1/22/04 This is a shorter version of a paper by the same title to appear in *Epistemology: New Essays*, Quentin Smith (ed), Oxford University Press

Vision, Knowledge, and the Mystery Link

John L. Pollock[1]
Department of Philosophy
University of Arizona
pollock@arizona.edu
http://www.u.arizona.edu/~pollock

Iris Oved[2]
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 primarily on visual perception. Figure one presents a crude diagram of the cognitive system of an agent capable of forming beliefs on the basis of visual perception. Cognition 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*.

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

[2] Supported by a grant from Rutgers University.

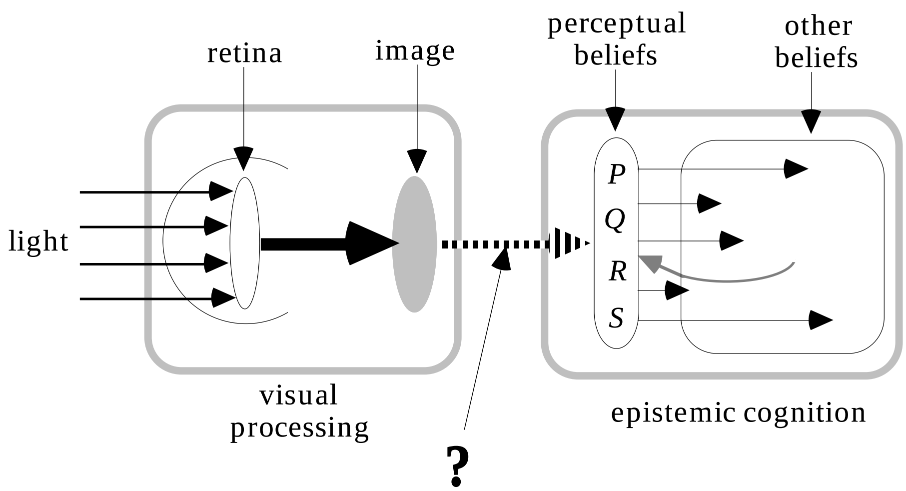

Figure 1. Knowledge, perception, and the mystery link

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

This paper will begin by looking briefly at epistemological theories of perceptual knowledge. We will present an argument for "direct realism", which we endorse, and then 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.

---

[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.

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

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

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 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 talking about precise shades of red or generic redness. The problems are basically the same for both, but they are more dramatic for the case of precise shades of red.

The principle (RED) relates the concept *red* to a way of looking — an *apparent color*. It tells us that something’s having that apparent color gives defeasible support for the conclusion that it is red. Defeasible support 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,

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

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 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. 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 numerous other kinds of color shifts to which human perception is subject. These include a blue-yellow shift caused by changes in illumination (the Bezold-Brücke effect), simultaneous color contrast, chromatic adaptation, and differences in the L-cone photopigment gene (Merbs & Nathans 1992). The upshot of this is that there is no way of looking — call it *looking red* — such that objects are typically red iff they look red. In fact, it is likely that for any apparent color we choose, objects that are red will typically *not* look that color. If our judgments of color were based on principles like (RED), we would almost always be led to conclude defeasibly that red objects are not red. Furthermore, it follows from direct realism that there would be no possible way for us to correct these judgments by discovering that they are unreliable, because any other source of knowledge about redness would have to be justified inductively by reference to objects judged red using the principle (RED). It seems apparent that the principle (RED) cannot be a correct account of how we judge the colors of objects. But (RED) also seems to be a stereotypical instance of direct realism. It is certainly the standard example that Pollock (1986) and Pollock & Cruz (1999) used throughout their defense of direct realism. Thus (DR) itself seems to be in doubt.

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

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

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 reasons that will become apparent later, we will reserve the term “representation” for higher-level mental items, including various constituents of the visual image.

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

Figure 2. Two retinal bitmaps

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 difference between the two bitmaps 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?[7]

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

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

[7] This is illustrated further in the longer version of this paper.

is viewed as an information processor that takes inputs from the rods and cones on the retinas and outputs the visual image as a structured array of mental representations. The visual system is understood as analyzing the stimulus in several stages, more or less following the original framework developed by Marr. 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 contianed in the retinal bitmap.

Representations

Computational theories of vision differ in their details, and no existing theory is able to handle all of the subtleties of the human visual system. But what we want to take away from these theories and use for our purposes is fairly general. First, all theories agree that there is a great deal of complex processing involved in getting from the pattern of stimulation on the retina to the introspectible visual image in terms of which we see the world. Not even the simplest parts of the visual image can be read off the retinal bitmaps directly. 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 epistemological problem begins with this image, not with an uninterpreted bitmap. As we will see, this makes the epistemological problem vastly simpler.

But why, the epistemologist might ask, does the visual system do all the dirty work for us? Is there something rationally suspect here? There are purely computational reasons that at least suggest that this could not be otherwise. There are 130 million rods and 7 million cones in each eye, so the number of possible patterns of stimulation on the two retinas is $2^{274,000,000}$, which is approximately $10^{82,482,219}$. That is an unbelievably large number. Compare it with the estimated number of elementary particles in the universe, which is 10^{78}. Could a real agent be built in such a way that it could respond differentially in some practically useful way between more patterns of retinal stimulation than there are elementary particles in the universe? 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. However, it is hard to construct an argument for the assumption that the visual system cannot respond differentially to more than 10^{78} different patterns of retinal stimulation without knowing more about how information is stored in the brain. It seems that with some kind of "distributed representation" it might be possible to build a system that is capable of responding to more patterns than there are neurons in the brain. Still, $10^{82,482,219}$ is an incredibly large number. It is hard to believe that any system could make sensible discriminations between more than a minute portion of these patterns. If so, this explains why the human visual system works by performing simple initial computations on the retinal bitmap, discarding the rest of the information, and then uses the results of those computations to compute the final image.

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, we do not usually have thoughts about our visual images. Usually, the first thoughts we get in response to perception are thoughts about the objects perceived, not thoughts about visual images. For a thought to be about something, it must contain a *representation* of the item it is about. In perceptual beliefs, physical objects can be represented by representations that are provided by perception itself.

We will call perceptual representations of objects *percepts*. The claim is then that the visual image provides the perceiver with percepts of the objects perceived, and those percepts can play the role of representations in perceptual beliefs.[8] That is, they can occupy the “subject position” in such thoughts.[9]

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 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

[8] 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).

[9] 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.

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 the following:

(7) The pot is from Amazonia.

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

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

When you *see that* something is true of an object, you are seeing that the object has a property. However, the different property attributions in these *see-that* claims have different statuses. Some are based directly on the presentations of the visual system, while others require considerable additional knowledge of the world. 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.

5.1 Some Perceptible Properties

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

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.[10] 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 impossible to see them that way in figure five, and it is impossible to see them as convex in figure six. So three-dimensional convexity and concavity have characteristic “looks” (but, of course, these looks are not always veridical). The percepts represent objects as having concave or convex features. It does not take much thought to realize that these looks are sui generis. The phenomenal quality that constitutes the look does not have an analysis. It is *caused by* (computed on the basis of) all sorts of lower-level features of the visual image, but it does not simply consist of those lower-level features. Once again, from a functional point of view this simply amounts to storing a tag of some sort in the appropriate field of the percept (viewed as a data structure). When the field is occupied by the appropriate tag, the object is perceived as convex. This is despite the fact that convexity has a logical analysis in terms of other kinds of spatial properties of objects.

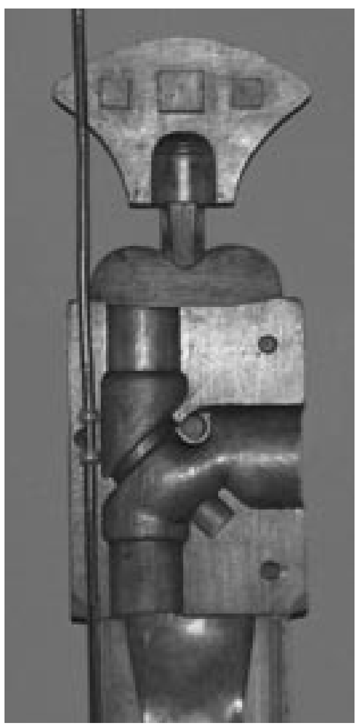

Figure 5. Convex

Figure 6. Concave

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

[10] The statue is the work of Tony de Castro, of Tiradentes, Brazil.

of the floor. Three-dimensional orientation is perceived partly on the basis of stereopsis, as illustrated by the stereograms in figure three, but it can also be perceived without the aid of stereopsis, as in figure ten.

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.

6.1 Reformulating Direct Realism

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

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 ⌜*O* has the property *F*⌝. Having the visual representation *O* that purports to represent an object as having the property *F* enables the cognizer to form the thought ⌜O has the property F⌝, 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.

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 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, and attention as a mechanism that selects items from that database and passes them to epistemic cognition for further processing. Attention is a complicated phenomenon. Some perceptual events, like abrupt motions or flashes of light, attract our attention automatically. Others are more cognitively driven. We have interests in particular matters. For instance, you might want to know the color of Joan's blouse. This interest leads you to direct your eyes in the appropriate direction, and to attend to the perceptual representation you get of Joan's blouse, and particularly to the representation of the color. The visual image you get by looking in Joan's direction contains much more than the information you are seeking, but attention allows you to extract just the information of use in answering the question you are interested in.[11] 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.

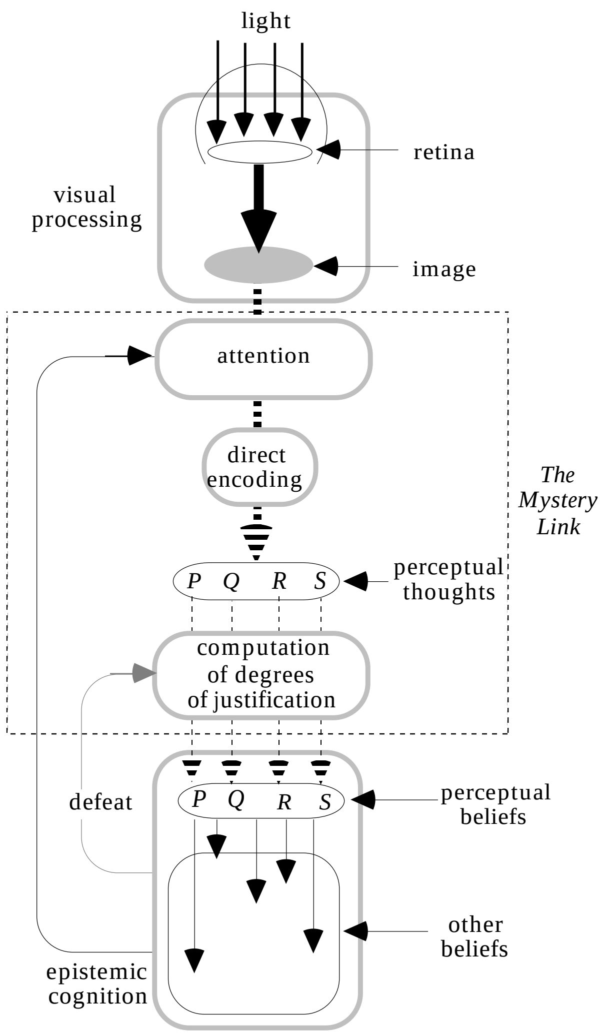

Figure 7. A first pass at explaining the mystery link

Figure seven also makes a distinction between perceptual thoughts and perceptual beliefs.

[11] 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.

On the basis of the visual image and driven by attention, we construct a thought employing the perceptual representation to think about the objects we are seeing and the properties we are attributing to them. However, we do 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.[12]

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.

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

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

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 seen, 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 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? Furthermore, we need not rely upon anything as exotic as brunescence to make this point. Simply stand in a room whose walls are painted some uniform color but unevenly illuminated by a bright window, and look at the color. A photograph of such a wall is displayed in figure 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 yellower than others. In fact, the effect is so pronounced in real life that we were unconvinced the wall really was a uniform color, and we tested it by moving a matching paint chip around the surface. It matched everywhere. So when we see this wall, we are seeing the same objective color everywhere — the same color universal is instantiated by every point on the wall. But in our visual image the representation of the wall is marked by widely varying color values. Which of these color values is the “right” color value? Does that make any sense? Surely not.

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.[13] There is no such thing as a "normal" perceiver.

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

[13] It seems obvious to us that colors do not 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.

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 no 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 nine. 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 nine appear again on both sides of the monolith in figure ten. 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.

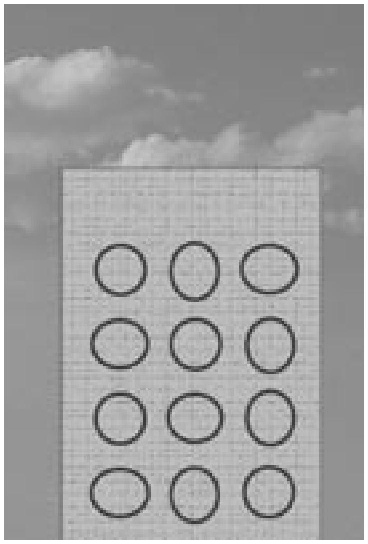

Figure 9. Circles at a right angle.

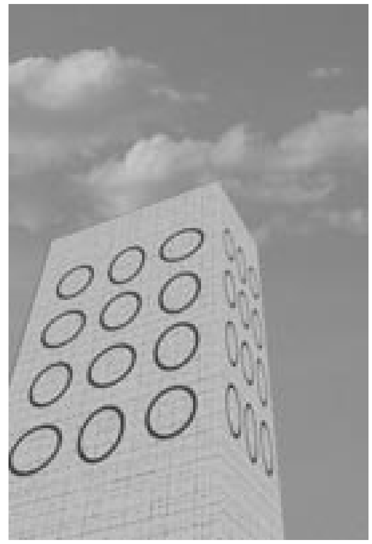

Figure 10. Circles at an oblique angle

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 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 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.[14] 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 will not pursue the details here, but our proposal is that something similar is involved in recognizing colors. We employ color detectors that respond primarily to the looks of things but are also sensitive to context. At least some color detectors are learned from experience. But there is some evidence to the effect that we are equipped innately with color detectors for the primary colors (Bornstein, Kessen & Weiskopf 1976). However, even if they are innate such color detectors must still be trainable so that we can learn to make use of context and correct for age-related changes to our perceptual apparatus.

The effect of this is to complicate the mystery link. The mystery link now represents two somewhat different ways to move from the visual image to beliefs about the world. First, we can do that by directly encoding some of the contents of the image into thoughts. Second, we can do this by acquiring visual detectors through learning and using them to attribute non-perceptible properties to the things we see. Thus we can expand figure sixteen as in figure seventeen.

---

[14] 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).

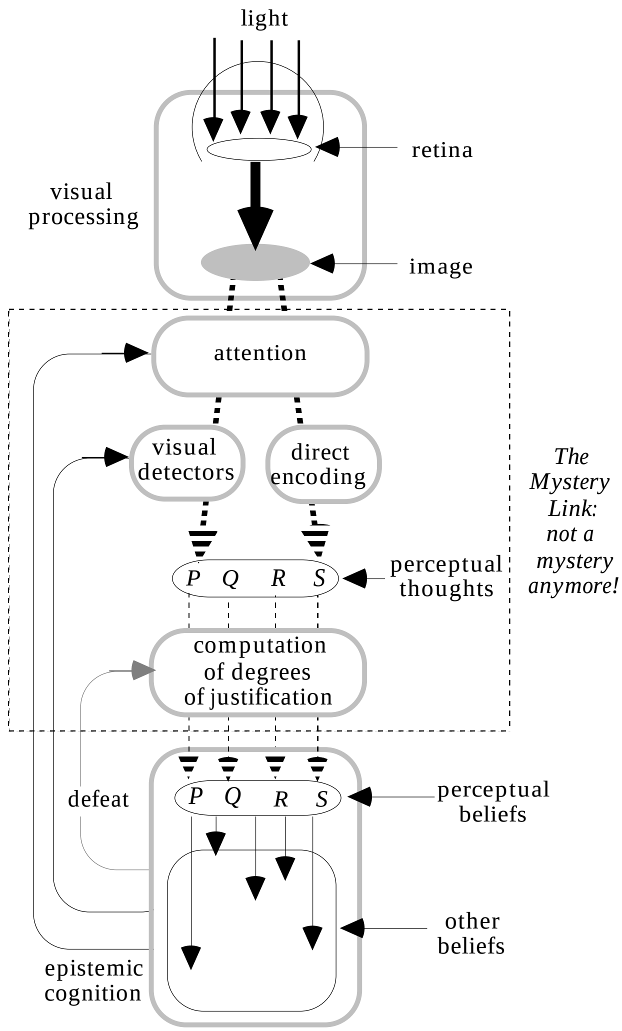

Figure 17. Explaining the mystery link

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.

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 melange of colors and shades corresponding directly to the bitmap of retinal stimulation. Contemporary scientific theories of perception insist instead that the visual image is the product of computational processing that 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.

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.

Biederman, I.

1985 "Human image understanding: Recent research and a theory." *Computer Vision, Graphics, and Image Processing* 32: 29-73.

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.

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*. Suttgart:

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.
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.
Marr, D. & E. Hildreth
1980 "Theory of Edge Detection", *Proc. R. Soc. London*, **B 207**, 187-217.
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.
Pylyshyn, Zenon
1999 "Is vision continuous with cognition? The case for Cognitive impenetrability of visual perception", *Behavioral and Brain Sciences* **22**, 341-423.
Pollock, John
1971 "Perceptual Knowledge", *Philosophical Review*, **80**, 287-319.
1974 *Knowledge and Justification*, Princeton University Press.
1986 *Contemporary Theories of Knowledge*, Rowman and Littlefield.
1995 *Cognitive Carpentry*, 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.
Quinton, Anthony
1973 *The Nature of Things*, London: Routledge and Kegan Paul.
Sellars, Wilfrid
1963 "Empiricism and the philosophy of mind." Reprinted in *Science, Perception, and Reality*, New York: Humanities Press; London: Routledge & Kegan Paul.
Smith, David Woodruff
1984 "Content and context of perception". *Synthese* **61**, 61-87.
1986 "The ins and outs of perception". *Philosophical Studies* **49**, 187-212.
Sosa, Ernest
1981 The raft and the pyramid: coherence versus foundations in the theory of knowledge. *Midwest Studies in Philosophy*, vol. 5, pp. 3-26. Minneapolis: University of Minnesota

Press.
Thompson, E. A., A. Palacios & F. J. Varela
1992 “Ways of coloring: Comparative color vision as a case study for cognitive science”. *Behavioral and Brain Sciences* **15**, 1-74.
Tye, Michael
2000 *Consciousness, Color, and Content*. Cambridge: MIT Press.
Tyler, H. R.
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.