

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.

We will focus 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 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. 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. The perceptual beliefs 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 think of this as an epistemological theory of reasoning. We will assume without argument that it 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

¹ Supported by NSF grants nos. IRI-IIS-080888 and IIS-0412791.

² Supported by a grant from Rutgers University.

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>.

about the mystery link, and a correct understanding of it severely constrains what kinds of epistemological theories of perceptual knowledge can be correct.

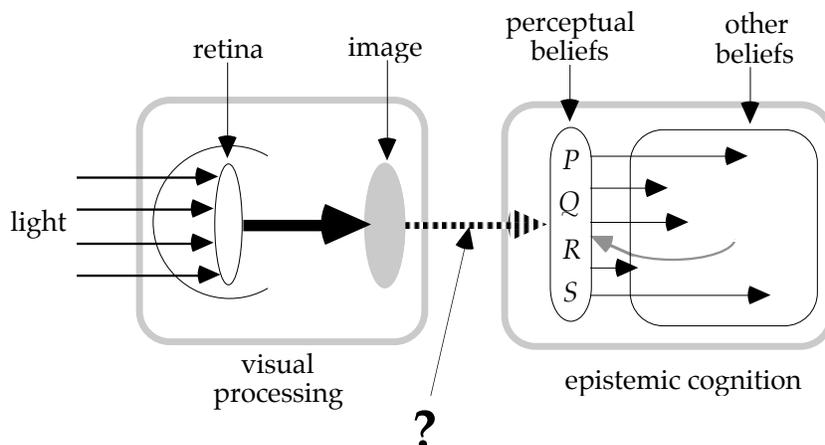


Figure 1. Knowledge, perception, and the mystery link

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 raise a difficulty for 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 be shown that this theory of the mystery link provides machinery for constructing a modified version of direct realism that avoids the difficulty 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. The doxastic assumption has unfortunate consequences when applied to perception. Perceptual beliefs — the first beliefs formed on the basis of perception — are by their very nature not obtained by inference from previously held beliefs. Perception gives us new information that we could not get by inference alone. As 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. Thus perceptual beliefs 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 nothing other than beliefs can be relevant.

Historical foundations theories tried to make this plausible by taking perceptual beliefs to be about the cognizer’s perceptual experience. The trouble is, 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. 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. These beliefs need support, so they are not self-justified.

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 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

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

following.⁵ The argument just given in its defense seems to us to be strong. However, in the next section we will raise a difficulty for the theory. 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 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 arises without requiring an independent argument. 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.” 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. 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? 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

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

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. However, one could maintain instead that the brain somehow compensates so that colors continue to look the same as one ages. Which is right? It turns out that there is hard scientific data supporting the conclusion that brunescence alters the way things look to us, even if we don't notice the effects. Brunescence lowers discrimination between blues and purples (Fairchild 1998). Consequently, people suffering from brunescence cannot discriminate as many different phenomenal appearances in that range of colors. 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. In *chromatic adaptation*, looking at one color and then looking at a contrasting color changes the second apparent color. This is illustrated by afterimages.

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? 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 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, for any apparent color we choose, it is likely that 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.

The classical picture of the visual image was, in effect, that it is a two-dimensional array of colored pixels — a bitmap image.⁶ 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 (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. Philosophers sometimes refer to this bitmap as “the retinal representation”, but for present purposes 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

⁶ This view has recently been endorsed again, at least tentatively, by Bonjour (2001), and also by Kosslyn (1983).

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.

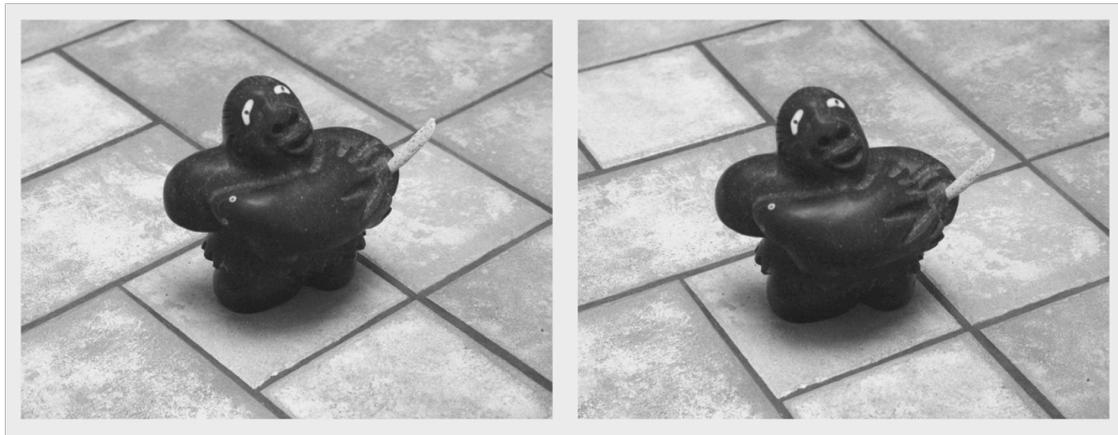


Figure 2. Two retinal bitmaps



Figure 3. Laying the bitmaps on top of one another.

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. This is illustrated in figure two. 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, as in figure three, 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 visual system uses the difference between the two bitmaps to compute depth, and this is an important part of why we can see three-dimensional relationships between the objects we see. This highlights 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 mimicking 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?

In fact, it does. A second illustration of this is that 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.

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. This is known as the "correspondence problem" of visual perception. 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 affects 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 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. Along the way representations are computed for edges, corners, motion, parts of objects, objects, etc. For our purposes, the most important idea these theories share is that visual processing produces *representations* of edges, corners, surfaces, objects, parts of objects, etc., and throws away most of the rest of the information contained in the retinal bitmap.

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. 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 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 to more patterns of retinal stimulation than there are elementary particles in the universe? That seems unlikely. 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. 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.

The crucial observation that will lead us to an account of the mystery link 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, people do not usually have thoughts about their 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.⁷ That is, they can occupy the “subject position” in such thoughts.⁸

5. Seeing Properties

The visual image does not just contain representations of objects. It also represents objects as having properties and as standing in relationships to one another. Consider the scene depicted in figure four, 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 that* a number of things are true of them. For

⁷ 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).

⁸ 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.

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.



Figure 4. Scene.

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

You *might* see that the following are true:

- (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. The visual system, by itself, cannot 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.

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 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 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 the representation of cats that you employ in thinking about them 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 have become bald. So the visual appearance of cats is not the representation we employ when thinking about cats. We have some other way of thinking about cats, and learn (or discover inductively) that things looking a certain way are generally cats.

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 further in section seven.

It is an empirical matter just what properties are represented by the visual system and recorded as properties of perceived objects. We cannot decide this *a priori*, but 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. One conclusion that vision scientists generally agree upon is 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. This is further illustrated by the “apparent motion” illusion. 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. From a functional perspective, your visual image 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.

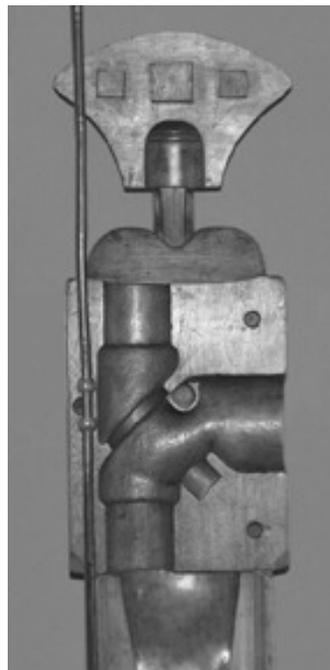


Figure 5. Convex

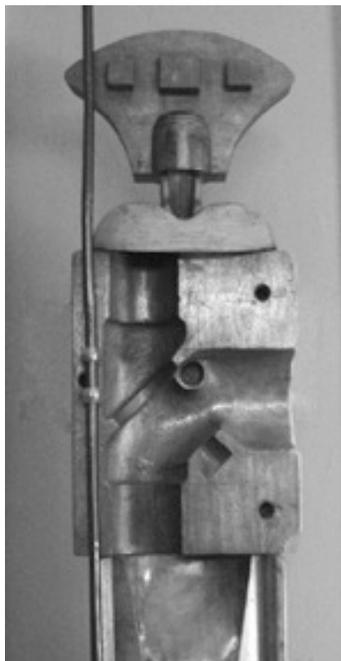


Figure 6. Concave

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 five appear.⁹ The convexity has a clear visual representation in your image. Contrast this with figure six, 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 almost impossible to see them that way in figure five, and it is very difficult to see them as convex in figure six. So three-dimensional convexity and concavity have characteristic “looks” (but, of course, these looks are not

⁹ The statue is the work of Tony de Castro, of Tiradentes, Brazil.

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 relation to the percept. We can usefully think of the percept as a data-structure. Data-structures have “fields” at which information is stored. When the appropriate field of the percept 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.

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 four. Note particularly the visual representation of the orientation of the floor. Three-dimensional orientation is perceived partly on the basis of stereopsis, but it can also be perceived without the aid of stereopsis, as in figure four.

6. 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.

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.

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 $\ulcorner O$ has the property $F \urcorner$. 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 $\ulcorner O$ has the property $F \urcorner$ 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 $\ulcorner O$ has the property $F \urcorner$, and our claim is that the perceptual experience defeasibly justifies the cognizer in doxastically endorsing this thought, i.e., believing it.

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, some 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 whatever it means, it doesn't seem to be relevant. The transformation of certain parts of the visual image into thoughts is a purely syntactical transformation. It takes a perceptual representation O , extracts a representation F of a property from one of the fields of the representation, and then constructs a thought by putting the O in the subject position and F in the predicate position. There is no mystery here.

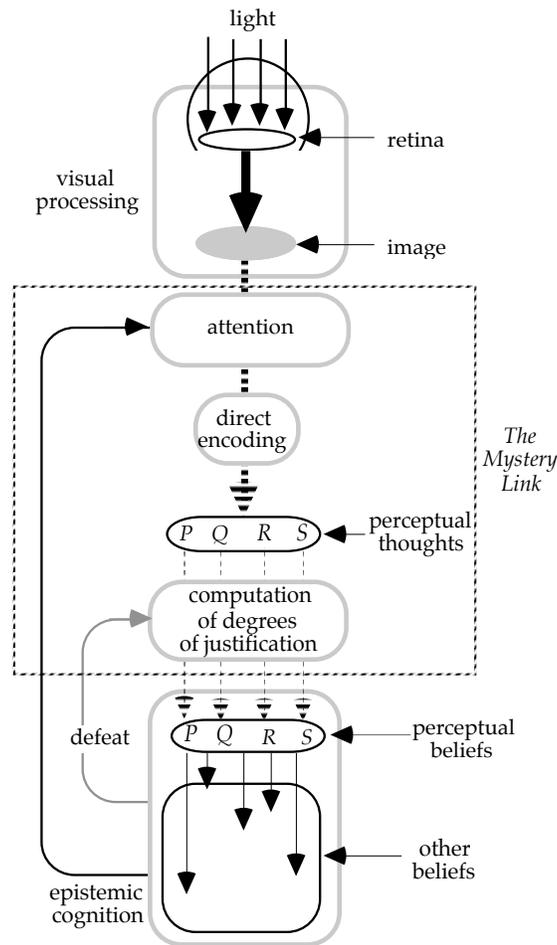


Figure 7. A first pass at explaining the mystery link

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*.

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 seven. Several comments are in order. First, note the role of attention in figure seven. 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 of perceptual representations, and we can abstractly regard 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. If we see that the apple is sitting on the table, 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.

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, 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. Some perceptual events, like abrupt motions or flashes of light, attract our attention automatically. Others are more cognitively driven and are susceptible to logical analysis deriving from the fact that our reasoning is “interest driven”, in the sense of Pollock (1995). Practical problems pose specific questions that 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. 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.¹⁰ 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.

Figure seven also makes a distinction between perceptual *thoughts* and perceptual *beliefs*. On the basis of the visual image and driven by attention, we construct a thought employing perceptual representations to think about the objects we are seeing and the properties we are attributing to

¹⁰ 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.

them. However, we need not endorse the thought that is produced in this way. 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. The visual experience leads you to entertain the thought, but the thought never gets endorsed.

This should not be surprising. Compare reasoning. 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.¹¹

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.

7. What Can't We See?

7.1 Seeing Colors

Now let us consider colors again. This was the original motivation for trying to make direct realism clearer. We see that objects have various colors. For example, when we see a London bus, we see that it is 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.

We certainly can see that a physical object is red. But as we have noted, this could either be a matter of directly seeing that it is red (in which case vision represents the object as red), or a matter of the cognizer visually recognizing that it is red. What is at issue is whether the visual system itself can provide the information that it is 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 part of their look.

We will refer to the points in this color space as *color values*. It is natural to suppose that the color values that are used to mark points on a perceived surface 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

¹¹ John Pollock has been pursuing this question for thirty years. For his most recent proposal, see Pollock (2002).

that colors have characteristic appearances that are partly constitutive of their being the colors they are.

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? For that matter, 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 eight. 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 more yellow 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.

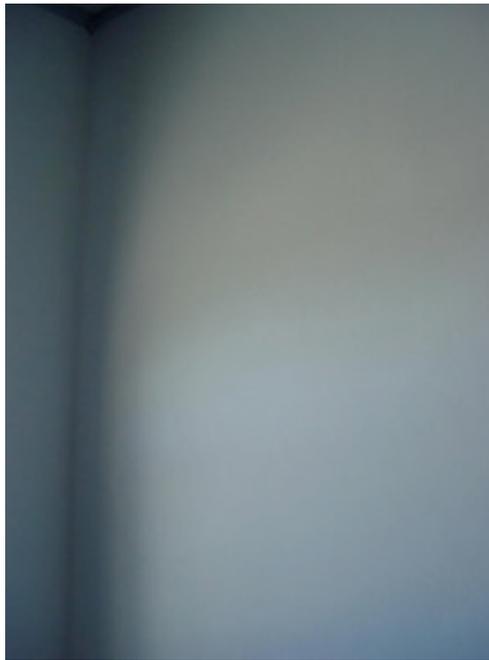


Figure 8. A wall painted a uniform color.

A single color looks very different under different circumstances. Which circumstances define “looking that color”? We imagine that some 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. It is affected by the brightness of the illumination (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.¹² There is no such thing as a “normal” perceiver.

¹² It seems obvious to us that colors do not have unique “correct” looks, 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.

Apparently we cannot directly see that objects have particular colors. We can visually recognize things as exemplifying specific color universals, but that is different from directly seeing the colors.

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. Because red does not have a characteristic look, there is no way the visual system can represent objects as being red. To see that an object is red must be to exercise a (partially) learned capacity to recognize that objects are red.

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.

7.2 Seeing Shapes

In response to the preceding, some may be tempted to say, "So what? We always knew colors were strange anyway. They are mere secondary qualities, maybe not real properties at all." It is interesting then that similar observations can be made about many shape properties. 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 8. 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 8 appear again on both sides of the monolith in figure 9. 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.

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 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.

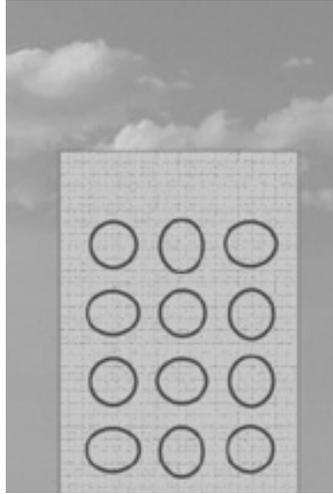


Figure 9. Circles at a right angle.



Figure 10. Circles at an oblique angle.

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.

8. Visual Recognition

8.1 Recognizing Cats and Kings

We observed above 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.

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

There may be nothing in common between two images of cats. 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*. 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). This is illustrated by perceptual illusions. Consider again the statue in figures five and six. 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. 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 there being 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.

How is it possible to recognize something as a cat 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.¹³ 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.

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. 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".

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.¹⁴ 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

¹³ 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).

¹⁴ 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.

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 seven as in figure ten.

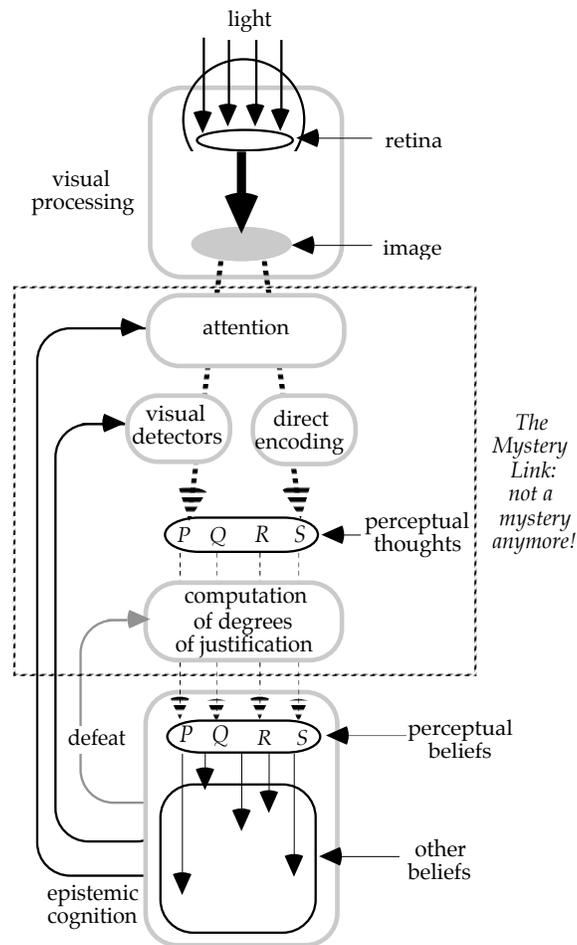


Figure 11. Explaining the mystery link

Direct realism must be extended to accommodate these observations and give visual recognition a central role in the formation of perceptual beliefs. 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.

8.2 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.

8.3 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, nor is there any sensible set of “normal viewing conditions”. 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, Lindsey and Brown (2002) indicates that many languages lack a word for “blue” because the adults in that linguistic community suffer from pronounced brunescence:

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.

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 color (but not the concept of any particular color) is innate. More generally, 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. If the concept of color plays this kind of role in our cognitive architecture, how 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 space) relatively unchanged. Hence *this* can be a perceptible relation, even though color categories themselves are not perceptible properties.

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 perception itself marks the percept of x and the percept of y 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 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 (albeit controversial) suggesting that we do have innate color-detectors specifically 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.

8.4 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 we 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 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.

9. Conclusions

The mystery link is the process by which information is moved from the visual image into epistemic cognition. A constraint on psychological theories of vision and philosophical theories of knowledge is that they must fit together via some account of the mystery link. 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 morass 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 produces visual representations of physical objects and some of their properties and relations. 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.

We have argued that this is done in two distinct ways. The simplest is direct encoding, wherein representations are retrieved from perception and inserted into thoughts. However, only the simplest perceptual beliefs can be produced in this way. More sophisticated cognition requires visual recognition, wherein we learn to recognize properties on the basis of their appearances. This is what is involved in seeing colors.

These have been remarks about how vision and the mystery link work. They are about the structure of visual cognition. What, one might ask, has this to do with the justifiedness of our perceptual beliefs? We do not have the space to discuss this issue at length, but a brief answer is that our interest is in the procedural concept of epistemic justification developed first in Pollock (1986) and further in Pollock and Cruz (1999). This concerns the epistemic norms that are built into our cognitive architecture, and their structure is to be elucidated by investigating how cognition works. See Pollock (2006) for the most recent incarnation of this view.

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.
- Bonjour, Lawrence
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**.
- Clark, Romane
1973 "Sensuous judgments". *Nous* **7**, 45-56.
- 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.
- 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.
- Hurvich, 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.
- Kosslyn, S. M.
1983 *Ghosts in the Mind's Machine*, New York: Norton.
- 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.
- 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.
- Pollock, John
1971 "Perceptual Knowledge", *Philosophical Review* **80**, 287-319.
1974 *Knowledge and Justification*, Princeton University Press.
1986 *Contemporary Theories of Knowledge*, Rowman and Littlefield.
1989 *How to Build a Person*, Cambridge, MA: Bradford/MIT 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.
2006 "Irrationality and cognition", in *Topics in Contemporary Philosophy*, ed. Joseph Campbell and Michael O'Rourke, MIT Press.
- 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.
- Pylyshyn, Zenon

- 1999 "Is vision continuous with cognition? The case for Cognitive impenetrability of visual perception", *Behavioral and Brain Sciences* **22**, 341-423.
- Quinton, Anthony
 1973 *The Nature of Things*, London: Routledge and Kegan Paul.
- 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.
- Sellars, Wilfrid
 1963 "Empiricism and the philosophy of mind." Reprinted in *Science, Perception, and Reality*, New York: Humanities Press; London: Routledge & Kegan Paul.
- 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.
- 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.