primary data to which theories of vision are expected to respond. On the other hand, the content of a person’s experience has also proven to be one of the most misleading sources of evidence, because it is not neutral with respect to the theories that the subject holds, be they scientific or folk theories. Moreover, which explanations appear most natural is highly sensitive to the way we describe our experiences, and conversely, the way we describe our experiences (even to ourselves) depends to a large extent on what tacit theories we hold.

The way we describe our perceptual experiences often carries with it the implication that the content of the experience itself explains observed phenomena—that the occurrence of experience X causes experience Y which then explains some ensuing behavior. There are more or less benign versions of this sort of what might be called intentional causation (also sometimes referred to as psychological determinism; Hochberg 1968). An important and essentially irreproachable version of this thesis is the appeal to the tight coupling that holds between how a part of a scene is experienced and how other parts of the scene tend to be experienced (see Epstein 1982; Rock 1997). For example, if you see the edge of a Necker cube marked x (in figure 4.1) as being the top front edge of the figure, then you are likely to see the face of which it is a part as the top face (as in the second panel); but if, instead, you see the edge marked y as being the top front edge, then you will see the face that it bounds as the top edge of the figure and the appearance of all the other edges will change so the interpretation of the figure remains coherent. When the percept changes (as it does in ambiguous figures such as in the first panel) the couplings force the interpretation of related parts to change accordingly. This fact has been the basis for a successful technique in computer vision called constraint propagation (see the
work on the “blocks world” that culminated in the successful system devised by David Waltz, described in, among other places, Pylyshyn 2003, chap. 3; Waltz 1975). (It is also the basis of an approach to models of reasoning using constraint satisfaction; Tsang 1993.)

When we speak of labels on the representation of a scene, or indeed when we speak of what a pattern is seen as, we are speaking of the contents of a perceptual representation. In most cases such contents are assumed to be conscious; hence we are speaking of the contents of a perceptual experience. But what exactly is the content of a conscious perceptual experience? The content of a belief is relatively clear, because beliefs are individuated in part by their contents—that is, we identify a belief by what it is about, or we may treat beliefs that are about the same thing as the same belief (I am ignoring for now the fact that beliefs can differ in ways other than in their content, e.g., their form). But what about the content of a conscious experience? The situation here is not at all straightforward.

4.2.1 The Contents Question

The role that conscious experience can play in vision science depends on our understanding of what such experience reveals. Our first impression is that what conscious experience reveals is both private and obvious to the person who experiences it: if you see something or other, then what the experience reveals is just what you see. But if you try to say what that is, you find it is far from straightforward. Indeed there is a long chapter in the

![Figure 4.1](image-url)

The way we see edges is intimately connected with the way we see the faces those edges bound. These appearances (or “ways of seeing”) form a tightly coupled system. If our percept of one part changes, the appearance of other parts change systematically in order to maintain the coherence of the whole. Panels 2 and 3 show the two versions of the ambiguous figure in the first panel. Notice how the interpretation of an edge is connected to the interpretation of the faces it bounds and that both interpretations change together when the ambiguous percept in the first panel “flips.”
history of psychology of the late nineteenth and early twentieth centuries, where that question was at the forefront of discussions of what psychology is and what it should be (Titchener 1912; Washburn 1922). The introspective method was taught as an objective way to study the contents of conscious experience by turning the mind on itself in a disciplined analytical manner, freeing it to report its own conscious contents, as opposed to reporting properties of the object of our perception (inadvertently being influenced by what one knew about the objects of perception was known as the stimulus error). In the end the method failed to provide the foundations for a science of conscious experience, although conscious experience itself continues to be a growing concern in both psychology and philosophy. It is not my purpose here to discuss introspection or to look at the fascinating history of the study of consciousness in psychology. I wish only to point out some of the problems raised by the use of conscious contents as a source of evidence for building theories of perception.

There are two sets of questions about our conscious awareness. One is what might be called the objective scientific question: What are we entitled to conclude about perception from certain perceptual experiences we have? The other, logically prior question is what the content of our perceptual experience is: What is the thing about which questions of interpretation can be raised? This question is independent of methodological issues concerned with how one should interpret reports of “how something looks.” The question even applies to one’s understanding of the content of one’s own conscious experiences. The question—what do I experience when I look at this stimulus?—is fraught with problems. One might reasonably take the position that to ask what we experience is already to take a theoretical stand, namely that the content of the experience is transparent to the person who experiences it; it is part of what Sellars (1956) called “the given.” There has been a considerable amount of philosophical discussion of this question. The assumption that one is the infallible arbiter of the content of one’s conscious experience has serious problems, particularly if one takes

1. For original writings from this era, and earlier, see the interesting website “Classics in the History of Psychology” maintained by Christopher D. Green at York University, Toronto, Canada: http://psychclassics.yorku.ca/index.htm/.
2. The Association for the Scientific Study of Consciousness (ASSC) has become a major scholarly society with a large annual meeting where, among other things, the relation between brain and consciousness is discussed and neuropsychologists report various fascinating brain damage syndromes in which patients exhibit disconnects between behavior and conscious contents. See http://assc.caltech.edu/index.htm/.
it to have conceptual content—that is, to be the experience as of something or other.

Consider what we experience when we look at a scene. Suppose I look at the wall of my room; what is the content of my conscious experience? If I try to describe what I am experiencing I find myself describing the things in the room and their visible properties (such as color, texture, or location). My conscious experience is the experience of the things that I perceive (possibly also what these things remind me of and what feelings they may arouse, but let’s confine ourselves to the perceptual experience itself). How exactly can I describe what I experience? Does it consist of all the properties that are in the scene, or only those that I notice? Do I experience only what is present in the incoming information, or does my experience also include what the visual system “fills in” and what I infer? For example, do I experience the uniformity of the color and lightness of the wall which, as it happens, I know is not in fact uniformly illuminated? Is the uniformity of lightness and color constancy that I am describing an inference or a direct experiential content? The lightness you perceive is known to depend on your perception of the location and arrangement of the surfaces in question (Gilchrist 1977). Also in a typical scene it is rare that I see all of any object because most things will be at least partially occluded by others, even though I do not notice these occlusions unless they are brought to my attention. It thus appears that the content of my experience includes postconstancy and post-filling-in information, and therefore, my experience relies on more than just the information coming from my eyes.

Let’s continue with this example. There is a picture on the wall. Do I experience it as a picture or do I experience what the picture depicts—or both? There is also a calendar on the wall which I see and which is therefore part of my experience. Since I forgot to change the page at the end of last month it shows the wrong month. Do I experience it as a calendar, and do I experience it as showing the wrong month? Ordinary informal talk is unclear on such issues. If I do experience it as a calendar showing the wrong month, then other people looking at the wall are unlikely to have the same experience as I do unless they know what I know about the calendar. I look down at my desk and see a sculpture that serves as a paperweight (or perhaps vice versa, depending on how you feel about such pieces). I see it as a three-dimensional object that has not only a front but also a back and sides and parts that are hidden or occluded by other parts. Do I experience only the front, or only the parts that are not occluded, or do I experience the back and the hidden parts as well? I do not see the back in the sense that I receive no optical information from it, so how can it be part of
my experience? Some writers in the Gibson “direct perception” tradition claim that we see the back of 3-D objects as well as the front because both are part of the experience of what Alva Noë (2004) calls the perceptual presence, which is different from our knowledge of the back of a perceived object. This may actually be the more common view. For example, Block (1995) says that if you are looking at a row of buildings and then find out that they are mere fronts of a movie set, the content of your conscious experience. According to Block, the visual experience as of a façade is different from the visual experience as of a building. If that is the case, then conscious content is clearly cognitively penetrable, which affects the role it can play in a perceptual theory. It cannot, for example, serve as the input to vision, as what some philosophers call the “given.”

Many also say that what we see, our perceptual experience, is viewpoint independent, which implies that we represent it as a solid without giving special status to the surface that faces the viewer. By contrast others insist that what we see is just the front of 3-D objects, and thus that what we see clearly depends on our viewpoint. David Marr’s theory of vision explicitly provides a middle ground by proposing what he called a “2½-D sketch,” which is a representation in depth of only the visible surfaces. Which of these is the content of our experience? For Gaetano Kanizsa (Kanizsa and Gerbino 1982), the perceptual reconstruction of occluded contours is an automatic and cognitively impenetrable stage in the process of seeing (I have also defended this view in Pylyshyn 1999). According to this view, what we experience when we see is not the incoming information but is a complex output of our early vision system together with some inferences, perhaps from other parts of the scene or perhaps from our knowledge and expectations of what is in the scene. Is all this part of our conscious perceptual content? It is certainly what we mean when we report what something “looks like,” so at least in the everyday sense it is part of our conscious content. Where do we draw the line? In Pylyshyn 2003 (chap. 1) I give examples to illustrate that the everyday nontechnical sense of “what something looks like” is very broad and includes visual puns of the sort popularized by Roger Price in what he called “droodles” (see http://www.droodles.com/).

In several thoughtful essays, Fred Dretske (e.g., Dretske 1993, 2006) adds to the perplexity for those who would appeal to the content of conscious experience in building theories of vision, by arguing that we may not always be aware of the content of our experience. That’s because, according to Dretske, there is a difference between being conscious of things and being conscious of facts. That one is conscious of something is itself a fact of which we may or may not be conscious. Dretske gives the example of
looking at a wall made of hundreds of orange bricks. Given enough time to scan the wall, does our experience include the experience of each of the bricks? Dretske claims it does, because if asked whether there was a blue brick among the orange ones we can confidently answer no. Dretske claims that this implies that we saw (and experienced) each of the bricks since the information that there was no blue brick depends on having been conscious of the properties of each brick. Yet if asked we might, quite reasonably, claim that we were not conscious of each of the bricks. According to Dretske that just shows that we need not be aware of the conscious content of our perception. Other philosophers have also spoken about the difference between phenomenal and nonphenomenal consciousness (e.g., Lormand 1996), thus further complicating the problem of using conscious contents for theory construction.

There are many examples of our being unaware of information that was readily perceivable and that, by other criteria, was in fact perceived. The question one might ask of each of them is whether they are cases in which we are not conscious of the information, or cases in which we are conscious of the information but were unaware that we were conscious of it. Examples include various cases of apparent functional “blindness.” One of the best-known examples is referred to as change blindness. In these demonstrations subjects are unable to report the change between two alternating briefly presented pictures even though the difference between the two pictures is clearly visible when attention is drawn to it (Simons and Levin 1997; Simons and Rensink 2005). Another example is inattentional blindness, in which subjects fail to see a clearly visible feature that occurs at precisely the point where they were visually fixated while they are attending to a more peripheral item (Mack and Rock 1998). Another such example that is extremely persuasive and puzzling involves watching a movie with several players who are passing a ball around while the subject is required to count the number of passes. In this example many subjects are unaware of a person dressed in a gorilla suit who walks right through the middle of the scene (Simons and Chabris 1999). These types of blindness appear to

3. In this example, however, judging that there were no blue bricks is likely an inference of the form: (1) If there had been one clearly visible blue brick I would have seen it; (2) I did not notice a blue brick; therefore (3) there was no blue brick. Thus it does not entail that information from each brick was perceptually (consciously) available, only that the perceiver believes that if there had been a blue brick he would have seen it. This is known in the computational inference field as negation as failure and is entailed by the closed world assumption that is part of the logic programming language Prolog. (See entry in Wikipedia.org.)
involve a failure of information to reach consciousness even though the information is in some sense clearly taken in since it is located on the fovea directly in view. In fact, Dretske (2006) claims that the lesson we should take from such examples is precisely that one can be conscious of something and at the same time be unaware that one is conscious of it.

The distinction between consciousness of things and consciousness of facts is similar in spirit, and might perhaps even be subsumed under, the distinction between phenomenal consciousness and access consciousness advocated by Ned Block (1995). According to Block, there are two functions of consciousness that should be distinguished: the purely phenomenal function (characterized as “what it is like to be in that state”) and the access function (characterized as states in which information is “poised to be used as a premise in reasoning, . . . the rational control of action . . . and speech,” Block 1995, p. 230). These are referred to as P-consciousness and A-consciousness, respectively. These two functions are not only conceptually distinct but also may involve different neural mechanisms (Block 2005). According to this view it is possible to have a vivid phenomenally conscious experience that does not “broadcast” information to other mental processes, and thus it is possible to have functional access to information that is accompanied by little or no phenomenal experience. Although Block often talks as though these were two forms of consciousness, he means a “phenomenal aspect” or “phenomenal content” in contrast to “representational aspect” or “representational content” of consciousness. Although these two aspects nearly always occur together they are conceptually separable, and in some cases empirically separated, as when one or the other is damaged by brain lesions. As we have already seen, various types of “blindness” demonstrate information access without phenomenal consciousness. Cases of phenomenal content without access (without representational content) are more difficult to find inasmuch as the best evidence for phenomenal content takes the form of verbal reports which ipso facto constitutes evidence for informational access. The best example may come from split-brain patients who are able to carry out certain tasks.

4. Rosenthal (2005, pp. 191–192) argues that it is hard to reconcile cases of P-consciousness without A-consciousness with the understanding of P-consciousness as “what it would be like” to be in that state, since there is no way to be in that state unless one is aware of it—i.e., without there being some A-conscious aspect we can use to identify those states. This debate is of interest for an understanding of the nature of conscious states, but it does not bear on the current point which is concerned with what we can learn about perception if we set “conscious content” as an explanandum.
Another quite different view of consciousness is provided by David Rosenthal (2005), who argues that consciousness consists of being the target of “higher-order thoughts” (HOTs) or of having noninferential (unmediated and typically unconscious) thoughts about one’s thoughts. This view, like other views about what it is to be conscious, deals with issues that are beyond the scope of the present discussion. I mention the HOT view, however, because although it is certainly very different from those of Dretske or Block, it does have room to encompass the distinction between conscious experience and awareness of conscious experience or between P-consciousness and A-consciousness. As long as you think that thoughts about thoughts are a real possibility and recognize that they (sometimes) underwrite conscious contents, you might consider cases where thoughts about thoughts do not yield conscious states, and also cases in which conscious states can arise from thoughts about other sorts of mental states besides thoughts (e.g., desires, acts of will). Because all three views allow for a certain degree of independence between qualitative experience and information-processing functions, all these options allow the possibility of being conscious of something without being aware of what you are conscious of, or of phenomenal consciousness without access consciousness and vice versa.

My point here is not to advocate a particular way of interpreting the notion of conscious experience, but merely to point out that while the content of experience is important for building theories of perception, it is encumbered with many problems. At the very least the examples above show that whether something is or is not part of the content of our experience is not self-evident, so experiential content is not something we can take at face value merely on the grounds that since it is your experience you alone are the authority on its content. In addition, there is no reason why you should be able to say what the theoretically relevant aspect of the experience is and, even worse, you can also get this wrong—as we will see in other examples I will provide below.

4.2.2 Conscious Experience and Public Report
Reports by subjects of what something looks like are even more problematic since what people report in an experimental setting is known to be affected by many factors, including what subjects think the experimenter wants (such compliance effects have been called experimenter demands), what they believe the task to be (which have been called task demands) as
well as subjects’ general beliefs and utilities. Every response requires making a decision that may involve weighing the costs (including embarrassment) of different sorts of errors, particularly errors of commission versus errors of omission. For example, work on subliminal perception or “perceptual defense” shows that accuracy in reporting whether one has “seen” a briefly presented word is different for taboo words than for neutral words equated for frequency of occurrence (Freeman 1955). These are typically not cases of subjects being disingenuous, but of making rational choices—choices that can be traced to processes described by models of decision making, such as utility theory and signal-detection theory. Sometimes signal-detection theory can separate contents from reporting biases in a fairly direct way by providing different measures for response criteria (the parameter usually written as $b$) and for sensitivity (the parameter $d'$) in experiments involving thresholds. This is done by taking into account not only the correct responses, but also the relative rate of errors of omission and errors of commission. If the subject has a bias to report seeing something independently of whether there was a signal, then both the hit rate and the rate of errors of commission will increase. Such a bias would be useful if the signal were present on most of the trials or if the utility of detecting the signal was high (e.g., if it signaled danger). If, on the other hand, the subject has a conservative bias, then the hit rate will be lowered but the rate of errors of omission will also increase. This sort of bias would be useful if the signal were present only rarely. These tendencies can be used to separate response bias from the availability of conscious contents (Snodgrass 2002). Experimental psychology has learned that sincere reports of conscious contents have to be evaluated in relation to other sources of evidence and in the light of developing theories.

Consider, for example, the problem of interpreting such findings as those reported by Wittreich (1959). A well-known illusion is that when people walk across the floor of a specially designed room called the Ames room (shown in figure 4.2) they appear to change in size.5 Wittreich confirmed

5. It does not reveal a magician’s proprietary trick to tell you that the room is actually distorted, having been constructed with one side much lower and shallower than the other. The design specifications are such that rays drawn from a peephole to every visual feature—i.e., corner and vertex (the room has windows)—bear the same visual angles to one another as they would have in a regular rectangular room. Thus when viewed through the peephole all the visual cues in the Ames room are identical to those that would have been available in its corresponding phantom rectangular room. Of course it is not possible to build a distorted room such that the illusion persists as the viewer moves inside the room, though this could (almost) be
this observation, but he also found that this did not happen when the people were well known to the observer—for example, the observer’s spouse—even if these people were accompanied by a stranger, whose size did appear to change! Notwithstanding the presumably sincere reports made by the subjects, there remains the question whether to interpret this finding as showing the malleability of judgments of the content of conscious experience, or of the operation of the visual system itself. The problem is not that subjects are disingenuous, but simply that the lines between what we report and what we believe with great conviction, as well as between what we report to others and what we report to ourselves, are not so clear. If, as many have supposed (e.g., Block 1995; Dennett 1991), part of conscious content is what vision (or imagination) reports to the rest of the mind,

done in an electronic virtual reality room (the reason for the qualifier is that VR displays cannot reproduce all cues exactly; in particular, since objects are not actually located at different depths but on the same 2-D surface, the eyes do not focus at different depths, which results in some conflicting cues).
then what it reports may be different from the information that it actually possesses. In other words, there may well be a partial dissociation between the content of our conscious experience and the information that is passed on to other stages in mental processing. Sometimes we can show this fairly directly by comparing measures from which the reporting bias has been mathematically factored out, as we do when we use the signal-detection measure $d'$ rather than percent correct. Such measures not only separate what information observers have from what they report to an experimenter, but also from what observers report to themselves—that is, what they are aware of. More often than not, general questions such as this are ultimately adjudicated according to whether a theory that takes certain observations at face value simply misses underlying (causal, functional) principles.

Focusing on the conscious contents of perception has also encouraged direct perception theories (such as those of Gibson 1979), which claim that perception allows us to directly access (“pick up”) information about properties of the world that are prominent in our experience, such as the property that things have of being suitable for certain purposes—from being eaten to being sat upon (suitability is referred to as having certain “affordances”). James J. Gibson has argued, quite reasonably, that we see not patterns of light and shadow and patches of color, but familiar things such as tables, chairs, and people. Moreover, we never see just the front surface of objects; we see entire objects and we see them as particular things, such as our car or our spouse, or as having certain affordances, such as being graspable or edible. Although the urge to shun visual representations led Gibson and his followers to embrace what is essentially a behaviorist position, they were right to claim that perception eventuates in the extraction of abstract properties rather than low-level sensory patterns (“sensations”). The moral of this observation should have been that what we see is a reconstruction of the properties of distal objects: we never experience the preconstancy proximal stimulus. But in direct realism theories (for various modern versions, see Smith 2003) this is not the moral that is drawn. Rather, these observations are taken to be an indictment of the view that perception begins with properties as described by physical science and constructs a representation of a scene (perhaps in some cases with the aid of inference from general knowledge). Instead, they are taken as support for the radical view that the world should be redescribed according to the categories of experience, which are assumed to be the starting point of perception; these are the categories to which perception is inherently attuned and which are “picked up” the way a tuning fork picks up the notes in its
immediate environment. To make this picture work Gibson also had to deal with the problem of misperception, which, in turn, led him the view that theories should be applied to perception in an “ecologically valid” environment (for a critical discussion of these ideas, see Fodor and Pylyshyn 1981; Pylyshyn 1984, chap. 6). Although it is not usually put in this way, it is the temptation to see the categories of conscious experience as the primitive bases for (or inputs to) perception that has been one of the siren calls of direct realism.

Notice that the position I have been describing in this book bears some similarity to Gibson’s. I too do not believe that we should take the starting point of vision (the nonconceptual first steps) to be sensations, if by sensations we mean consciously experienced colors, shapes, textures, and so on (or whatever the primitive sensations turn out to be). Rather, the starting point should be nonconceptual, in particular it should be nonconceptual demonstrative references to proto-objects or FINGS.

4.3 What Subjective Experience Reveals about Psychological Processes

In this section I will move quickly through some evidence showing that attempts to infer the nature of psychological processes from the evidence of conscious experience have led us into blind alleys in a number of areas of psychology. I will conclude that the problem arises when one views the content of conscious experience as anything but fallible evidence, which has to be assessed in comparison to evidence from psychophysics and neuroscience.

4.3.1 The Illusion of Conscious Will

The conscious experience of deciding and of willing an action has been called “the mind’s best trick” (Wegner 2002, 2003). Daniel Wegner has reviewed a great deal of evidence that points to there being large and frequent discrepancies between how and when we have the experience of willing some action and the actual causal antecedents of the action. The research comes from many different phenomena and reveals such things as the following.

(1) **The experience of willing an action and the actual decision to act can be dissociated**  It has been shown by neurophysiological evidence that the experience of willing an action comes at least 0.3 seconds and maybe even longer after the effective decision for the action has occurred (these experiments are reviewed in Libet 2004).
(2) The experience of personal agenthood or authorship of actions can occur when actions are controlled by someone else. This is shown in a variety of experiments but can be seen most dramatically in the so-called rubber-hand illusion, in which by the use of mirrors, the experimenter’s (or someone else’s) hand is optically located where the subject feels his or her hand to be, and the manipulation of the seen hand’s fingers is done by the experimenter. In this dramatic illusion, the movements of the hand appear to the observer to be his or her own actions. The illusion persists until some major discrepancy occurs (e.g., the faux hand is withdrawn while the subject’s hand remains in place).

(3) The experience of other-controlled action can occur when the action is that of the experiencer. This is the converse of the rubber hand illusion and has been demonstrated in many controlled experiments, but also occurs frequently in such settings as dowsing (“water witching”), Ouija boards, and other “spirit” manifestations where it has been shown that the subject is unwittingly doing the controlling.

A critical aspect of perceived agenthood comes from the timing of events. Just as we experience causality between objects in such demonstrations as Michotte’s tunnel, so we experience ourselves as the cause of some action when the timing is appropriate. Recall that in the Michotte demonstrations, if an object disappears behind an opaque surface (an occluder) and an object appears on the other side at an appropriate time, the experience is one of a single object disappearing and then reappearing on the other side of the occluder; or if an object collides with a stationary object which begins to move and the timing is appropriate, the experience is that the collision by the first object causes the previously stationary object to begin to move. The same sort of perception of causality appears to work where the first event is the experience of willing and the second is some visible action; then, with the appropriate timing, the subjective experience of will is perceived as being the cause of the action.

The conscious experience of will is not exactly the same as the conscious experience that arises in visual or auditory perception—it is not a sensory experience or a sensation—but it is a conscious experience nonetheless. The person who has the experience reports the clear perception that he or she has initiated an action (or in some cases that he or she did not initiate an action and therefore that someone or something else had done so). These are just the sorts of experiences that make their way into the corpus of data that lend support to one or another theory of perception; they are the “experience that” something or other has occurred, or the “experience
as” of something or other. So the point here is the same as it was in other cases where the contents of conscious experience are used in building theories of perception: The opportunity for being misled by illusion remains.

4.3.2 Conscious Experience, Interpretation, and Confabulation
Closely related to the illusion of conscious will are cases where observers falsely report the reasons for their observed behaviors or the steps they go through in reasoning. The answers people often give to why and how questions are based on their conscious experience of their mental processes. The most egregious cases of mistaken reports of psychological processes arise in the case of reports of reasoning with the aid of mental images, and I will spend some time on this special case later in this chapter. Other cases arise when people are asked to report why they said or did something. Among those investigators who made the most of reports of how and why subjects made certain moves in playing a game like chess, or in solving slow and deliberate problems such as problems in logic, were Allan Newell and Herb Simon, whose work on problem solving appeared in an important book (Newell and Simon 1972). In those studies they made a great deal of use of “thinking out loud” protocols, in which subjects indicated what they were thinking as they attempted to solve a problem, as well as why they were considering various options. Even though the problems chosen for analysis were ones that were solved slowly and deliberately and made little use of prior knowledge, Newell and Simon still found that they had to fill in and refine the recorded protocols in various ways. One problem was that subjects rarely disclosed all the moves they considered or the reasons they had for considering and rejecting them. A large number of these intermediate “states of knowledge” went unreported and had to be inferred from other states that were mentioned and from the rational demands of the problem-solving process. Even among the states that were reported, many had to be discounted because they played no obvious role in the reasoning path (called a “problem behavior graph”) but seemed rather after-the-fact reconstructions (much the way that recollections are typically reconstructions, as shown in the classical work by Bartlett 1932). The best such problem-behavior paths were inferred by including additional sources of evidence, such as eye movements, that proved to be more reliable indicators of what the subject was focused on at various points in time. Thus even under the rather favorable conditions of slow, deliberate, and frequently conscious problem solving, the reports of conscious states required
a great deal of reconstruction by the theorist. In other words, the reports were treated as fallible sources of evidence.

In social psychology, the idea that we are extremely poor at expressing the processes and causes of our behavior by introspecting our conscious thoughts is well known (Nisbett and Valins 1987; Nisbett and Wilson 1977). Although we think we know why we do things or why we make the choices we do, the evidence shows the contrary (as we already saw in the studies of the experience of conscious will in section 4.3.1 above). The reasons we give ourselves and others are more often than not fabrications based on intuitive folk psychology theories. In addition, the methodology of asking people what they are aware of thinking in the course of planning some action is clearly unsuited for studying such processes as understanding a sentence, where almost none of the process is available to conscious scrutiny. In very many cases subjects have no information or conflicting information about their mental processes, and when forced to provide reports they simply manufacture (“confabulate”) explanations and rationalizations as best they can.

There has been a great deal of interest in recent years in widespread observations of confabulation, in which people provide descriptions of their mental processes when they do not have the relevant information. Particularly relevant cases are those in which people do not have access to information about why they made a particular choice (verbally or manually) yet they nonetheless provide a coherent story for why they did what they did (see, e.g., Hirstein 2005). Confabulation is quite frequently reported among patients with dysfunctions that prevent them from accessing the correct information for one reason or another. For example, it is often found in patients with large scotomas (blind regions in their vision) that prevent them from receiving information from large parts of the scene. People with these scotomas are often unaware of having blind spots, yet they (incorrectly) report patterns in the region of the scotoma (in fact everyone has a blind spot where retinal fibers leave the eye, though most are unaware of it). There are also some remarkable cases of blind people who insist that they are not blind (denial of obvious impairments is known as anosognosia and the special case of blindness is sometimes called Anton’s syndrome). These patients guess at what they are shown, and then confabulate elaborate explanations of why they misidentify things by sight or why they bump into things (Hirstein 2005; McDaniel and McDaniel 1991). Conversely, there have been patients who exhibit an even more astounding capacity to make some correct judgments of the location and shape of
patterns in their ostensibly “blind” field, while insisting that they can see nothing there. These are the famous cases of what is called blindsight (Bornstein and Pittman 1992; Weiskrantz 1995). Blindsight and other types of agnosias have attracted the interest of students of consciousness because they demonstrate the dissociation of functional vision from conscious visual experience.

An important point in all these examples is that subjects are perfectly sincere in what they report; they do not feel that they are making up answers even though they could not have known the true (and rational) basis for their answers. There are very many things that we do not know but are not aware that we do not know. The contemporary study of what is called metacognition is in part about that. Just as we have the tip-of-the-tongue phenomenon where we feel we almost have the word we are looking for, so there is the feeling-of-knowing (or the feeling-of-not-knowing) which often convinces us that we either know something that we do not, or that we do not know something that is just below the conscious horizon (for examples of this sort of phenomenon, see the collection of papers in Metcalfe and Shimamura 1994). When we think we know something that we do not know, we often engage in confabulation—we make up a plausible story.

4.3.3 Failures of Conscious Access: Split Brains and Split Visual Systems
Confabulation is most clearly illustrated in so-called split-brain patients—patients in whom nerve fibers (called the corpus callosum) that normally connect the right and left half of their brain are either congenitally missing or were surgically severed to ameliorate severe epileptic symptoms. In these patients, experiences that occur in one hemisphere are not available to processes in the other hemisphere. Since information from the right half of each retina goes to the left hemisphere, control of the right hand is from the left hemisphere, and most language functions are in the left hemisphere, it is possible to set up experiments in which half the brain has the information and the other half has to make a response. Michael Gazzaniga has studied these patients extensively and has reported cases where information is presented to the mute right cerebral hemisphere where it is used to make a right-hemisphere controlled response (say, a pointing with the left hand). The patients are then asked why they made the response they did. In these cases the left hemisphere that has language must respond—but it does not have access to the relevant information, since it was the right hemisphere that received the visual information and made the re-
In such cases the left (linguistic) hemisphere generally confabulates an answer. Confabulation in split-brain patients has been described extensively (see the summary in Gazzaniga 1995; Gazzaniga 2000).

One example that Gazzaniga gives is the case in which different pictures were shown to the two hemispheres of a split-brain patient (referred to as patient PS). The left hemisphere was shown a picture of a chicken claw while the (mute) right hemisphere was flashed a picture of a snow scene. Then an array of pictures was shown to the subject that included a chicken and a shovel, and the patient was asked to choose one related to the pictures he had seen earlier, making one choice with each hand. The patient chose the shovel with his left hand (controlled by the mute right hemisphere which saw the snow scene) and the chicken with his right hand (controlled by his linguistic left hemisphere which saw the chicken claw), even though the patient could not report seeing the snow scene (since it had been shown to the mute right hemisphere). When asked why he chose the chicken and shovel he (or rather his left hemisphere) replied, “Oh that’s simple. The chicken claw goes with the chicken and a shovel is needed to clean out the chicken shed.” To account for the way that the speaking hemisphere takes on the task of providing a rationale for the apparent inconsistency, Michael Gazzaniga has proposed a theory that credits the left hemisphere with the task of integrating information, both information that it possesses directly (if it is a left-hemisphere function) and information it gathers indirectly by observing some of the behavior controlled by the mute right hemisphere. There were also puzzling examples in which the linguistic left hemisphere was able to give the right answer to something that had been shown to the right hemisphere. Upon careful analysis, it turned out that the left hemisphere had observed the response made by the right hemisphere (in one case it had heard a sound that allowed it to figure out that the left hand, controlled by the right hemisphere, had touched a brush) and inferred what it had seen, but was unaware that the verbal response was related to this information.

These examples are relevant to the question of what reports of conscious states are about and what they imply with respect to their use in building theories of visual processing. In the split-brain cases, the left hemisphere (which has language and therefore answers the why question) faces a conflict between what it experiences and what it infers from watching the actions of the left hand (controlled by the mute right hemisphere), leading to a guess of what might have happened—that is, a confabulation. Yet patients report that their phenomenal experience in answering the
questions in these strange cases is the same as their experience in cases where things are normal (i.e., when both hemispheres have the relevant information).

Many of the split-brain examples involve a conflict between two different sources of knowledge (in the two cerebral hemispheres), but there are also many cases where there is no conflict, just the failure of consciousness to access the information that in fact determined an action. This arises frequently under conditions where the motor system is able to act on the basis of information that is unavailable to the conscious recognition system, because of a brain injury that prevents the information from reaching the part of the visual system responsible for sensory consciousness. An outstanding example of this is the famous case, studied by Milner and Goodale (Goodale and Milner 2004), of patient DF, who, because of severe bilateral damage in the ventral part of her visual-motor cortex, could not recognize the simplest patterns but could react appropriately and accurately to the same information when executing actions such as adjusting her hand orientation and grasp size while reaching for the article that she was unable to identify or even describe.

This independence of vision for conscious experience and vision for action occurs because the visuomotor system resides largely in the dorsal part of the brain—the part that feeds information from the eye through posterior parietal cortex to the motor system—whereas the conscious recognition system consists primarily of activity in the ventral part of the visual system (which routes information through inferotemporal cortex). In many experiments reported by David Milner and Melvyn Goodale (Milner and Goodale 1995) it was shown that in both animals and humans, the part of the nervous system that is in the dorsal visual pathway works differently from the part that is in the ventral pathway. Dorsal processing works rapidly, is more responsive to magnitudes (size, distance, location), and is relatively insensitive to the sorts of visual illusions in which visual context results in an inaccurate experience of size, distance, or motion. For example, if a subject reaches to grasp a circle whose apparent size is altered so it appears larger than it really is by virtue of being surrounded by smaller circles (or made to look smaller by being surrounded by large circles), which occurs in the Ebbinghaus or “size contrast” illusion, the grasp-control process is not fooled by the illusion but sets the grasp to the correct size (Aglioti, DeSouza, and Goodale 1995). In another example, a subject reaches for an object that is displaced during the saccadic eye movement that precedes the arm movement. Because of saccadic suppression, the sub-
ject is unaware of seeing any change in the object’s location, yet the (dorsal) reaching system seamlessly corrects for the displacement (Goodale, Pelisson, and Prablanc 1986).

Studies by Wong and Mack (1981), subsequently confirmed by Bridge- man (1992) using a different methodology, showed that the information available to consciousness can be put in direct conflict with the information used by the motor system. The Wong and Mack study involved stroboscopically induced motion. A target and frame both jumped in the same direction, although the target did not jump as far as the frame. Because of induced motion, the target appeared to jump in the opposite direction to the frame. Wong and Mack found that the saccadic eye movements resulting from subjects’ attempts to follow the target were in the actual direction of the target’s motion, even though the perceived motion was in the opposite direction. However, if the response was delayed, the tracking saccade followed the perceived (illusory) direction of movement, showing that the motor-control system could use only immediate visual information, even though the conscious experience is the same in the two cases. In all these demonstrations it seems clear that the conscious percept differs from the information that the motor system uses in determining actions.

4.4 The Phenomenal Experience of Seeing

A note about organization: In the remainder of this chapter I address the general problem of what to make of our conscious experience of space and of other properties of mental images. Since these topics constitute a central application of the ideas on selection and perceptual demonstratives (FINSTs), I treat them in some detail. However, for expository purposes I have divided these topics into two parts. The present chapter emphasizes the role of conscious experience in driving theories in these two areas. The next chapter addresses the same problems from the perspective of spatial representation, discusses some conditions that a theory of spatial representation should meet, and offers an alternative account to the one generally given in the mental imagery literature. The account I offer is not a general theory of spatial representation, but an account that deals only with the spatial properties of one sort of spatial representation, namely, the representation we construct when we reason about spatial patterns and relations, which I call active spatial representation or ASPAR. Consequently the discussion of representation of space and other properties of mental images is split between the two chapters.
The conscious experience we have when we imagine something (as when we have a “mental image”) is strikingly like that of seeing something. It is this aspect of the experience that makes it problematic as a source of evidence about the nature of our mental representation. That’s because the experience we have is that of seeing a perceived world and not of our mental state. As with other conscious contents discussed earlier, our visual image is actually the result of many different mental processes, including our perceptual-motor skills, our concurrent perception of things located around us, and inferences we draw from our beliefs about the properties, location, and likely behavior of objects we are imagining. Our experience is typically of a stable panoramic layout of spatial locations, some of which are empty while others are filled with objects, surfaces, and features that stand in some spatial relation to one another. This is the very phenomenology that leads people to postulate an inner replica of the perceived world and to suppose that this replica constitutes the experiential content of our mental image—a panoramic display that fills the world around us (Fred Attneave called it “cycloramic” since it appears to cover 360 degrees of view; Attneave and Farrar 1977). If we assume that the content of experience must somehow arise from a representation that encodes that content, and that the representation is constructed from the information we receive through our senses, then there is a problem about how such a representation could possibly come about, given the poverty of the incoming information. The incoming information consists of a small peephole view from the fovea (no more than 2 degrees of visual angle or about the width of your thumb at arm’s length) that jumps about several times a second, during which we are essentially blind (the information available to the brain is a familiar story and has been described in detail; see, e.g., O’Regan 1992). So the gap between our visual experience and the available visual information requires some explanation. There are many ways to try to fill the gap (some of which will be discussed in the next chapter), but the natural way, given the form of the experience, is to try to build an internal facsimile that corresponds to the contents of the experience. In other words, we find ourselves postulating a process that takes account of eye movements and constructs an inner picture in synchrony with these eye movements, along the lines shown in figure 4.3.

But as we now know, this theory is patently false—there is no inner replica or picture of any kind in our head, neither literally nor in any other nonvacuous sense capable of explaining how we represent spatial information in perception and thought. The mistake of reifying the spatial experience in this case is reminiscent of Kepler’s worry (mentioned in chapter 1).
about how we can perceive the world veridically when the retinal image is upside down. Just as Kepler and his contemporaries spent many years looking for a place in the brain where the image was reinverted, so also have many vision scientists searched for a place in the brain where the fragmentary incoming visual information is completed or filled in. The experience of visual perception suggests that vision provides a dense manifold of panoramic information, so theorists have searched for where such a detailed representation might occur in the brain. The answer is *nowhere*: There is no reconstructed detailed representation such as shown in figure 4.3. (The difficulty of accepting this conclusion has gone hand in hand with the difficulty of casting both a theory of vision and a theory of mental imagery in terms other than some form of pictorial or iconic representation—but more on this later.)

What has led so many people to succumb to the “picture” story depicted in figure 4.3? What has gone wrong is that we have been attributing the content of the experience to certain intrinsic properties of a representation (or, more precisely, of the structure or medium in which the scene is represented). But this makes two untenable assumptions. First, it assumes that the content of experience reflects the content of some mental representation that plays a role in the process of perception and imagination. Second, it assumes that the content of thoughts or imaginings reveals the structure and properties of the format or medium in which the mental representations are expressed. I have been discussing the first assumption,
which gives conscious content a special status over other sources of evidence that I have argued it does not merit. The second assumption is the result of the well-known intentional fallacy, the fallacy of attributing properties of what is being represented to the representation itself (as if our representation of a red square were itself red and square). Yet so long as we assume that the form of some mental representation must account for the content of the perceptual experience we are inevitably led to postulate a picture-like representation to match the picture-like experience. Should we, then, discount the experience and start afresh from psychophysical data alone? I will return to this topic below, as well as in the next chapter where I consider what a theory of spatial cognition needs to explain—and where I will in fact appeal to some phenomenology to motivate the conditions that need to be met.

4.4.1 Nonconceptual Representation in Visual Perception

As we have seen in previous chapters, a theory of perception and cognition needs an ultimate link of some sort with the perceived world in order to ground perception-based mental representations. Furthermore, this link must ultimately be causal on pain of infinite regress. Or, more precisely, the link must not be conceptual and must not rely solely on the semantic relation of satisfaction. In the previous chapter I offered a proposal for a particular mechanism of selection and reference based on FINST indexing, which we have seen works very much like demonstrative identification. I ended that discussion with the open question of what happens to the rest of the information in a scene—the information about objects and properties that are not indexed. I hinted that we may have to live with the uncomfortable idea that it is unavailable to the mind, at least at that instant. Such a conclusion seems particularly implausible with respect to visual perception, and in particular with respect to the visual perception of space, because it is here that our phenomenal experience most strongly insists that we have a grasp on space in some sort of bulk manner that is very different from the punctate index-based account I have been offering. This intuition is that we have what people have called a nonconceptual representation of space. The experience of space has been the subject of extensive analysis by philosophers, psychologists, and neuroscientists because it offers so much scope for exploring the idea of a different kind of nonconceptual representation—one that departs from the sort of format that seems appropriate for representing propositional attitudes (see, e.g., the essays in Eilan, McCarthy, and Brewer 1993; Gunther 2003; Luce, D’Zmura, Hoffman, Iver-
son, and Romney 1995; Paillard 1991). The question of how we cognize and represent space will be dealt with in the next chapter. For now my concern is with perception more generally.

As I mentioned above, characterizing the experience of visual perception is a deep and interesting problem on its own. For example, the purely phenomenal content of experience may be relevant to understanding certain distinctions we experience, such as the qualitative difference between vision and mental imagery (Dalla Barba, Rosenthal, and Visetti 2002) or perhaps that between clear perceptions that fail to be convincingly real and vague perceptions that seem very real (a distinction that is orthogonal to perceptual content, as Gestalt psychologists recognized). Beyond such qualitative observations, it is not clear how cognitive science can build on these ideas, because it is not clear how the detailed phenomenological experience of vision or imagery captures the distinctions and the mental structures required by a causal/functional theory.

Perhaps there is a parallel here to the relation between generative theories of grammar and theories of language learning and parsing. Even though it is clear that the rules of grammar characterize what a speaker implicitly knows, the form of the rules required for characterizing the grammar do not appear to be suited for direct application to parsing or language learning (for more on this issue as it pertains to language, see Pylyshyn 1973a). In fact it is plausible that the rules (expressed in some generative formalism such as rewrite rules and transformations) are not themselves explicitly represented (Pylyshyn 1991). Similarly it is not clear that any description of the experience of perception or imagery can be taken as constituting a form of (nonconceptual) representation that is functional in perception.

Even if our perceptual experience were correctly characterized (e.g., in terms of something like Peacocke’s scenario content discussed in Peacocke 1992), it need not correspond to some representation that figures in an explanation of how perception or thought works in an information-processing account. We have seen many examples of this sort of disconnect between phenomenal experience and functional states in the previous section. But there are even more problems with misinterpretations of experience in discussions of mental imagery. Even when the descriptions of the phenomenology are correct, their functional significance is at best problematic and often simply irrelevant. An example is the explanation offered by a numerical “savant” of how he multiplies two four-digit numbers. According to a story was originally reported by Terry Moran and Lenny
Bourin on “World News Tonight,” June 5, 2005, and reported again in the *Guardian* (Feb. 12, 2005), the savant Daniel Tammett explains how he multiplies two four-digit numbers as follows: “When I multiply numbers together, I see two shapes. The image starts to change and evolve, and a third shape emerges. That’s the answer. It’s mental imagery.” The claim that he “reads off” the answer from the resulting shape is not very different from explaining why it takes longer to report details in a “small” image than in a “large” image by saying that the details are harder to see in the smaller image so one has to “zoom in” to see them (Kosslyn 1980). These examples of appealing to the contents of one’s image to provide an explanation illustrate the seduction of the intentional fallacy.

Philosophers have another reason to appeal to conscious experience in characterizing perception; conscious experience is thought to provide a justification for our perceptual beliefs. The idea is that you are justified in believing $F(x)$ if you can see that $x$ is $F$ just by looking, where “see” is taken to mean “consciously experience.” This may be what is behind John Campbell’s claim (mentioned earlier) that without consciousness we would not know what our demonstrative reference refers to (see section 3.5). But there are many ways to justify a belief (even to justify the belief to ourselves), and, given the examples reviewed earlier, our conscious experience may not be the most reliable. As Jerry Fodor (2007) remarks in a footnote:

The doctrine [that one's justification of a perceptual claim that $P$, is typically its seeming to one that $P$], though venerable, strikes me as confused; in particular, as confusing offering a justification for a perceptual claim with offering a justification for making that claim. Compare: My sincerely believing that $P$ generally justifies my claiming that $P$; but it’s not a reason to believe $P$ is true (or, anyhow, it’s not much of one. Surely it can’t be *my* reason for believing that $P$ is true). Why suppose that the epistemology of perception differs, in this respect, from the epistemology of other sorts of belief fixation.

Discussion of the content of perceptual experience brings us to the question of the nature of mental imagery, which I raise in the next section. In the final chapter I will revisit the question of spatial representation, which many people believe is at the heart of what is special about mental imagery, and offer a suggestion for how spatial “representation” might arise without any internalizing of spatial properties—without an “inner space” of any kind. But for now I will focus on the way that theories of mental imagery are informed (or I should say, misinformed) by phenomenal experience. The pull of subjective experience is so powerful and has so thoroughly misled the majority of cognitive scientists (and cognitive neuroscientists) that even patently obvious fallacies often go unnoticed.
4.5 The Phenomenal Experience of Mental Imagery

If we are tempted by the model of visual perception shown in figure 4.3 above, then we will be equally, if not more, tempted by the view that in the absence of input from the eyes, the inner display in that figure can also be filled from memory or from reasoning (since according to that view there is top-down involvement in painting the inner picture, even in vision). According to that view of visual perception, we have a display surface with the nonconceptual content corresponding to our experience; so it would be logical that we might use that display to imagine as well as to see. This is indeed the received view in much of cognitive psychology (Kosslyn 1980, 1994), neuroscience (see the commentaries appearing with my article in Behavioral and Brain Sciences, Pylyshyn 2002a), and even a fair amount of philosophy (see, e.g., the essays reprinted in Block 1981; Tye 1991). Although it is not generally acknowledged (in fact it tends to be vehemently denied), the driving force behind this sort of theorizing is the desire to account for the experience we have when we entertain mental images. The writings on mental imagery typically begin with the assumption that because the experience of having a mental image is very much like the experience of seeing something, entertaining an image must also involve seeing something. And if imagining is seeing something, then there must be something that one is seeing—there must be something in the head that plays a role analogous to that played by an actually perceived scene (and of course there must be something playing the role of the eye, though that is less often mentioned).

Why a picture and not something else? The only other possibility is that there is a replica of the world to be perceived. As Nelson Goodman (1968, p. 3) said about art (quoting from an unnamed source), “Art is not a copy of the real world . . . one of the damned things is enough.” This is even truer when applied to mental representations. If it seems unreasonable that there is a replica of a world in the head, perhaps it is possible that there may be at least a picture of the world instead.6 When I am imagining a visual scene it certainly feels like I am looking at something and the thing

6. An actual 3-D replica of the world is no more egregious than a picture of the world, given that there is nothing 2-D about the experienced image. In fact our images are distinctly 3-D rather than 2-D, both in their phenomenology and in their psychophysical properties, as I will point out later when I discuss the experiments. For some reason it seems less fantastic to ask for a 2-D replica even though all the problems with 3-D replicas appear with 2-D pictures.
I am looking at something in the world. A picture also looks like something in the world; so maybe what we have is a picture. That brings us to the assumption that what we have in our heads (or brains) is something that shares the essential properties of a picture, namely, it is a structure that is *depictive* (where the latter is defined in the quotation below).

But it is not enough that we have some structure that looks to be depictive when pictured on paper. The structure itself must be implemented in neural tissue in such a way as to impose constraints like those we find in the world, or at least in a picture, such as requiring that when you scan your eye (or your attention) from place A to place B you must pass through the intermediate (possibly empty) places. Many other properties that we discover in mental imagery experiments must also be determined by the structures that underlie the depictive representation. Since most of the constraints that the medium is alleged to impose concern properties of space, I will leave those for the next chapter where I consider the larger question of how we represent space. What I will not do in the present section is revisit aspects of the imagery debate. That debate goes back a long way (at least since Berkeley and Locke quarreled over it) and in its modern form is now nearly thirty-five years old (if we date it from the first salvo in Pylyshyn 1973b). It has changed in emphasis during that time (and has incorporated neural image data), but the basic disagreements remain essentially the same. The debate relevant here is fundamentally about whether postulating certain kinds of mental/cortical constructs that are consonant with our conscious experience constrains the hypothesized mechanisms in any way. If it does not, then these assumptions simply appease our intuitions, derived from our conscious experience, without serving any explanatory purpose. This, I claim, is indeed the case for “picture theories” (or “depictive” theories) of mental imagery.

Images are said to be depictive. This is a well-chosen word because it suggests that the relation between mental images and the world is not a semantic one—as understood in linguistics and logic—but one closer to what one might call “resemblance.” Resemblance has a long history in philosophy of mind. It served for Hume as one of the three fundamental principles of association (along with contiguity and causation). But it failed in the end for reasons that are well known—thoughts can use symbols that do not resemble their referents (e.g., words), and if there is a resemblance the resemblance itself cannot be what determines the reference or meaning (this is not the place to rehearse these ideas; see Fodor 1965, 2003; Pylyshyn 1984).
One of the people who has tried to be explicit about what it means for a mental image to be depictive is Stephen Kosslyn, as expressed in the following quotation (Kosslyn 1994, p. 5):

A depictive representation is a type of picture, which specifies the locations and values of configurations of points in a space. For example, a drawing of a ball on a box would be a depictive representation. . . . The space in which the points appear need not be physical, such as on this page, but can be like an array in a computer, which specifies spatial relations purely functionally. That is, the physical locations in the computer of each point in an array are not themselves arranged in an array; it is only by virtue of how this information is “read” and processed that it comes to function as if it were arranged into an array (with some points being close, some far, some falling along a diagonal, and so on). . . . In a depictive representation, each part of an object is represented by a pattern of points, and the spatial relations among these patterns in the functional space correspond to the spatial relations among the parts themselves. Depictive representations convey meaning via their resemblance to an object, with parts of the representation corresponding to parts of the object. . . . When a depictive representation is used, not only is the shape of the represented parts immediately available to appropriate processes, but so is the shape of the empty space. . . . Moreover, one cannot represent a shape in a depictive representation without also specifying a size and orientation. . . .

I don’t know whether this view is universally received (or even whether it is still Kosslyn’s view—see chap. 5, note 5, page 111), but it will serve as a basis for my comments because it has the merit of being explicit. What it defines are the constraints that are assumed to hold by virtue of something’s being an image rather than, say, a representation in a compositional system of symbols—that is, a language of thought. Notice right off that what it describes is unabashedly a picture—a 2-D object laid out in space. True, it says that the space need not be physical; it might be only functional. We will see in the next chapter that this idea is a ruse: there is no such thing as a functional space that is capable of explaining the apparent spatial properties of mental images—it is a blank check that be used to explain any property you wish the idea of depiction does, however, come close to corresponding with one’s phenomenal experience of looking at a picture, which, I suppose, is why we call such experiences “images.”

But explanatory adequacy requires that one specify why the depictive structure has the properties it has. In particular, there are two very different possible reasons why the representation has such properties. One is that this is the nature of the mind–brain—it is part of the relatively fixed architecture of mind or of an encapsulated vision/image module. If the space mentioned in the quotation above were real space such an account would
be explanatory: real space has certain properties, including the properties described by the metrical axioms and possibly also the Euclidean axioms that constrain the things that can be represented and the types of transformations or processes that can take place (e.g., moving attention through real space requires that attention “pass over” all the empty places along the way). The second possible reason why the properties described in the quotation hold is that people (i.e., subjects in the experiments) have certain beliefs about what things look like, how they change (e.g., how they move), and how events happen in space and time, and they can use these beliefs to predict or to mimic what would happen in a real situation (e.g., it would take longer to move a greater distance; it would be harder—and so take longer—to see small features than large ones). If the phenomenon holds because a person believes that this is how things would unfold in the world (because of a folk theory or because of recollecting something similar happening in the past) then the phenomenon reveals not a property of the mind–brain but only a property of the person’s beliefs or knowledge (often tacit) of how things work in the world. The distinction between a regularity attributed to the nature of the architecture of mind and one that is attributed to tacit beliefs or knowledge is about as fundamental a distinction as there is in this field. If the phenomenon is attributed not to some property of the architecture, but to tacit beliefs or knowledge, then it is in principle changeable by rational means (being told, being shown, etc., any appropriate rationally connected belief-changing information). The notion of tacit knowledge is one of the fundamental ideas in cognitive science (Fodor 1968; Pylyshyn 1981).

4.5.1 Phenomenal Experience and Explanation: The Role of Tacit Knowledge

One major problem with relying on introspective evidence (even if one is not aware that one is doing so) is that, as in the case of illusions of will and other types of confabulation discussed in section 4.3, conscious experience is powerless to tell us why something happens. And to the extent that it matters why, we cannot get the requisite answers from our conscious experience. Here are a few examples, intended solely to clarify the difference between an architecturally based property and one based on tacit knowledge.

Imagine that you are watching Galileo’s (apocryphal, as it turns out) experiment atop the leaning tower of Pisa. A large, heavy cannonball is released at the same time as a tennis ball. You watch what happens to the two objects as they fall in your image. You must press one button when
the heavier ball hits the ground and another button when the light ball
hits the ground. What do you think will happen? In all likelihood, unless
you have studied physics, you will press the button for the heavier ball
before that for the light ball.7 But the critical question is: Why did you press
the buttons when you did? Does it matter whether it was because of prop-
erties of the mind–brain, or properties that you learned in school or
believed for other reasons (e.g., watching balls fall in various field games)?

Now imagine a person on a bicycle traveling down a hill and then turn-
ing around and pedaling back up the hill. Which took longer in your imag-
ining, the downhill portion or the uphill portion? Again, the important
question is: Why did those time intervals appear in your imaginings? It
should not be hard to think up innumerable such imaginings involving
time, in which you are likely to agree that the reason one event takes
longer than another is not because of how your mind–brain is constituted,
but because of what you know, even if you did not know that you knew
(i.e., even if you gave a different answer when asked on a written question-
naire). The use of reaction time in psychological experiments has been a
major boon to information-processing theories because it has enabled us
to compare the computational complexity (typically interpreted as an indi-
cation of the number of operations performed) of processes under different
inputs. And yet in this case it seems that it tells us little about the process
and its underlying architecture, except that that architecture is capable
of storing beliefs and drawing inferences from those beliefs and that it is
capable of generating time intervals based on independently computed
estimates.

Now try another task using mental imagery. Imagine a beam of blue light
and a beam of yellow light producing two patches of light on a white surface
side by side. Move the patches closer together until they overlap. What

7. As it happens, this is also what was found in Pisa, not by Galileo but by oppo-
ents of the Galilean theory (see Kuhn 1957). Incidentally in this experiment you
are also likely to press the buttons after a delay that is a linear function of the dis-
tance fallen, which means a constant velocity and not the Newtonian constant accel-
eration. The dynamics of your mental image have not, it seems, incorporated
Galilean physics but remain stuck in the Aristotelian/medieval worldview (McClos-
key and Kargon 1988; McCloskey, Washburn, and Felch 1983). On the other hand,
modern sophisticated observers appear to be stuck in Galilean idealizations and have
erroneous expectations about motion in air (Oberle, McBeath, Madigan, and Sugar
2005). Thus what you will imagine in the present example will very much depend
on factors that are not related to the architecture of the imagery system (even if there
were such a thing).
color do you see in your image at the overlap? People who have vivid imagery have no problem providing a quick answer. Once again the question of interest is: why do they answer as they do? Here's another example. Imagine two identical glasses, one half full of sugar and another nearly full of water. Imagine slowly pouring the water from the water glass into the glass containing the sugar. Does the water in the sugar-glass overflow in your image? The right thing to say in this case is probably, “How should I know?” That answer captures not only one’s state of mind, but also acknowledges that what happens in your image depends on what you think would happen. In this case the correct answer depends on some sophisticated knowledge about what happens to a solid in solution in a liquid so the correct answer (under the right conditions it does not overflow) would depend on such knowledge or on recollections or on informal folk theory. But whatever answer you give surely depends not on properties of your mind–brain, but on what you believe. Now the reader may well ask how I can be sure that in the above examples the outcome depends on what you believe? The answer is easy: Because you could easily make the outcome different by willing it! It is your image, so you can make it do whatever you like. If you don’t believe me try the above examples making the outcome something quite different: you can make the balls dropped from the leaning tower of Pisa fly away, or the fluid miss the container and pour on the floor, or the colored light beams mix to form chartreuse or no color at all, if you wished.

I present these examples to illustrate the difference between the two types of causes of imagery processes. Although these are not published experiments, very similar experiments have been published and discussed in the literature on mental imagery. For example, there is a notion of rep-

8. The chances are good that you gave the wrong answer because few people know the difference between additive and subtractive color mixing. The example here involves light, which results in additive color mixing, so combining a blue light with yellow light should result in a white patch. But if you were looking though two filters, one blue and the other yellow, you would see a patch of green. For more on this complex but well-studied phenomena see Rossotti 1983.

9. Kosslyn (1981) discusses the color-mixing example and actually cites empirical data to show that many people give different answers when asked to imagine than when they are merely asked to provide a verbal response. It's not clear what this is supposed to show, over and above my present claim that people can make their image have whatever properties they wish. I suspect one can get different answers if one asks the question in different settings or in different ways (in a lab vs. a paint studio, inside vs. outside, in a timed task vs. an untimed condition, in a purported IQ
resentational momentum, hypothesized to account for why in tracking a moving object one generally makes errors in indicating its final position, where it disappears. The idea is that the imagined motion has momentum the way a real moving object does, and like a real object, it does not stop suddenly but continues to move after it disappears. Other such examples are discussed in Pylyshyn 2002a.

Having described the two different sorts of causes involved in imagined processes, we can now look at examples of imagery that have been discussed in the literature. These generally do not depend on knowledge of physical principles, but on geometrical-optical properties, which seem more likely to be built in to the visual system. Consider, for example, experiments involving image size (Kosslyn 1975). In these experiments, subjects are asked to imagine, say, a mouse under two size conditions: (1) imagined next to an elephant so that both are present and fill the entire “mental screen” (in which case the mouse has to be visualized as small—i.e. it has to occupy a small visual angle in the “mental display”), and (2) imagined by itself in your hand and close up, so it occupies a large visual angle in the mental display. Subjects are asked to report details in the image of the mouse (e.g., does it have whiskers?) under these two conditions. It was found that it takes longer to report such details when the image is small than when it is large. Subjects feel that with the small image they can’t “see” the details and have to “zoom in” to see them (and the picture theory actually postulates a “zoom” operation). So the question is: Is the increased time attributable to a property of the architecture or to subjects’ beliefs of what it is like to see a small mouse? The pictorialists claim it is the former. The argument given (Kosslyn, Thompson, and Ganis 2006, p. 148) is that “the inhibitory connections in topographically organized areas are typically short, and thus when a lot of spatial variation is packed into a small region strong input is required to overcome the inhibition.” In other words, the visual cortex is limited in its resolving power so you can’t get all the information in if the image is small. But a “larger” phenomenal image is not larger on the cortex. The cortical activity that shows up in PET scans shows at most that a mental image experienced as being larger may be accompanied by activity that is anterior of activity accompanying images experienced as small (though even that is not without some question).
Thus it may be located in areas where larger retinal images would project from the retina. But even if the locus of cortical activity is shifted when the image size changes, the area of activity for small mental images is not smaller in visual cortex, and so an explanation based on limited cortical resolution is irrelevant. Cortical resolution applies only to the resolution of information originating on the retina, not to information originating internally (from memory) and projected onto the surface of the cortex. For more on the futility of appealing to the neural properties of visual cortex to explain imagery phenomena, see section 4.5.2 below.

But once again I am willing to give the pictorialists all their claims, even though the actual data are problematic in ways discussed in Pylyshyn 2002a, 2003, because the problems here are conceptual. What does it mean to make one’s image small or large? Can one distinguish between an image being small and imagining something as being small? What do you know about seeing details on a small or a large object? If the size is what makes it easier or harder to see, consider what would happen if we kept the size fixed and manipulated the amount of detail. Imagine a medium-sized mouse viewed through a pair of steamed or scratched glasses so it is fuzzy. Now imagine the same mouse viewed clearly (you can substitute low-definition and high-definition TV if you like). In which version is it harder to “see” the whiskers? Do we even need to do the experiment? What if we did the experiment and it took more time to see the whiskers on the high-definition image? Would we conclude that the architecture of the visual system has these strange properties? I doubt it. We are more likely to conclude that the subject misunderstood the instructions because what it means to be fuzzy is that you can’t see the details, and what it means to do the reaction time test is that one is supposed to re-create as closely as possible the phenomena that would occur if one were looking at and seeing the large/small mouse. The details of the neurology of V1 are interesting on their own—for example, that inhibitory processes may explain the limited resolution of vision—but they do not clarify the problem of the resolution of mental images. That problem arises because one is attempting to match the experience of having a mental image with properties of the architecture of vision when the facts at hand have nothing to do with that architecture, but have everything to do with what it means to have a small (or large) image.

The same can be said for the widely cited study of the “visual angle of the mind’s eye” (Kosslyn 1978). If you ask people to stand close enough to various objects so that the objects fill their field of view, you can compute the visual angle of the eye. If you then ask people to use their imagination and
tell you how close they should stand to a car, a horse, a toaster, and so on, so that it fills their field of view, they also give reliable answers, which establishes that the visual angle subtended by images is similar to that subtended by vision, allegedly showing that the two share a common display. But of course it might also show that subjects know how close they should stand to a car, a horse, a toaster, so that it fills their field of view. The fact that they can’t tell you those distances if you ask the direct question is irrelevant here as it is in all the other such cases: One thing that psychologists have learned is that how you ask is critical in determining the answer you get. In this case, if you use a different way of measuring the visual angle subtended by images—one that does not invite subjects to imagine that they are getting closer to some object until it overflows their visual field—you get quite different answers. If, for example, you simply provide a task that requires subjects to recall where things are by using their mental image of a room, you find that the visual angle is $360^\circ$—what Fred Attneave, whose sympathies have tended toward the picture theory, called a cyclo-ramic display (Attneave and Farrar 1977). The pattern is clear: If you ask subjects to pretend that they are looking at some particular display, or if you present them with a display that they memorize, then you tend to get parallels between seeing and imagining, but generally not otherwise. The tacit knowledge explanation should be treated as the default explanation, barring evidence to the contrary, since that is the way we pretheoretically understand what imagining something means: by default it is an invitation to put yourself in the position of watching something unfold before your eyes.

Here is another example, to which I will return in the last chapter. One of the most widely quoted and replicated results cited in support of the depictive nature of mental images is mental scanning. The finding so impressed the pictorialists that they refer to it as a “window on the mind” (Denis and Kosslyn 1999). The typical experiment goes like this. Subjects are asked to memorize a display—usually a map of some fictitious island—until they can reproduce it to within some margin of error. They are then asked to imagine the map and to focus their attention on a particular (named) place on it. Next they are asked to move their attention to another named place. This is done in different ways. In the early experiments subjects were asked to imagine a spot moving from the initial focus to the second named place. In subsequent experiments they were asked to switch their attention or to simply look for the second named place, and in some cases they were asked to report on parts of the map that were off the to side of the imagined region (beyond the visual angle of the mind’s eye). What
was found is that the time it took to arrive at the second place was a linear function of the distance between the two places on the map. So the question this raises is Why? The pictorialist has a ready answer—the image is actually laid out in real space in visual cortex so attention (or gaze) travels across it just as it does across real external space, and therefore the relation \( \text{time} = \frac{\text{distance}}{\text{velocity}} \) holds. The account for the case where the item being scanned to is off the edge of the image ought to be an embarrassment but it is not viewed that way: there is a story there too (it involves an “image transformation” process—though it’s not clear why that should yield a linear reaction time effect; see Kosslyn, Thompson, and Ganis 2006). Now if you imagine a spot moving across your image of your favorite scene you will notice that it takes more time to go further. But you might also notice that you can make the spot speed up, slow down, back up, hop around, disappear from the scene and do any sort of trick you like. Not only is it your image but it is your spot to do with as you choose—the motion of the spot is not constrained in any way by properties of your imagery system. So why do you choose to make it take longer when the distance is greater? Surely it is because that’s what it would do if there was no reason for it to do anything else—because moving in a straight line at a fixed velocity is what physical things do in real space.

But wait, you say, the increase in time with distance occurs under many other conditions. It occurs if the subjects are not told to scan, but only to look for or to notice the other named place. But isn’t that the same as asking them to pretend that they are looking at a map? And who would fail to know (or remember) that noticing something further away takes more time—because it takes time to move your line of sight or your attention? But why do I claim that subjects know that? If you ask subjects what would happen in such an experiment, for example, they often say they don’t know. And that’s the right answer—they don’t know what would happen in an experiment. They might even not know what they themselves would do. But that’s just the perennial finding that asking subjects an outright question—especially one about what would happen in an experiment—is the worst way to find out what they believe (witness all the cases covered in section 4.3 above). You need to see whether different beliefs would yield different results. And you don’t do that by trying to induce strange expectations by telling unrealistic stories about object movements (such as telling subjects that scanning times would be long for short distances because of some sort of crowding effect or that it would be different for different colored items; see Jolicoeur and Kosslyn 1985).
Another pictorialist defense is to cite an experiment reported by Finke and Pinker (1982), in which no scanning instructions were given. Subjects were shown a display of points on a screen. The points disappeared and an arrow appeared. Subjects simply had to extrapolate the arrow to see whether the line would intersect with one of the points that had been there before. Here too, time increased with distance. I leave it to the reader to explain why this does not entail a “picture-in-the-head” view (just put yourself in the position of a subject and ask what you would do and why). They may not have been given instructions to imagine, but they were given a visual memory task which is much the same thing. It should be clear where this sort of altercation is headed.

Interestingly, pictorialists do not cite the mental scanning experiments we did (actually part of a Ph.D. dissertation by Liam Bannon at the University of Western Ontario) and which I reported in Pylyshyn 1981. In one experiment (which was scarcely worth doing since the outcome is pretty obvious given a moment’s thought) we showed subjects a board with a map mounted on it that contained lights and switches. When the appropriate switch was toggled the light that was currently on went off and another went on immediately. Subjects played with this board for a while then were asked to imagine the board and, as in the scanning experiments, to press a button when they saw, in their mind’s eye, the light come on at the second named place. As you might have guessed, there was no time increase with increasing distance. Why? Because there was none in the situation they were imagining. Notwithstanding such findings, the claim was made (Kosslyn 1994, p. 11) that research “showed that imagery is highly constrained. Subjects can control some aspects of processing, such as the speed of scanning, but not others; in particular they could not eliminate the effects of distance on time to shift attention across an imagined object.” Yet so far nobody has shown any phenomena that are constrained by the depictive nature of the display and which are not better explained by appeal to tacit knowledge.

It’s also interesting that the scanning effect can be made to disappear simply by playing down the importance of the task of moving one’s attention over the image. For example, we asked subjects to use their imagined map to do the following. Start by focusing attention on some specified place. Then when a second place is named, they were to say what direction the first place would be from that second place (using a clock face as the way of specifying direction). This is a task which really does require that one focus on the second place in order to use it as the reference point in
giving the direction to the first place. But no movement is mentioned. And in that case we found no distance effect on reaction time (Pylyshyn 1981).

But even if you accept the tacit knowledge view, namely that the scanning effect is the result of subjects' recreating an imagined state of affairs where attention or gaze is scanned over a scene, there is still one remaining question. How does this simulation create the appropriate observed time delays? Surely it is not the case that a subject simply counts the seconds until the right amount of time has gone by! Why do you get an approximately linear time when (and, I assume, only when) subjects imagine the scanning taking place and not, for example, when they are asked to wait a certain amount of time and then press the button? The answer depends on one's theory of what goes on in the interval. My assumption has been that what goes on is that people imagine the spot or focus of attention being here and then here and so on until it gets to the target. But does this not require some place for the demonstratives here to refer? And does this not require a depictive display? As far as I know this question has not been asked, and yet it deserves an answer. Precisely this question will be the focus of the next chapter (although the allusion to demonstratives should give the reader a hint that FINSTs are going to play a role in the story).

4.5.2 Does the Architecture of Visual Cortex Matter to Explanations of Imagery?

I said that the use of tacit knowledge (to simulate what would happen if the event were actually witnessed) is the first line of explanation, but it is not always the last, for there may be other factors involved—some architectural properties often reveal themselves, though not necessarily the architectural properties of the display that are postulated in the pictorialist's canonical story. This sometimes arises in imagery experiments where the task done with imagery is compared with the same task performed with actual vision. Such experiments tend to introduce architectural properties into the picture. I consider one such example because it nicely illustrates three important points: (a) Not all imagery phenomena can be subsumed under the tacit knowledge explanation—many are a mixture of tacit knowledge interacting with some architectural constraint not obviously related to the task; (b) here, as elsewhere, the details matter and we have to look at the postulated mechanisms to see what explanatory work they do; and (c) it illustrates how deeply committed some people are to the literal picture theory, so much so that they are willing to ignore obvious problems in their account in order to salvage the picture theory of mental imagery.
The example I have in mind concerns a low-level psychophysical property known as the oblique effect.

The original oblique effect finding in vision is that it is easier to resolve closely packed lines when the lines are vertical or horizontal than when they are oblique (i.e., at 45° from horizontal). This general phenomenon has also been found to be true when the stripes are imagined at various orientations (or, rather, when memorized bars of different orientations are used in an image recall task such as mentally comparing pairs of such memorized bars for properties like width and spacing; see Kosslyn, Sukel, and Bly 1999). The explanation given for why performance on such imagery tasks is worse when the stripes are oblique is that the imagined stripes are displayed on the visual cortex, and it is known that there are more cells with horizontally and vertically tuned receptive fields than oblique ones in visual area V1. Now, I agree that this does not sound on first hearing like a phenomenon traceable to tacit knowledge being used to mimic perception, but we have already seen examples where the exact wording of the question or of the task made a major difference in the results obtained. The major problem with the tacit knowledge account in this case is that very few people know about the oblique effect even in some informal guise. Of course the fact that people don’t explicitly know about the oblique effect does not mean that they cannot recognize cases of it—it does not mean that there is no familiarity with how things look when they are oblique, especially since oblique contours are far less common in our world than are vertical or horizontal ones (Hansen and Essock 2004). We also need to keep in mind that there is a lot we don’t know about the oblique effect in general and the (probably many) reasons for it. Notice, however, that the oblique effect is found in the haptic modality as well as in vision (Gentaz and Hatwell 1998; Kappers and Koenderink 1999); there are many cases where oblique lines are perceived better than horizontal or vertical ones (e.g., when the spacing is variable and broadband or when the measure is adaptation rather than discrimination, Essock et al. 2003; Heeley, Buchanan-Smith, Cromwell, and Wright 1997; McMahon and MacLeod 2003; Wilson, Lofler, Wilkinson, and Thistlethwaite 2001); and it seems that the frame of reference for classifying orientation depends on gravity, and so on—all in all not a strong argument for connecting the visual and imaginal oblique effect via properties of a common display in cortical area V1.

What explanation does the pictorialist offer for this effect? Here is their most recent explanation (Kosslyn, Thompson, and Ganis 2006, p. 69):
if the result emerges from the neurophysiology of the visual buffer, it is easily
explained.... neurons in topographically organized areas are known to have
orientation-tuning ... and to be less sensitive to distinctions along the diagonal. In
addition, at least in area V1 in the cat brain, so-called simple cells (which fire when
the animal sees edges and not to complex combinations of features) not only fire
more vigorously when horizontal and vertical lines are shown, but also have sharper
tuning for horizontal and vertical lines. ... These results underscore the fact that the
oblique effect reflects properties of the neurons that populate the early visual cortex.

So there is the story. There are more finely tuned receptors for vertical
and horizontal lines in V1, where mental images are projected as a pattern
of activity, and so imagining horizontal and vertical bars get preference in
imaginical tasks. But there is a critical assumption in this story which reveals
how the seductive picture theory can blind us to assumptions that are es-
sential to the explanation but remain unstated and unquestioned. The
assumption is that a pattern of activity projected onto the surface of V1 by
higher cognitive functions, as assumed by the picture theory, is equivalent
to the same pattern of activity applied at the retina. But cells that are sen-
titive to orientation are sensitive to the orientation of patterns on the retina—as
picked up by photoreceptors—not to patterns imposed on the cells them-
selves (i.e., patterns depicted on the surface of V1). The orientation sensitiv-
ity of cells in V1 is the result of the arrangement of photosensitive cells on
the retina and how they are connected to the simple cells in visual cortex.
The exact form of this arrangement is uncertain—it could be a simple tem-
plate, as postulated by the organization of simple and complex cells
reported by Hubel and Weisel (1968), or it could be a more complex ar-
rangement such as the wiring of a perceptron-like mechanism—but what-
ever it is, it is clear that activating an oriented pattern of cells in V1 will not
selectively activate orientation-tuned cells. If a pattern of activation, such
as a grid of parallel stripes, were imposed ("depicted") on the surface of V1
it would activate all sorts of cells in its path and would not favor different
orientation-specific cells depending on the orientation of those parallel
stripes. Although parallel stripes on the retina might create parallel stripes
of activity on V1 (assuming the retinotopic mapping is accurate), the con-
verse does not hold: activating stripes on the surface of V1 does not pro-
duce striped activity on the retina or anywhere that serves as input to
orientation-tuned cells.

I have belabored this point because it is a recurring theme in recent
picture-theory writings (including the recent overview in Kosslyn, Thomp-
son, and Ganis 2006). When certain phenomena of mental imagery, such
as the apparent lower resolution of imagery relative to vision reported in
Kosslyn, Sukel, and Bly 1999, or the longer time it takes to report fine details from a small image, or the “visual angle of the mind’s eye” result reported in Kosslyn 1978, the explanation always alludes to properties of cells in V1, in order to support the view that images are projected onto V1. But the properties of cells in V1 could not account for such patterns since these properties arise from the way those cells are activated from the retinal photoreceptors; the alleged top-down activation from memory or imaging instructions would not produce the same effect as activation from photoreceptors. The only way this sort of explanation would work is if when we imagine something, images were projected onto the retina, which so far nobody has had the audacity to propose.

4.5.3 Problems in Accounting for Phenomenal Space by Appealing to Brain-Space

In addition to the problems raised by attributing properties of imagery to properties of cells in V1, there are even more serious problems with the basic premise behind the evidence cited in favor of the picture-in-V1 view of imagery, namely the assumption that the phenomenal experience of looking at a picture arises from the activation of a pattern on the corresponding display in the brain (specifically on the surface of visual cortex). Many of these reasons are discussed in Pylyshyn 2002a, 2003. Here is a quick listing.

1. **Displays in visual cortex (V1) are retinotopic** Fibers run from the receptors on the retina to cells in the early visual cortex and map spatial patterns in a fairly direct manner (as shown by the activity on monkey cortex when it was made to stare at a flashing pattern—see the photograph of the unfolded visual cortex in Tootell et al. 1982, in figure 4.4). This photograph shows that there is a continuous mapping from activity on the retina to activity on the occipital cortex of the macaque monkey.

   Being retinotopic means that, as is the case with patterns on the retina, the patterns move with eye movements and have a small area of high resolution. If patterns were projected onto the visual cortex in the course of mental imagery and there were eye movements, the interpretation of the patterns would be garbled (and there generally are spontaneous eye movements during imagery; see Brandt and Stark 1997). Moreover, the mental image is fixed in allocentric space (see section 5.5.1)—its natural inclination is to remain fixed in extrapersonal space when you turn your head or your body and even when you walk around it. Also as noted earlier, one of the purposes of the depictive image display is to provide a place where the
experienced information could be depicted as a panorama with gaps filled in, and where mental scanning might take place (since it seems that "mental scanning" can occur even off the central "foveal" region of the mental image; see Kosslyn, Ball, and Reiser 1978). None of these phenomena is compatible with the retinotopic nature of activity in V1.

(2) Displays in visual cortex (V1) are two-dimensional Mental images are three-dimensional, not only in their phenomenology, but also because all objective mental imagery phenomena involving distances or angles produce the same results if they are done in depth as in the plane.10 To suggest that the display in cortex depicts depth the way we might depict it in a drawing (by using an isometric or perspectival convention) is to miss the basic fact that the depictive image is supposed to be the interpretation as ex-

10. Oddly enough, the dimensionality of the display has even been used to support the pictorial view. It has been suggested that the display is able to represent 3-D shapes in some unspecified way (orthographic or isometric projection?), but that it is incapable of representing 4-D shapes and that this is an inherent constraint of the display itself, which accounts for our inability to imagine 4-D objects (Kosslyn 1981). This ignores the simple fact that we do not know what a 4-D object would look like, which by itself is all we need to explain why we can't imagine 4-D objects. Some physicists think that objects in fact are 4-D (or 6-D or higher according to string theory)—in which case there is no problem imagining them, since they look like the ordinary things we see around us!
experienced, not a figure from which such an interpretation is to be derived. If it were not, then we would require yet another form of representation for the interpreted mental image. Since mental scanning and rotation experiments can be done in three dimensions, this new form of representation would have to be depictive as well, so the 3-D problem would still be with us.

(3) The same cortical display cannot be used by vision and imagery If the same V1 display area were used for vision and imagery then it should be possible to superimpose visual and imagined images to get a composite. There have been a few claims of this sort, though none of them withstands scrutiny (see Pylyshyn 2003, section 6.5). If images and percepts both involved patterns on V1 they would be indistinguishable, or at least interpretable in similar ways. But images painted on the retina (e.g., afterimages) work very differently from images created by mental imagery. For example, Emmert’s law holds of afterimages but not of mental images. Emmert’s law says that the apparent size of a pattern projected onto the retina depends on how far away the background is. For a given retinal image, if you look far into the distance the image will appear very large, but if you look at a surface close by it will look small. Although pictorialists are always quick to deny the literal interpretation of a picture displayed in V1, Emmert’s law constitutes a serious challenge to any homeomorphic mapping of retinal onto cortical topography. As long as the image preserves relative size of objects (which has been the explicit assumption of image-scanning experiments), then it should be equivalent to a retinal image of a certain retinal size and Emmert’s law ought to hold. It’s puzzling why pictorialists ignore this particular problem.

(4) Images are not (re)perceived by early vision If images are projected onto V1, then perceiving information from an image should not be very different from reading it from a display. But images cannot be scanned freely; they cannot be freely reinterpreted visually; they do not show signature properties of vision (ambiguity, bistability, visual illusions, apparent motion). Most important (and probably the main reason behind all the above problems) is that images are intentional objects—they are conceptual interpretations, not raw sensory signals. If an image of a line drawing of an ambiguous figure is imagined, it does not switch between interpretations, because it already is an interpretation—it is a representation of a 3-D object and it does not change as you look at it (at least it does not change for the same reason that corresponding visual displays change—i.e., because
If I ask you to imagine two identical parallelograms, one directly above the other, and to connect each vertex of one with the corresponding vertex of the other by a vertical line, no amount of gazing at this in your mind’s eye will enable you to see what you would automatically see if you drew it. In addition it seems clear that although you can carry out certain kinds of reinterpretations of geometrical patterns in your image (e.g., you can tell that if you rotate an upper-case D by 90° counterclockwise and attach it to a J the result will look somewhat like an umbrella), you cannot perform a visual reinterpretation (reperception), as Peter Slezak has shown (Slezak 1992, 1995; for further discussion of this issue and other related experiments, see Pylyshyn 2002a).

There is also evidence (some of it described in chapter 1 of Pylyshyn 2003) that images retrieved from memory or created from descriptions do not function the way displays on the fovea do. Even with visual patterns that are too large to be accommodated in one fixation (e.g., if they exceed the visual angle of the fovea—just under 2°, or about the angle covered by your thumb at arms’ length) the part that is off-fovea shows signs of being already interpreted. For example, if you see a reversing figure such as a Necker cube or an impossible drawing (e.g., the “devil’s pitchfork”) that is elongated so that some of it is off-fovea, the spontaneous reversals are not observed nor is the conflict caused by the mismatch of two local views that occurs with “impossible figures.” Similarly, the famous eye-of-the-needle or anorthoscope presentation, in which a slit is moved back and forth in front of a figure (or a figure is moved back and forth behind a stationary slit), does not yield a true percept with the signature properties of visual perception. Moreover, the ability to recognize the shape in the anorthoscope is sensitive to the way the figure can be decomposed. If the figure is one that requires a larger number of line labels to be held in memory during the traverse of the slit then it will not be readily seen (see figure 4.5). In fact quite a few phenomena suggest that off-retinal figures are treated differently than those on the retina—they show more general-memory properties (e.g., they depend on the number of items held in memory) that are the hallmarks of constructed and interpreted figures.

11. What this describes is a Necker cube such as the one in the left panel of figure 4.1. Even if you guessed that it was a Necker cube it would not automatically turn into a 3-D object and involuntarily switch between its two perceptual interpretations (although if you knew about Necker cubes you could switch interpretations voluntarily).
4.6 Does Phenomenal Appearance Correspond to a Level of Representation?

As the examples listed earlier show, what is typically referred to as the “appearance” of the perceived world includes not only the operation of the perceptual system, but also of our beliefs and expectations and folk theories, and it incorporates these to a much more profound degree than generally believed. Our perceptual experience is the experience of seeing familiar people and things, not of seeing surfaces, textures, shapes, and colors. Our experience of our own actions reflects what we believe about the agency (or authorship) of the actions, and so on. In other words, experiences are generally experiences of interpreted sensory information. A dramatic illustration of this is what happens under hypnosis, during which it seems that even the experience of pain (or at least one’s involuntary

12. There are also intransitive conscious experiences that are not experiences of something or other, but merely floating experiences, such as sadness, dizziness, or free-floating anxiety. Since I am here concerned with the use of the content of experiences in developing theories of perception I confine myself to transitive experiences (experiences as of some sensory input).
reactions to it) can be altered. Thus we have every reason to be skeptical about what our subjective experience reveals about the information that is functional in the perceptual process. An even more serious problem with the use of conscious contents for inferring the processes underlying perception is that there is no room in phenomenology-based theories for the growing evidence of vision-without-awareness mentioned in section 4.2.1, including change blindness, inattentional blindness, visuomotor control without conscious awareness, blindsight, visual agnosia, and disorders of visual-motor coordination, as well as other sources of behavioral and neuroscience data.

I am not suggesting that the perception itself is contaminated by expectations and beliefs—as proposed by many writers in the past, including those of the “New Look” movement that dominated perceptual theorizing in the 1960s and later (Bruner 1957), and by linguist Benjamin Lee Whorf (Whorf 1958) and his latter-day followers (Gumperz and Levinson 1996) who claimed that language and culture determine how we perceive the world (the Sapir-Whorf Hypothesis). Quite the contrary, I have argued repeatedly that a major part of what we call visual perception is cognitively impenetrable (Pylyshyn 1999, 2003). The claim I am making now is that the content of one’s conscious experience is the result of these manifold influences, which is a very different claim. Of course if you think that one’s phenomenal experience in seeing the world is constitutive of visual perception, then this does mean that seeing is cognitively penetrable to the extent that the content of one’s visual experience is penetrable. The widespread assumption that what we see is given by how things look, or that how we perceptually experience the world defines what we mean by perception, may explain why the recent evidence of unconscious perception or of inconsistencies between how things look and what information is made available to other aspects of cognition (such as motor control) has attracted so much attention. I am not suggesting that we ignore perceptual experience (as behaviorists advocated). Not only would this be impossible, given that our conscious experiences of seeing, hearing, touching, smelling, and so on present deep scientific puzzles themselves; perceptual experience is also (and will continue to be) one of the main sources of evidence about perception. The alternative, rather, is to take conscious experience as one of many fallible sources of evidence concerning perception, which may in fact indicate not only what perception yields, but also the process by which we interpret incoming information. This also suggests that we may need to develop methodologies to take cognizance of the multifaceted origins of perceptual experience, the way that signal detection theory provides
a methodology for separating sensitivity and response bias in certain decision-making situations (Swets 1964).

Although there are some differences between consciously held beliefs and perceptions and those that are not conscious, these differences appear to be contingent rather than principled, in the sense that most of the unconscious ones could have been conscious and in any case function in similar ways in perceptual information processing. The failure to find a specifically distinct role for the content of conscious experiences, as opposed to the information content of unconscious or unreported experiences in information-processing theories of perception, raises the question of whether the experienced content corresponds to a level or to a type of representation. Representations play a central role in explaining how cognitive processes work, and why they lead to the behaviors they do. If there were a type of representation that had a different role, that did not contribute to capturing systematic features of behavior, the question would naturally arise what role it does play in our theories. It may be that conscious contents do not constitute a distinct level of representation because they are a mixture of levels.

Consider, for example, that the inverted retinal image is not part of our experience (at least not as an inverted image), although inversion produced by special glasses is. The proximal (retinal) size of a tree, before it is adjusted for the perceived distance away, the whiteness or color of a surface before it is adjusted for the perceived ambient light (see Gilchrist 1977), and other preconstancy properties are not part of our experience, but they nevertheless function in information processing the same way as features that we are conscious of (in fact they would continue to function the same way if you did notice them and they became conscious, as happens when one is taught to draw). It is also possible that preconstancy unconscious states such as those just alluded to are not representations at all, or if they are representations they may be conceptual but subpersonal, in that they involve concepts that do not enter into general reasoning because they remain inside encapsulated modules (see section 3.2.2). It seems that what we experience is a mixture of sensory information, subpersonal representations, and some high-level cognitive recognitions (i.e., familiar people, places, things, and events), so it is likely that experiential content draws on many types and levels of representation.

Although conscious perceptual experience cannot be discounted in the study of perception, neither can one assume that the experience itself is to be taken at face value as an indicator of the nature of a functional mental state—a state that plays a role in the explanation of how perception works.
The question of how to interpret a particular observation can be resolved only as we build more successful theories. The situation here is very similar to that which linguists faced some sixty years ago. Intuitions of grammatical structure, which play a central role in linguistics, similar to the role played by conscious contents in perception, resulted in many disagreements early on. Take Chomsky’s famous sentence “Colorless green ideas sleep furiously,” which was meant to illustrate the distinction between grammaticality and acceptability judgments. The question of whether this sentence was grammatical led to arguments in which people provided interpretations of the sentence (you can always interpret a sentence, no matter how bizarre it is). This and other such linguistic intuitions were debated because what constitutes grammaticality, as opposed to acceptability, is not given by intuition alone but must await the development of the theory itself. As generative linguistics became able to capture a wider range of generalizations, it found itself relying just as much on linguistic intuitions. What changed is that the use of the intuitions was now under the control the evolving theories. Even such general questions as whether a particular intuitive judgment was relevant to linguistics became conditioned by the theory itself.13 So, as theories of vision formulate general principles, the theories themselves will direct us to the interpretation of evidence from conscious experience.

13. For example, it was once suggested that grammaticality may not be effectively computable because the judgment of which of the two sentences “I am having trouble choosing among/between $P$” is grammatical is undecidable (since for a general number-theoretic predicate $P$ it is undecidable which numbers satisfy it). One answer to this apparent conundrum was that this just shows that although the choice between “among” and “between” may be taught in grammar classes, it is not a syntactic distinction after all.