

Vision, Knowledge, and the Mystery Link

John L. Pollock¹
Department of Philosophy
University of Arizona
pollock@arizona.edu
<http://www.u.arizona.edu/~pollock>

Iris Oved²
Department of Philosophy
Rutgers University
irisoved@eden.rutgers.edu

1. Perceptual Knowledge

Imagine yourself sitting on your front porch, sipping your morning coffee and admiring the scene before you. You see trees, houses, people, automobiles; you see a cat running across the road, and a bee buzzing among the flowers. You see that the flowers are yellow, and blowing in the wind. You see that the people are moving about, many of them on bicycles. You see that the houses are painted different colors, mostly earth tones, and most are one story but a few are two story. It is a beautiful morning. Thus the world interfaces with your mind through your senses.

There is a strong intuition that we are not disconnected from the world. We and the other things we see around us are part of a continuous whole, and we have direct access to them through vision, touch, etc. However, the philosophical tradition tries to drive a wedge between us and the world by insisting that the information we get from perception is the result of inference from indirect evidence that is about how things look and feel to us. The philosophical problem of perception is then to explain what justifies these inferences.

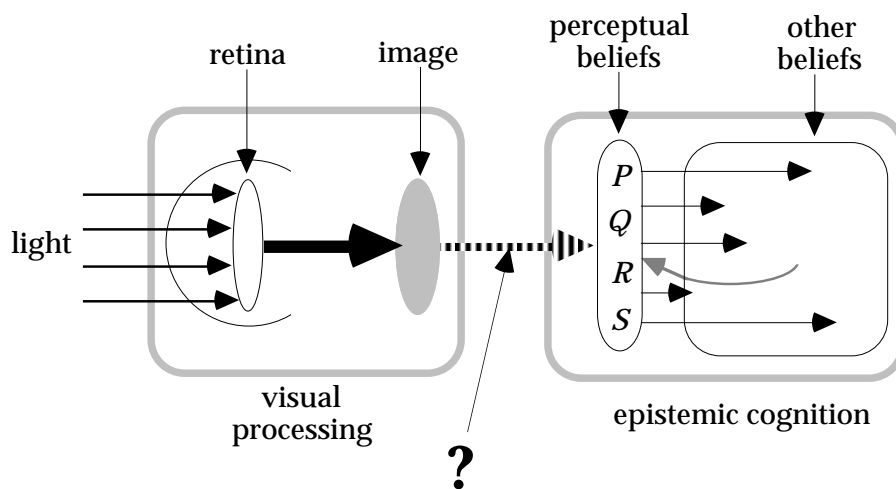


Figure 1. Knowledge, perception, and the mystery link

We will focus primarily on visual perception. Figure one presents a crude diagram of the cognitive system of an agent capable of forming beliefs on the basis of visual perception. Cognition

¹ Supported by NSF grant no. IRI-IIS-0080888.

² Supported by a grant from Rutgers University.

begins with the stimulation of the rods and cones on the retina. From that physical input, some kind of visual processing produces an introspectible visual image. Note that when we talk about the visual image, we are following the standard philosophical usage. It is common for psychologists to use the term “visual image” to mean the image projected on the retina. We use the term instead to mean the introspectible (mental) image from which epistemic cognition derives beliefs. In response to the production of the visual image, the cognizer forms beliefs about his or her surroundings. Some beliefs — the *perceptual beliefs* — are formed as direct responses to the visual input, and other beliefs are inferred from the perceptual beliefs. We have drawn inferential arrows from the perceptual beliefs to other beliefs, but there are also arrows coming back the other way, as signified by the grey arrow. We will discuss these arrows in sections eight and nine. In addition, it is incontrovertible that there is some kind of link between the visual image and the perceptual beliefs. The latter are, at the very least, caused or causally influenced by having the image. This is signified by the dashed arrow marked with a large question mark. We will refer to this as the *mystery link*.

Figure one makes it apparent that in order to fully understand how knowledge is based on perception, we need three different theories. First, we need a psychological theory of visual processing that explains how the introspectible visual image is produced from the stimulation by light of the rods and cones on the retina. Second, we need a philosophical theory of higher-level epistemic cognition, explaining how beliefs influence each other rationally. We will identify this with an epistemological theory of reasoning. We will assume without argument that this involves some kind of defeasible reasoning.³ These first two theories are familiar sorts of theories. To these we must add a third theory — a theory of the mystery link that connects visual processing to epistemic cognition. Philosophers have usually had little to say about the mystery link, contenting themselves with waving their hands and pronouncing that it is a causal process producing input to epistemic cognition. However, the main contention of this paper will be that there is much more to be said about the mystery link, and a correct understanding of it severely constrains what kinds of epistemological theories of perceptual knowledge can be correct.

This paper will begin by looking briefly at epistemological theories of perceptual knowledge. We will present an argument for “direct realism”, which we endorse, and then present what seems to be a clear counterexample to direct realism. This will lead us into a closer examination of vision and the way it encodes information. From that we will derive an account of the mystery link. It will then be shown that this theory of the mystery link provides machinery for constructing a modified version of direct realism that avoids the counterexample and makes visual knowledge of the world explicable.

2. Direct Realism

Historically, most epistemological theories were *doxastic theories*, in the sense that they endorsed the *doxastic assumption*. That is the assumption that the justifiability of a cognizer’s belief is a function exclusively of what beliefs she holds. Nothing but beliefs can enter into the determination of justification. This leads to either a *foundations theory*, according to which some beliefs are basic and do not depend for their justification on other beliefs, or a *coherence theory* according to which no beliefs have a privileged status and every belief is potentially dependent on every other belief for its justification. The basic beliefs of a foundations theory must be *self-justified* in the sense that they are justified (at least defeasibly) by the mere fact that the cognizer holds them. On a doxastic theory, this is the only alternative to their being inferred from other beliefs, because

³ See Pollock (1995) and (2002) for accounts of our preferred theory of defeasible reasoning. One version of this theory has been implemented in OSCAR. For up to date information on OSCAR and the implementation of the architecture, go to <http://www.u.arizona.edu/~pollock>.

nothing other than beliefs can be relevant. As a coherence theory accords no beliefs a privileged status, it must either take all beliefs to be (defeasibly) self-justified, or deny that any are.

In filling out this picture, a foundations theory typically regards the perceptual beliefs as being about the perceiver's visual image. They are about qualia, or sense-data, or at the very least about how things look to the cognizer. On a coherence theory, the perceptual beliefs may instead be regarded as ordinary everyday beliefs about the physical objects the cognizer is seeing. Regardless of your theoretical inclinations, perceptual beliefs play a novel role in epistemic cognition, because they represent the introduction of new information. Although they may be influenced by inferences from other beliefs, they cannot be *inferred from* beliefs the cognizer already has. When you see a new object, you cannot infer beforehand what its color is going to be. You have to look and see. So perceptual beliefs are not adopted on the basis of inference from previously held beliefs.

If perceptual beliefs are not inferred from other beliefs, that cannot be the source of their justification. But on a doxastic theory, the justification of a belief cannot depend on anything other than the cognizer's beliefs. If the justification of a perceptual belief does not depend on the agent's other beliefs, then it can only depend on itself. In other words, the belief must be self-justified, in the sense that merely having the belief constitutes a source of justification. Such beliefs could be *incorrigibly justified*, in the sense that other beliefs are not even negatively relevant to the justification of the beliefs, or they could be *prima facie justified*, in the sense that in the absence of other beliefs that provide defeaters, the perceptual beliefs are justified. The logical geography of these distinctions is explored in depth in Pollock & Cruz (1999).

Let us consider the proposal that perceptual beliefs are self-justified. How reasonable is this claim? That depends on what kinds of beliefs are taken to be perceptual beliefs. Historical foundations theories typically took perceptual beliefs to be about the cognizer's perceptual experience. Let us call these *appearance beliefs*. It was often argued that appearance beliefs are incorrigible (Ayer, Carnap, C.I. Lewis), but more moderate foundationalists might claim that they are only *prima facie* justified (Pollock 1974). Pollock (1986) and Pollock & Cruz (1999) argued that they are not self-justified in any way, but we need not rest our rejection of these theories on that basis.

The simple problem for the foundationalist is that perceptual beliefs, as the first beliefs the agent forms on the basis of perception, are not generally about appearances. It is rare to have any beliefs at all about how things look to you. You normally just form beliefs about ordinary physical objects. When you look around the room you see people sitting at a table writing on notepads or on personal computers. When you look out the window you see buildings and trees and students milling about. It never occurs to you to think, "There is an oblong red blob in the upper right hand corner of my visual field." You *can* form such beliefs, but it requires a deliberate shift of attention. So even if beliefs with such peculiar contents should turn out to be self-justified (we believe they are not), this would not explain how the more ordinary perceptual beliefs that are about physical objects get justified when we have no appearance beliefs that support them inferentially.

In the normal case, perceptual beliefs — the first beliefs formed on the basis of perception — are about the physical objects we see around us. You see a table and judge that it is round, you see an apple on the table and judge that it is red, etc. Can such beliefs be self-justified? They cannot. The difficulty is that the very same beliefs can be held for non-perceptual reasons. While blindfolded, you can believe there is a red apple on a round table before you because someone tells you that there is, or because you looked in other rooms before entering this one and saw tables with apples on them. Worse, you can hold such beliefs unjustifiably by believing them for inadequate reasons. Wishful thinking might lead to such a belief, or hasty generalization. These are not cases in which you have good reasons that are defeated. These are cases in which you lack good reasons from the start. If, in the absence of defeaters, these beliefs can be unjustified, it follows that they are not self-justified.

It seems clear that what makes perceptual beliefs justified in the absence of inferential support

from other beliefs is that they *are* perceptual beliefs. That is, they are believed on the basis of perceptual input. The same belief can be held on the basis of perceptual input or on the basis of inference from other beliefs. When it is held on the basis of perceptual input, that makes it justified unless the agent has a reason for regarding the input as non-veridical or otherwise dubious in this particular case. But this is not the same thing as being self-justified. Self-justified beliefs are justified without any support at all, perceptual or inferential. But these beliefs need support, so they are not self-justified.

It is tempting to insist that these beliefs are self-justified *when they are held on the basis of perception*. But that is not something that a doxastic theory can say. According to a doxastic theory, justification cannot be a function of whether the beliefs are held on the basis of perception. It can only be a function of the agent's other beliefs. We want to say that the perceptual experience itself is what justifies your belief that you see a red apple on a round table, but to do that we must reject the doxastic assumption.

What is it about my perceptual experience that justifies me in believing, for example, that the apple is red? It seems clear that the belief is justified by the fact that the apple looks red to me. In general, there are various states of affairs *P* for which visual experience gives us direct evidence. Let us say that the relevant visual experience is that of *being appeared to as if P*. Then *direct realism* is the following principle:

(DR) For appropriate *P*'s, if *S* believes *P* on the basis of being appeared to as if *P*, *S* is defeasibly justified in doing so.

Direct realism is "direct" in the sense that our beliefs about our physical surroundings are the first beliefs produced by cognition in response to perceptual input, and they are not inferred from lower-level beliefs about the perceptual input itself. But, according to direct realism, these beliefs are not self-justified either. Their justification depends upon having the appropriate perceptual experiences. Thus the doxastic assumption is false.

Direct realism is, in part, a theory about the mystery link. It tells us, first, that perceptual beliefs are ordinary physical-object beliefs, and second that the mystery link is not just a causal connection — it conveys justification to the perceptual beliefs. It does not, however, tell us how the latter is accomplished. For the most part, it leaves the mystery link as mysterious as it ever was. This gives rise to an objection that is often leveled at direct realism. The objection is that perceptual beliefs involve concepts, but the visual image is non-conceptual, so how can the image give support to the perceptual belief?⁴ We are not sure what it means to say that the image is or is not conceptual, but this objection can be met in a preliminary way without addressing that question. If there is a problem here, it is not a problem specifically for direct realism. It is really a problem about the mystery link. If it is correct to say that the image is non-conceptual but beliefs are conceptual, then on *every* theory of perceptual knowledge, what is on the left of the mystery link is non-conceptual and what is on the right is conceptual. The problem is then, how does the mystery link work to get us from the one to the other? This is just as much a problem for the foundationalist who thinks that perceptual beliefs are about the image, because we still want an explanation for how cognition gets us from the image to the beliefs, be they about the image or about objects in the world. Clearly it does, so this cannot be a decisive objection to any theory of perceptual knowledge, and it has nothing particular to do with direct realism. It is instead a puzzle about how the mystery link works. Hopefully, it will be less puzzling by the end of the paper.

Direct realism has had occasional supporters in the history of philosophy, perhaps most notably Peter John Olivi in the 13th century and Thomas Reid in the 18th century. But the theory was largely ignored by contemporary epistemologists until Pollock (1971, 1974, 1986) resurrected

⁴ This objection is often associated with Sellars (1963). See also Sosa (1981) and Davidson (1983).

it on the basis of the preceding argument. The name of the theory was suggested by Anthony Quinton (1973), although he did not endorse the theory. In recent years, direct realism has gained a small following.⁵ The argument just given in its defense seems to us to be strong. However, in the next section we will present a counterexample that appears initially devastating. That will lead us to a closer examination of the mystery link, and ultimately to a formulation of direct realism that avoids the difficulty.

3. A Problem for Direct Realism

The argument for direct realism seems quite compelling. Surely it is true that perceptual beliefs are justified *by being* perceptual beliefs. That is, they are justified by being beliefs that are held on the basis of appropriately related perceptual experiences. And it appears that this is what (DR) says. (DR) has most commonly been illustrated by appealing to the following instance:

(RED) If *S* believes that *x* is red on the basis of its looking to *S* as if *x* is red, *S* is defeasibly justified in doing so.

However, it now appears to us that the principle (RED) cannot possibly be true. Let us begin by distinguishing between precise shades of red (“color determinates”) and the generic color red (“color determinables”, composed of a disjunction of color determinates). The principle (RED), if true, should be true regardless of whether we take it to be talking about precise shades of red or generic redness. The problems are basically the same for both, but they are more dramatic for the case of precise shades of red.

The principle (RED) relates the concept *red* to a way of looking — an *apparent color*. It tells us that something’s having that apparent color gives defeasible support for the conclusion that it is red. Defeasible support is normally understood as support that arises without requiring an independent argument. Thus, for instance, the principle (RED) enables us to avoid having to discover inductively that objects with this apparent color tend to be red. Indeed, the claim of the direct realist has been that it would be in principle impossible to discover this (Pollock 1986, Pollock & Cruz 1999), because to do that we would have to have some other way of determining that things are red and then compare red things and things with the apparent color red. However, (RED) is supposed to describe our fundamental access to the redness of things. We do not have independent access to whether something is red. Thus if (RED) is to be a correct description of our epistemological access to whether objects are red, it must describe an essential feature of the concept *red*. That is, there must be an apparent color (a way of looking) that is logically or essentially connected to the concept *red*.

To most philosophers, this will not seem to be a surprising requirement. It is quite common for philosophers to think that the concept *red* has as an essential feature a specification of how red things look. For instance, Colin McGinn (1983) writes, “To grasp the concept of red it is necessary to know what it is for something to look red, since this latter constitutes the satisfaction condition for an object’s being red.” However, for reasons now to be given, this seems to us to be false.

In the philosophy of mind there has been much discussion of the so-called “inverted spectrum problem”, and debate about whether it is possible. We want to call attention here to a variant of this that is not only possible but common. This is the “sliding spectrum”. Some years ago, one of us (not Iris) underwent cataract surgery. In this surgery, the clouded lens is surgically removed from the eye and replaced by an implanted silicon lens similar to a contact lens. When the operation was performed on the right eye, the subject was amazed to discover that everything looked blue through that eye. Upon questioning the surgeon, it was learned that this is normal.

⁵ See, for example, Pryor (2000) and Huemer (2001).

In everyone, the lens yellows with the passage of time. In effect, people grow a brownish-yellow filter in their eye, which affects all apparent colors, shifting them towards yellow. This phenomenon is so common that it has a name in vision research. It is called “photoxic lens brunescence” (Lindsey and Brown 2002). For a while after the surgery, everything looked blue through the right eye and, by contrast, yellow through the left eye. Then when the cataract-clouded lens was removed from the second eye a few weeks later, everything looked blue through both eyes. But now, with the passage of time, everything seems normal.

Immediately following surgery, white things look blue and red things look purple to a cataract patient. After the passage of time, the patient no longer notices anything out of the ordinary. What has happened? There are two possibilities. The simplest explanation is that the subject has simply become used to the change, and now takes things to look red when they look the way red things now look to him. On this account, in everyone, the way red things look changes slowly over time as the eye tissues yellow, but because the change is slow, the subject does not notice it. Then if the subject undergoes cataract surgery, the way red things look changes back abruptly, and the subject notices that. But after a while he gets used to it, and forgets how red things looked before the operation.

There is another possibility, however. Perhaps the brain compensates for the shift so that as the eye tissues slowly yellow, red things continue to look the same, and although red things look different after the cataract operation, with the passage of time they go back to looking the way they did before. This would be a form of *color constancy*. However, color constancy is a much misunderstood phenomenon. It is popularly alleged that because of color constancy, colors look the same when viewed under different lighting conditions, e.g., tungsten light, fluorescent light, daylight, and shadow. For instance, Yullie & Ullman (1990) claim, “without this effect the perceived color of a red London bus would change strongly whenever the bus turns the corner from a shady side street into the sun.” Note that this claim is ambiguous. When we talk about the perceived color of the bus, we could be talking about the color we *judge* the bus to be on the basis of perception, or we could be talking about how the bus *looks to us*. Only the latter would have the result that brunescence does not affect the way colors look. So consider the suggestion that the way a red London bus *looks* does not change when it passes from shade into sunshine. Surely this is wrong. Figure 2 is a photograph of some London buses moving in and out of the shade. Does anyone seriously think the colors look the same throughout? It is important to distinguish between our epistemic judgment about what color the bus actually is and the phenomenal experience of how it looks to us. The buses look very different, but we are able to compensate for that difference epistemically and judge that the objective color has not changed.

What figure two illustrates is that there is a difference between being able to judge that colors are the same and their having the same phenomenal appearance. The evidence cited in scientific studies of color constancy is typically that people make the same color judgments regarding familiar objects even when illumination varies.⁶ But this does not show that there is no phenomenal difference. Furthermore, even if it is sometimes true that there is no phenomenal difference, it is certainly not universally true. For instance, have you ever tried picking out paint colors by looking at paint chips in a store with fluorescent lighting? It cannot be done. You have to take the paint chips back to the room to be painted and see what they look like there. It is simply false that they look the same under both lighting conditions. To take another example, when you don or remove sunglasses with colored lenses, you can certainly notice the way the colors of things seem to change.

⁶ This underlies Land’s experiments with *mondrians* — pictures made up of rectangular patches of different colors. Land & McGann (1971) argued that the perceived colors of the patches do not change much as the illumination changes.



Figure 2. London buses. (Photo by Matthew Wharmby)

The upshot is that even if it were true that changes in illumination sometimes leave phenomenal color appearances unchanged (and it isn't clear that they ever do), they do not always do so. Consequently, this gives us no reason to think that although eye tissues yellow slowly over time, this has no effect on the way colors look to us. On the other hand, the only reason given so far for thinking that brunescence does alter the way colors look is common sense. Can we do better? It turns out that there is hard scientific data that can be used to support the conclusion that brunescence has lasting effects on the way things look to us, even if we don't notice the effects. Fairchild (1998) writes,

However, we are all looking at the world through a yellow filter that not only changes with age, but that significantly differs from observer to observer. The effect is most noticeable when performing critical color matching or comparing color matches with other observers. It is particularly apparent with purple objects. An older lens absorbs most of the blue energy reflected from a purple object but does not affect the reflected red energy, so older observers tend to report that the object is significantly redder than reported by younger observers.

Brunescence has some surprising side-effects. Lindsey and Brown (2002) write,

Many languages have no basic color term for "blue". Instead, they call short-wavelength stimuli "green" or "dark". The article shows that this cultural, linguistic phenomenon could result from accelerated aging of the eye because of high, chronic exposure to ultraviolet-B (UV-B) in sunlight (e.g., phototoxic lens brunescence). Reviewing 203 world languages, a significant relationship was found between UV dosage and color naming: In low-UV localities, languages generally have the word "blue"; in high-UV areas, languages without "blue" prevail. Furthermore, speakers of these non-"blue" languages often show blue-yellow color vision deficiency.

The observations about language are interesting, but the main thing to take away from these quotes is that brunescence affects observers' abilities to discriminate between colors. People suffering from brunescence cannot discriminate as many different phenomenal appearances. But this means that their phenomenal experience is different from what it was before brunescence. Hence, the phenomenal appearance of colors has changed.

There are other kinds of color shifts to which human perception is subject. In what is called

the *Bezold-Brücke effect*, when levels of illumination are increased, there is a shift of perceived hues such that most colors appear less red or green and more blue or yellow. The result is that the apparent colors of red things differ in different light even when the relative energy distribution across the spectrum remains unchanged.

There are numerous other well known phenomena. In what is known as *simultaneous color contrast*, the apparent colors of objects vary as the color of the background changes. For example, one of us reports with confidence that most women have noticed how a shade of lipstick looks different on different people. In *chromatic adaptation*, looking at one color and then looking at a contrasting color changes the second apparent color. This is illustrated by afterimages.

All of these phenomena illustrate that there is no single phenomenal color that red things normally elicit. Apparent colors undergo pervasive and systematic variations, and in some cases, e.g., brunescence, the changes can be dramatic. After cataract surgery, no matter how broad your generic concept of *red*, red things may not look red to you. (Even more obviously, white things do not look white — they look blue.) If there is a phenomenal color that is logically connected to the concept *red*, most red objects will fail to elicit it most of the time. If we consider a precise shade of red — call it *red**, and *looking red** consists of eliciting a particular phenomenal color, then even for normal subjects *red** things rarely look *red**. Byrne & Hilbert (2003) acknowledge that this holds for color determinates, but deny that it holds for color determinables. And it is true that for most subjects, we may be able to avoid this result if we consider broad generic color categories like *red*. But for subjects with advanced brunescence, there isn't even a broad generic phenomenal color such that both (1) things tend to be red iff they look that color and (2) it was also true before the onset of brunescence that things tended to be red iff they looked that color. Furthermore, we all suffer from varying degrees of brunescence, and this does not affect our ability to perceive colors.

These psychological phenomena produce variations within a single subject. But just thinking about all the things that can affect how colors look makes it extremely unlikely that red things will normally look the same to different subjects. Between-subject variations seem likely if for no other reason than that there are individual differences between different people's perceptual hardware and neural wiring. No two cognizers are exactly the same, so why should we think things are going to look exactly the same to them?⁷ We need not merely speculate. There is experimental data that strongly suggests they do not. This turns upon the notion of a *unique hue*. Byrne and Hilbert (2003) observe,

“There is a shade of red (“unique red”) that is neither yellowish nor bluish, and similarly for the three other *unique hues* — yellow, green, and blue. This is nicely shown in experiments summarized by Hurvich (1981, Ch. 5): a normal observer looking at a stimulus produced by two monochromators is able to adjust one of them until he reports seeing a yellow stimulus that is not at all reddish or greenish. In contrast, every shade of purple is both reddish and bluish, and similarly for the other three *binary hues* (orange, olive, and turquoise).

But what is more interesting for our purposes is that different people classify different colors in this way. As Byrne and Hilbert go on to observe:

There is a surprising amount of variation in the color vision of people classified on standard tests ... as having “normal” color vision. Hurvich et al. (1968) found that the location of “unique green” for spectral lights among 50 subjects varied from 590 to 520nm. This is a large range: 15nm either side of unique green looks distinctly bluish or yellowish. ... A more recent study of color matching results among 50 males discovered that they divided into two broad groups, with the difference between the groups traceable to a

⁷ Block (1990) speculates similarly.

polymorphism in the L-cone photopigment gene (Merbs & Nathans 1992). Because the L-cone photopigment genes are on the X chromosome, the distribution of the two photopigments varies significantly between men and women (Neitz & Neitz 1998).

The upshot of the preceding observations is that there is no way of looking — call it *looking red* — such that objects are typically red iff they look red. In fact, it is likely that for any apparent color we choose, objects that are red will typically *not* look that color. If our judgments of color were based on principles like (RED), we would almost always be led to conclude defeasibly that red objects are not red. Furthermore, it follows from direct realism that there would be no possible way for us to correct these judgments by discovering that they are unreliable, because any other source of knowledge about redness would have to be justified inductively by reference to objects judged red using the principle (RED). It seems apparent that the principle (RED) cannot be a correct account of how we judge the colors of objects. But (RED) also seems to be a stereotypical instance of direct realism. It is certainly the standard example that Pollock (1986) and Pollock & Cruz (1999) used throughout their defense of direct realism. Thus (DR) itself seems to be in doubt.

It might be supposed that there is something funny about color concepts, and these problems will not recur if we consider some of the other kinds of properties about which we make perceptual judgments. These would include shapes, spatial orientations, the straightness of lines, the relative lengths of lines, etc. But in fact, analogues of the above problems arise for all of these supposedly perceivable properties. For example, anyone who was very nearsighted as a child and whose eyes were changing rapidly has probably had the experience of getting new glasses and finding that straight lines looked curved and when they walked forwards it looked to them like they were stepping into a hole. This is a geometric analogue of brunescence. Less dramatically, most of us suffer from varying degrees of astigmatism, which has the result that straight lines are projected unevenly onto the surface of the retina, with presumed consequences for our phenomenal experience. Furthermore, the severity of the astigmatism changes over time. In addition, the lenses in our eyes are not very good lenses from an optical point of view. They would flunk as camera lenses. In particular, they suffer from large amounts of barrel distortion, where parallel lines are projected onto the retina as curved lines that are farther apart close to the center of the eye. The amount of barrel distortion varies from subject to subject, so very likely the looks of geometric figures, straight lines, etc., vary as well.

4. The Visual Image

Our solution to this problem is going to be that there is a way of understanding the principle (DR) of direct realism that makes both it and (RED) true. The above problem arises from a misunderstanding of what it is to be “appeared to as if *P*”, and in particular what it is for it to look to one as if an object is red. To defend this claim, we turn to an examination of the visual image. We will investigate what the visual image actually consists of, and how it can give rise to perceptual beliefs. In effect we are investigating the mystery link.

The classical picture of the visual image (exemplified, for example, by C. D. Broad, Nelson Goodman, C. I. Lewis, G. E. Moore, Bertrand Russell, Rudolph Carnap and A. J. Ayer at various stages) was, in effect, that it is a two-dimensional array of colored pixels — a bitmap image.⁸ Then the epistemological problem of perception was conceived as being that of justifying inferences from this image to beliefs about the way the world is.

We can, in fact, think of the input to the visual system in this way. The input consists of the stimulation of the individual rods and cones arrayed on the retina. Each rod or cone is a binary

⁸ This view has recently been endorsed again, at least tentatively, by Bonjour (2001), and also by Kosslyn (1983).

(on/off) device responding to light of the appropriate intensity (and in the case of cones, light of the appropriate color). A bitmap image represents the pattern of stimulation. The human eye contains approximately 130 million rods and 7 million cones. The pixels in the bitmap are mostly uncolored (only the cones are sensitive to color). Most of the cones are located close to the center of the retina, with rods proliferating further out. Philosophers sometimes refer to this bitmap as “the retinal representation”, but for reasons that will become apparent later, we will reserve the term “representation” for higher-level mental items, including various constituents of the visual image.

Although this may be a good way to think of the input to the visual system, it does not follow that its output — the introspectible visual image — has the same form. Early twentieth century philosophers (and indeed, most early twentieth century psychologists) thought of the optic nerve as simply passing the pattern of stimulation on the retina down a line of synaptic connections to a “mental screen” where it is redisplayed for the perusal of epistemic cognition. Of course, this makes no literal sense. How does epistemic cognition peruse the mental screen — using a mental “eye” inside the brain? This picture is really just a reflection of the fact that people had no idea how vision works. It is the mystery link that takes us from the visual image to epistemic cognition, so what they were doing was packing all the interesting stuff into the mystery link and leaving its operation a complete mystery.

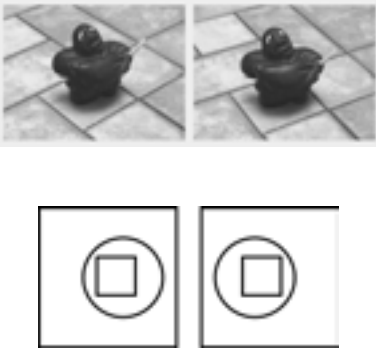


Figure 3. Stereograms. To see a three-dimensional image, hold your nose to the paper between the figures and let your eyes relax. You will see four images. Slowly move the page away from you until the middle two images merge. Then focus on them. The line-drawing stereogram can be viewed either by relaxing your eyes or by crossing your eyes.

The inadequacy of the “pass-through” conception of the visual image is obvious when we reflect on the fact that we have just one visual image but two eyes. The single image is constructed on the basis of the two separate retinal bitmaps. The two bitmaps cannot simply be laid on top of one another, because by virtue of being from different vantage points they are not quite the same. Nor can they be laid side by side in the mind. Then we would have two images. In fact, the difference between the two bitmaps is an important part of why we can see three-dimensional relationships between the objects we see. This is what underlies stereograms. Consider the two separate photographs (two-dimensional images) at the top of figure three. They are taken from slightly different angles to mimic the separation of the eyes. If you focus just right so that the images merge, you will see a three-dimensional image. The two line drawings in figure three also form a stereogram, and may be easier to see. These stereograms highlight the fact that the visual image is not a two-dimensional pattern at all. It is three-dimensional. On the one hand, it has to be, because there would be no way to merge the bitmaps from the two retinas into a single two-dimensional image. But on the other hand, in order to get a three-dimensional image out of

two two-dimensional bitmaps, a great deal of sophisticated computation is required. So far be it from mimicing the retinal bitmap, the visual image is the result of sophisticated computations that take the two separate retinal bitmaps as input. If the visual image is more sophisticated than the retinal bitmaps in this way, why shouldn't it profit from other sorts of computational massaging of the input data?

The study of vision has developed into an interdisciplinary field combining work in psychology, computer science, and neuroscience. Contemporary vision scientists now know a great deal about how vision works. Most contemporary theories of vision are examples of what are called "computational theories of vision", an approach first suggested in the work of J. J. Gibson (1966) and David Marr (1982), and developed in more recent literature by Irving Biederman (1985), Oliver Faugeras (1993), Shimon Ullman (1996), and others. On this approach, the visual system is viewed as an information processor that takes inputs from the rods and cones on the retinas and outputs the visual image as a structured array of mental representations. The visual system is understood as analyzing the stimulus in several stages, more or less following the original framework developed by Marr. For our purposes, the most important idea these theories share is that there are representations formed at each stage of visual processing, and the early representations are used in the production of the later representations. The following is a rough sketch of the stages and representations understood to be involved.

Input

The visual process is understood as starting with stimulation of the rods and cones on the retinas. These respond to light reflecting from objects in the world. However, the visual system does not compute the visual image on the basis of a single momentary retinal bitmap. We have high visual acuity over only a small region in the center of the retina called the *fovea*. In your visual field, the region of high visual acuity is the size of your thumbnail held at arm's length. To see this, fix your eyes on a single word on this page, and notice how fuzzy the other words on the page look. Now allow your eyes to roam around the page, as when looking at it normally, and notice how much richer and sharper your whole visual image becomes. In normal vision your eyes rarely remain still. The eyes make tiny movements (saccades) as the viewer scans the scene, and multiple saccades are the input for a single visual image. For another example of this, attend to your own eye movements as you are standing face to face with someone and talking to them. You will find your eyes roaming around your interlocutor's face, and you will have a sharp image of the face. Now focus on the tip of the nose and force your eyes to remain still. You will have a sharp image of the nose, but the rest of the face will be very fuzzy.

Apparently the visual image is the product of a number of retinal bitmaps resulting from multiple saccades. Think about what this involves. The visual system must somehow merge the information contained in these multiple bitmaps. The bitmaps cannot simply be laid on top of each other, because they are the result of pointing the eyes in different directions. Saccades involve movements of the eyes, which means that the oculomotor system that detects and controls these movements must provide spatial information to the visual system. Without such input, the visual system would not be able to merge the bitmaps. To illustrate this, jiggle one of your eyes by pulling on the outer corner of your eyelid with your finger, and notice that your visual image is shaky and blurry. Since the visual system is not taking into account the eye motions that result from your manual jiggling, the resulting motion across the retina is processed into the visual representation.

The visual system must make use of multiple saccades to get high resolution over more than a minute portion of the visual image, which means that the image is not computed on the basis of the momentary retinal bitmap. This is the reason we do not notice the retinal "blind spot" (the spot on the retina that carries no information because it contains the opening of the optic nerve). Another dramatic illustration of this occurs when you are riding in a car alongside a fence consisting of vertical slats with small openings between them. When you are stationary you

cannot see what is behind the fence, but when you are moving you may have a very clear image of the scene behind it. Momentary states of the retinal bitmap are the same whether you are stationary or moving. The fact that you can see through the fence when you are moving indicates that your visual representations are computed on the basis of a stream of retinal bitmaps extending over some interval of time.

The Primal Sketch

After merging the stream of retinal bitmaps, the next stage is to compute what Marr (1982) called *the primal sketch*. This represents basic features in the scene like edges, lines, line terminations, contours, and blobs. The system extracts these from the retinal bitmaps by computing patterns in the changes of activation across the retinal bitmaps. Much of this low-level processing is well understood, and can be performed by existing computer programs for image processing. For example, figure four illustrates the results of applying an edge detection algorithm using Gaussian filtering to a bitmap image (Marr & Hildreth 1980). The features computed at this level seem to be extracted quickly and easily, and the rest of the bitmap information seems to be thrown away.

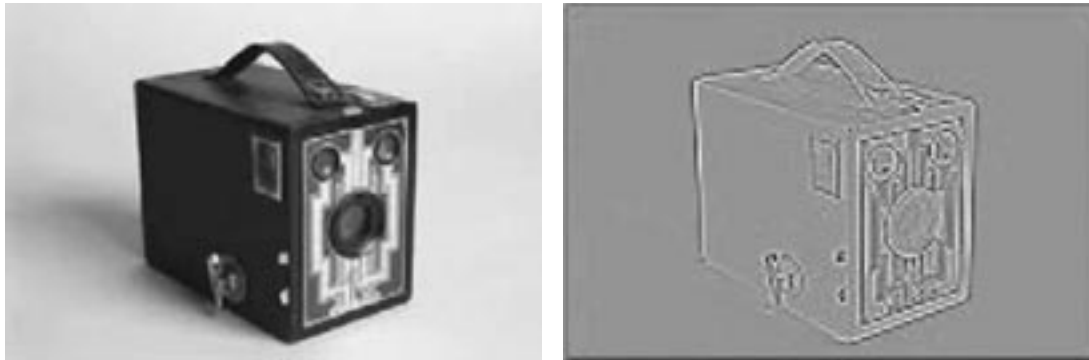


Figure 4. Edge detection by Gaussian filtering

The $2^{1/2}$ -D sketch

The representations of edges, lines, corners, curves, etc., computed in the primal sketch are passed to the next stage of processing. What Marr called the $2^{1/2}$ -D sketch encodes the orientation and depth of the lines, edges, and corners that have been detected. These representations are in terms of a viewer-centered coordinate frame, and include representations of distances and orientations of the features, as well as discontinuities in depth and surface orientation.

A major contribution to the perception of depth comes from the disparity between the two eyes, which take in light from slightly different perspectives. The visual system uses the lines, corners, etc., computed in the primal sketch to match the points in the images produced by the two retinas. This enables it to compute a representation of the distances of the points from the viewer. This is illustrated by the stereograms in figure three. In a stereogram, the photo on the left is identical to the one on the right except that it has been slightly displaced, roughly matching the displacement of light hitting the two eyes. Each image looks flat on its own, but if you focus your eyes past them so that they fuse into one, you see the objects as three-dimensional.

There are other cues used in computing representations of depth. Motion across the retina is one. Points that move across the retina quickly are represented as nearby, whereas points that move slowly are represented as further away. This works because of *motion parallax*. Texture gradients are also used — when a surface has a uniform texture or pattern, the patterns at more distant points project smaller images on the retina. The visual system interprets this as an indication of depth. Occlusion, relative size, and perspective are other sorts of information that is taken into account in the representation of depth. These are all illustrated in figure five.



Figure 5. House occluded by trees.

Objects and their parts

From feature and depth representations, the visual system computes representations of objects. A lot of information about the nature of this computation can be gleaned from the consideration of patients who have suffered neurological damage (e.g., strokes) that has rendered them unable to perform specific parts of the computation. Most of the objects we see are complex organizations of simpler objects — their parts. Apparently the visual system begins by computing representations of parts of objects, and then combines them to construct representations of objects. Evidence for this is provided by patients suffering from *dorsal simultanagnosia*. They have normal vision in most respects, and are able to see parts of objects, but are unable to see complex objects. Tyler (1968) describes one such patient. When she was shown an American flag she described her experience as follows:

I see lots of lines. Now I see some stars. When I see things like this, I see a lot of parts. It's like you have one part here and one part there, and you put them together to see what they make.

We all see objects as having parts. When we see a chair, we also see the legs, back, and seat. We have visual representations of these *as* parts of the chair. It is generally supposed that the representation of the whole object is computed out of representations of its parts. In this connection, it is noteworthy that the parts may not have well-defined physical boundaries. For instance, consider the staircase depicted in figure six. We see the steps as parts of the staircase. It is clear where the part boundaries lie on the front of the staircase, but where are the part boundaries on the other side? How much of what lies to the left and below the step is included in the step? There does not seem to be any way to answer that question. We see the step as a part, but with an indeterminate boundary.

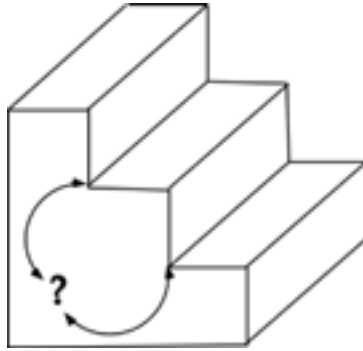


Figure 6. Staircase and steps.

This phenomenon is not unique to parts of objects. The same thing is true of complete objects. Look at the beach scene from Rio de Janeiro in figure seven. Enumerate the things you see. They include the beach, the ocean, some waves, the mountain range, individual mountains, the sky, and the clouds. None of these have well-defined three-dimensional boundaries. You also see “the city” — you often cannot make out the individual buildings — and this too has indeterminate boundaries. The only things you can see in the picture that have determinate boundaries are some of the closer buildings, and the people and birds on the beach.



Figure 7. Ipanema beach in Rio de Janeiro

Even when we see an object having well-defined spatial boundaries, we may not see those boundaries. Consider the house in figure five. We see it as a complete house — a three-dimensional object with well-defined spatial boundaries, but we do not see those boundaries. They are largely occluded by the plants. And, of course, we never see the back boundaries.

The fact that we can often see an object without seeing much of it has an important consequence.

The house in figure five has a certain “look”, but that look is not the same thing as our representation of the house. We can distill the look out of the picture as in figure eight, which includes all visible parts of the house. That is the way the house looks, but that is clearly not our representation of the house — the way we think of the house. Rather, vision represents the house as a complete object *having that look*. This is a very important observation about vision. Vision parses up the colors, textures, etc., and assigns them to particular objects, but the representation of an object is something over and above its look — it is something to which the look is attached. Similarly, in the case of the steps, vision provides a representation of each step and assigns some information to it, including the boundaries of the front side of the step, but it just does not include any information about the boundaries of the back side of the step. The representations computed by the visual system are more like abstract data structures in which a variety of information is stored.



Figure 8. The look of the house.

Representations

Computational theories of vision differ in their details, and no existing theory is able to handle all of the subtleties of the human visual system. But what we want to take away from these theories and use for our purposes is fairly general. First, all theories agree that there is a great deal of complex processing involved in getting from the pattern of stimulation on the retina to the introspectible visual image in terms of which we see the world. Not even the simplest parts of the visual image can be read off the retinal bitmaps directly. In particular, the image is the product of multiple saccades, not a single instantaneous bitmap. Second, the things we see have looks, but these looks are not the same as their visual representations. Vision parses up what we see into whole three-dimensional objects, but the look of an object is often quite impoverished. We can never see all of an object (including the back side), and we are often unable to see much of even the front side of the object. Instead of being the look of the object, the visual representation of an object is literally a mental representation — a way of thinking of the object seen — and it represents the object as looking a certain way.

For epistemological purposes, what is most important about these theories is that the end product is an image that is parsed into representations of objects exemplifying various properties and standing in various spatial relations to one another. The image is not just an uninterpreted bitmap — a swirling morass of colors and shades. The hard work of picking out objects and their properties is already done by the visual system before anything even gets to the system of epistemic cognition. The workings of the visual system itself are impenetrable to introspection. They constitute a computational blackbox, and the first thing the agent has introspective access to is the fully parsed image. The epistemological problem begins with this image, not with an uninterpreted bitmap. As we will see, this makes the epistemological problem vastly simpler.

But why, the epistemologist might ask, does the visual system do all the dirty work for us? Is there something rationally suspect here? There are purely computational reasons that at least suggest that this could not be otherwise. There are 130 million rods and 7 million cones in each

eye, so the number of possible patterns of stimulation on the two retinas is $2^{274,000,000}$, which is approximately $10^{82,482,219}$. That is an unbelievably large number. Compare it with the estimated number of elementary particles in the universe, which is 10^{78} . Could a real agent be built in such a way that it could respond differentially in some practically useful way between more patterns of retinal stimulation than there are elementary particles in the universe? Suppose not. If you divide $10^{82,482,219}$ by 10^x for any $x < 78$, the result is still greater than $10^{82,482,141}$. So less than 1 out of $10^{82,482,141}$ differences between patterns can make any difference to the visual processing system. In other words, almost all the information in the initial bitmap *must be* ignored by the visual system. However, it is hard to construct an argument for the assumption that the visual system cannot respond differentially to more than 10^{78} different patterns of retinal stimulation without knowing more about how information is stored in the brain. It seems that with some kind of “distributed representation” it might be possible to build a system that is capable of responding to more patterns than there are neurons in the brain. Still, $10^{82,482,219}$ is an incredibly large number. It is hard to believe that any system could make sensible discriminations between more than a minute portion of these patterns. If so, this explains why the human visual system works by performing simple initial computations on the retinal bitmap, discarding the rest of the information, and then uses the results of those computations to compute the final image.

5. Visual Representation

The conclusion of the previous section is that the visual image is the result of considerable computational massaging that retrieves some simple information from the retinal bitmap, discarding everything else (which is almost everything), and then producing the final visual image. The visual image comes to us parsed into objects and some of their properties. But this does not yet tell us the exact nature of the visual image, or its functional role in subsequent cognition.

The crucial observation is that when we perceive a scene replete with objects and their perceivable properties and interrelationships, perception itself gives us a way of thinking of these objects and properties. For instance, if you see an apple on a table, you can look at it and think to yourself, “That is red”. The apple is represented in your thought by the visual image of the apple. You do not think of the apple under a description like “the thing this is an image of”, because that would require your thought to be about the image, and as we remarked above, we do not usually have thoughts about our visual images. Usually, the first thoughts we get in response to perception are thoughts about the objects perceived, not thoughts about visual images. For a thought to be about something, it must contain a *representation* of the item it is about. In perceptual beliefs, physical objects can be represented by representations that are provided by perception itself.

We will call perceptual representations of objects *percepts*. The claim is then that the visual image provides the perceiver with percepts of the objects perceived, and those percepts can play the role of representations in perceptual beliefs.⁹ That is, they can occupy the “subject position” in such thoughts.¹⁰

5.1 The Visual Encoding of Information

Our working hypothesis is that the visual image provides the cognizer with a structured array of percepts (object representations) together with representations of some of the properties of the perceived objects, and representations of their spatial relationships to one another, and

⁹ This view was advanced in Pollock (1986). For some related earlier accounts, see Kent Bach (1982), Romane Clark (1973), and David Woodruff Smith (1984,1986).

¹⁰ Here we are unabashedly assuming at least a weak version of the language of thought hypothesis (Fodor 1975), according to which thoughts can be viewed as having syntactic structure. We will say more about this in section six.

perhaps representations of other kinds of relationships as well. Perception provides our initial access to the world, and our claim is that these perceptual representations provide our (logically) initial way of thinking about the perceived objects and their perceived properties and relationships. Normally, only after perceiving objects and properties, which requires thinking of them in this way, can we come to think of them in other ways.

The percept does not give us a *permanent* way of thinking of an object. Once we are no longer seeing the object, we no longer have the percept, so insofar as we can still think of the object we must do so in some other way. Our view would be that having the percept initiates a *de re* representation, in the sense of Pollock (1983), and we normally continue thinking of the object in terms of the *de re* representation. But that is another story and we need not go into it here.

The visual image does not just contain representations of the objects and properties — it represents the objects *as having* those properties and *as standing in* those relationships to one another. We see the apple, we see its color, and we see the color *as being* the color of the apple. That is, our visual representation links these two representations as being about a common object. Similarly, we see the shape of the apple, and we see it *as being* the shape of the apple. What is it that links the representation of the apple and the representations of its properties? Introspection does not provide us with any informative answer to this question. All we can say is that they are linked. When we perceive the apple, we perceive it as an object *having* those properties. You cannot perceive the apple without perceiving some of its (putative) properties. We might think of the percept as a data structure with fields for various perceivable properties, including colors, shapes, sizes, spatial relationships to other objects, etc. It is the values in these fields that make up the look associated with the percept. Visual processing produces the percept by creating such a data structure with certain property representations in its fields. This is an abstract characterization of visual representations and their interconnections.

5.2 The Mechanical Representation Hypothesis

The preceding account of the role of perceptual representation in thought implies that when we think of perceived objects and properties in terms of visual representations of them, there can be no willful “act of interpretation” whereby we decide to use a representation to represent a particular representatum. In this respect, visual representation contrasts with linguistic representation in a public language. In using a public language, we first decide what we want to talk about, and then we decide to represent it in a certain way. It is our intention to be talking about certain things that determines what we mean, even if we violate the rules of language in our attempt to convey that meaning. And to intend to be talking about a particular thing, we have to have a way of thinking about it. So linguistic representation requires a prior mental representation of the objects we are talking about. Visual representation cannot work similarly. Perception provides our initial access to the things perceived, so it cannot be required that we first have another way of thinking about them which we employ in deciding that a particular percept is to be used to represent a particular thing. Perceptual representation must be automatic — not deliberate. What a perceptual representation represents must be determined by intrinsic properties of the representation (e.g., what occupies various fields in the data structure) together with causal relationships between the cognizer’s perceptual apparatus and the world. The agent’s intentions cannot be relevant here. We will refer to this as *the mechanical representation hypothesis*.

There are, however, some examples that seem to conflict with the mechanical representation hypothesis. Imagine seeing an old stone church, while standing directly in front of it. You see the church, and think of it in terms of a percept. You might then focus your attention on just the front wall, which is an interesting object in itself because of its intricate construction out of the massive blocks of stone. You can think of either the church or the front wall perceptually, so you must have a perceptual representation of each, but are they different percepts? Your visual experience seems to be the same, regardless of which you are attending to. This suggests that the same percept can represent either the church or the wall. If this is right then it seems that what

determines the representatum of the percept at any given time is the cognizer's intentions. But if that is right then the mechanical representation hypothesis is wrong.¹¹

It is tempting to try to avoid this problem by supposing you never really see the church when you can only see the front wall of it. You only infer the existence of the church from the percept of the wall. So the percept is unequivocally a percept of the wall — never a percept of the church. But this cannot be right. If it were then it would follow that you can never really see the church even when you see it from an angle. You can see at most two sides of it (and part of the roof). Similar reasoning would lead you to conclude that you cannot even see the wall — only the front surface of it. You infer the existence of the solid block wall from your percept of the front surface. But this all seems wrong. If the percept is really the representation employed in your thought, it is normally of the church, because that is all you normally have a thought about. You *can* have thoughts about walls or surfaces of walls, but that is unusual. That is not what goes on in normal perception. You have to explicitly change your focus of attention for that.

5.3 What is Really Represented in Perception?

To disentangle the above example, let us distinguish between two cases. First, consider the easiest case, which is the case in which it turns out that the wall is all that is left of an old church — there is nothing behind it. When we take ourselves to be seeing a church we are making a mistake, but it is not a perceptual error. Our percept purports to be a percept of a three-dimensional object. Regardless of whether the church is there or just the wall, we are seeing a three-dimensional object. After all, the free-standing wall is itself a three-dimensional object — just a relatively thin one. Perception represents the perceived object as a three-dimensional one, but it does not contain any information about how deep it is, or about whether it is a free-standing wall or a church. Those are judgments made by epistemic cognition after perception presents it with the information it encodes. The information encoded by perception is not wrong — it is the judgment we make on the basis of it that is wrong.

This illustrates a general point about percepts. It was remarked in section four that they represent objects as three-dimensional, but they are in a certain sense incomplete. They encode the *look* of the front of the object, but they never encode the look of the back of the object. They can encode a very limited amount of information about the back. For example, concave curves on the visual contour of the object are generally interpreted as saddle points on the three-dimensional surface of the object. But, strictly speaking, this tells us about only infinitesimally much of the back of the object. The incompleteness of percepts shows how unlike photographs they are. For instance, while standing before a tall building you may have a percept of the building without bothering to look up. The encoded look of the building just fades into obscurity in the upper part of our visual image, but nonetheless our percept represents the building as a complete object. Similarly, you can have a percept of a passing train without being able to see either end.

Without being able to see beyond the wall, we normally judge that we are seeing a church because we have certain contingent expectations. We can imagine situations in which we lack those expectations, and then we will not make any inferences about the perceived object being a church. For instance, if you are wandering around through some old ruins, and all that is left standing is individual walls — no intact structures — and you are aware of all this, then you will naturally take the object perceived to be a wall rather than a church. But it is still a three-dimensional object.

This handles the case in which you are mistaken about whether you are seeing a church or a free-standing wall, but it does not yet explain what is happening when there is a church before you but, through a deliberate change of attention, you can look at the wall rather than the

¹¹ We owe this example to Joe Cruz (in conversation).

church. Then it seems you have a single percept that you can use either to think about the church or to think about the front wall of the church, and that conflicts with the mechanical representation hypothesis.

However, the only reason for thinking that you are employing the same percept to think about both the church and the wall is that your visual image does not change as a result of your change of attention. Thus any visual representations that are present when you are attending to the church are also present when you are attending to the wall. But it does not follow from this that you are employing the same representation to think about both. It is our contention that your visual image contains two different representations — one representing the church and the other representing the front wall — and both representations are present in the image regardless of whether you are attending to the church or to the wall. When you think of the wall you are using one representation, and when you think of the church you are using the other representation.

To make this plausible, let us make two separate observations. First, our visual image is very rich and contains much more information than we actually use on any one occasion. We can think of the visual image as a transient database, and attention as a mechanism that selects items from that database and passes them to epistemic cognition for further processing. Attention is a complicated phenomenon. Some perceptual events, like abrupt motions or flashes of light, attract our attention automatically. Others are more cognitively driven. We have interests in particular matters. For instance, you might want to know the color of Joan's blouse. This interest leads you to direct your eyes in the appropriate direction, and to attend to the perceptual representation you get of Joan's blouse, and particularly to the representation of the color. The visual image you get by looking in Joan's direction contains much more than the information you are seeking, but attention allows you to extract just the information of use in answering the question you are interested in.¹² The use of attention to select specific bits of information encoded in the visual image is a mechanism for avoiding swamping our cognition with too much information. Perceptual processing is a feedforward process that starts with the retinal bitmap and automatically produces much of the very rich visual image. Making it automatic makes it more efficient. But it also results in its producing more information than we need. So epistemic cognition needs a mechanism for selecting which bits of that information should be passed on for further processing, and that is the role of attention.

The second observation we need to sort out the church/wall example is that we can see many different kinds of things — not just physical objects. For example, you can see the edge (line) formed where two walls of a building intersect. This is literally something you can see and think about. You can, for example, notice that it is not quite straight. As such, you must have a visual representation of it. In light of the way computational theories of vision have developed, this should not be surprising. All such theories suppose that early in visual processing we form representations of edges, and these are used in later stages for computing representations of physical objects. Of course, visual processing could have been organized in such a way that those early representations of edges don't make it into the final visual image, but introspection makes it pretty clear that they do. This should not be surprising. Edges can be important — for instance, walking off the edge of a cliff can be a bad thing. Similarly we can see corners, and for pretty much the same reason. They are used in computing representations of physical objects, and are useful for getting around in the world. What is important about this is that edges and corners are not themselves physical objects. It follows that the visual image contains perceptual representations of more things than physical objects. At least, it contains representations of edges and corners.

¹² This "cognitive" view of attention contrasts strongly with the familiar "mental spotlight" view, according to which attention picks out a region of the visual field and we attend to everything in it. That is not a correct account of attention in general because, for example, I can attend to the color of the apple without attending to its shape.

Armed with the observation that our visual image contains perceptual representations of things other than physical objects, consider the church/wall example again. The wall is definitely something you can see, and hence something for which you have a visual representation, but it does not follow from this that the kind of representation that represents the wall is the same kind of representation that represents the church. Walls are things that we can see, but we can see them even in cases in which there is no possibility of mistaking their representations for representations of complete objects. Consider looking at one of the *inside* walls of a room. You can look at the wall and judge, for example, that its two vertical sides are not exactly parallel. The wall is not a complete physical object. It does not even purport to be a complete physical object.

Note that there are two senses in which you can look at and see the wall. You can see the wall as a physical part of a larger structure, but you can also see the surface of the three-dimensional wall. When you look at the surface of the wall, you can observe, for instance, that it has an interesting granular texture, or that it is very smooth, or that it undulates. Surfaces are things we can see, and hence things for which we have perceptual representations. Again, this should not be surprising. Presumably, the representation of an object is computed in part out of representations of surfaces, so there is no reason why the representations of surfaces should not find their way into our visual image. And being able to make visual judgments about properties of surfaces can be very useful. Seeing a paw print in the mud may warn you of the presence of a grizzly bear.

But ordinarily, when you look at the wall and have thoughts about it, you are not thinking about the surface of the wall. You are thinking about the wall as a physical part of a larger whole. This is an instance of a more general phenomenon. We can often see, and perceptually represent, parts of physical objects. When you look at a cup, you see only one object. You can see the handle, but you do not see it as a separate physical object. You see it as part of the cup. But if you break the handle off the cup, then you see two objects. You see the handle as a separate object rather than as part of the cup. That we can see parts of physical objects, and see them *as* parts rather than complete physical objects in their own right, fits naturally with computational theories according to which perceptual representations of complex objects are constructed out of (computed on the basis of) perceptual representations of their geometrically simpler parts. Thus we can see the parts, and see them *as* parts without seeing them as physical objects in their own right.

There are obvious difficulties involved in drawing a precise logical or metaphysical distinction between objects and parts of objects. On the other hand, this is a distinction we do draw all the time. We suggest that the distinction has its origin in our perceptual processing system, and reflects how object representations are computed. As such, there may be no metaphysically precise way of making sense of the distinction. At the same time, this may be the explanation for why mereology seems so odd to most people. It is tempting to say that, *logically*, mereology ought to be true — but that just isn't the way we think of objects. There simply isn't a physical object that has both my wristwatch and the planet Mars as parts. Our concept of physical objects and their parts derives from our system of perceptual representation, and such "quasi-objects" cannot be represented as single physical objects.

We are being led to a long catalogue of the different kinds of things we can see, and accordingly of the different kinds of perceptual representations that go into making up the visual image. Thus far the catalogue includes edges, corners, surfaces, parts of objects, and objects. We will see below that it includes more besides.

Now let us return to the church/wall example. It should be clear what is going on. When we look at the wall, while regarding it as just the front wall of a larger church, we are seeing it as a part of the larger object that is the church. Thus it is represented by an "object part" representation rather than an object representation. Both representations are present in the visual image, regardless of what we are attending to, just as representations of edges, corners, and surfaces are present without being attended to. To substantiate this claim, first consider the case in which the wall is

part of a larger church. We can think of either the wall or the church in terms of visual representations. The wall is represented by a perceptual representation R_1 , and the church by a perceptual representation R_2 . The question is whether R_1 and R_2 are the same representation. The important observation to make here is that it is possible for one to have a single thought in which one thinks of both the church and the front wall perceptually. Using these representations, the cognizer may have the thought $\lceil R_1 \text{ is just the front wall of } R_2 \rceil$, and be right. If prodded, he might agree (and be right) that it is false that $\lceil R_1 \text{ is all that is left of } R_2 \rceil$. (Of course, asking him this might make him suspicious, but suppose it does not.) However, if R_1 and R_2 were the same representation, then it could not be false that $\lceil R_1 \text{ is all that is left of } R_2 \rceil$. This would be the same thought as that involved in believing it false that $\lceil R_1 \text{ is all that is left of } R_1 \rceil$. Thus R_1 and R_2 must be different representations. It follows that there are two different percepts operative in the church/wall example, and hence the example poses no threat to the mechanical representation hypothesis.

5.4 What Can We See?

If you can literally see something, your visual system must be able to construct a visual representation of it and incorporate it into your visual image. We can see a number of different kinds of things, and see them as different kinds, indicating that they are represented by different kinds of perceptual representations. The sense in which they are different is simply that they are phenomenologically distinguishable. For example, having an edge representation is phenomenologically distinguishable from having a corner representation, which is phenomenologically different from having a surface representation, which is phenomenologically different from having an object representation. All it takes for them to be phenomenologically distinguishable is that we can introspect which kind of representation we have, and that in turn requires no more than that the data structure that is the representation contains information indicating its type. That is going to be essential for the use of these different kinds of representations in computational vision because, for example, edges are used differently than corners and so the system must be able to tell which are which.

In addition to edges, corners, surfaces, objects, and parts of objects, we also see some of the properties of these things. But let us put that aside for now. We will examine the visual apperception of properties more closely in section six.

It seems undeniable that visual representations of edges, corners, surfaces, objects, etc., *can* be present in the visual image, but one might question whether they are present even when the cognizer is not attending to them. That some of the information is there without our attending to it follows from the fact that we can, for a short period of time, remember it even if we did not originally attend to it. There is a short-term memory buffer associated with each sense modality, and that enables the cognizer to attend to an aspect of the sensory input that was at first ignored. For a simple illustration of this phenomenon, stand in a light-tight room. First, have the light on. Then as you turn the light off, turn your head. You will note that the room seems to turn with you. What is happening is that, in the darkness, no new visual image is being produced and so your visual memory buffer briefly maintains the visual image as it was when the lights went off.

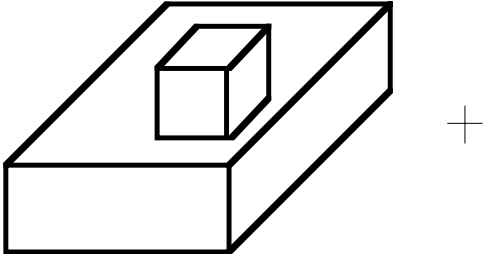


Figure 9. Two boxes.

On the other hand, attention does affect the content of the image. Some visual representations may be computed only as a result of attention. For example, consider figure nine, devised by Jepson & Richard (1993). If you attend to the bottom front edge of the smaller box, you see it as sitting atop the larger box. But if you attend to the right front edge of the smaller box, you see it floating in the air above the larger box, and if you attend to the top rear edge of the small box, you see it as inside the larger box. Furthermore, this is not a matter of where you are focusing your eyes. The same phenomenon occurs if you keep your eyes focused on the cross to the right and attend to the different edges without moving your eyes. So it appears that the introspectible contents of the visual image can be affected by attention, but this does not mean that you have to be attending to a visual representation in order for it to be contained in your visual image.

In addition to what we have already noticed, humans can see a number of more esoteric things. For instance, we can literally see movement. Vision researchers have long known that movement is not (just) inferred from sequentially perceived positions. There are neurons in the visual cortex that respond to the sequential firing of spatially oriented pairs of rods. These neurons only respond when the rods are stimulated in the right order and with the right time interval between stimulations. This gives rise to the phenomenal impression of movement. The phenomenal impression can arise even when nothing is moving. For example, after strenuous aerobic exercise, if you look at a blank blue sky it is common for it to seem to be moving, despite the fact that there are no object representations in the visual field that are changing apparent position. This is a well-known phenomenon. Conversely, there is a condition known as *akinetopsia* in which brain damage to a region between the temporal and occipital lobes renders patients unable to see movement. Zihl et. al. (1983) described one such patient:

The visual disorder complained of by the patient was a loss of movement vision in all three dimensions. ... She could not cross the street because of her inability to judge the speed of a car, but she could identify the car itself without difficulty. "When I'm looking at the car first, it seems far away. But then, when I want to cross the road, suddenly the car is very near." She gradually learned to "estimate" the distance of moving vehicles by means of the sound becoming louder.

The reason the perception of motion is not tightly linked to the perception of the positions of objects is that motion representations are computed at an earlier stage than the object representations and are used in computing the object representations. That we can literally see movement is another illustration of the fact that our visual image is not a function of the momentary state of our retinal bitmap. It takes account of information drawn from that retinal bitmap over some time interval.

We can sometimes see groups of objects. For example, looking down on a crowd of people from a high window, you might observe a group of identically green-shirted individuals that are dispersed throughout the crowd but moving systematically in a certain direction. Even more interesting, it has often been noted that when you have a perceptual representation of a small group (of six or fewer objects), you can also perceive the cardinality of the group without counting. If there are three plates on a table, you can tell by glancing at the table that the number of plates is three. Thus small cardinalities are perceivable properties of perceivable groups. It follows that we have perceptual representations of those properties.

A very interesting category of things that we can see includes patterns, designs, pictures, paintings, printed words, and simple sentences. It is tempting to suppose that when we see a pattern, we are just perceiving a property of a surface. But this is wrong. Imagine a pattern displayed on the wall by a slide projector. You can move the pattern around by moving the projector. The same pattern can be displayed on different surfaces, so it is not perceived as a

property of the surface. It is a distinct kind of perceivable thing.¹³ Interestingly, most researchers in animal cognition are agreed that only higher primates are capable of seeing patterns. It is claimed that the visual processing systems of other animals simply do not compute visual representations of patterns. It seems that the ability to perceive patterns, designs, etc., is necessary for language, so this is a very important perceptual ability.



Figure 10. Scene.

6. Seeing Properties

Obviously, we can see many of the properties of the things we can see. Consider the scene depicted in figure ten, and imagine seeing it in real life with both eyes. What do you see? Presumably you see the Marajó pot (from Marajó island at the mouth of the Amazon River), the Costa Rican dancer, the Inuit soapstone statue, and the ruler. You also see various lines, corners, surfaces, contours, and parts of objects (e.g., the handles on the pot). You also *see that* a number of things are true of them. For instance you see that the dancer is behind the ruler. In saying this, we do not mean to imply that you see that the dancer is a dancer or that the ruler is a ruler (although you might). We are using the referential terms in “see that” indirectly, so that what we mean by saying “You see that the dancer is behind the ruler” is “You see what is in fact the dancer, and you see what is in fact the ruler, and you see that the first is behind the second.” So it is only the property attributions we are interested in here. With this understanding, you might *see that* any of the following are true:

- (1) The pot is to the left of the soapstone statue.
- (2) The dancer is behind the ruler.
- (3) The end of the line marked “6” on the ruler coincides with the point on the base of the dancer.
- (4) The contour on the top of the dancer’s skirt is concave.
- (5) The pot has two handles.

¹³ We owe this observation to Bill Ittleson.

(6) The base of the pot is roughly spherical.

You *might* see the following:

(7) The pot is from Amazonia.

(8) The point on the base of the dancer is six inches to the right of the pot.

(9) The soapstone figure depicts a boy holding a seal.

When you *see that* something is true of an object, you are seeing that the object has a property. However, the different property attributions in these *see-that* claims have different statuses. Some are based directly on the presentations of the visual system, while others require considerable additional knowledge of the world. Presumably no one believes that the visual system, by itself, can provide you with the information that the pot is from Amazonia, or that the dancer is six inches from the pot. If you are an expert on such matters, you might *recognize* that the pot is from Amazonia, but the visual system does not represent the pot *as being* from Amazonia. This is an important distinction. Much of what we know from vision is a matter of recognizing, on the basis of visual clues, that things we see have or lack various properties, where such recognition normally involves a skill that the cognizer acquires. The skill is not a simple exercise of built-in features of the visual system, and may depend upon having contextual information that is not provided directly by vision. By contrast, it is very plausible to suppose that the visual system itself *represents* the pot *as being* to the left of the soapstone statue, represents the contour on the top of the soapstone statue's head *as being* convex, etc.

Let us make it clear that when we talk about mental representations, we take them to be mental items that (1) can be constituents of thoughts, and (2) make the thoughts be about their representata by virtue of containing the representation. When something that we see is represented as having a certain property, the visual system computes that information and stores it as part of the perceptual representation of the item seen. It is part of the specification of how the perceived object looks. When this happens, the visual system is computing a representation of the object as having the property. We might put this by saying that the visual system is computing a representation of the property (a universal) and storing it with the representation of the object. This is not quite the right way to put it though, because we don't literally think of the property. That is, the property is not an object of our thought. Rather, it is what we think *about* something else. So vision provides us with a way of thinking the property of something. We can take that to be what we mean when talking about vision providing a representation of the property. The representation is still a constituent of our thought, but it plays a predicate role rather than a subject role. We will call such properties *perceptible properties*. Just to have some convenient terminology, we will call this kind of seeing-that *direct seeing-that*, as opposed to *recognizing-that*.

The distinction between direct-seeing and visual recognition turns on whether the visual system itself provides the representation of the property or the visual system merely provides the evidence on the basis of which we come to ascribe a property that we think about in some other way. For instance, we can recognize cats visually. However, cats look many ways. They can be curled up in a ball, or stretched full length across a bed, they can be crouched for pouncing, or running high speed after a bird, they can have long hair or short, they can have vastly different markings, etc. There is no such thing as *the* look of a cat. Cats have many looks. Furthermore, these looks are all only contingently related to being a cat. As you learn more about cats, you learn more ways they can look and you acquire the ability to visually recognize them in more ways. In acquiring such knowledge, you are relying upon having a prior way of thinking of cats. Nor does your representation of cats change as a result of your learning to recognize cats visually. This follows from the fact that the appearance of cats can change and that does not make it impossible for you to continue thinking about them. Imagine a virus that spread world-wide and resulted in all cats losing their fur. Furthermore, it affects their genetic

material so all future cats will be born bald. One who was not around cats while this was occurring would probably find it impossible to visually recognize the newly bald cats as cats, although this would not affect her ability to think about cats, or to subsequently learn that cats had become bald. So the visual appearance of cats is not our representation of cats. We have some other way of thinking about cats, and learn (or discover inductively) that things looking a certain way are generally cats.

Although most authors have agreed that we cannot directly see that something is a cat, a few have denied the distinction between directly seeing and recognizing. For example, Peacocke (1992) observes that there is a marked difference between “the experience of a perceiver completely unfamiliar with Cyrillic script seeing a sentence in that script and the experience of one who understands a language written in that script”. Peacocke observes that the second perceiver “recognizes symbols as of particular orthographic kinds, and sequences of the symbols as particular semantic kinds.” He suggests that there is no difference between this and what we are calling “direct seeing that”.¹⁴ Similar observations are made by Charles Siewert (1998). However, the fact that recognizing something changes one’s phenomenal experience doesn’t indicate that you are thinking of the kind in terms of a mental representation computed by the visual system itself. This seems clear, for example, in the above case of cat recognition.

Direct realism has traditionally been focused on direct-seeing, and that will be the focus of the rest of this section. However, we feel that visual recognition is of more fundamental importance than has generally been realized. Accordingly, it will be discussed at length in section nine.

6.1 Seeing Spatial Properties

Among the properties we can see, it is useful to distinguish spatial properties from colors. In this section, we will discuss spatial properties, and in the next section we will discuss colors.

It is an empirical matter just what properties are represented by the visual system and recorded as properties of perceived objects. It might be urged that (1) – (6) are all instances of this, while (7) – (9) are not. For the most part, we do not have to decide this matter here. We can leave this up to the vision researchers to determine. Nevertheless, we can suggest some constraints. When properties are represented by the visual system and stored as properties of perceived objects, they form part of the look of the object. As such, there must be a characteristic look that objects with these properties can be (defeasibly) expected to have. This rules out such properties as “being from Amazonia”, but it is less clear what to say about some other properties.

In deciding whether the visual system has a way of encoding a property, we must not assume that the encoding itself has a structure that mirrors what we may think of as the logical analysis of the property. Perhaps the clearest instance of this is motion. It would be natural to suppose that we *infer* motion by sequential observation of objects in different spatial locations, but perception does not work that way. It was remarked in section four that representations of motion are computed prior to computing object representations. This is because information about motion is used in computing object representations. For example, motion parallax plays an important role in parsing the visual image into objects. Motion parallax consists of nearby objects traversing your visual field faster than distant objects. The importance of motion in perceiving objects is easily illustrated. Consider looking for a well-camouflaged insect on a tree leaf surrounded by other tree leaves. You may be looking directly at it but be unable to see it until it moves. The motion is indispensable to your visual system parsing the image so that the insect is represented as a separate object. The important thing about this example is that although motion can be logically analyzed in terms of sequential positions, the representation of motion is not similarly compound. It does not have a structure. As noted above, you can sometimes “see” motion while looking into a blank blue sky in which there are no object representations at all.

¹⁴ This conclusion is drawn explicitly by Susanna Siegel (2004).

From a functional perspective, your visual image, viewed as a data structure, simply stores a tag “motion” at a certain location. There is no way to take that tag apart into logical or functional components. It is just a tag. It may be *caused* by other components of the visual image, but it does not *consist* of them.

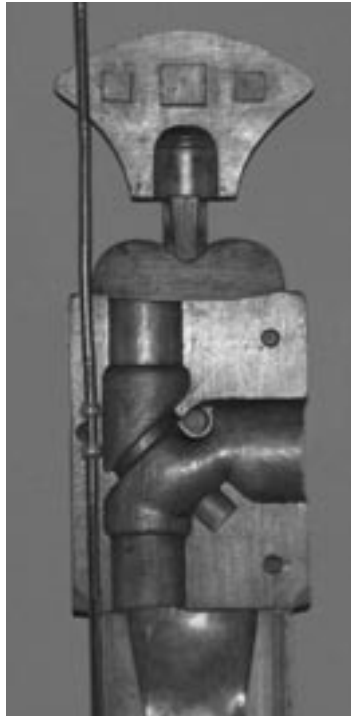


Figure 11. Convex



Figure 12. Concave

Consider another example. According to most computational theories of vision, representations of convexity and concavity are computed fairly early and play an important role in the computation of other representations. We can talk about convexity and concavity in either two dimensions or three dimensions. In two dimensions, the convexity of a part of the visual contour (the outline) of a perceived object plays an important role in computing shape representations, but for present purposes it is more illuminating to consider three-dimensional convexity and concavity. For example, note how obvious the convexity of the “plumbing fixtures” on the front of the statue in figure eleven appear.¹⁵ The convexity has a clear visual representation in your image. Contrast this with figure twelve, which is a photograph of the same statue in different light. Now the plumbing fixtures appear clearly concave. In fact, they really are concave, but it is impossible to see them that way in figure eleven, and it is impossible to see them as convex in figure twelve. So three-dimensional convexity and concavity have characteristic “looks” (but, of course, these looks are not always veridical). The percepts represent objects as having concave or convex features. It does not take much thought to realize that these looks are *sui generis*. The phenomenal quality that constitutes the look does not have an analysis. It is *caused by* (computed on the basis of) all sorts of lower-level features of the visual image, but it does not simply consist of those lower-level features. Once again, from a functional point of view this simply amounts to storing a tag of some sort in the appropriate field of the percept (viewed as a data structure). When the field is occupied by the appropriate tag, the object is perceived as convex. This is despite the fact that convexity has a logical analysis in terms of other kinds of spatial properties of objects.

¹⁵ The statue is the work of Tony de Castro, of Tiradentes, Brazil.

Obviously, we can perceive relative spatial positions and the juxtaposition of objects we see, and it is generally acknowledged that we can see the orientation of surfaces in three dimensions. All of this is illustrated by figure ten. Note particularly the visual representation of the orientation of the floor. Three-dimensional orientation is perceived partly on the basis of stereopsis, as illustrated by the stereograms in figure three, but it can also be perceived without the aid of stereopsis, as in figure ten.

What about shapes? This is harder. We can certainly recognize shapes visually, but it is less clear that we see them directly. For example, consider the circles and ellipses on the side of the monolith in figure thirteen. When we are looking at them from a perpendicular angle, it is easy to tell which are which. This might suggest that circularity is represented in the visual image much as convexity is. It is popularly alleged that circles look like circles and not like ellipses even when seen from an angle. If this were right, it would support the suggestion that circularity is seen directly. But we doubt that it is right. The same circles and ellipses that appear in figure thirteen appear again on both sides of the monolith in figure fourteen. Do some of them look circular and others elliptical? That does not seem to us to be the case. This suggests that one cannot see directly that a shape is circular.

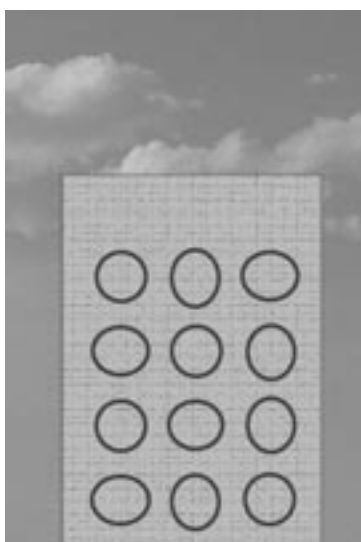


Figure 13. Circles at a right angle.



Figure 14. Circles at an oblique angle

It might be suggested that there is no need for us to be able to directly see that objects have particular shapes, because shape properties have definitions in terms of simpler perceptible spatial properties. The thinking would be that we can judge shapes in terms of those definitions. But two considerations suggest that this is not an adequate account of our ability to judge shapes. First, children can judge shapes without knowing the definition of *square* or *circle*. In fact, these definitions were only discovered fairly late in human history by the ancient Greek geometers. Second, the standard definitions presuppose Euclidean geometry, but our best current physics tells us that space is not Euclidean. In a non-Euclidean space (such as the one we actually reside in), you cannot employ the familiar Euclidean definitions of *square* or *circle*. So it seems clear that they do not provide the basis for our judgments.

How then do we judge shapes? One suggestion is that, because one can directly see the orientation of surfaces and one can easily see that a shape is circular when it is viewed from a perpendicular angle, the property of being a circle oriented at a right angle to us is a perceptible property. Similarly for squares. Then other shape judgments could be parasitic on the perceptual judgments made at a right angle. This is just a tentative suggestion, however. These are issues

that require further investigation.

6.2 Seeing Colors

We see that objects have various colors. For example, we see that the London buses in figure two are red. Is *red* a perceptible property? Does the visual system represent perceived objects as being red? Most philosophers have thought so. For example, Thompson et al (1992) claim, “That color should be the content of chromatic perceptual states is a criterion of adequacy for any theory of perceptual content.” But let us consider this matter more carefully.

People see colors, but what does this amount to? There are two senses in which people might be able to see colors. First, one can see an expanse of color *as an object*. For example, you can look at the color of an apple, and compare it with the color of another apple, without attending to the fact that the colors are the colors of apples. Here you are just attending to a “bit of color” — what philosophers call *tropes*. We can make judgments about tropes that are in many cases similar to the judgments we make about physical objects. E.g., we can make judgments about where they are and what their shape is. From a metaphysical point of view, tropes may seem mysterious, but what we are observing now is that, regardless of the metaphysics, the visual system computes representations of tropes. Furthermore, it can do so without computing representations for the objects to which the tropes presumably attach. A number of years ago, one of us had the experience of walking across the Berkeley campus at night in a heavy fog and seeing “colored shapes” looming out of the fog without being able to tell what they were or even how close they were. Basically, what was being seen were tropes, and the visual system was having trouble parsing the image into object representations. Suddenly it succeeded, and then it was observed that the tropes attached to distant buildings.

Philosophers sometimes talk as if tropes must be expanses of uniform color. If that is a necessary property of tropes, then the phenomenon of seeing colors as objects is not properly described as seeing tropes, because one can look at a surface (say a wall) and see the color at every point, but the color may vary continuously as one scans across the surface. One still sees an expanse of color, but there is no expanse of uniform color. However, it is useful to have a word for what we are seeing, so we will continue to talk about seeing tropes in these cases.

When one sees colors in the sense of seeing tropes, one sees colors as objects. Furthermore, one can look at a physical object, see its color in the sense of seeing a trope, and see that the trope is the color of that object. We might say that the object is the “physical substratum” of the trope. In this sense, the object is represented as having that trope as its color. This is different from representing the object as red. Representing an object as having a certain trope as its color is to represent a relationship between two perceived items. On the other hand, representing an object as red is to represent it as having a certain property — a color universal. Note that tropes also have colors in this sense. So let us not confuse the visual representation of the trope with a representation of a color universal. They are of logically different sorts.

We certainly can *see that* both a trope and a physical object are red.¹⁶ But as we have seen, this could either be a matter of directly seeing that they are red (in which case vision represents the object and the trope as red), or a matter of the cognizer visually recognizing that they are red. What is at issue is whether the visual system itself can provide the information that they are red, or if the recognition of colors is a learned skill making use of a lot of contextual information over and above that provided by the visual system. For the visual system to provide the information that something is red, it must have a way of representing the color universal. If vision provides representations of color universals, how does it do that? We have a mental color space, and different points on a perceived surface are “marked” with points from that color space. This is

¹⁶ We do not mean to imply that we are ascribing literally the same color property to the trope and the physical object.

part of their look. We will refer to the points in this color space as *color values*. The color space is a metric space in the sense that we can compare the color values in terms of their similarity or closeness. This metric is built into our system of visual cognition, although, of course, it does not allow us to introspect real numbers as values of the metric. It is natural to suppose that the color values that are used to mark points on a perceived space represent color universals, and hence marking a surface or patch of surface with such a color value amounts to perceiving it as having that color. This is probably the standard philosophical preconception, and is responsible for the idea that colors have characteristic appearances that are partly constitutive of their being the color they are.



Figure 15. A wall painted a uniform color.

However, the discussion of brunescence in section three strongly suggests that this philosophical preconception is mistaken. If color values (points in color space) represented color universals, it would turn out that objects having a particular objective color will hardly ever be represented by percepts marked with the corresponding color value. What then would make it the case that a particular color value represents a particular color universal? Furthermore, we need not rely upon anything as exotic as brunescence to make this point. Simply stand in a room whose walls are painted some uniform color but unevenly illuminated by a bright window, and look at the color. A photograph of such a wall is displayed in figure fifteen. Notice how much variation there is in the apparent color. The differences are not just differences of shading. The wall is a pale blue, but some areas look distinctly yellower than others. In fact, the effect is so pronounced in real life that we were unconvinced the wall really was a uniform color, and we tested it by moving a matching paint chip around the surface. It matched everywhere. So when we see this wall, we are seeing the same objective color everywhere — the same color universal is instantiated by every point on the wall. But in our visual image the representation of the wall is marked by widely varying color values. Which of these color values is the “right” color value? Does that make any sense? Surely not. A single color looks very different under different circumstances. Which circumstances define “looking that color”? We imagine that many philosophers will be tempted to say that the right color value is the one we experience when we view the color in

white light. But this just exhibits ignorance about the wide variety of things that can affect how colors look. It is not just the color of the light that affects how a color looks. In section three we enumerated the Bezold-Brücke effect, simultaneous color contrast, chromatic adaptation, brunescence, and the sensitivity of one's photopigments, and there are probably many other factors that affect how colors look as well.¹⁷

We take this to indicate that the relationship between a color and its look on any particular occasion is not one of representation. To get around these difficulties, Cohen (2003) suggests relativizing to perceivers what phenomenal color value represents what objective color. But this would have to be relativized to more than the perceiver. First, it would have to be relativized to a perceiver *at a time*, because colors look different at different times due to phenomena like brunescence. Second, it would have to be relativized to a location in the perceiver's visual image, because different objective colors seen in different parts of the image can look the same. The latter, we submit, makes nonsense of the notion of mental representation. If the look of the color is a mental representation of it, but two different colors look the same, then a single mental representation would be representing two different things at one and the same time. That is impossible, because what we are thinking is determined by what mental representations are incorporated into our thoughts. It follows that if we employ the same property representation we must be thinking the same thing about the objects to which we are ascribing the properties, and hence ascribing the same property (i.e., color) to what we are seeing.

Note also that if the look of a color always represented that color, then it would be impossible for us to be wrong when we make a perceptual judgment about the color of something. Obviously, we can be wrong, so the most we can say is that the look *sometimes* represents the color. But then we are back to the problem of deciding which looks represent a color, and that question doesn't seem to have an answer.

Apparently we cannot directly see that objects have particular colors. But surely we can see colors, can't we? Yes, but the sense in which we see colors is that we see tropes — expanses of color — not that we see color universals. To see a trope is to see a color as an object, not to categorize a perceived object using a color universal. We can visually *recognize* things as exemplifying specific color universals, but that is different from seeing the color universals. In the same sense we can visually recognize cats or oak trees (or newborn female chicks if you are a chicken sexer), but this does not mean that the visual system can single-handedly compute a representation of an object as being a cat or oak tree. So the fact that we can visually recognize that something is red gives us no reason for thinking that the visual system has a way of representing the color universal *red*.

If color values are not representations of color universals, what are they for? The answer is simple. They are part of what makes up the look of the object. The look of the object does not contain a representation of a color universal, but it is nevertheless a large part of the basis upon which we recognize what color the object is. Such recognition would be particularly simple if each color value represented a unique color universal, but as we have seen, because of the variability in how colors look under different circumstances, there is no way for the visual system to achieve that. So color recognition must instead take account of both the look of the object and the context in which the object looks that way. The point we are making here is similar to the point we made about *circle* in the previous section. Neither *circle* nor *red* have characteristic looks, so there is no way the visual system can represent objects as having these properties. To see that an object has one of these properties must be to exercise a learned capacity to *recognize* that objects have these properties. In section nine, we will consider more carefully how such recognition works.

¹⁷ This seems to us to be obvious, but Michael Tye (2000) commits himself to the view that if a color looks different to different people then at most one of them can have it right.

To recapitulate, percepts do three things. First, the percept is a mental representation of the object perceived. Second, it represents the object as having certain perceptible properties or as standing in certain perceptible relations to objects represented by other percepts. Third, it encodes the look of the object perceived. The latter is different from representing the object as having perceptible properties (unless we want to count looking that way as a perceptible property). Looks are important, because they can provide the evidence on the basis of which we ascribe non-perceptible properties to objects. That is what goes on in visual recognition.

7. The Visual Image

The contemporary epistemological problem of perception has been strongly conditioned by the view of the visual image that was prevalent at the start of the 20th century. That view took the visual image to be an undifferentiated melange of colors and shades corresponding directly to the bitmap of retinal stimulation. Contemporary scientific theories of perception insist instead that the visual image is the product of computational processing that dispenses with most of the information potentially in the retinal bitmap and produces visual representations of physical objects, some of their properties and relations, and numerous other kinds of things as well, including edges, corners, surfaces, parts of objects, colors, shapes, spatial relations, motion, patterns, etc. This rich array of preprocessed visual information is the input to the kind of epistemic cognition that is the topic of epistemological theorizing. Epistemology begins with the visual image, not the retinal bitmap.

The visual image encodes all of these kinds of visual representations. The question arises just how this information is encoded.¹⁸ Earlier philosophers tended to think of the visual image as being like a picture of objects and their properties, with properties depicted as they are in pictures. E.g., the redness of the apple is represented by making the image of the apple red, and the spatial relationship between the apple and the table is represented by the spatial relationship in the overall visual image between the image of the apple and the image of the table. But it should be clear that this leads to an infinite regress. In order to make use of a picture as an encoding of information, we require an agent (a homunculus?) that can look at the picture and interpret the encoded information. That requires the agent to form an image of the picture and retrieve information from that image. But then the infinite regress is off and running.

Instead of saying that the visual image arranges its objects in a spatial configuration and colors them red, we should say that the visual image *represents* the objects as standing in a certain spatial configuration and as looking a certain way. It represents them in this way by some method of encoding, but we cannot informatively describe that encoding by ascribing the same properties and relationships to the constituents of the image. How then does the image encode the information? This is not something we can tell by introspection. There is no reason we should be able to. Our cognitive architecture enables us to introspect various kinds of mental goings-on only when it has a use for the results. We can introspect our thoughts and reasoning because we must be able to do so in order to correct bad reasoning, to learn that a certain kind of reasoning is unreliable under specific circumstances, etc. The information encoded in the visual image is channeled directly into thoughts about the world. That is the leading idea behind direct realism, and will be the essential idea behind our account of the mystery link. As such, that information is introspectible because thoughts must be introspectible. But for the purposes for

¹⁸ This should be clearly distinguished from the question of how the information is encoded in the brain (as opposed to the image). It is well known that there are numerous retinotopic mappings, where the pattern of stimulation in the brain has the same geometric patterns as the perceived object. This suggests that at least spatial relations are encoded "somewhat spatially" in the brain, but this has no implications for how they are encoded or accessed in the visual image.

which we use introspection, it makes no difference how the information is encoded in the perceptual image, so that is not something we can introspect.

Although we cannot, on the basis of introspection, tell how information is encoded in perceptual representations, it was observed above that we can abstractly regard a perceptual representation as a data structure. We can think of a data structure as having fields in which various information is encoded. In the case of a perceptual representation, there must be a field recording the kind of representation (e.g., edge, line, surface, object-part, object, etc.) and there must be fields recording the information about the perceived object that is provided directly by perception. When we perceive the apple, we perceive it as having an apparent color and shape properties, so the visual image contains a perceptual representation of the apple — a percept — and the apple percept contains fields for color and shape information. We also perceive the apple's spatial relationship to the table, so that information must be stored in both the apple percept and the table percept, each making reference to the other percept.

Some philosophers go so far as to deny that there is an introspectible visual image. For example, Harman (1996) writes,

You have no conscious access to the qualities of your experience by which it represents the redness of the tomato. You are aware of the redness of the represented tomato. You are not and cannot become consciously aware of the mental “paint” by virtue of which your experience represents the red tomato.

But this is just false. Consider figure fifteen again. The wall is one uniform color, but it looks different to us in different places. This is something we can introspect about ourselves.

There may remain some temptation to insist that although the look of the object is introspectible, the percept — the mental representation of the object — is not. The claim might be that when we try to attend to the percept, all we succeed in doing is attending to the object. Percepts are “transparent” (Shoemaker 1994). Although there seems to be something right about this observation, it does not imply that we cannot introspect that we have the visual representation of the object. Consider the case of hallucinating a pink elephant floating five feet above the seminar table. If you enter the room and experience such an hallucination, you will not form the belief that there actually is an elephant in that position. But you can attend to your representation and think, “How odd.” You can tell introspectively that you have a visual representation of an elephant, and you can attend to the representation while knowing full well that there is no object to attend to. Conversely, when you are in the presence of a tomato, you can tell introspectively whether you see it (have a percept of it) or fail to see it because, for instance, your eyes are closed.¹⁹

So on our view, the visual image is a transient database of data structures — visual representations. It is transient because it changes continuously as the things we see change and move about. It is produced automatically by our perceptual system,²⁰ and contains much more information than the agent has any use for at any one time. However, any of the information in the visual image is, presumably, of potential use. So our cognitive architecture provides attention mechanisms for dipping into this rich database and retrieving specific bits of information to be put to higher cognitive uses. Attention is a complex phenomenon, part of it being susceptible to logical analysis, but other parts of it succumbing only to a psychological description. The aspects of attention that are susceptible to logical analysis derive from the fact that our reasoning is “interest driven”, in the sense of Pollock (1995). Practical problems pose specific questions that

¹⁹ Shoemaker (1986) has argued that we do not literally introspect the percept. What we introspect is that we have a percept, and that it has certain properties. This may be right. We do not mean to deny it.

²⁰ We do not mean to rule out the possibility that attention plays a role in this production.

the cognizer tries to answer. Various kinds of backward reasoning lead the cognizer to become interested in questions that can potentially be answered on the basis of perception, and this in turn leads the cognizer, through low-level practical cognition, to put herself in a position and direct her eyes in such a way that her visual system will produce visual representations relevant to the questions at issue. Interest in the earlier questions then leads the cognizer to extract information from the visual image and form thoughts that provide answers to the questions. In addition to these high-level attention mechanisms, there are low-level mechanisms that function automatically. Certain kinds of perceptual representations automatically capture our attention and lead to the production of thoughts. These include things like bright flashes, loud noises, and sudden movements. Presumably evolution has decided that these are things we ought to be interested in even if we are not initially trying to answer questions to which they provide answers.

8. Direct-Seeing and the Mystery Link

Now let us return to direct realism. Direct realism is intended to capture the intuition that our perceptual apparatus connects us to the world “directly”, without our having to think about our visual image and make inferences from it. The key to understanding this is to realize that the visual image is representational. Perception constructs perceptual representations of our surroundings, and these are passed to our system of epistemic cognition to produce beliefs about the world. The latter are our “perceptual beliefs” — the first beliefs produced in response to perceptual input — and they are about physical objects and their properties, not about appearances, qualia, and the like.

8.1 Reformulating Direct Realism

These observations can be used to make a first pass at explaining the mystery link. In broad outline, our proposal will be that perception computes representations of objects and their perceivable properties. The objects are represented *as having* those properties. This is the information that is passed to epistemic cognition. The belief that is constructed in response to the perceptual input is built out of those perceptual representations of the object and of the property attributed to it. Let us see if we can make this a bit more precise.

It is useful to distinguish between beliefs and thoughts. A cognizer can entertain a thought without endorsing it as true. When she does the latter, she has a belief. We will say that she *doxastically endorses* the thought. This is just a way of expressing the familiar observation that belief is a propositional attitude. You can desire, fear, or hope for the same things you can believe. Thoughts are the neutral information encodings that can be believed, desired, feared, or hoped for.

We assume that thoughts can be regarded as having syntactic structure. This need not be understood in terms of a physical part/whole relationship, as it is for sentences of public language. The assumption is just that thoughts encode information in a productive manner — complex thoughts can be built out of simpler thoughts by some kind of compositionality. We can minimally conceive of thoughts as data structures that encode information. Some thoughts are data structures that encode their information by making reference to other thoughts. For example, a thought can be a conjunction, and then we can represent it as a data structure with a field representing its type (“conjunction”) and fields for the conjuncts. In this way, the information encoded in logically complex thoughts can be specified recursively. But not all thoughts are logical compounds of other complete thoughts. Some have subject/predicate or relational form. They can be regarded as data structures containing fields filled by mental representations of objects and their properties.

In section one we formulated direct realism as follows:

(DR) For appropriate P 's, if S believes P on the basis of being appeared to as if P , S is defeasibly justified in doing so.

We still want to endorse a principle of this form, but now we are in a position to say what counts as an appropriate P . Our suggestion is that P should simply be a reformulation of the information computed by the visual system. More precisely, suppose the cognizer sees a physical object. Then his visual system computes a visual representation O of the object — a percept. The visual system may represent the object *as having* a perceptible property. This means that the visual system also constructs a representation F of the property and stores it in the appropriate field of the percept. O and F are visual representations, and hence mental representations. As mental representations, the cognizer can use them in thinking about the object and the property. In other words, the cognizer can have the thought $\lceil O$ has the property $F \rceil$. This is not a thought about O and F — it is a thought about what O and F represent. We are using the corner quotes here much as they are used in ordinary logic. So to say that S has the thought $\lceil O$ has the property $F \rceil$ is to say that S has a thought of the form “ x has the property y ” in which “ x ” is replaced by O and “ y ” is replaced by F . Having the visual representation O that purports to represent an object as having the property F enables the cognizer to form the thought $\lceil O$ has the property $F \rceil$, and our claim is that the perceptual experience defeasibly justifies the cognizer in doxastically endorsing this thought, i.e., believing it.

In addition to perceptible properties, there are also perceptible relations (e.g., spatial relations). The same account works there. That is, where R is a visual representation of an n -place perceptible relation and the cognizer has percepts O_1, \dots, O_n putatively representing n objects as standing in the perceptible relation to one another, this enables the cognizer to form the thought $\lceil O_1, \dots, O_n$ stand in the relation R to one another \rceil , and the perceptual experience defeasibly justifies the cognizer in doxastically endorsing this thought.

This account removes the veil of mystery from the mystery link. The mystery link is the process by which a thought is constructed out of a visual image. It appeared mysterious because the thought and the visual image are logically different kinds of things. In particular, many philosophers have been tempted to say that the thought is conceptual but the visual image is not. So how can you get from the one to the other? But now we can see that this puzzlement derives from an inadequate appreciation of the structure of the visual image. It has a very rich representational structure. We are not sure what to say about whether it is conceptual. We are not sure what that means. But that doesn't seem to be relevant. The transformation of certain parts of the visual image into thoughts is a purely syntactical transformation. It takes one or more perceptual representations O_1, \dots, O_n , extracts a representation R of a property or relation from one of the fields of those representations, and then constructs a thought by putting the O_1, \dots, O_n in the subject position and R in the predicate position. There is no mystery here.

The key to understanding this aspect of the mystery link is the observation that our thought about a perceived object can be about that object by virtue of containing the percept of the object that is contained in the visual image. That is what the percept is — a mental representation of an object — and as its role is to represent an object, it can do so in thought as well as in perception. To have a thought about a perceived object, we need not somehow construct a different representation out of the perceptual representation. We can just reuse the perceptual representation. There is no “mysterious inference” involved in the mystery link. It is a simple matter of constructing one type of mental object out of another. We will refer to this process as the *direct encoding of visual information*.

There do remain interesting questions about which thoughts are constructed out of the visual image. The visual image contains much more information than we can use at any one time, so only a bit of it is extracted and moved into epistemic cognition, the rest being quickly forgotten. It is attention that determines what thoughts are extracted from the image. As remarked above, attention is a complex phenomenon, part of it being susceptible to logical analysis, but other parts of it succumbing only to a psychological description.

8.2 When Do We Form Perceptual Beliefs?

We can divide the mystery link into two parts. First, on the basis of the visual image and driven by attention, we construct a thought employing the perceptual representation to think about the objects we are seeing and the properties we are attributing to them. That is the first half of the mystery link. Second, we doxastically endorse the thought, turning it into a belief. But it remains to be said when we turn the thought into a belief. When do we doxastically endorse the thought?

It might be supposed that we automatically endorse the thought as soon as we construct it, but sometimes retract the belief immediately in the face of readily available defeating information. But that seems wrong. For instance, if you walk into the seminar room and have the visual experience of seeming to see a six foot tall transparent pink elephant floating in the air five feet above the seminar table, you are not apt to form the belief that such a thing is really there and then subsequently retract it on the basis of your knowledge of elephants. The visual experience leads you to entertain the thought, but the thought never gets endorsed.

This should not be surprising. When we form beliefs on the basis of reasoning (rather than perception), the reasoning is a process of mechanical manipulation. However, beliefs come in degrees. Some beliefs are better justified than others, and this is relevant to what beliefs we should form on the basis of the reasoning. The mere fact that a conclusion can be drawn on the basis of reasoning from beliefs we already hold does not ensure that we should believe the conclusion. After all, we might also have an argument for its negation. For instance, Jones may tell us that it is raining outside, and Smith may tell us that it is not. Both of these conclusions are inferred from things we believe, viz., that Jones said that it is raining and Smith said that it is not. But we do not want to endorse both conclusions. Having done the reasoning, we still have to decide what to believe and how strongly to believe it. Getting this right is the hardest part of constructing a theory of defeasible reasoning.²¹

It is apparent that when we engage in reasoning, forming beliefs is a two step process. First, we construct the conclusions (thoughts) that are candidates for doxastic endorsement, and then we decide whether to endorse them and how strongly to endorse them. The same thing should be true when we form beliefs on the basis of perception. First we construct the thoughts by extracting them from the perceptual image, and then we decide whether to endorse them and how strongly to endorse them. This process should be at least very similar to the evaluation step employed when we form beliefs on the basis of reasoning. Thus when you seem to see a pink elephant, you have the thought that there is a pink elephant floating in the air before you, but when that thought is evaluated, background information provides defeaters that prevent it from being endorsed. These defeaters were represented by the backward-directed grey arrows in figure one.

It will turn out that this is not yet a complete account of the mystery link, but it is useful to diagram what we have so far as in figure sixteen.

²¹ John Pollock has been pursuing this question for thirty years. For his most recent proposal, see Pollock (2002).

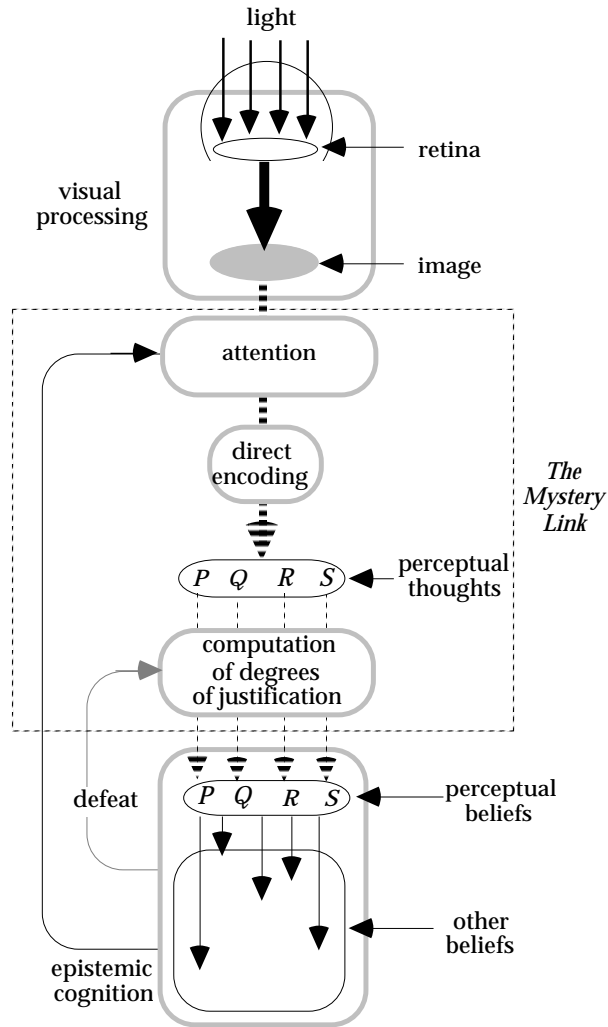


Figure 16. A first pass at explaining the mystery link

9. Visual Recognition

9.1 Recognizing Cats and Kings

In section six, it was observed that there is an important distinction between a cognizer being able to visually recognize that an object has a property and his being able to directly see that it does. The latter requires that the object be visually represented as having the property. For this to be possible, the property must have a visual representation — a characteristic look that can be encoded in the percept. Such properties have been called “perceptible properties”. As thus far explained, direct realism can only accommodate judgments attributing perceptible properties to perceived objects. However, most of our visual judgments are not like that. When you walk into the room you see that your cat is sprawled out on your easy chair. In seeing this you see an object and recognize it as a cat, and as your cat. You see a second object and recognize it as a chair, and as a particular chair. The spatial relationship consisting of the first object being on the second is something you can see directly.

How does visual recognition work? Consider recognizing something as a cat. What is extremely interesting about this is that it does not seem to be the result of an explicit inference from other simpler beliefs. We defined *perceptual beliefs* to be the initial beliefs that we form on the basis of perception. Your belief that what you see before you is a cat is, in this sense, a perceptual belief

just as much as the belief that it is on the second object (the chair). But this is not a belief attributing a perceptible property to an object.

It seems clear that most of the perceptual judgments we make are more like judgments that something is or is not a cat than like judgments that a surface is convex. The difference is that to look convex is to look a particular way. There is just one way of looking convex. It is a *sui generis* way of looking. If you doubt this, consider figures eleven and twelve again. There are other lower level aspects of the image on the basis of which the visual system decides to apply the tag “convex”, but the tag does not simply consist of those features, and the tag is introspectible and common to all cases of perceived three-dimensional convexity. By contrast, when we see a cat, there may be nothing in common between two images of cats. Vision does not supply an introspectible “cat” tag and attach it to every percept you recognize as being of a cat. Rather, cats have many different looks, and these are used evidentially in deciding you are seeing a cat. Different cats, seen in different circumstances, can look very different, but by virtue of having learned to recognize cats visually, we can relate them all to the category *cat*. However, the first point at which there is something common to all cases of recognizing cats is when our recognition issues in the *thought* “That is a cat”. By contrast, in all cases of seeing movement or seeing three-dimensional convexity, there is something common already at the level of the introspectible image that is responsible for our having the thought “That is moving” or “That is convex”.

Recognition is also context dependent. Consider recognizing a person that you know only slightly, e.g., a student in one of your classes. In the context of the class, you can recognize him reliably. But if you run into him in the grocery store, you may have no idea who he is. So recognition is cognitively penetrable, i.e., it is influenced by our beliefs. On the other hand, a vast amount of psychological evidence strongly supports the thesis that the production of the visual image is not cognitively penetrable (Pylyshyn 1999). For example, in figure eleven, knowing that the front of the statue is actually concave does not enable you to see it that way. Your other beliefs can prevent your perceptually derived thought from being endorsed as a belief, but they cannot affect what thought you entertain as the product of perception.

So visually recognizing and directly seeing are quite different in some ways, but alike in other epistemologically important respects. It is beliefs based on recognition rather than beliefs based on directly seeing that usually provide our initial epistemological access to our surroundings. Direct realism was originally defended by observing that the beliefs we get directly from perception are usually about the physical world around us and not about our own inner states. Philosophers inclined to endorse this observation have nevertheless tended to assume that the beliefs we get on the basis of perception involve only perceptible properties. The assumption has been that you cannot literally see that something is a cat or a table — only that it is shaped and colored in various ways. Now we are going one step further and noticing that perceptual beliefs are not usually beliefs attributing perceptible properties to perceived objects. Normally, we don’t form beliefs about colors and shapes much more often than we form beliefs about apparent colors and apparent shapes. What we believe on the basis of perception is, for example, that the cat is sitting on the dinner table licking the dirty plates. It never occurs to us to believe that an object with a certain highly complex shape and mottled pattern of colors is spatially juxtaposed with and above an object with a different somewhat simpler shape and pattern of colors.

If one doubts that cognition proceeds directly from perception to beliefs about cats, plates, and dinner tables, they must suppose that we first form beliefs about objects having complex shapes and colors and then make inferences to beliefs about cats. Perhaps we just make the transition so rapidly that we do not notice the first belief. But the implausibility of this hypothesis is manifest when we realize that we have no precise idea what it is about the look of a cat that makes us think it is a cat. We can say things like, “It is about a foot long, furry, with pointy ears and a long tail, and has a mottled brown color”. But note, first, that these are not perceptible properties any more than *cat* is. They are still too high level. And second, even if they were perceptible properties, they would not suffice to distinguish cats from a host of other small furry

creatures. We can *recognize* cats, but we cannot say how we do it.

Consider an even simpler example — the infamous chicken sexers. These are people who learn to identify the gender of newborn chicks on the basis of their visual appearance, the purpose being to keep only the females for their future egg-laying capabilities. Some people can learn to do this reliably, but they generally have no idea how they do it. Newborn male and female chicks do not look very different. There must be a difference, but the chicken sexers themselves are unsure what it is, and so they are certainly not first forming a belief about that difference and then inferring that what they are seeing is a female chick. As they do not know what the difference is, they do not have a belief to the effect that the chick they are observing displays that difference.

An example of this that should be familiar to everyone is face recognition. We are very good at recognizing people on the basis of their faces, but imagine trying to say what it is about a person's face that makes you think it is them. Face recognition turns on very subtle visual clues, and we often have no idea what they are.

It is useful to contrast these examples with another example. A cognizer can directly see the orientation of surfaces with respect to himself. The visual system computes a representation of this by appealing to a variety of clues, like stereopsis. However, there is an ophthalmological condition known as *aniseikonia* in which the size of the retinal image differs between the two eyes (Bennett & Rabbetts 1998). As Byrne & Hilbert (2003) report, "One effect of aniseikonia is that the orientation of surfaces in the horizontal plane is misperceived because of the binocular distance errors introduced by the difference in the magnification". What is important about this example is that perceivers cannot learn to correct for this. Even though they know better, perpendicular surfaces will never come to look perpendicular to them. By contrast, we have no difficulty learning to compensate for brunescence. The difference is that we directly see that surfaces are perpendicular to us but only visually recognize that something is red.

How is it possible to recognize something as a cat, or a female chick, without inferring that from something simpler you can see directly? We do not have a complete answer to give, but we can propose the beginnings of an answer. Consider connectionist networks (so-called "neural nets"). In their infancy, these were proposed as models of human neurons, but it is now generally recognized that existing connectionist networks are only remotely similar to systems of neurons. Nevertheless, they exhibit impressive performance on some kinds of tasks. They are perhaps most impressive in pattern recognition. A rather small network can be trained to recognize crude cat silhouettes and distinguish them from crude dog silhouettes. Larger connectionist networks have proven to be impressive pattern recognizers in a number of applications.

What does this show? It doesn't show that we are chock full of little connectionist networks. What connectionist networks are is efficient statistical analysis machines. They do a very good job of finding and encoding statistical regularities. Although it is not plausible to suppose that we are full of little connectionist networks, it is eminently plausible to suppose that our neurological structure is able to implement something with similar capabilities.²² And such capabilities are what are involved in visual recognition. Experience has the effect of training category recognizers in us. These can, in principle, take anything accessible to the system as input. In particular, they can be sensitive to data from the visual image and also to beliefs about the current context. Thus there can be many different looks that, in different contexts, fire the "cat-detector". We aren't built with cat-detectors — we acquire them through learning, much as a connectionist network learns to recognize cat silhouettes. Furthermore, different people may learn to recognize cats differently. A person who has never seen a manx cat may not recognize one as a cat because it has no tail.

²² For some recent work along these lines, see Duygulu et al (2002), Barnard et al (2003), Barnard et al (2003a), Belongie et al (2002), Yu et al (2001).

What makes visual recognition possible is the fact that cat-detectors can be sensitive to facts about the visual image and not just to the cognizer's beliefs. We do not have to have beliefs about how the cat looks in order to recognize it as a cat. The move from the image to the judgment that it is a cat can be just as direct as the move from the image to the belief that one object is on top of another. The difference is just that the latter move is built-in rather than learned, while the ability to recognize cats is learned from experience.

We assume that the look of a cat is incidental to being a cat. It follows that we have some way of identifying cats other than recognizing them visually. Then from a logical point of view, learning how cats look could be a simple matter of statistical induction. We could discover inductively that things that look a certain way in a certain context tend to be cats. This information could then be used to identify cats by applying the statistical syllogism. Roughly, the statistical syllogism licenses a defeasible inference from "This looks such-and-such, and the probability is high that if something looks such-and-such then it is a cat" to "This is a cat".²³

Inductive reasoning is difficult for a cognizer with human-like resource constraints. In particular, to engage in explicit inductive reasoning we would have to remember a huge amount of data. Outside of science, humans rarely do that. Instead, we are equipped with special purpose modules that summarize the data as we go along, without our having to keep track of it, and do induction at the same time.²⁴ Such modules can occasionally lead us astray, but on the whole it is essential for real cognitive agents to employ such short-cut procedures for inductive reasoning.

We can imagine cognitive agents that do not have cat-detectors, relying instead upon such induction modules to learn generalizations of the form "Things that look this way under these circumstances tend to be cats". They could then use those generalizations to detect cats. But there would remain an important difference between the way they detect cats and the way humans detect cats. Those agents would have to form beliefs about how objects look and then explicitly infer that they are cats. As we have noted, although humans *can* reason that way, they don't have to. Humans can recognize cats simply by having appropriate perceptual experiences, without forming beliefs about those perceptual experiences.

Logically, visual detectors should work like an explicit appeal to statistical induction and the statistical syllogism, but they make cognition more efficient by simplifying the inductive reasoning and shortcutting the need to form beliefs about appearances. The effect of this is to complicate the mystery link. The mystery link now represents two somewhat different ways to move from the visual image to beliefs about the world. First, we can do that by directly encoding some of the contents of the image into thoughts. Second, we can do this by acquiring visual detectors through learning and using them to attribute non-perceptible properties to the things we see. Thus we can expand figure sixteen as in figure seventeen.

²³ For a more careful discussion of the statistical syllogism, see Pollock (1990).

²⁴ These are Q&I ("quick and inflexible") modules, in the sense of Pollock (1989), (1995), and Pollock & Cruz (1999). Their role in induction has been discussed in all these places.

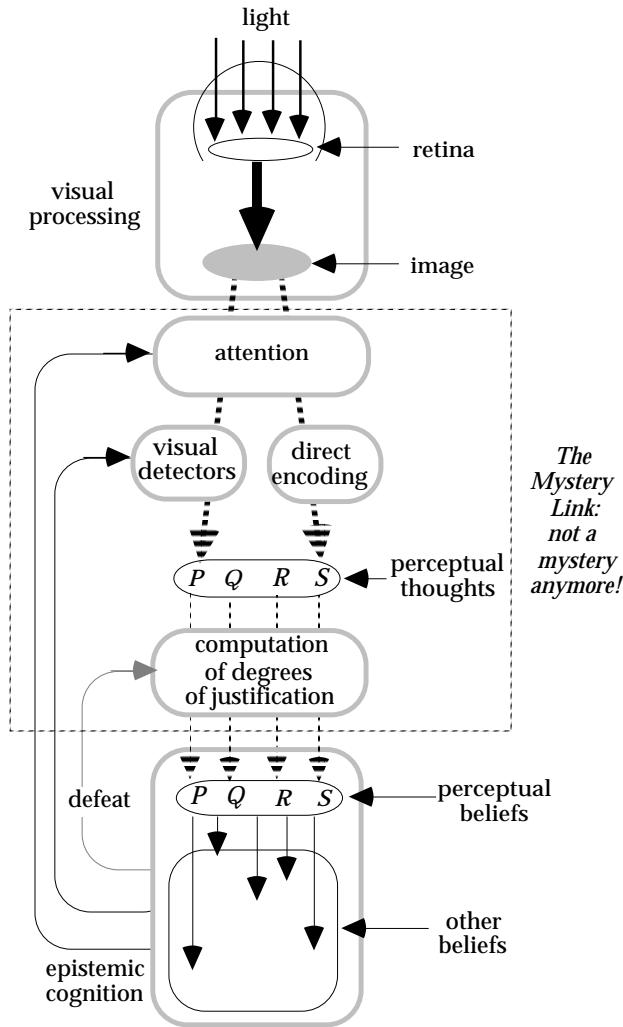


Figure 17. Explaining the mystery link

The ability to employ visual detectors must be a built-in feature of the human cognitive architecture. As remarked above, *logically*, it should work as a replacement for introspecting appearances and doing statistical reasoning about them. Cognition must be designed so that this replacement adequately mirrors the processes it is replacing. Both direct encoding and visual detectors produce thoughts — what we are calling “perceptual thoughts” to indicate their genesis — and then it remains for the cognitive system to decide whether to doxastically endorse them, i.e., turn them into beliefs. This decision must be based on the availability of defeaters, and if defeaters became available later that must lead the cognizer to retract beliefs adopted on this basis. We can capture this by understanding (DR) in the right way. Recall that (DR) was formulated as follows:

(DR) For appropriate P 's, if S believes P on the basis of being appeared to as if P , S is defeasibly justified in doing so.

We will henceforth interpret “being appeared to as if P ” as a matter of either (1) having a visual image part of which can be directly encoded to produce the thought that P , or (2) having a visual image or sequence of visual images that fires a P -detector. Appropriate P 's are simply those that can either result from direct encoding or for which the cognizer can learn a P -detector. We will

thus understand direct realism as embracing both direct encoding and visual detection.

In the previous paragraph we implicitly noted that visual detectors can appeal to a sequence of visual images rather than a single momentary image. Thus, for example, in identifying something as a cat when it is curled up in a furry ball and sound asleep, it may help a lot to walk around it and view it from different angles. Note also that if it is purring, this may decide the issue. This indicates that it is a bit simplistic to talk about exclusively *visual* detectors. We should really be talking about *perceptual* detectors, or perhaps something even broader, because the input can be cross-modal. However, for simplicity we will continue to use our present terminology.

Visual detectors can produce recognition with varying degrees of conviction. If you see the curled-up cat from across the room, you may recognize it only tentatively as a cat. If you examine it up close and from different angles, you may be fairly confident that it is a cat. If you also hear it purring, you may be certain. So we should think of visual detectors as producing outputs with differing degrees of defeasible justification depending upon just what inputs are being employed.

The incorporation of visual detection into (DR) enables it to handle perceptual beliefs that are not the product of direct encoding of the visual image, but this also changes the character of (DR) in an important respect. In Pollock (1986) and Pollock & Cruz (1999), (DR) was taken to formulate a logical relationship between *P* and being appeared to as if *P*. The claim was that this defeasible reasoning scheme is partly constitutive of the conceptual role of *P*, and hence it is a necessary truth that (DR) holds for *P*. However, *P*-detectors are only contingently connected with *P*. We generally have to learn what it looks like for *P* to be true, and different people can learn different *P*-detectors. The acquisition of a *P*-detector depends on a (generally implicit) statistical analysis. In effect, *P*-detectors are based on inductive reasoning. Thus (DR) itself remains a necessary truth, but for many *P*'s, what it is to be appeared to as if *P* is something that an agent must discover inductively.

9.2 The Mechanics of Visual Detectors

We have been intentionally vague about the way visual detectors work, just saying that they involve a statistical analysis of how things look. The human cognitive architecture may include a lot of built-in structure that facilitates this process. For example, Biederman (1985) makes the observation that many kinds of things may be recognized in terms of their parts. Consider cats. They have a number of movable parts, including legs, tails, heads, ears, eyes, etc. Because these parts can stand in many different spatial relations to one another, cats can have many different looks. But the parts have more stereotyped looks. So Biederman suggests that we may first recognize the parts in terms of their appearances and then recognize cats in terms of how the parts fit together. If Biederman is right, this amounts to saying that the human cognitive architecture makes preferential use of certain kinds of regularities in learning visual detectors. This seems very plausible.

There is at least one case in which it seems undeniable that humans employ special purpose cognitive machinery in learning visual detectors. Humans and most higher mammals are very good at recognizing other members of their species on the basis of their faces. This seems to involve skills that go beyond ordinary object recognition. Some of the evidence for this derives from the fact that damage to the medial occipitotemporal cortex can cause people to lose this ability without losing other recognitional abilities. The resulting disability is known as *prosopagnosia* or “face blindness”. Hoffman (1998) reports an example from Pallis (1995):

After his stroke, Mr. P still had outstanding memory and intelligence. He could still read and talk, and mixed well with the other patients on his ward. His vision was in most respects normal — with one notable exception: he couldn't recognize the faces of people or animals. As he put it himself, “I can see the eyes, nose, and mouth quite clearly, but they just don't add up. They all seem chalked in, like on a blackboard. ... I have to tell by the clothes or by the voice whether it is a man or a woman. ... The hair may help a lot, or if

there is a mustache ...” Even his own face, seen in a mirror, looked to him strange and unfamiliar.

The internet is a rich source of first-person accounts of prosopagnosia.

Vision scientists often fail to distinguish between visual recognition and the computation of the visual image. For example, both Marr and Biederman include the recognition of objects as the final stage of visual processing. But this is somewhat misleading. There is little that is explicitly visual about visual recognition. We noted above that the process should be viewed as multi-modal because it can be responsive to non-visual perceptual information too, like the purr of a cat. It may in fact be best to strike “visual” from “visual recognition” altogether, and just talk about recognition. Recognition is a general cognitive process that can take any information at all as input. Much of it employs visual information, but not all of it, and it seems to be essentially the same process whether visual information is used or not. For example, we often recognize people by their voices. This is a case of auditory recognition, with no visual input. For a more complex case of auditory recognition, note that it is common for a person to be able to recognize the composer of a piece of music even when they have never heard the music before. Mozart, for instance, is easy to recognize. Furthermore, there are cases of recognition that are completely non-perceptual. Just as you may recognize the composer of a piece of music, you may recognize the author of a literary work without having previously read it. There is nothing perceptual about this case at all, but it does not seem to work differently from other more perceptual varieties of recognition. The general cognitive process is recognition. Visual recognition is just recognition that is based partly on visual input. Vision provides the input, but vision does not do the recognizing.

9.3 Recognizing Colors

In section one we observed that direct realism has often been illustrated by appealing to the following putative instance of (DR):

(RED) If *S* believes that *x* is red on the basis of its looking to *S* as if *x* is red, *S* is defeasibly justified in doing so.

We noted, however, that the principle (RED) seemed to assume that there is a way of looking — *looking red* — that is logically connected with being red. It was argued that unless red objects generally look that way, (RED) will usually lead us to conclude that red objects are not red. We assume that this result would be unacceptable. However, the sliding spectrum argument shows that there is no way that red objects characteristically look to all persons and at all times. This was illustrated by a variety of phenomena, including brunescence, the Bezold-Brücke effect, simultaneous color contrast, and chromatic adaptation, plus the observation that individual differences in perceptual hardware and neural wiring (e.g., the sensitivity of photopigments) lead colored objects to look different to different people. To have a collective name for all of these phenomena, let us call them cases of *color variability*.

The preceding considerations seemed initially to constitute a counterexample to direct realism in general and to (RED) in particular. However, in light of the preceding section, we are now understanding (DR) in a more liberal way — as including reference to *P*-detectors. And we can understand (RED) analogously. What color variability indicates is that *red* is not a perceptible property. These psychological phenomena illustrate that there is no fixed way of looking that is associated, for all people and all time, with being red. Hence there is no way for the visual system to compute a fixed representation for the property of being red, i.e., it cannot represent a perceived object *as being* red. So we cannot directly see that something is red. However, we clearly do see that things are red, so this must be a case of visual recognition. *Red* is like *cat* in that we can visually recognize things as red, but this is not a matter of directly encoding aspects of the visual image. Because red things look different to different people, each person has to learn for herself how red things look, and hence how to visually identify red things. For each

cognizer, red things do have typical looks, described in terms of the cognizer's mental color space. That is what the color space is for — to enable the cognizer to identify colors. But the identification cannot be direct, because we cannot fix beforehand what region of the cognizer's color space corresponds to a particular objective color universal.

How would it be possible to acquire red-detectors through learning? Compare them to cat-detectors. Cat-detectors are just doing a statistical analysis of the looks of a pre-existing category of things, viz., cats. For red-detectors to work similarly, there must be a pre-existing category — *red* — that we can access independently of red-detectors and investigate statistically. But how can we access color categories without color-detectors? One possibility is to appeal to the fact that *red* is an interpersonal category, enshrined in public language. It denotes a range of color universals. Many ranges of color universals might constitute useful categories, and there is no reason to expect that every linguistic culture will have words for the same ranges. For instance, the quote in section three from Lindsey and Brown (2002) indicates that many languages lack a word for “blue”. Such a word is not very useful in a society in which most people suffer from pronounced brunescence.

Children certainly learn the *word* “red” from other members of their culture. This *might* be the way they learn the category *red*. This would make it entirely conventional. However, this is not the only way to learn new color categories. For example, many cars are now painted with a metallic paint that is roughly the color of old pewter. This has become a fairly familiar color, although the authors do not know a name for it. It probably has a name, but we did not learn to identify this color by having it pointed out by name. For current purposes, let us call it “pewter”. Instead of having this color pointed out by name, we observed a number of cars painted that color, and mentally constructed that category and began thinking in terms of it.

How is it possible to learn new color categories in this way? To do this we must be able to think about colors and ranges of colors and we must be able to tell when an object has a color falling within a particular range so that we can inductively generalize about what such objects look like. We assume that the concept of a color, as a kind of thing, is innate. Our cognitive architecture equips us with a number of innate concepts like *object*, *edge*, *part*, *color*, etc. These concepts do not have logical analyses. They are characterized by the role they play in cognition. How then do we reason about colors, and how does that enable us to learn color categories?

Color categories pick out continuous ranges of color — *color continua*. We have the concept of two colors being more or less similar (another innate concept), and a color continuum is a range of colors satisfying the condition that if x and y are in the range, and z is a color that is more similar to both x and y than they are to each other, then z is also in the range. We typically judge the similarity of two colors on the basis of their looking similar, i.e., their eliciting nearby color values in our phenomenal color space. This amounts to saying that color similarity is a perceptible relation, and we can make judgments about it by employing the following defeasible inference scheme, which is an instance of (DR):

(COLOR) If S believes that two simultaneously perceived objects x and y have the same (or similar) colors on the basis of their looking to S as if they have the same (or similar) colors, S is defeasibly justified in doing so.

Color variability prevents colors from having predetermined fixed looks, but it does not similarly prevent color similarity from having a predetermined fixed look. Changes in lighting, or physical changes like brunescence may change how a color looks to a person, but they will tend to change all colors in the same way. Hence they will leave perceived similarity (closeness in the color metric space) relatively unchanged. Hence *this* can be a perceptible relation, even though color categories themselves are not perceptible properties.²⁵

²⁵ Of course, to say that this is a perceptible relation is not to say that perception always represents it veridically.

If we can pick out a set O of similarly colored objects, we can then pick out a range of colors — a color category C — by stipulating that a color falls in the category C iff it is sufficiently similar to the color of one of the objects in O . So color categories get defined by reference to sets of objects. For this reasoning to be possible, we must be able to pick a set O in such a way that it consists of similarly colored objects. For this approach to be non-circular, we must be able to tell that objects are similarly colored without categorizing them as, e.g., pewter. Our proposal is that this can be done by appealing to two epistemic principles. The first is the aforementioned principle (COLOR). By appealing to (COLOR) we can judge whether two objects seen at the same time are similarly colored. But we need more than that to acquire a color category like *pewter* and learn a color-detector for it. A statistical analysis of how pewter-colored things look under different circumstances will require identifying an object as being of that color and then examining it under varying conditions. To do that we have to know that it remains the same color while we vary other conditions. This is not something that we can determine by using (COLOR), because it only applies to objects seen at the same time — not to temporally separated objects or temporally separated stages of the same object.

Can we simply generalize the principle (COLOR) to apply to objects perceived at different times? Not exactly, because (COLOR) appeals to its being the case that the percepts of x and y represent x and y as having similar colors, and that requires that the percept of x stores a reference to a percept of y and marks them as similarly colored. For that to be possible, the cognitive agent must have simultaneous percepts of x and y . But we might propose a somewhat similar principle according to which, if we can remember how x and y looked when we are no longer perceiving one or both, then if their apparent colors were similar, that gives us defeasible justification for believing that they are (objectively) similarly colored. One difficulty for applying such a principle is that we are not very good at remembering how things looked. But even if we could remember apparent colors reliably, this principle would lead us to make incorrect judgments in cases of color variability. In doing a statistical analysis of the look of a color under various circumstances, our objective is to discover when the same color looks different. But if we reasoned about colors using a principle like this one, we would be led to conclude that the colors are changing rather than that a single color is looking different.

What we need instead is the assumption that the colors of objects tend to be fairly stable, so that when we vary the lighting conditions, although the look of the color changes the color does not. This is an instance of a more general principle that is known as *temporal projection*. Temporal projection is a familiar principle in the artificial intelligence literature (Sandewall 1972, McDermott 1982, McCarthy 1986, Pollock 1998), but it has been largely overlooked in philosophical epistemology. It is basically a defeasible assumption that most “logically simple” properties of objects are relatively stable. If an object has such a property at one time, this gives us a defeasible reason for thinking it will continue to have it at a later time, although the strength of the reason decreases as the time interval increases. Pollock (1998) formulates this more precisely as follows:

If P is temporally projectible, then P 's being true at time t_0 gives us a defeasible reason for expecting P to be true at a later time t , the strength of the reason being a monotonic decreasing function of the time difference $(t - t_0)$.

This is also known as the “commonsense law of inertia”. The temporal projectibility constraint reflects the fact that this principle does not hold for some choices of P . For instance, its being 3 o'clock at one time does not give us a reason for expecting it to be 3 o'clock ten minutes later.

What justifies temporal projection? It is not an empirical discovery about the world. Any cognitive agent must employ some such principle if it is to be able to use perception to build a

The judgment is still defeasible. One place in which the representation is non-veridical is when variations are due to shadows, as, e.g., on the buses in figure two. However, this variation is the same from person to person and time to time, and is handled by temporal projection. See below.

comprehensive account of the world. The problem is that perception is essentially a sampling technique. It samples disparate bits and pieces of the world at different times, and if we are to be able to make use of multiple facts obtained from perception, we must be able to assume that they remain true for a while. For instance, suppose your task is to read two meters and record which has the higher reading, but the meters are separated so that you cannot see both at once. You look at one and note that it reads “4.7”. Then you look at the other and note that it reads “3.4”. This does not yet allow you to complete your task, because you do not know that after you turned away from the first meter it continued to read “4.7”. Obviously, humans solve this problem by assuming that things stay fixed for awhile. The logical credentials of this assumption are discussed further in Pollock (1998).

By employing (COLOR) and temporal projection in unison we can create new color categories. For instance, by employing (COLOR) we can observe that two cars are similarly colored. Later we might note that one of them is similarly colored to a third car. By temporal projection we can infer that the color of the first has not changed, so the third car is similarly colored to both of the original two cars. In this way, the set of similarly colored cars can grow, and we can stipulate that something has the color *pewter* iff it is similarly colored to the cars in this set. On the assumption that the cars tend to retain their colors, we can go on to investigate how the color looks under various circumstances, thus establishing a pewter-detector. Or more realistically, having observed the set of similarly colored cars, a pewter-detector will be learned automatically.

Note that we appeal to the set of similarly colored cars to fix the reference of the term “pewter”, not to define its meaning. It could turn out that we were mistaken about one of the cars being the same color as the others — we saw it in peculiar lighting conditions. Thus it cannot be a necessary truth that the cars in the set are pewter colored. We appeal to them simply as a way of fixing our thought on a color range we take them to exemplify.

So it seems that it should be possible to learn all of our color-detectors from experience. And it is clear that we have to learn, to some extent, how red things look to us. They look different under different circumstances, and they may well look different to each person. However, this does not imply that red-detectors cannot be innate. There are three kinds of considerations that affect how a red thing looks to a person. First, there are transitory variations in illumination, background contrast, etc. Second, there are interpersonal variations resulting from differences in perceptual hardware and neural wiring that are present “right out of the box”. Third, there are long-term variations resulting from changes to perceptual hardware and neural wiring. Brunescence is an example of the latter. Transitory variations can be accommodated by, at least initially, having a red-detector attach only low degrees of justification to conclusions based on momentary perceptual experiences. It seems likely that red-detectors could be constructed so that they are unaffected by out-of-the-box hardware variations. The neural pathways leading from red-sensitive cones to the look of the image are determined by the initial neurological structure of the cognizer, and a color-detector could be innately designed to respond to whatever range of looks is in fact wired to those cones. Thus the first two classes of variations could be accommodated reasonably well using an innately configured red-detector. However, the third class of variations has the effect that things may not stay the way they were originally. To accommodate these variations, even if the cognizer has an innate red-detector, she must be able to modify its detection properties in light of experience. Even without long term changes, this will be desirable to allow people to become more sophisticated in detecting colors so that they can take account of things like the difference between tungsten and fluorescent lighting (or even firelight). This is also going to be desirable because, although we cannot always place great reliance on color judgments based on momentary visual images, it seems that often we can. We are able to learn that under some circumstances, a glance is enough to tell the color of something, but under other circumstances we must examine things more carefully to judge their colors. So even if there are innate color-detectors, they should be tunable by experience.

We don’t need innate color-detectors, but it might be useful to have them. Do we? This can

only be answered by empirical investigation. There is, however, some evidence suggesting that we do have innate color-detectors for the primary colors. It turns out that infants can discriminate primary colors by four months of age (Bornstein, Kessen & Weiskopf 1976). It is pretty unlikely that they have learned these categories in that short amount of time, so they probably have innate detectors for the primary colors. On the other hand, it is also clear that we can acquire color-detectors for new color categories like *pewter* through learning.

9.4 Color Variability

This entire paper was motivated by the inability of traditional philosophical views about color to handle the way we reason about phenomena like brunescence. Let us see if the present account can do better. Consider four cases:

- (1) We see a white cube against a black background. Then a blue filter is put over our eyes and the cube looks blue. Then the filter is removed, and the cube looks white again. In this case we want to be able to conclude that the cube did not change color — it just looked different.
- (2) Unbeknownst to us, scientists have discovered a new electromagnetic phenomenon. By imposing an oscillating electromagnetic field of a certain frequency, they can permanently change the color of an object. The effect is analogous to painting the object, but it alters the surface properties of the material rather than covering it with a different material. We observe the field being applied to the white cube in the laboratory. Then the scientists go home for the day, leaving the cube sitting on its pedestal unattended. When the field was applied, the cube began to look blue, and it continues to do so forever more. In this case we want to conclude that the cube really did change color. This is just like painting it.
- (3) Brunescence: a membrane in our eye yellows very slowly — so slowly that we do not notice it. The look of all objects slowly changes, but we cannot remember how they looked earlier, so are unaware of the change. Then the membrane is surgically removed, and objects look bluer than they did before the surgery. We look at the white cube, and it now looks blue. It continues to do so forever more. In this case we want to conclude that things did not really change color. Rather, colors changed appearance.
- (4) A blue filter is surgically implanted in our eyes, so everything comes to look bluer than it did before. We look at the white cube, and it now looks blue. It continues to do so forever more. This is a cleaner version of case (3). The difference between this and case (1) is that, like case (2), the change is permanent. But unlike case (2), we want to conclude that things did not really change color.

In case (1) we want to judge that the cube did not change color. This seems to be based on temporal projection. Suppose we have a strong reason for thinking that the cube is white at t_0 . By temporal projection, we can infer that it is still white at a later time t , although the strength of the reason decreases as the interval $(t - t_0)$ increases. Meanwhile, the cube's looking blue at time t gives us a reason for thinking it is blue. If we assume, as discussed above, that momentary perception gives us a weaker reason for thinking it is blue than we originally had for thinking it was white, then when $(t - t_0)$ is small, temporal projection gives us a stronger reason for thinking the cube is still white than momentary perception gives us for thinking it is now blue. But when $(t - t_0)$ becomes large enough, perception overwhelms temporal projection. In case (1), we can make the appearance switch quickly by placing the filter over our eyes and removing it. So temporal projection swamps momentary perception, and we conclude that the cube has not changed color.

To make this reasoning work, it must be the case that momentary perception does not give us as good a reason for thinking that the cube is blue as we had originally for thinking it was white. It was remarked above that it seems to be a general characteristic of visual detectors that recognition based on more observations gives us more confidence (a higher degree of justification) in the

recognition. For instance, if your cat is moving about, or if you move around it and see it from different angles, this increases your confidence that you are seeing a cat. Similarly, if you observe the color of an object over an extended period, in various lighting conditions and from different angles, you will be more confident about its color. So case (1) seems to be unproblematic on the current theory.

In case (2), when the cube appears to change color, temporal projection initially gives us a rebutting defeater, just as in case (1). But the fact that the change is permanent will eventually make us reconsider. This is explained by noting that as the interval $(t - t_0)$ increases temporal projection gives us a systematically weaker reason for thinking that the cube is still white. At some point that reason becomes significantly weaker than the reason momentary perception gives us for thinking it is blue, so at that point the application of (DR) overwhelms temporal projection. Thus it becomes reasonable to think that the cube is blue. We can then inductively generalize about the change, learning that subjecting an object to such an electromagnetic field changes its color. In this way we can acquire an undercutting defeater for future applications of temporal projection in such cases.

The crucial difference between cases (3) and (4) and case (2) seems to be that *everything* changes apparent color. To handle cases (3) and (4) we have to think about how the color detector works. Acquiring a blue-detector through learning is logically analogous to confirming by statistical induction that when things look a certain way they tend to be blue. Let us pretend for the moment that this is how we acquire the blue-detector. Then it is based on an initial evidence set E . E will consist of many blue things that look a certain way, many non-blue things that do not look that way, and perhaps a few non-blue things that do look that way. Suppose, suddenly, everything looks bluer than it did, and in particular things that were previously white all look blue. By temporal projection we can conclude that those white things are still white but just look blue — this is like case (1). From this we can infer inductively that white things now look blue. We can also infer by temporal projection that most of the non-blue things in E are still non-blue. But we can also conclude, from our inductive generalization, that many of them now look blue. This undercuts the earlier statistical induction supporting the establishment of the generalization that when things look a certain way they tend to be blue. Thus the fact that everything looks bluer now than it did before gives us no reason for thinking that things really are bluer, and temporal projection gives us a reason for thinking they are not. Hence it is reasonable to conclude that colored things no longer look the same way they did. In fact, we have confirmed inductively that the way they look has shifted towards the blue. This provides the basis for new (or modified) color-detectors.

Of course, the preceding is based on the pretense that color detectors are established by explicit statistical confirmation. They aren't really. They are the product of a psychological learning process that goes on without any direction from the cognizer. But the logic of the use of such color detectors should be the same as if they were the product of statistical confirmation. This indicates that if we notice that everything seems to have changed apparent color, this should defeat color judgments based on the application of (DR) and our color-detectors. Initially, we will conclude by induction that apparent colors have changed, and so we will reason about colors by thinking about apparent colors — something we ordinarily avoid. As one of us can attest from personal experience, this is in fact exactly what patients do after cataract surgery. But after a while our color detectors will adjust, and we will no longer have to think about how things look. We can go back to making automatic judgments on the basis of having the visual image, unmediated by beliefs about the visual image.

Our conclusion is that the present theory — a refined version of direct realism — is able to handle variations in the appearance of colors in ways that earlier theories could not. Earlier theories assumed that every color has a characteristic appearance that is an essential and unchangeable feature of it. The connection between colors and appearances must be looser, and the present theory accommodates that by understanding principles like (RED) in terms of color-

detectors rather than direct encoding. So these examples do not constitute a threat to direct realism.

10. Whence Justification?

We began this paper with a familiar philosophical question. When we look around and see the world surrounding us, what justifies the beliefs we form? Traditionally, many epistemologists tried to answer this question by appealing to “the given” and postulating some kind of “direct apprehension” of the given. The claim was then that our perceptual beliefs are based on inference from the “basic beliefs” resulting from this direct apprehension. However, there is no appropriate class of basic beliefs from which our perceptual beliefs can be inferred. Direct realism proposes to accommodate this by retaining the given, but denying that perceptual beliefs are inferred from beliefs about the given. We don’t directly apprehend the given — we just *have* it, in the form of visual images. According to direct realism, it is the having of an appropriate visual image that justifies one in holding perceptual beliefs about one’s immediate surroundings.

Direct realism avoids the problems concerning direct apprehension, but only at the expense of encountering another problem. What is the link between having the image and having the justified perceptual belief? This is what we have called “the mystery link”, and the bulk of the paper has been concerned with explaining the mystery link. Our answer has appealed to a wealth of empirical knowledge about the human visual system, and then used that to explain how “higher-level” epistemic cognition can interface with it. This explains the “cognitive mechanics” of visual cognition. But by virtue of what is our answer an answer to the philosophical question? We have talked about how visual cognition works, but what we want to know is why it *justifies* us in forming beliefs about the world.

To answer this question, we must consider what exactly we are asking when we ask about the justification of a belief. Epistemologists of an internalist persuasion have traditionally sought to answer epistemological questions by describing how we actually reason about various subject matters and describing what we regard as good reasoning. Our discussion is in the same vein, except that unlike most traditional epistemologists we do not approach this as an a priori question. Vision is a complex matter, and one cannot construct a credible epistemological theory of perceptual knowledge without having a clear idea of what the visual system provides to us as input to higher-level cognition.

Externalists will applaud our appeal to the empirical facts about how vision works, but they will object to the idea that we can throw light on the structure of epistemic justification just by looking at how we in fact reason. Externalists typically insist that to show that our epistemological procedures produce justified beliefs we must show that they are reliable — that they tend to produce true beliefs. We have no doubt that visual cognition is highly reliable, but that seems largely beside the point in evaluating reasoning, and much too crude an instrument of assessment for investigating the fine structure of the reasoning. For example, can you justify the use of cat-detectors on the grounds of reliability? Presumably they are fairly reliable, but it seems likely that a careful statistical analysis of the appearance of cats would produce even more reliable recognition of cats. Does that mean we should not employ cat-detectors? Certainly not. More than reliability must be involved in the design of a cognitive architecture. At the very least it is important that it can provide useful information in a timely fashion. Careful statistical analysis will be more reliable than visual detection, but it will be very slow and consume much too much of our limited cognitive resources. If we want to know whether that is a tiger (a *big* cat) stalking us, we do not have time for statistical analysis.

Reliability is important, but it is our conviction that externalists have misplaced its significance in rational cognition. A distinction must be made between assessing a cognitive architecture and assessing individual cognitive performances by agents possessing that architecture. We assess a cognitive architecture in terms of its contribution to the achievement of various design goals (of

either artificial or evolutionary origin). A propensity to produce true beliefs will contribute to the achievement of most design goals a cognitive architecture might be deemed to pursue. However, for familiar reasons, merely producing true beliefs will not make an architecture successful. At the very least, it must produce useful beliefs, and in many cases it must produce them fairly quickly.²⁶

Epistemic justification is about the assessment of individual beliefs, not cognitive architectures. Considerations of reliability are useful for assessing cognitive architectures, but not individual beliefs.²⁷ How then can individual beliefs be assessed? This raises a puzzling question. It is clear why we might want to assess an architecture, but why assess individual beliefs? Of course, we might assess them in terms of truth, but assessing the justifiedness of a belief is to assess it in a different way. What are the assessments of epistemic justification all about?

The human cognitive architecture exhibits an interesting characteristic. Human beings can sometimes flaunt the dictates of rationality. They can engage in wishful thinking, hasty generalization, inadequate searches for conflicting evidence, and so on. Why is this possible? In designing a cognitive architecture, the most straightforward procedure would be to lay down rules for how we want the agent to cognize and then build the agent so that it invariably does cognize in that way. For such an agent, irrationality would be impossible. There would be no distinction between rational and irrational thought. But humans are not like this. The explanation for this aspect of human cognition is at least partly that humans are “reflexive cognizers”. That is, they can think about their own cognition and, to some extent, redirect it. At the very least they can make deliberate choices about what to think about. Thus they can decide what questions to address next, whether to look for more evidence for a hypothesis, whether to think about certain considerations that might lead them to the solution to a problem, and so on. This makes it possible for them to tailor their problem solving activities in light of what they have learned about how to solve various kinds of problems and the likelihood of being able to solve a particular problem given the information currently at their disposal. A cognitive architecture must implement a set of default rules for how and when to engage in various cognitive activities, but by enabling an agent to override these default rules we make the agent more flexible and we give it the ability to tune its problem solving procedures in light of experience. On the other hand, this also opens the door to various kinds of irrationality. For example, when one has a cherished view and it is suggested that certain considerations may show it wrong, there is a temptation to think about other things. Ignoring possible difficulties for one’s views is a prime example of irrationality, but it is made possible by giving the agent some control over the course of its cognition.

If an agent has voluntary control over some aspects of its own cognition, then it becomes possible for it to make choices based on desires having nothing to do with efficient problem solving. The agent might, for example, value not being proven wrong, and on that basis decide not to think about possible difficulties for a cherished theory. A well-designed cognitive architecture will incorporate features aimed at avoiding this. In human beings, this is accomplished by an appeal to what can be regarded as a competence/performance distinction. The cognitive architecture incorporates rules for how to cognize, but does not absolutely insist that the agent conform to those rules for cognition. On the other hand, it is desirable from the point of view of agent design to make sure that the agent can tell when it is not conforming to the rules, and build in some kind of disposition to modify one’s cognitive behavior when one realizes that one is not conforming to the rules.

²⁶ See Pollock & Cruz (1999) for a more sustained discussion of this point.

²⁷ Of course, reliabilists have tried to use them for that purpose, but the generality problems shows that cannot be done. See Pollock & Cruz (1999), chapter four.

This can be regarded as a competence/performance distinction because it works like competence/performance distinctions elsewhere.²⁸ It is a general characteristic of human beings that we are able to learn complex behaviors for performing various activities, and internalize them in the sense that we do not have to think about how to perform the activities in order to employ those learned behaviors. When we do this, we have acquired *procedural knowledge*. However, having the procedural knowledge means only that we *know how* to engage in the activity. It does not ensure that we will successfully pull it off. The learned behavior may be sufficiently complex that it is hard to do. Think of swinging a golf club or a tennis racket, or riding a mountain bike over difficult terrain. When a cognizer is engaging in one of these activities, it is helpful if he can tell when he is not managing to do it in conformance with the way he has learned to do it. The learning provides a target for his performance, but it can be a hard target to hit. Thus, in human beings, acquiring procedural knowledge for how to do something carries with it the ability to tell by a kind of introspection that you are not doing it the way your procedural knowledge tells you to do it.

Chomsky (1957) introduced the distinction between competence theories and performance theories, applying it specifically to linguistics, but the distinction can be applied wherever we have procedural knowledge for how to do something. A performance theory describes what we actually do under various circumstances, while a competence theory describes what we have learned to do but may not successfully pull off. In other words, a competence theory articulates our procedural knowledge. Our procedural knowledge is about what to do, so a competence theory can be formulated as a set of rules for behavior — a set of *norms*. For most kinds of procedural knowledge, there is nothing binding about these norms. They are normative in form only. They describe how we have learned to do something, but we might very well have learned to do it differently.

Cognizing is something we know how to do. Of course, we have no control over many aspects of cognition, like the computation of the visual image. These are not things that *we* do — our cognitive system does them, but not us. But by the same score, they do not fall under the purview of rationality. A person is not irrational because her cognitive system computes a representation of the statue in figure eleven that represents the front as convex. Considerations of rationality do not apply there, because these aspects of cognition are not under her control. But other aspects of cognition are under her control. If she has prior knowledge that the front of the statue is concave, but for some reason ignores that and accepts the visual representation as veridical (for instance, because for some reason she wants it to be the case that the good fairy has transformed the shape of the statue), then she is being irrational. Assessments of rationality only apply in cases in which she has control over her own cognitive behavior. The norms describing an agent's built-in rules for epistemic cognition are her *epistemic norms*. An agent's epistemic norms are a built-in feature of her predetermined cognitive architecture.

The term “epistemic justification” is a term of art, but one way it has often been applied is in evaluating whether the cognition supporting a belief conforms to the cognizer's epistemic norms. Following Pollock (1986), we will call this *procedural justification*. Epistemic norms are built in. There can also be learned rules for how to cognize, but conforming to the learned rules is only justified in this sense insofar as the learned rules themselves can be justified on the basis of the agent's built-in rules.

It was remarked above that it is natural to describe procedural knowledge in terms of a set of norms for how to engage in the target activity, but for most kinds of procedural knowledge there is nothing binding about the norms. One could have learned to do things differently. The norms are “normative” in name only. However, there is more to the normativity of cognitive norms.

²⁸ This was first observed by Pollock (1987), and forms the basis of the theory developed in Pollock (1986) and Pollock & Cruz (1999).

When we discover that our cognitive behavior does not conform to our cognitive norms, we have a built-in disposition to try to correct our cognitive behavior to make it conform to the norms. As we remarked above, this is simply part of being a reflexive cognizer. By giving an agent the power to diverge from its default rules for cognition, we are apt to give it too much power. One way to bring its cognitive behavior back into line with the design goals of the cognitive architecture is to build in a pressure to conform to a set of cognitive norms, and that is just what happens in human beings. For this to work, cognitive agents must have a way of telling that they are not conforming to their cognitive norms. This requires that they be able to detect cases in which their behavior diverges from the norms. This is something that humans can do. Note that they can do this without being able to articulate the norms themselves.

When we (the authors of this paper) talk about justified beliefs, it is procedural justification that interests us. We want to know what rules for rational cognition are built into the human cognitive architecture. In this paper we are specifically interested in the rules governing belief formation in response to visual input. It is a contingent fact that the human cognitive architecture has the features it does. Thus the investigation of our built-in rules for rational cognition is an empirical investigation. However, it is one that we, as human beings, can perform more easily than alien psychologists could. To do this we can trade upon our ability to detect when we are not conforming to the norms. This general methodology is the same as that employed by linguists and first articulated by Chomsky. Linguists attempt to articulate their linguistic norms by taking as their data their “linguistic intuitions” about whether they are conforming to those norms in specific cases. Similarly, human philosophers can attempt to articulate their cognitive norms by taking as their data their “cognitive intuitions” about whether they are conforming to their cognitive norms. This is, we suggest, the best way to understand much familiar philosophical methodology, with its appeal to “philosophical intuitions”. What is important here is that our cognitive intuitions concern individual cognitive performances. Discovering the norms underlying these intuitions is a matter of constructing a theory that explains the data, and this is no different from scientific theorizing elsewhere.²⁹

For investigating some aspects of epistemic cognition, this methodology of “cognitive intuitions” may be all we need. For example, this may give us all the data we need for trying to construct a theory of inductive reasoning. However, this methodology is too limited to enable us to construct an adequate theory of the epistemic norms governing perceptual beliefs. We will probably all agree about which perceptual beliefs are justified and which are not, but this is not enough to enable us to figure out what the norms are that give rise to these cognitive intuitions. For that we have to know what the input is that produces perceptual beliefs. That sounds like a straightforward psychological problem, but there is a converse constraint on a psychological theory of vision. This is that it must produce the kind of output that higher-level cognition needs for belief formation. This is a logical and computational constraint.

We remarked that to completely understand perceptual knowledge, we need three theories. First, we need a psychological theory of visual processing and the production of the visual image. Second, we need a philosophical theory of higher-level cognition telling us how perceptual beliefs are to be evaluated in light of our other beliefs and how they can be used to support other beliefs. But third, we also need a theory of the connection between the first two theories — a theory of the mystery link. This cannot be a purely philosophical theory, because it must take account of what the output of the visual processing system is, i.e., the nature of the visual image. And it cannot be a purely psychological theory because it must take account of what the visual image is to be used for in higher-level cognition. These three theories are inextricably interlinked. A constraint on any one of them must be that it fits together properly with the others. A theory of visual processing cannot ignore the problem of making the mystery link work, and a philo-

²⁹ These observations derive from Pollock (1986).

sophical theory of perceptual reasoning cannot ignore what kind of input it is possible for the visual processing system to supply to higher-level cognition. The crux of all of this is a theory of the mystery link, and that has been our topic. It is thus simultaneously a theory of epistemic justification as it pertains to perceptual beliefs, a theory of the logical structure of the visual image, and a theory of how perceptual beliefs are constructed out of visual images with such a logical structure.

This explains why, to give a philosophical theory of the conditions under which a person is justified in forming beliefs on the basis of perception, we must construct a more comprehensive theory of visual cognition. Our theory is encapsulated in figure seventeen. Assessments of epistemic justification turn on assessing the correctness of those cognitive acts over which the cognizer has some voluntary control. It is useful to look at figure seventeen from that perspective. Much of visual cognition is carried out automatically and without the cognizer having any voluntary control, however the cognizer's decisions about how to proceed have more effect than one might initially suppose. The visual image is a transitory database from which information is extracted by either direct encoding or visual recognition. Most of the information in the image is not extracted. It is attention that drives the extraction, and attention is to a large extent under voluntary control. Some aspects of attention are automatic, but most are deliberate and are affected by prior epistemic cognition. Epistemic cognition is "interest-driven" and bidirectional (Pollock 1995). Epistemic cognition does not aim at forming beliefs randomly. It tries to answer specific queries posed in pursuit of the cognizer's goals. Both epistemic interests and beliefs affect how the cognizer directs its attention in retrieving information from the visual image. Thus prior rational transgressions can have adverse effects even at the level of extracting information from the visual image.

Perhaps most aspects of the visual image are computed automatically, but as we saw in section three, at least some aspects of it are affected by attention. Furthermore, the contents of the visual image are profoundly influenced by where the cognizer directs his eyes, and this is something the cognizer can do deliberately. In pursuing epistemic interests, a cognizer often has to engage in practical reasoning about how to acquire various kinds of information. Some of it can be acquired by looking in the right direction, and a cognizer can be criticized for irrationality if he does not do so. The visual image is also the product of multiple saccades, and these too are under a certain amount of voluntary control.

Everything in figure seventeen between attention and the computation of degrees of justification proceeds automatically, and perceptual beliefs are the automatic product of this. However, the computation of degrees of justification appeals to what defeating beliefs the cognizer already has, and that is a matter of prior reasoning. So even though a perceptual belief is produced automatically, the cognizer can be unjustified in holding it because he has not formed defeating beliefs he should have formed, and he can be unjustified in not endorsing a perceptual thought because he believes a defeater unjustifiably.

The question whether a person is justified in holding perceptual beliefs about his surroundings is thus answered by appeal to the rules governing the voluntary aspects of visual cognition. The hard part of getting those rules right is understanding how visual cognition is able to encode information and pass it to epistemic cognition. That is a matter of explaining the mystery link, and that is what we claim to have done in this paper.

References

- Bach, Kent
1982 *De re* belief and methodological solipsism. In *Thought and Object: Essays on Intentionality*, ed. Andrew Woodfield. Oxford: Oxford University Press.
- Barnard, Kobus, Pinar Duygulu, Nando de Freitas, David Forsyth, David Blei, and Michael I.

- Jordan
 2003 "Matching Words and Pictures," *Journal of Machine Learning Research*, Vol 3, pp 1107-1135.
 Barnard, Kobus, Pinar Duygulu, Raghavendra Guru, Prasad Gabbur, and David Forsyth
 2003a "The effects of segmentation and feature choice in a translation model of object recognition",
 in Danielle Martin (ed), *Computer Vision and Pattern Recognition*, vol II, 675-684. IEEE
 Press.
- Belongie, Serge, Jitendra Malik, and Jan Puzicha
 2002 "Shape matching and object recognition using shape contexts", *IEEE Trans. on Pattern
 Analysis and Machine Intelligence* **24**, 509-522.
- Biederman, I.
 1985 "Human image understanding: Recent research and a theory." *Computer Vision, Graphics,
 and Image Processing* **32**: 29-73.
- Block, Ned
 1990 "Inverted earth", *Philosophical Perspectives* **4**, 53-79.
 2001 "Toward a defense of empirical foundationalism", in Michael R. DePaul (ed.), *Resurrecting
 Old-Fashioned Foundationalism*. Lanham, Maryland: Rowman and Littlefield.
- Bornstein, M. H., Kessen, W., & Weiskopf, S.
 1976 "Color vision and hue categorization in young human infants". *Journal of Experimental
 Psychology: Human Perception & Performance*, **2**(1), 115-129.
- Byrne, Alex and David R. Hilbert
 2003 "Color realism and color science", *Behavioral and Brain Sciences* **26**.
- Chomsky, Noam
 1957 *Syntactic Structures*, The Hague: Mouton and Co.
- Clark, Romane
 1973 "Sensuous judgments". *Nous* **7**, 45-56.
- Cohen, Jonathan
 2003 "Perceptual variation, realism, and relativization, Or: How I learned to stop worrying and
 love variations in color vision", *Behavioral and Brain Sciences* **26**.
- Davidson, Donald
 1983 "A coherentist theory of truth and knowledge". D. Henrich (ed.), *Kant oder Hegel*. Stuttgart:
 Klett-Cotta.
- Duygulu, Pinar, Kobus Barnard, Nando de Freitas, and David Forsyth
 2002 "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary",
Seventh European Conference on Computer Vision, pp IV:97-112.
- Fairchild, Mark D.
 1998 *Color Appearance Models*. Reading, Mass: Addison-Wesley.
- Faugeras, O.
 1993 *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, MA: MIT Press.
- Fodor, J. A.
 1975 *The Language of Thought*. Cambridge: Harvard University Press.
- Gibson, J.
 1966 *The Senses considered as Perceptual Systems*. Boston: Houghton Mifflin.
- Harman, Gilbert
 1996 "Explaining objective color in terms of subjective reactions", *Philosophical Issues* **7**, 21-38.
- Hoffman, D.
 1998 *Visual Intelligence: How We Create What We See*. New York. W.W. Norton & Company.
- Huemer, Michael
 2001 *Skepticism and the Veil of Perception*. Lanham, Maryland: Rowman and Littlefield.
- Hurvitch, L M.
 1981 *Color Vision*. Sinauer Associates Inc.
- Hurvich, L. M., D. Jameson & J. Cohen

- 1968 "The experimental determination of unique green in the spectrum". *Perception and Psychophysics* **4**, 65-68.
- Jepson, A. D., and W. A. Richards
1993 *Perceiving events and objects*. Hillsdale, N. J.: Lawrence Erlbaum.
- Kosslyn, S. M.
1983 *Ghosts in the Mind's Machine*, New York: Norton.
- Land, E. H. & McCann, J. J.
1971 "Lightness and retinex theory". *American Journal of Optical Society of America* **61**, 1-11.
- Lindsey, D. T., & Brown, A. M.
2002 "Color naming and the phototoxic effects of sunlight on the eye". *Psychological Science*, **13**(6), 506-512.
- Marr, D.
1982 *Vision*. San Francisco: Freeman.
- Marr, D. & E. Hildreth
1980 "Theory of Edge Detection", *Proc. R. Soc. London*, **B 207**, 187-217.
- McCarthy, John
1986 "Applications of circumscription to formalizing common sense knowledge." *Artificial Intelligence* **26**, 89-116.
- McDermott, Drew
1982 "A temporal logic for reasoning about processes and plans", *Cognitive Science* **6**, 101-155.
- McGinn, C.
1983 *The Subjective View: Secondary Qualities and Indexical Thoughts*. Oxford University Press: Oxford.
- Merbs, S. L. and J. Nathans
1992 "Absorption spectra of human cone pigments" *Nature* **356**, 433-35.
- Neitz, M. and J. Neitz
1998 "Molecular genetics and the biological basis of color vision". In *Color Vision: Perspectives from Different Disciplines*, (eds) W. Backhaus, R. Kliegl and J. S. Werner. Walter de Gruyter.
- Pallis, C. A.
1995 "Impaired identification of faces and places with agnosia for colors". *Journal of Neurology, Neurosurgery and Psychiatry* **18**, 218-224.
- Peacocke, Christopher
1992 *A Study of Concepts*. Cambridge: MIT Press.
- Pylyshyn, Zenon
1999 "Is vision continuous with cognition? The case for Cognitive impenetrability of visual perception", *Behavioral and Brain Sciences* **22**, 341-423.
- Pollock, John
1971 "Perceptual Knowledge", *Philosophical Review*, **80**, 287-319.
1974 *Knowledge and Justification*, Princeton University Press.
1983 *Language and Thought*, Princeton University Press.
1986 *Contemporary Theories of Knowledge*, Rowman and Littlefield.
1990 *Nomic Probability and the Foundations of Induction*, Oxford University Press.
1995 *Cognitive Carpentry*, MIT Press.
1998 "Perceiving and reasoning about a changing world", *Computational Intelligence*. **14**, 498-562.
2002 "Defeasible reasoning with variable degrees of justification", *Artificial Intelligence* **133**, 233-282.
- Pollock, John, and Joseph Cruz
1999 *Contemporary Theories of Knowledge*, 2nd edition, Lanham, Maryland: Rowman and Littlefield.
- Pryor, James
2000 "The skeptic and the dogmatist", *Nous* **34**, 517-549.
- Sandewall, Erik

- 1972 "An approach to the frame problem and its implementation". In B. Metzger & D. Michie (eds.), *Machine Intelligence 7*. Edinburgh: Edinburgh University Press.
- Quinton, Anthony
1973 *The Nature of Things*, London: Routledge and Kegan Paul.
- Sellars, Wilfrid
1963 "Empiricism and the philosophy of mind." Reprinted in *Science, Perception, and Reality*, New York: Humanities Press; London: Routledge & Kegan Paul.
- Shoemaker, Sydney
1986 "Introspection and the self". *Midwest Studies in Philosophy* **10**, 101-120.
1994 "Self-Knowledge and 'Inner Sense': Lecture III: The Phenomenal Character of Experience", *Philosophy and Phenomenological Research* **54**: 291-314.
- Siegel, Susanna
2004 "Which properties are represented in perception?", in Tamar Gendler-Szabo and John Hawthorne (eds.), *Perceptual Experience*, New York: Oxford University Press.
- Siewert, Charles
1998 *The Significance of Consciousness*. Princeton: Princeton University Press.
- Smith, David Woodruff
1984 "Content and context of perception". *Synthese* **61**, 61-87.
1986 "The ins and outs of perception". *Philosophical Studies* **49**, 187-212.
- Sosa, Ernest
1981 The raft and the pyramid: coherence versus foundations in the theory of knowledge. *Midwest Studies in Philosophy*, vol. 5, pp. 3-26. Minneapolis: University of Minnesota Press.
- Thompson, E. A., A. Palacios & F. J. Varela
1992 "Ways of coloring: Comparative color vision as a case study for cognitive science". *Behavioral and Brain Sciences* **15**, 1-74.
- Tye, Michael
2000 *Consciousness, Color, and Content*. Cambridge: MIT Press.
- Tyler, H. R.
1968 "Abnormalities of perception with defective eye movements (Balint's syndrome)", *Cortex* **3**, 154-171.
- Ullman, S.
1996 *High-level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: MIT Press.
- Yu, Yizhou, Andras Ferencz, and Jitendra Malik
2001 *IEEE Trans. on Visualization and Computer Graphics* **7**, 351-364.
- Yullie, A. L. & S. Ullman
1990 "Computational theories of low-level vision", chapter one of *Visual Cognition and Action*, vol. 2 of *An Invitation to Cognitive Science*, ed. D. Osherson, S. Kosslyn, and J. Hollerbach, Cambridge, MA: MIT Press.
- Zihl, J., D. von Cramon & N. Mai
1983 "Selective disturbance of movement vision after bilateral brain damage". *Brain* **106**, 313-340.