Flint Schier, Selective Commitment and Frame Knowledge

Abstract: Both Flint Schier and Robert Hopkins invoke the concept of frame knowledge when explaining how drawings can be selectively committed to how and what they depict. Both claim that our background knowledge of the world enables us to determine which elements of a sketch refer and which don't. It is curious that despite the fact that these writers espouse radically different accounts of how depiction works, that they both need to resort to the concept of frame knowledge – effectively common-sense - to resolve a central issue in the theory of how we interpret depictions. This essay evaluates the role of common-sense in theories of depiction.

The sketch depicts a naked woman. She is delineated with only a dozen or so lines but the viewer is easily able to discern her posture and gender – we can guess her age and speculate on her mood. The square, boyish shoulders and full hips tell us that it is a young woman. The sketch lacks details in crucial areas - there is no face, no breasts and no hands, even the chin is missing. Nevertheless, our impression is not a vague one – her weight is on her right leg, she might be in the act of swaying, but she is balanced. A few marks on paper convey all this. The sketch is quite specific about the gender, posture and build, yet it is vague about other attributes. This capacity of a drawing to pick out certain attributes over others is an important characteristic of depiction – it is a selective activity. The artist has chosen to highlight certain features and merely intimate others. Some aspects of the lines may be arbitrary, such as thickness or colour. Other aspects, such as the indistinctness of the hair on the model's head, are deliberate attempts by the artist to convey the concept of hair but not any particular kind of hair. If the hair

1

was depicted with elaborate detail we would wonder why the rest of the picture was so sketchy.

A sketch is necessarily selective in what it depicts. It is constrained by materials, time and, in this case, the skill of the artist (myself). However, as Goodman, and many commentators since have, pointed out, no picture can depict every aspect of a scene. A picture selects the things and the properties of things that it represents.

Both Flint Schier and Robert Hopkins invoke the concept of frame knowledge when explaining how drawings can be selectively committed to how and what they depict.[1] Both Schier and Hopkins claim that the reason we do not interpret sketches, such as the one above as depicting someone with thick black lines around them, or having no face, is that our background knowledge of the world tells us that people always have faces and never have black lines around them. It is curious that despite the fact that these writers espouse radically different accounts of how depiction works, that they both need to resort to the concept of frame knowledge to resolve a central issue in the theory of how we interpret depictions.

Schier also refers to this frame knowldge as "background knowledge" and "common knowledge about the world".[2] Hopkins refers to it as "widespread knowledge of a very general nature."[3] It is clear that Hopkins' definition includes:-

- Knowledge of what sort of items the world contains and their properties;
- Knowledge of the sort of items in general depicted;
- Knowledge of the various means for producing depictions.

It is clear that to some extent the kind of knowledge both these writers are appealing to explain selective commitment is what we normally call common-sense. [Need to invoke Dennett here] What is not clear is how deep the epistemological problem is with regard to depiction and selective commitment.

If we are able to ascertain which elements of a drawing are representational, and which are not, using common-sense, it raises the question "How important are the other mechanisms proposed by Schier and Hopkins in explaining the mechanism of pictorial interpretation?" That is, do we really need to subscribe to the sophisticated mechanisms

---

[1] See {Hopkins, 1995 #288}and {Schier, 1986 #32}
[2] {Schier, 1986 #32@167-8}
[3] {Hopkins, 1998 #290@137}

of resemblance theory or recognition theory if the explanation for the central question in depiction – selective commitment - is that we use common sense to filter out the noise. If we can tell which elements of a picture refer and which don't using common sense why do we need theory of depiction? It looks as if the most significant part of the process is resolved using frame knowledge of the world and the kind of things that pictures depict. Could it be that they are both simply wrong and that it is implicit and explicit knowledge of the conventions of depiction which enables us to distinguish between noise and depictive elements?

There are two reasons to suspect that the conventionalist view does not help here. Firstly, if knowledge of drawing techniques enabled people to discount the significance of thick outlines and missing and sketchy elements, naïve viewers would have more trouble seeing content in line drawings than in colour photographs. In fact, naïve viewers have no trouble identifying content in line drawings and neither do certain kinds of monkey and ape. It follows that viewers who have never seen a line drawing before cannot be using knowledge of drawing techniques to gauge what is to be taken literally in a picture and what is an accident of technique.

Secondly there is the famous Hochberg and Brooks experiment which conclusively demonstrates that [4] pictorial recognition is an unlearned ability. [account of Hockhberg/Brooks her with pictures]

In the case of the sketch above child might say "She's got no face." But that does not mean the child has no grasp of the techniques of drawing, it means that the child doesn't know whether the drawing is unfinished or that the face is deliberately omitted. Sometimes that's a tricky call. What the child does know is that the person *should* have a face, and that is what Schier and Hopkins call frame knowledge.

It is further clear that when Schier invokes common background knowledge he is also saying that we know that artists have the same visual system and similar background knowledge of the world to us. We therefore assume that when artist depict things they want us to see them as things which can occur in the world as opposed to surreal, stylized or impossible objects. Schier comments:

---

[4] {Hochberg, 1962 #521}

> One could – if one were extremely naïve or stupid – take the drawing by Hockney to depict the boy as if he were a set of lines. One could, that is, take the lines at 'face value'. One could take the black-and-white photograph to depict Nijinsky as black and white (and grey) all over. One *could* take a heavily outlined figure by Van Gogh to depict the man as literally have dark black bands running along the edges of his body. One *could* read the red-figure vase as ascribing the colour of the pottery to the depicted figure. But we don't do any of these things.[5]

We assume a standard of what Schier calls "conversational cooperativeness between ourselves and the artist or photographer.

> The content of the black-and-white photograph is constrained by our common knowledge of the world and by our common knowledge that photographer and beholder are relying on the assumptions of conversational cooperativeness. Because of our common knowledge that the world is not black-and-white and because of our common knowledge that we are adhering to the maxims of conversational co-operativeness ……we assume that the picture is non-committal with respect to colour since if we thought otherwise we would have to suppose that he artist was either a fool or a knave.[6]

Thus, for Schier, the crucial aspect of this background knowledge is that it is common between us and the artist/photographer. The other crucial feature of Schier's theory is that it invokes the intention of the artist to depict something recognizable to people with similar recognition abilities to his or her own. One of the strengths of Schier's position is that it makes a virtue of the very issue which defeats alternative theories of depiction – the artist's intention. Twentieth-century criticism wasn't kind to the notion of artists' or authors' intentions. In fact, it is generally been thought that the idea that we can ascertain intention is at best naïve and at worst nonsensical. The death of the author in the 1960s and, by extension, the death of the artist, put the artist's intention firmly out of bounds for the serious theorist.

rewrite following on H***

Schier bucks this trend, but Hopkins' does not have that option. Hopkins' theory of "outline shape resemblance" hinges on his assertion that mechanism whereby we recognize what is depicted triggered by actual properties of objects not through subjective factors such as the experience of resemblance. He argues:

---

[5] {Schier, 1986 #32@168}
[6] {Schier, 1986 #32@169}

> Although outline shape is a property of objects, not of experiences, an experience may be one of perception of outline shape.[7]

The invocation of common background knowledge has different implications for the two theories.

The recognition account maintains:

We recognize S in P because P triggers S recognizing abilities.

The resemblance account maintains:

We recognize S because P has properties in common with S [***this needs work look at Lopes account***]

Clearly these positions are very different, but only one has the right to invoke the background knowledge of the viewer because in the resemblance case it would seem that the congruence of properties is doing the work and not knowledge of what kinds of properties pictures and things may have. Hopkins' jumps aboard the "widespread general knowledge" raft at the point where he tries to extend the mechanism of resemblance to pictures which are sketchy and unfinished. His hope is that such knowledge will somehow be able to account for how we can see the girl in the drawing above despite the fact that she has no face and that the lines are remarkably sketchy. However discounting these factors does not actually let resemblance off the hook. Hopkins still needs to maintain that there is some resemblance, no matter how slight, and that we recognize the girl and her posture because that resemblance mechanism triggers recognition.

****

Flint Schier comments in Deeper into Pictures:

> All representations crop and detach a slice of the world and forego the rest. A picture which obeys the Albertian rules of perspective – or which approximates to such obedience – must of necessity select the angle from which it is to view a given facet of its subject; the rest it must deny itself. When Constable conjures up a cloud with a pencil stub he must relinquish any hope of capturing the colour which the clouds pick up from the sky.[8]

---

[7] Robert Hopkins, *Picture, Image and Experience : A Philosophical Inquiry* (Cambridge ; New York: Cambridge University Press, 1998), 68.
[8] {Schier, 1986 #32}

Schier goes on to enumerate various other ways in which pictures exhibit selective commitment. Amongst them he considers: black and white photographs, sketches and stick figure drawings.

Schier resolves the "selective commitment" problem by considering the kind of "frame knowledge" we bring to bear when we look at a picture, and invoking Grice's concept of conversational co-operation.

Schier observes that when we look at a black-and-white photograph we don't assume that picture depicts objects and people that are actually black and white. The thought never crosses our mind. It is part of our deep background knowledge of the world – our frame knowledge – that the world is not a colourless place. He argues that the same principle applies when we look at drawings and engravings.

> One could – if one were extremely naïve or stupid – take the drawing by Hockney to depict the boy as if he were a set of lines. One could, that is, take the lines at 'face value'. One could take the black-and-white photograph to depict Nijinsky as black and white (and grey) all over. One *could* take a heavily outlined figure by Van Gogh to depict the man as literally have dark black bands running along the edges of his body. One *could* read the red-figure vase as ascribing the colour of the pottery to the depicted figure. But we don't do any of these things.[9]

We are selective about which aspects of the drawing depict which aspects of an object, not because we have prior knowledge of Hockney's unique technique, but because we have commonsense. Furthermore, we assume that the artist has commonsense, and that the artist knows that his viewer will see the lines as demarcating boundaries and disregard the significance of the pencil colour in the pencil drawing of a boy [or cloud].

Our natural recognition abilities cause us to recognise an object or scene and the conventions of depiction suggest that if we can see it in the picture we were probably meant to see it in the picture. If we can't see anything it probably isn't a depiction. If we see something that is radically contrary to our experience of the world, such as people with green outlines, we discount these things as accidents of the technique of depiction. Furthermore, you can't teach someone how to naturally generate an interpretation of a depiction – so knowledge of depictive techniques is not going to help with the initial recognition. Schier's important point against the conventionalist view of depiction is that even if one did have the kind prior knowledge of depictive techniques and conventions

---

[9] {Schier, 1986 #32@168}

which conventionalists claim is necessary, one would still not be able to interpret the picture without recognition abilities.

> For a linguistic system, if you know the conventions which govern the meaning of the parts and the way they are to be composed, you have sufficient credentials for interpreting the sentence's meaning. Not so in the case of the conventions relevant to pictorial interpretation. Knowledge of these would not enable you to concoct a *pictorial* interpretation in the first place – only recognitional ability can do that.[10]

The selective commitment issue is crucial to an account of depiction.

Whichever theory you espouse you need a mechanism whereby a person can decide which aspects of a picture can be taken literally and which aspects are accidents of technique, or a feature of the medium.

Even a Gombrichian, who believes that the improvements in depictive techniques and variations in style have influenced both our perception of pictures and our perception of the world, must account for how selective commitment actually works.

It is one things to say that it is commonsense that people are not black and white and do not have thick black outlines, it is another to say that

What do Schier and Hopkins mean by generalized frame knowledge?
Do they mean commonsense?
Does Schier mean that we have a 'folk psychology' understanding of the kinds of things a person might depict in a picture?
Is the frame knowledge which they invoke to solve the selective commitment problem related to the kind of epistemological problem which Dennett refers to as "the frame problem"?
Are Schier and Hopkins invoking a larger epistemological problem to mask the inadequacy of their explanation and slipping out the epistemological back-door?
Can we characterize different kinds of 'frame knowledge"?
Is there a kind of frame knowledge appropriate for looking at pictures?
What is conversational cooperativeness when it comes to pictures?
What is the frame?
What is not the frame?
Where does frame knowledge end and specific knowledge begin?
Do we need such a major epistemological solution to solve the selective commitment problem?

---

[10] {Schier, 1986 #32@171-172}

Do we notice the outlines?

On any philosophical journey, like any trek into the unknown, you need to know what to pack. Of course you can't always be sure what you will meet, but usually you can decide whether you need to bring a hand-gun or a howitzer, a list of styles of depiction or five-hundred years of epistemological theory. We need to know whether we can solve the selective commitment problem with a knowledge of artistic techniques or a deep knowledge of the epistemological issues which are raised by the concept of frame knowledge.

Let's imagine that the conventional view is correct and we somehow assimilate a knowledge of depictive techniques. What kind of list of techniques would we need to assimilate in order to be able to correctly recognize the different kinds of pictures we encounter every day?

Willat's taxonomy indicates that the list is long. Projection systems, denotational systems. Optical vs denotational.

I don't believe that the distortion of different projection systems, and accidents of technique such as outlines, monochrome, exaggeration, simplification etc are enough to hinder recognition. Thus we are not arguing about the raw ability to see what is depicted we are arguing about what parts of the depiction we take as literally depicting and which parts we discount.

Of course, frame knowledge, is just another term for common sense. Could it be that the central issue in theories of depiction is really just a matter of common sense?

****

Goodman, Resemblance and Selective Commitment

Nelson Goodman's objections to resemblance, in his book *Languages of Art,* have had remarkable longevity and are still used to attack naïve accounts of pictorial resemblance.[11] Goodman presents a number of arguments against resemblance being either a *necessary* or *sufficient* condition for a picture to refer to its subject. These arguments, in one form or another, have been used by generations of critics and philosophers to diminish the significance of resemblance in depiction. In fact, the opposition between Goodmanesque conventionalist accounts and resemblance accounts has largely set the agenda in the debate about how depiction works. One of the dominant issues has been how does the viewer know which aspects of a picture refer to the subject and which are accidents of the picture's style or features of the projection technique? This is known as the "selective commitment" problem.

Figure 4 John Constable, *Salisbury Cathedral*, 1811. Drawing, black and white chalk on grey paper, V&A, London

---

[11] For example, the majority of aesthetics courses in US universities use Goodman's arguments against resemblance and his conventionalist account to focus issues of how pictorial representation works.

For example, the conventionalist account suggests that a viewer would need prior knowledge of the system of symbolization used by Constable in his drawing of Salisbury Cathedral (fig. 4) in order to realize that he did not intend to depict black trees against a grey/brown sky. By extension, the viewer would also need to know that the rough cross-hatching which represents the trees nearest the Cathedral is supposed to represent leaves. Thus, the conventionalist account suggests that our interpretation of the drawing is largely determined by our knowledge of the symbol system which it employs. The resemblance view, on the other hand, has no problem with how we ascertain that the drawing is of trees and a Cathedral – we simply see a resemblance. On the issue of how we discount the accidents of the drawing technique, such as the colour of the paper and the cross-hatching of the trees, resemblists are divided. The implication is that even if there were some resemblance between the marks on the paper and a cathedral surrounded by trees which initiated an interpretation, this reading of the picture would need to be supplemented by knowledge of drawing techniques.

Recognition accounts of depiction, such as Flint Schier's, which hold that we see content in pictures by virtue of our natural recognition abilities, are not necessarily immune from such objections.[12] Indeed, it would seem that in order for the kind of thoroughgoing cognitive account of depiction, which he envisages, to be an improvement on resemblance it still needs to account for how our perceptual system is able to filter out the significance of properties of the picture which are irrelevant.

Cognitive accounts of depiction do not exclude the possibility that we are deciphering a symbol system when we look at a picture. Indeed, Goodman has been hailed as a pioneer of cognitive accounts of aesthetics by many commentators.[13] On the other hand, his account of how depiction actually works is rather sketchy. In *Languages of Art,* Goodman concentrates on demolishing the resemblance account of depiction, but says very little about how the interpretive process of recognising the content in a picture actually gets started.

We can see from the above Constable drawing example that Goodman's objections to resemblance might apply to natural recognition capacities no matter whether they are

---

[12] {Schier, 1986 #32}
[13] See {Giovannelli, 2005 #446}, {Elgin, 2000 #477} and {Freeland #447}

resemblance based or have their origin in some other native perceptual ability. If some kind of native ability, other than resemblance, enables us to see content in the Constable drawing, we still have the residual problem of accounting for how this native ability assesses the cross-hatching and grey/brown sky. Thus, although Goodman's account fails to address the crucial issue of how a pictorial interpretation gets going, the success or failure of his attack on resemblance theories has implications for any theory of depiction where the initial interpretation of the picture is predicated on natural perceptual abilities (i.e. where the interpretation of the picture does not rely on having prior knowledge of pictorial symbol systems).

SECTION TWO
It is a stronger version of the selective commitment problem. A sketchy line drawing, such as this drawing of a dog (fig. 17), easily evokes its subject. Yet, when pressed you might find it hard to say which part of the sketch actually resembles a dog. Sketches like this raise the suspicion that even though Goodman is wrong about the role of resemblance in more thoroughly realist depiction, he may be right about the conventions behind line drawings. That is, it would seem that knowledge of the conventions of line drawing would seem to be necessary to interpret this sketch. When one examines the sketch there are so many properties associated with a dog that are missing, that one is hard-pressed to pin-down the manner in which it resembles a dog. It is almost always the case that when one dwells on how a picture resembles its subject, the nature of the resemblance becomes more elusive.

Goodman's explanation of how we see the dog is that we have an acquired knowledge of line drawing conventions. His premise is that somehow during our upbringing we have soaked up knowledge of the appropriate depictive conventions and apply them without realising it when we interpret the picture. Thus, the conventionalist explanation of selective commitment proposes that somewhere during our visual education we have been exposed to pictures, have absorbed some knowledge of depictive techniques and apply this knowledge in our selective interpretation of elements of pictures.

Natural Generativity

A picture has "natural generativity" precisely because our recognitional abilities are called upon to interpret the picture. Other factors may facilitate an interpretation, such as a caption, or prior knowledge of the content of the picture, but these ways of identifying the content of a picture do not invoke our natural recognition abilities, and are therefore not *pictorial* interpretations. The fact that we can see a cathedral, trees and cows in the Constable painting (Fig. 5) is indisputably linked to the fact that we can see cathedrals, trees and cows in real life. Some viewers may not have seen real cathedrals before, only pictures of cathedrals, and that is how they recognise the cathedral, but we will find that natural generativity easily accounts for these cases. Schier comments:

> The claim is that a truly pictorial interpretation is in some sense caused by or brought about by the relevant recognition ability. If S depicts O that is because an ability to recognise O could be enough, given an initiation into the relevant symbol system, to explain P's getting his interpretation of S right.
> If this is right, then the logic of our theory resembles various causal theories of action, perception, knowledge and so on. We are characterising icons as those symbols whose interpretation can be causally explained by relevant recognition abilities. An icon is iconic because its interpretation can be explained in a certain way: an interpretation that *is* explained in this way is iconic.[14]

Schier mentions here a kind of pictorial initiation, where someone sees a picture, perhaps for the first time, and interprets pictorially for the first time – the Ur-icon moment. According to Schier once someone has identified pictorial content pictorially, as opposed to identifying it by some other means, such as a caption, they have demonstrated pictorial competence and can go on to recognise novel pictorial content *ad infinitum*.[15] If the Constable painting was shown to someone who had never seen a picture before, it is likely that they would identify trees, cows, people and a building. No prior knowledge of depictive techniques would be required because the interpretation is naturally generated.

So as Schier suggests, natural generativity is a power to be reckoned with:

---

[14] {Schier, 1986 #32@49}

[15] Pictorially identifying the Constable painting as a picture of Salisbury Cathedral, and not some other cathedral, can only happen if you either know what Salisbury Cathedral looks like in real life, or have seen other pictures of it. Natural generativity is therefore not necessarily the key to identifying it as Salisbury Cathedral. On the other hand, you don't need prior knowledge to see it as a picture of a building surrounded by trees; that interpretation is caused through natural generativity.

it tells us what counts as a picture, what counts as a pictorial system and what counts as pictorial competence. Who could ask for more?[16]

In a later chapter entitled "Resemblance Strikes Back", Schier offers an account of pictorial resemblance:

> I want to claim that the theory of natural generativity can solve the fundamental problems of the resemblance theory: it tells us what kind of resemblance between S and O is required for S's depicting O and it gives us an idea of the 'amount' of resemblance required (or rather, it gives us a criterion for telling whether the resemblance is sufficient). The respect in which S resembles its depictum O is this: there is an overlap between the recognitional abilities triggered by S and O. [17]

Schier is saying that we see a resemblance between picture and depictum by virtue of the fact that some of the recognitional abilities triggered by the picture are also triggered by the object in real life. Schier argues that the painting's capacity to trigger some of the recognition abilities which may be triggered when we encounter these objects in real life is the essence of depiction and *is* resemblance. This move by Schier effectively redefines resemblance as a phenomenon engendered by the triggering of overlapping recognition cues. This is an important move by Schier because:

1. He is proposing the mechanism that underpins the resemblance experience.

2. He is claiming that seeing a picture of X exploits similar perceptual abilities to seeing X face-to-face.

This move also addresses one of the essential criteria of a theory of depiction which I enumerated in my introduction: It should be able to indicate what kinds of knowledge and perceptual abilities are required to kick-start a pictorial interpretation. Schier is suggesting that the perceptual abilities which we use when we look at Constable's painting overlap with those that we would use to view the actual scene. On this reading, our initial interpretation of the Constable painting as a cathedral surrounded by trees is triggered by the fact that some of the same perceptual abilities that we use when we see actual cathedrals and trees are triggered by seeing the picture.

This conclusion raises the spectre that the role of recognition in natural generativity is prey to the same objection as resemblance theory – the **selective commitment problem**. The problem, as it might be posed for Recognition theory, goes something like this:

---

[16] {Schier, 1986 #32@46}
[17] {Schier, 1986 #32@186-187}

13

How do we know that the aspects of the picture which trigger our recognitional capacities are the aspects which the picture is intended to trigger? For example, if the picture of Salisbury Cathedral is a line drawing in green ink and thereby triggers green line recognition abilities – how do we know that the Cathedral isn't meant to have green outlines?

Schier's theory also raises a second problem - **The illusion problem:**

If both picture and depictum trigger the same recognitional capacities how do we tell the experiences apart? For example, how do we know we are not looking at real trees and a real church?[18]

These kinds of objections undermine both recognition and resemblance because, although commonsense tells us that we can't mistake a picture of Salisbury Cathedral for the real thing and we don't really believe that objects have green outlines, they highlight the fact that in order for a picture to resemble its depictum the viewer has to select which aspects he/she construes as similar. If resemblance is generated by overlapping recognition abilities the viewer has to select which overlapping recognition triggers are relevant. This opens the door to the conventionalist argument that we are primed by habit and convention to select the aspects of the picture which we have become accustomed to associating with tree and churches – that is, we see "resemblances" between features of the painting and reality because we are habituated to associating certain ways of depicting objects and scenes with features of reality. The conventionalist would argue that the Constable painting does not trigger natural recognition abilities, but that Constable's depictive techniques trigger symbol interpretation abilities drilled into us by habit and custom. The conventionalist accounts for the fact that we don't mistake Constable's painting for the real cathedral by virtue of the fact that it looks nothing like the real cathedral. For example, the picture is small and flat, and the cathedral is an enormous three dimensional object.

The other objection (selective commitment) to the resemblance theory of depiction also seems to apply to recognition theory of depiction – how do we know which aspects of the picture are the significant ones exhibiting resemblance or triggering recognition?

---

[18] Gombrich claims we are aware of both at the same time.

How does the viewer of the Ottawa Constable study (Fig. 1) know that the cathedral hasn't really got walls and windows inclined at 5 degrees from the vertical? What leads our eye to the brown paint smudges in the foreground that we have assumed are cows?
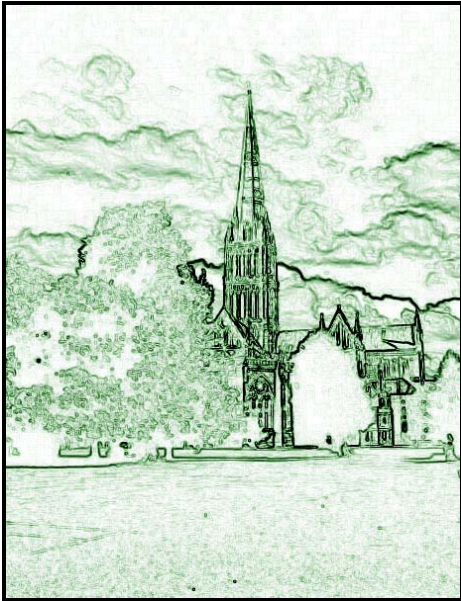

Figure 20 Photograph of Salisbury Cathedral computer rendered as a sketch

How does the viewer know that the sketch above (fig. 20) doesn't depict a cathedral and trees which actually have green outlines? The viewer needs prior knowledge of how to separate aspects of the picture which are accidents of the depiction technique (e.g. green lines) from those which have depictive content (the shape of the spire). In short, the viewer needs to understand the conventions which the depiction is using in order to know in which respects the picture resembles its depictum. That information needs to be communicated by the picture itself at the moment of viewing for it to be a truly pictorial interpretation. If the viewer knows which aspects of the picture to discount because of knowledge acquired prior to looking at the picture then it *seems* to be the case that:

- some aspects of the interpretation are conventional (e.g. we know that outline drawings are not pictures of monochromatic things with outlines, therefore we know that we are not looking for resemblance in this respect)

and

- the interpretation of the picture relies on habituation to objects being depicted using such and such a technique (e.g. we habitually see trees depicted in this way so we know that Constable has depicted trees).

15

According to this line of reasoning we can tell which aspects of a picture resemble its depictum because we already know that features of the picture are accidents of the depictive technique and must be discounted. Thus, we interpret a wavy brown line as a tree branch because we are accustomed to seeing trees depicted like this, not because the wavy brown line resembles a tree branch. Despite our intuition that we are able to interpret a picture because of its resemblance to its depictum, it seems that, on reflection, this cannot be the case. The lines depicting the cathedral walls in the Ottawa painting (fig. 1) are crooked, whereas we know that cathedral walls need to be vertical. The leaves on the trees are undifferentiated smudges. There are so many features of the painting which need to be discounted that, on reflection, it is hard to argue that the picture resembles its depictum. More importantly, it seems that we need to be aware of the vagaries of depictive techniques in order to be able to generate an uncluttered interpretation of what the artist is intending us to see.

It is considerations, such as the above, which have fuelled the conventionalist case against the resemblance theorist. It seems that the picture needs to communicate to us in what respects it resembles its depictum and in what respects it does not – thus the resemblance theorist (and by extension the recognition theorist) must concede that it is not pure resemblance that enables us to interpret the content of a picture, but depictive conventions. Given the vast chasm between what a line drawing of Salisbury Cathedral looks like and what the actual cathedral looks like, it is hard to see how we could see a resemblance unless pictorial resemblance is the product of habituation to pictorial technique. It seems that in order to eliminate the irrelevant aspects (green outlines) and regard the relevant (branch shapes) we must invoke our familiarity with depictive techniques. Factors other than the actual look of the picture (resemblance) must be the basis for our interpretation.

These obstacles to resemblance theory seem to apply with equal force to Schier's natural recognition theory. How can the visual cues in a painting be naturally recognised when, on reflection, it seems that there are precious few cues in the painting that would match those that trigger perception of the real object? Even if the colour of the leaves in the painting cues colour recognition in much the way that leaves trigger colour recognition when we see a real tree, it is clear that the range of greens, brightness and

contrast which the painting can achieve is no where close to that which we can experience on a sunny day in the Bishop's Garden in Salisbury. On reflection, it seems there can be nothing very natural about the way that Constable's paint-smudges trigger an interpretation. We will see that Schier's hypothesis takes these objections in its stride and, by suggesting that visual cues in the picture trigger a natural interpretation, he provides a causal explanation for the power of depiction which avoids many of the pitfalls of other theories of depiction. Notably he resolves the tangled issue of how we ascertain the artist's intention, and avoids having to formulate a string of depictive conventions to account for the variety of techniques available to the artist.

Schier's Two-Stage Theory

Schier is happy to concede that factors other than our natural capacity to recognise objects inform our interpretation of a picture. Clearly a sophisticated cultural practice like depiction must be governed by rules and conventions as well as natural perceptual abilities. But, the fact that there are conventions involved in interpreting a picture does not mean that knowledge of the conventions entirely determines how the viewer interprets the picture. Many conventions govern musical composition, but the listener needs no inkling of them to hear and appreciate music. Understanding the conventions of music may enhance the experience, but clearly, our innate capacity to hear rhythms and distinguish notes is the key to listening to music – not understanding conventions.[19] Schier proposes that our natural visual recognition ability is the key to our capacity to recognise pictorial content and that the conventions of depiction play a following role in refining our interpretation. He maintains that iconic interpretation has at least two levels:

> One level involves the generation of the interpretation; the other involves the validation or confirmation of it.[20]

Effectively, Schier's account of pictorial interpretation is a two-stage process. In the first stage, natural generativity "tells us what kind of resemblance between S and O is

---

[19] Schier alludes to the debate concerning the innateness or otherwise of language. He notes that Chomsky's theory of universal grammar allocates a role for an innate propensity for language acquisition and understanding, despite the fact that the surface grammatical rules are entirely convention dependent.
[20] {Schier, 1986 #32@167}

required for S's depicting O".[21] This interpretation is naturally generated by our recognition abilities. The second stage invokes conventional rules about depiction to determine whether the naturally generated interpretation is the correct one. In fact, Schier maintains that there is only one convention which we need to bear in mind when we consider what is happening in this second stage: Convention C.

> C: Given that S is of O, it is intended that those who are able to recognise O should be able, on that basis, to interpret S.[22]

Schier maintains that conventions have a "central role in the theory of iconicity" but adds a crucial qualification:

> Conventions do not explain the interpretations we generate; the iconic convention merely determines whether the naturally generated interpretation is the right one.[23]

Let us see how this works with our Constable painting. We look at the picture and the first stage (the raw-recognition stage) triggers some of our natural tree, cow and building recognition abilities. In the second stage (the validation stage) Convention C comes into play, and something like this is the case:

> Given that the painting is of trees, it is intended that those who are able to recognise trees should be able, on that basis, to interpret the painting as of trees.

And so on for the other features of the painting.

It is notable that Convention C invokes the intention of the maker of the picture. It is also notable that the only thing Convention C is saying about intention is that if we are looking at a depiction (a drawing, painting or photograph) and we can recognise a tree then it is overwhelmingly likely that whoever created the picture intended us to see a tree – it is a kind of tacit contract between artist and beholder. Convention C does not require us to have some preternatural insight into the intention of the artist. After all, it may have been Constable's intention to depict elms, whereas, due to some taxonomic deficiency he has depicted beeches.

---

[21] {Schier, 1986 #32@186} – This reference to resemblance seems rather controversial in this context, perhaps we might say that natural generativity ensures that any overlapping natural recognition abilities are triggered and some kind of resemblance between S and O is experienced.
[22] {Schier, 1986 #32@137}
[23] {Schier, 1986 #32@132}

He has made his painting in accord with Convention C and anyone who can recognise beeches and tell elms and beeches apart will say that it is a picture of a beech, and they are right: it is a picture of a beech.[24]

So, we can see that it is not the artist's intention *per se* that Convention C invokes, it is the artist's intention, however fallible, to paint something that will be recognisable to those with normal human visual recognition capacities. When Constable paints his tree he paints it such that, for him, it looks like a tree (it triggers his tree recognition abilities) and he reasons that if it looks like a tree to him it will look like a tree to other human beings with similar recognition abilities.

The experience of resemblance (Constable's tree looking like a tree) is precisely the triggering of natural recognition abilities in Schier's first stage of pictorial interpretation. Recognition and resemblance are part of the first "raw-recognition" stage and not the second "validation" stage, which determines whether depictive conventions are in play. The question this thesis addresses is how we recognise depicted content in the primary recognition moment – in that first half-a-second of sighting a picture. The issue is "How does the interpretation process get started?" Schier's theory offers an elegant solution. The recognition moment is segregated from issues of depictive conventions because recognizing content and noticing the style and method of depiction are different processes. We recognise the depicted content of a picture at a glance using our natural recognition abilities. After this initial recognition moment we validate our initial raw-interpretation and begin to refine it by scanning the picture more closely. This move, by Schier, promises to dissolve the selective commitment and illusion problems and provide a plausible mechanism – recognition triggers - whereby a pictorial interpretation gets a foothold.

Recognition Triggers and Illusions

So, how does Schier explain the so-called "selective commitment" problem? Let's look at the facts:

- The green outline drawing of Salisbury Cathedral (fig. 20) doesn't lead us to think that the artist was depicting a cathedral with green edges;
- We *don't* assume that the walls of the cathedral in Constable's Ottawa study (fig. 1)

---

[24] {Schier, 1986 #32@135}

are meant to be at strange angles;

- We *do* assume that some of the paint smudges are meant to be cows.

The conventionalist argues that depictive conventions govern how these features of depictive technique are discounted from the literal interpretation of the picture. However, the only convention Schier allows is Convention C, which is merely an invocation of depictive conventions to validation of the first stage of interpretation. In Schier's view, conventions play no role in the initial interpretation.

Hopkins and Resemblance Theory

Hopkins maintains that three serious problems beset any version of the resemblance view:-

1. ***The particular vs. the general problem***: How can a picture which represents a thing as having particular features represent something in general e.g. a horse (but no particular horse).
2. ***Pictures don't really resemble what they depict***: The differences between picture and object depicted usually outweigh the similarities.
3. ***The caricature/stick figure problem***: Even if one concedes that a realist picture looks something like the object it depicts, it is hard to maintain that caricatures, stick figures and cubist paintings resemble their depicta.

These three problems to some extent recapitulate some of the issues we noted with *selective commitment.* That problem can be stated thus:

> *How do we know which features of the picture are supposed to resemble X when there are so many features which don't resemble X and when there isn't an object to compare?*

Schier's deals with these problems by arguing that "natural generativity" and common sense tell us what aspects of a picture are depicted content and which are accidents of the depiction technique. He also argues that if we know we are looking at a flat surface but can see what appears to be a three dimensional object, it is likely we were meant to see

that surface as a three-dimensional object. So Schier deals with the Leonardo stains on the wall, and the Virgin Mary toast issue, by arguing that evidence of artisanship or some other such cue will tell us that the surface has been designed to trigger the object recognition that has actually been triggered.

SECTION four
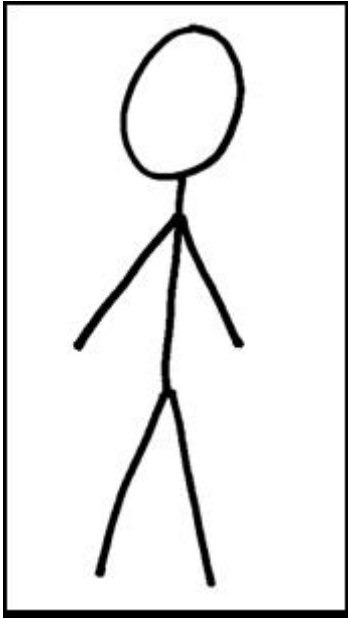

Figure 51 Hopkins' stick-figure

Stick figures and Seeing-in

It should be noted here that Hopkins believes that stick figures are not like caricatures. Caricatures are an example of misrepresentation within limits – a kind of exaggeration.[25] Stick-figures, he maintains, are a case of indeterminacy in representation. The Leonardo picture is more determinate about the features of the girl's head than the Klee picture.

The silhouette outline drawings, which we looked at in the previous section, suggest that we often don't notice shape resemblance in circumstances where we have every reason to believe that the silhouette is true and the "outline shape" phenomenon ought to be present. With the stick figure we might be reluctant to say that the figure resembles a person at all, but Hopkins perseveres and attempts to stretch the "outline shape" concept to cover our capacity to interpret stick figures. It is hard to imagine that we experience any resemblance in outline shape between a stick figure and a real person, because it is

---

[25] Schier also distinguishes caricature from 'selective commitment', but it is notable that he explains both by an appeal to frame knowledge and Gricean rules of conversational co-operativeness.

precisely the outline that the stick figure omits. Hopkins' strategy is to invoke the indeterminacy principle – "pictorial content can in various ways be indeterminate."[26] He points out that there are at least two main sources of indeterminacy:

- **Resemblance is limited to properties ascribed**: The picture fails to ascribe a given property e.g. some pictures omit background;

- **Resemblance is limited by the degree of determinacy of the properties ascribed**: The picture ascribes a property but within limits of determinacy expected in a depiction e.g. we can tell that Leonardo's girl has long wavy hair but that is indicated with a few lines.

So a picture may, for example, have a relatively indeterminate point of view and consist of sketchy lines and therefore have an indeterminate outline shape; thus the resemblance experience is correspondingly reduced.

Hopkins doubts that either of these strategies will work with the stick figure and comments:

> The claim would be that we see the picture as resembling in outline shape something highly indeterminate: a person standing, with his arms at his side, but whose shape is pretty much unspecified. This requires the outline shape of the item resembled to be very indeterminate indeed. Is there any reason to think that the experiences of resemblance can be this indeterminate? It is hard to convince oneself that there could certainly be such experiences.[27]

The indeterminacy in what is depicted is traceable to indeterminacies in what is seen in the surface. There doesn't seem to be enough in the surface marks to support resemblance to anything other than a figure made of sticks. However, it is not our experience of the marks in the stick figure drawing which is indeterminate – we can see exactly what shape the stick-legs are. We can assume that this is not a realist depiction of a person and conclude that the person who drew the picture wished to be non-specific about whether the person is male or female, fat or skinny, clothed or unclothed etc. The person who drew the picture has omitted most properties that we might ascribe to a person except having a torso, head and four limbs in an upright configuration. In fact, if I described to you such a drawing you would have no trouble identifying the drawing as a person. There doesn't appear to be an issue here unless one is trying to claim that the

---

[26] {Hopkins, 1998 #290@123}
[27] {Hopkins, 1998 #290@124}

reason we *see* the stick figure as a person is because of a resemblance between the straggly lines and a real person. In fact, that is the line Hopkins takes. He argues that:

> We see an oddly shaped person in the stick figure picture, without making the implausible claim that that is what is depicted. Now, if that is what is seen in the surface, there is no need to postulate experiences of resemblance of an extremely indeterminate nature. The marks are seen as resembling something more determinate than merely a person standing with his hands at his side, since that person is also of an odd, very straggly shape. There is no more reason to be skeptical about the possibility of this experience of resemblance than there is about many of those we have previously accepted as unproblematic.[28]
>
> If what a picture depicts is indeterminate in a certain respect, either this is because what we see in the surface is indeterminate in that way, or it is despite the fact that we see something more determinate therein.[29]
>
> The difference between the two situations is that in one of them we do not conclude that what we see in the surface is what it depicts.[30]

If I understand this reasoning correctly, we assume that Leonardo has depicted a girl with slightly fat eyelids, a slightly long nose, a rather pronounced chin and long wavy hair. With the stick figure, although we see a resemblance in the picture to a person who has a round head and no eyes, nose or mouth, and who is very skinny, we do not assume that the artist was depicting such a person. Similarly, we do not assume that Klee has depicted a person whose eyes are vacant circles or who has no nose or mouth etc. In both the Leonardo and the Klee pictures we see a resemblance to people's heads but in the Klee we see these rather odd looking heads but assume that the artist was not expecting us to read these heads as horribly distorted. Hopkins' point is that with caricature we *do* want people to read the drawing as horribly distorted, with the stick figure we *don't*. The stick figure distortion is an accident of the technique or the lack of skill of the artist. The distorted stick figure requires us to fill-in the properties omitted.

It is interesting to note here that in the Leonardo drawing the level of realistic detail is not consistent through the picture. Some parts of the girl's head are superbly rendered and others are indicated with sketchy pencil lines e.g. her hair and shoulders. Despite its realist basis, however, we do not assume that the girl's complexion is grey or that her hair is made of a few transparent strands. We see this varying selective commitment to detail

---

[28] {Hopkins, 1998 #290@124-125}
[29] {Hopkins, 1998 #290@125}
[30] {Hopkins, 1998 #290@125}

throughout the Leonardo drawing and it can also be found in the majority of drawings and paintings in just about any culture. In fact, selective commitment is more of a rule than an exception in non-photographic depiction[31] and leads Hopkins to suggest that the indeterminacy in what we see-in the picture can be accounted for in two ways which he calls Marriage and Separation:

1. **Marriage:** We see-in the content exactly what was intended but with the indeterminacy in our seeing-in caused by indeterminacy in pictorial content;

2. **Separation**: The content of the picture does not match what we see in it.

The best way to understand these two modes of seeing-in is using Hopkins' own example of two drawings of an oblong table. [32]

Hopkins imagines two drawings of oblong tables, one in perspective the other not:

> While the two pictures do not disagree on the table's properties, one is simply indeterminate with respect to features which the other represents. Perhaps the one in perspective conveys highly detailed information about the table's shape, whereas the other is just a hasty line drawing, conveying the rough outline of the whole, and no more……
>
>   It may be that the less determinate picture is so through Separation. Perhaps we see a rather awkwardly shaped table in the sketch, some of its legs longer than others, its top a distorted parallelogram. For all that, we need not take it to depict, and it need not really depict, such a table. The other sub-possibility is that indeterminacy exploits Marriage, our seeing in the sketch nothing more determinate than a roughly oblong table. Either way, there is again no difficulty for the resemblance view. The two pictures both depict, and both depict an oblong table, for all that they differ in the marks their surfaces bear.[33]

In Separation mode one avoids taking literally the distortions in the picture – part of the content one sees in the picture is bracketed. In Marriage mode we see in the picture what we see in the picture, even though what we see is indeterminate. Given that most pictures partake of some degree of indeterminacy this is a major issue for seeing-in and experienced resemblance theories. In particular, the Separation account leaves Hopkins

---

[31] I suspect that we have gotten used to seeing paintings reproduced in books and sometimes forget how prominent the brush strokes are in paintings by artist we would characterise as 'realist' e.g. Velasquez, and Rembrandt.

[32] Hopkins is aware that his account of resemblance is open to the criticism that if depiction essentially requires experienced resemblance in outline shape then only pictures in something like Alberti perspective can truly depict. Such pictures treat the drawing surface as a window pane on which is traced the shape of what is before it. Hopkins must allow that drawings not in perspective can depict, or he must credit Renaissance artists with inventing depiction when they discovered perspective.

[33] {Hopkins, 1998 #290@153-154}

open to the problem of how we can establish "standards of correctness" for what is seen in a picture when with Separation the picture itself is no longer the guide. Once we give the viewer the latitude to see things that are patently not in the picture we abandon standards of correctness about what can be seen in the marks on the surface. It is hard to see how this account amounts to an explanation of how we interpret a picture. Hopkins had hoped to explain how "seeing-in is experiencing resemblance in outline shape" but now asks us to believe that with certain pictures we must discount the depictive significance of this outline shape, so that what we see in the picture is more than it actually depicts. He may in fact be right about our seeing more in the picture than the marks depict, but the manoeuvres he has had to make to include stick figures in his theory has rendered "resemblance in outline shape" a redundant concept. Because if we can conclude that what we see in pictures is not what is depicted, we really don't need shape resemblance to explain what we see. This seems like a familiar problem – how do we know which marks to discount as accidents of style and how do we know which parts of the picture depict things which we should interpret literally? What Hopkins is struggling to resolve here is the selective commitment problem. It seems remarkable that Hopkins' theory fails to adequately address this most fundamental issue. It is to some extent the core issue of a theory of pictorial representation. Roughly stated the issue is: "How do we know what the marks represent?" This question gets harder to answer for resemblance theorists the more indeterminate the features of the picture are. Having arrived with his stick figure at a position with indeterminate depiction which seemingly undermines any possible role for resemblance, Hopkins concludes that the interpreter of a picture has a "novel" problem:

> She must not merely decide whether to take her experience of the surface as a guide to its content at all; she must also decide which aspects of what she sees in the surface to take as aspects of what it depicts, and which to ignore. But then on what basis is she able to make this new decision? The experience of seeing-in can clearly be of no further help, since the issue is precisely which aspects of that experience to take at face value, and which to discount. To what, then, can the viewer appeal?[34]

---

[34] {Hopkins, 1998 #290@137}.

His answer is that she must appeal to "widespread knowledge of a very general nature."[35]He characterises this knowledge as:

- Knowledge of what sort of items the world contains and their properties;

- Knowledge of the sort of items in general depicted;

- Knowledge of the various means for producing depictions.

These three forms of knowledge are equivalent to the kind of "frame knowledge" which Schier claims is brought to bear when Convention C is applied to a naturally generated interpretation of a picture. Schier summarises:

> Thus far I have entertained a picture of iconic interpretation as having two levels (levels which are conceptually distinct). One level involves the generation of an interpretation; the other involves the validation or confirmation of it. If S is a picture of O, then an ability to recognize O would suffice to give one the ability to generate the interpretation that S 'refers to' or 'is partially about' O. Such an interpretation is 'naturally generated'. In addition, there is a convention, Convention C, that says the naturally generated interpretation is the correct one. As we have seen this convention is essential. Now we are faced with the fact that Convention C may apply selectively even within the frame of a single picture. We may take our ability to generate naturally an interpretation of S as *criterial* for the application of Convention C. In other words, there is, by and large, a prima facie but defeasible assumption that the naturally generated assumption is correct – that is, there is an assumption that Convention C is in force for our naturally generated interpretations. However, these interpretations may be negated or constrained if they would go against elements of our frame knowledge and of the knowledge which it is common knowledge is part of all men's frame knowledge.
> For example, it is part of our 'background' knowledge of the world that people, flowers, churches and so on are not colourless studies in black and white.[36]

Thus, the reason that we don't read Leonardo's drawing as a girl with a grey pallor and wispy hair is because we know that girls don't look like that and because we also know that Leonardo didn't mean for us to understand his picture in that way. Schier comments on how we understand what is depicted in a black and white photograph:

> The content of the black-and-white photograph is constrained by our common knowledge of the world and by our common knowledge that photographer and beholder are relying on the assumptions of conversational cooperativeness. Because of our common knowledge that the world is not black-and-white and because of our common knowledge that we are adhering to the maxims of conversational co-operativeness (extended to pictorial one-way communication), we take it that Convention C does not apply to the potentially color-iconifying

---

[35] {Hopkins, 1998 #290@137}
[36] {Schier, 1986 #32@167}

aspects of the black-and-white snapshot. That is, we assume that the picture is non-committal with respect to colour since if we thought otherwise we would have to suppose that he artist was either a fool or a knave.[37]

This last comment on the intention of the artist is crucial to Schier's account. It is notable that the artist's intention that what she has depicted is naturally recognizable, far from being a problem for Schier's theory, is a central feature of 'natural generativity'. Convention C states:

> C: Given that S is of O, it is *intended* that those who are able to recognise O should be able, on that basis, to interpret S.[my italics] [38]

Thus Schier makes a virtue of the very issues which defeat alternative theories of depiction – the artist's intention and the problem of what aspects of the marks on the surface should be taken as literally depicting content and which should be ignored. We are beginning to see that Schier's account has a certain appeal purely on grounds of parsimony.

SECTION WILLATS

> When thinking about how vision works it is best not to think of a camera. The visual system is nothing like a camera. Ron Gallagher

> It should be clear that the theory of natural generativity is soaked in causation. Essential to that theory are two causal claims: that an interpretation of S as being of O is iconic or pictorial in so far as it has been prompted by the interpreter's O-recognising abilities and that a picture of O is precisely something which can trigger the interpreter's O-recognising abilities. Iconic interpretation and iconicity are thus functionally defined. However, I have stressed the S's engaging my O-recognising abilities does not entail my recognising O in S or my being conscious that my O-recognising abilities have been engaged. Typically, all I am conscious of is that I am seeing a picture of O. Consequently, our causal or functional analysis of depiction seems to force upon us the conclusion that our pictorial experience is the result of prior cognitive 'processing'; and this conclusion is equivalent to the homuncularist hypothesis that there are sub-personal centres of cognitive activity in the visual system. If I am right, the analysis of depiction properly understood forces us to accept some computational or cognitive theory of mental activity. Pleasant or unpleasant, this conclusion is certainly a surprise. Flint Schier [39]

---

[37] {Schier, 1986 #32@169}
[38] {Schier, 1986 #32@137}
[39] {Schier, 1986 #32@195}

How Do Techniques of Depiction Differ?

Schier's Challenge

In the above quotation Schier excuses us the task of identifying overlapping recognition cues but challenges us to determine the cognitive *character* of the overlapping natural recognition abilities which enable us to recognize both objects in the world and depicted content in pictures. In the brief survey of vision research and recognition theory in the previous section we saw that the concept of 'overlapping recognition abilities' could be interpreted in a number of ways. If we take a cue from Marr we should distinguish the computational problem facing us, from the algorithms which could be used to solve the problem and the actual implementation in terms of the physiology of our visual system. There could be overlaps at all, some or none of these levels.

The explanation of depiction we are looking for will involve more than a description.[40] It will be an account of the kinds of information processing which happen when we look at a picture and, if Schier is right, the kinds of processes we identify will be parallel to those identified in some of the theoretical work reviewed in the last section. This account may also involve psychological and neurological explanations of how picture recognition could work in the human organism.

We know that people can see the kind of content in a picture that they can see in real life – they see tables and chairs, dogs and cats, faces with smiles. No one is doubting the brute fact that even in the worst drawings these things can be seen. But when we consider the high level question (the computational problem) we need to allow for the possibility that the mechanisms whereby people see content in bad line drawings are different from those whereby they see content in full-colour photographs. If we adopt a Marr-type categorization of the three levels of the problem, we might state it like this:-

1. The information processing required for a colour photograph and a sketchy black and white line drawing must be quite different. The colour photograph involves the task of extracting contours and details from rich

---

[40] {McIntosh, 2003 #218}

information. The sketchy line drawing involves extracting contours and detail from sparse information.

2. Consequently the algorithm which can do the extracting must be different in each case. Therefore the processes are different

3. If the task and the algorithm are different then the neurophysiological hardware implementation and underlying mechanics will be different.

The foregoing reasoning may be fallacious because it embodies the assumption that a colour photograph contains less information than a sketch – however even if we allow that objection, it is hard to argue that they could contain the same level of information embodied in the same way. However, it is looking as if classifying depictions on a sliding scale of information carrying is not going to be useful.

Driving a wedge between various forms of depiction highlights how tricky Schier's challenge really is. We have arrived at a challenge which is a variant on Schier's: to determine whether interpreting different kinds of depictions involves overlapping recognition abilities. Could we prove that colour photographs and sketches engage the same or similar recognition abilities? What would count as evidence here? Do any of the theories and experiments reviewed in the last chapter prove that various kinds of depictions are functionally equivalent? In fact, the theories disagreed on fundamental aspects of the visual process, yet Kathleen Akins notes in her introduction to her volume on *Perception*, there is a convergence of opinion on the fundamental questions about the nature of visual perception.

> Taken together they ask "what is the nature of vision – what does vision do?" and more specifically "what are the form and content of visual representations – both of those representations involved in unconscious (or preconscious) visual processing and those that support visual phenomenology, our conscious visual perceptions of the world?"[41]

Atkins sees convergence where others see divergence. She concludes that the fundamental question for vision researchers involves discovering how visual stimuli are represented in the brain. Many advocates of direct perception would take issue with that assessment of the goal of vision research. Not only do they largely eschew the practice of

---

[41] {Akins, 1996 #1@3}

correlating our visual phenomenology with neural representations, but they are also the only ones who don't believe that a picture is an adequate substitute for a 3D object.

Thus, our methodological problem is threefold.

1. The bulk of vision researchers do not distinguish using a picture from using an object. Typically a psychophysical experiment will involve measuring how long subjects take to categorize or identify an image. This typically takes between 100ms and 200ms – less that 1/5 second. It takes roughly the same amount of time to categorize and identify a real object. From a classic psychophysical perspective recognizing pictures is a good guide to how we recognize things in the world. Thus, psychophysical experiments are not telling us much about the *difference* between looking at depictions and looking at objects.

2. There is very little evidence that vision researchers distinguish different techniques and styles of depiction – drawings, photographs and CAD generated images seem to be used interchangeably. However, when we think of the problem from a Marr-ian perspective we have no reason to assume that these different kinds of depiction either have the same psychological effects or have the same underlying mechanisms.

3. We have yet to find a way of establishing that interpreting a drawing and interpreting a sentence invoke different cognitive processes. We have every reason to believe that they do, but we have not established criteria which would settle the question. This would help us towards establishing how we could determine if an image was being interpreted symbolically or using natural recognition abilities.

Before we can make any progress we need to:-

- Look more carefully at the theoretical grounds on which vision researchers base their use of 2D pictures as substitutes for 3D objects.

- Tease out the differences between various techniques and styles of depiction in order to isolate possible information processing issues.

- Establish whether it is possible in principle to distinguish a picture of O which can "trigger the interpreter's O-recognising abilities"[42] and one

---

[42] {Schier, 1986 #32@195}

which triggers some other, more conventional, ability such as symbol decoding. That is, can we cognitively distinguish radically different forms of representation?

I will tackle these in reverse order.

Is it possible in principle to distinguish a picture of O which can "trigger the interpreter's O-recognising abilities" and one which triggers some other, more conventional, ability such as symbol decoding?

The appeal of Conventionalism, in both its hard and soft forms, hinges on the intuition that if our ability to interpret pictures is a natural recognition ability then it should be harder to recognize line drawings, fuzzy drawing and kids drawings than it is to recognize full-colour photographs or realist paintings. Of course it isn't. We can recognize a child's drawing just as readily as a photograph So, if you're hard conventionalist or a structuralist you might argue that we learn the symbology of pictures much like we learn language – starting with simple shapes like stick figures and progressing to the full-colour-realist images of western culture. Thus not only is it logical that we can see stick figures as easily as photographs, but it reflects the ontogeny of how we supposedly learn visual symbol systems. A soft conventionalist, who has a sneaking suspicion that some of the content of photographs is recognized using natural recognition abilities would still argue that line drawings are a symbol system and use Goodman's arguments against resemblance to clinch the argument. In fact John Willats' argument in *Art and Representation* concerning line drawing amounts to such a claim.[43]

The conventionalist argues that stick figure does not resemble a person therefore our ability to recognize a stick figure cannot be based on resemblance or natural recognition.

In the last chapter we saw that Hopkins' needed to almost turn his theory of solid shape resemblance inside out to accommodate stick figures. The reason he fails is because, not only does he argue that there a real resemblance in terms of the shape discerned in the visual field between a stick figure and a person, but he also maintains we experience them as resembling each other. Clearly there is no shape matching taking place and it is

---

[43] {Willats, 1997 #36 36}

hard to maintain that there is a resemblance. If you are inclined to argue that there is a low-level resemblance between the stick figure and a person – configuration of limbs maybe - you are conceding that property resemblance is a very vague notion. On the other hand, you may be right. The Gestaltists believed that configuration was a crucial element in image identification.

There is no objective measure of how recognizable something is. Psychophysical experiments with reaction times are not a good measure, but they provide good indicators for the kinds of properties of images that make a difference. In terms of information content, most artists would argue that a drawing provides more information than a photograph. This again is almost impossible to substantiate. However, in the last chapter I suggested that the artist did much of the work of disentangling the object from its background and emphasizing salient features. This factor alone could explain why it is hard to distinguish line drawing recognition and photograph recognition in psychophysical experiments.

One of the strengths of the conventionalist view stems from the observation that even with spoken language, which has a strong conventional element, introspection does not enable us to separate our innate response (in the Chomskian sense) from the rule-based response. If our ability to see a lion, a person and a landscape in the Rembrandt's study for *St Jerome Reading in an Italian Landscape* (Fig #) is enabled by some drawing conventions, it is difficult to isolate these conventions. It is worth bearing in mind at this point that there is no consensus that language is either innate or conventional or both, despite the staggering amount of work that has been done in linguistics and endless studies of the brain. But the weakness of the conventionalist argument is that it is a classic sceptical argument. To argue 'For all you know all what seems to be a perfectly natural process of seeing a landscape in a painting could be interpreting a symbol system.' when all the evidence points the other way is a fairly weak position. It's a position that works for language because we know that each language has its own conventions and we remember learning them. The evidence, as it applies to pictures, supports the view that we do naturally recognise the lion, and we know we weren't taught how to see lions in pictures. Furthermore, we weren't taught that different lines represent different features – some lines represent the lion's straggly mane, others represent its

body outline against the background. There is a puzzle to be solved here and that puzzle concerns how the ink lines are able to evoke the lion so powerfully.

Thus, although it is probably not always possible to distinguish symbol decoding from natural recognition, we can firmly say that given the strength of the poverty of stimulus argument and the fact that we don't need to learn to see depictions, that the onus is on the conventionalist to establish if, where and when the symbol decoding takes place.

Are there significant differences between various techniques and styles of depiction which might amount to different explanation for their effects and their underlying processes?

Line Drawings: Refined or Degraded

Resemblance theory seems to founder on an intuition that feeds conventionalism – that stick figures don't really look like people and people don't have arms made of squiggly lines. If a recognition theory of depiction could undermine this intuition and explain how it is that a stick figure is so easily seen as a person, or how a pen drawing can so powerfully evoke objects, animals, persons and scenes with just a few lines, then it would undermine the attraction of conventionalism and justify that feeling of resemblance we sometimes get from sparse but evocative line drawing. When we look at Rembrandt's sketch, the lion and the landscape are powerfully evoked, yet it is hard to believe that just a few sketched pen lines can trigger natural lion recognition cues. Furthermore; we are always aware that we are looking at an ink drawing of a lion and a person and a landscape. If there is any overlap between visual processes which enable us to see real persons, landscapes and lions and the processes which enable us to see them in this drawing they must coexist with our awareness of the pen lines.

61 Rembrandt, *St Jerome Reading in an Italian Landscape*, c.1651. Pen and brown ink brown and white wash. Kunsthalle, Hamburg.

When we look at a photograph of a lion, we are not necessarily aware of the marks which are doing the work. We are aware that we are looking at a picture, but unlike with the case of the pen drawing, we do not entertain the dual experience of visually tracing the line of the lion's tail whilst considering the artist's skill with a pen. Thus when we talk of the duality of the pictorial experience we need to distinguish:-

- Being aware that we are looking at a picture.
- Being aware of the marks which constitute the picture.

Do these differences in the way we look at the drawing and the photograph amount to fundamental differences in the way they depict? A fuzzy photograph is a degraded image, but it doesn't have distinct lines. Perhaps the visual system interprets line drawings, such as Rembrandt's, as refined rather than degraded. Perhaps line drawings are using visual pathways that fuzzy photographs do not use or vice versa.

Do we have any *grounds* to assume that different kinds of depictions, using different techniques, refer to their subjects in the same way? Perhaps 'depictions' is just the name we give to range of visual representation methods which are as different in the way they invoke visual processes as they are in the way that are structured. After all, if the two visual experiences, looking at a pen drawing and looking at a photograph, are so different on the surface they could be different all the way down to basic visual processing.

When you interpret the word 'lion' you are using a visual channel but only Goodman is claiming that the interpretation process has something in common with looking at the Rembrandt drawing. Perhaps the way that a photograph works is as different from how a pen drawing works as words are.

We have at least arrived at one way of separating depictions:-

In some depictions, such as line drawings, we are aware of the marks which are doing the depicting and in others, such as photographs, we are not.

Projection systems and other taxonomies: Willats

Of course, this is not the only way of slicing the depiction cake, there is a long-standing practice in the philosophy of art of discerning styles of depiction through their methods of projection (eg orthogonal perspective) and production techniques (eg mosaic). John Willats in *Art and Representation* identifies two representation systems and in the glossary defines them thus:-

> **Drawing systems** The representational systems that map spatial relations in the scene to corresponding relations on the picture surface.

> **Denotation systems** The representational systems that map scene primitives into corresponding picture primitives.[44]

Drawing systems include projection systems such as perspective, oblique projection and orthogonal projection, as well as non-projection systems such as caricatures, and non-depicitve representation such as circuit diagrams and route maps such as the London Underground map.

Denotation systems are silhouettes, line drawings and "optical denotation systems such as Pointillism"[45]

It is clear, if you look at how Willats allocates techniques of depiction to each system, that the systems aren't exclusive. Caricatures, which in his view belong to denotational systems, are also line drawings. Photographs adhere to a perspective projection system but are composed of 'picture primitives' such as dots or pixels.

However, one could crudely characterize his taxonomy as a split between object-oriented systems, which *model* objects, and view-oriented systems, which *approximate* the view of an object. Willats comments:-

> Whereas drawing systems map spatial relations in the scene into corresponding spatial relations in the picture, the denotation systems map *scene primitives* into *picture primitives*. That is, whereas drawing systems say where the picture primitives *go*, the denotation systems say what the picture primitive stand for, refer to, or *denote*.[46]

---

[44] {Willats, 1997 #36 36 @367}
[45] {Willats, 1997 #36 36 @4}
[46] {Willats, 1997 #36 36 @15-16}

Willats toys with the idea that one could "build up a classification system for pictures corresponding to Marr's account of the main stages of the visual system".[47]

> This would still involve considering the drawing and denotation systems separately, because the image and the 2½-D sketch both provide viewer-centered descriptions (that is, representations whose geometry would correspond to a picture in perspective) but are based on different kinds of primitives.

Willats in the above quotation refers to the primal sketch stage as "the image". This is because he defines this stage as "a description of the intensity changes and local geometry of the image"[48] on the retina. He suggests that this first class, the "image state, might contain photographs, TV pictures, Impressionist paintings and Pointillist paintings."[49] A second class (the 2½-D sketch type), he suggests, could be various types of line drawing and the third class (the 3-D model stage) could be Cubist paintings and some children's drawings.

The Three Graces: Being 'optically faithful'

Willats' attempt to develop a taxonomy of pictures based on his two types goes nowhere because most pictures, even pictures which seem to be in a fairly realist style, such as Raphael's *Study of the Three Graces*, (Fig #) unravel the classification.

He asks,

> Is it possible to place such pictures, in terms of the denotation systems on which they are based, along a smooth continuum running from optically faithful pictures such as TV pictures at one end to pictures that are less optically faithful at the other.[50]

But when he considers Raphael's drawing he concludes that the most of the lines represent occluding contours whereas there are also optical elements such as tonal modeling and cast shadows. Thus, by his definition, Raphael uses optical and non-optical devices.

---

[47] {Willats, 1997 #36 36 @153}
[48] {Willats, 1997 #36 36 @152}
[49] {Willats, 1997 #36 36 @153}
[50] {Willats, 1997 #36 36 @132}

Fig # Raphael, Study of the three Graces, before 1518. red Chalk over stylus underdrawing, The Royal Collection, Her Majesty Queen Elizabeth

He also notes that for tonal modeling Raphael uses a cross-hatching technique called "bracelet shading" where the lines follow the direction of the surface contours of the figures. It can be seen in the shading on the breasts of the leftmost Grace (Fig #) and in the hatching between her breast and neck. He concludes, after considering a series of devices in work of other artists that appear "at first glance, to be more or less realistic":-

The pictorial devices employed in most artists' pictures, as distinct from photographs, are just that: pictorial devices, rather than direct facsimiles of the array of light from a scene. The effect of light and shade in a drawing like Raphael's *The Three Graces* is so convincing that it is all too easy to suppose that it has been produced by merely copying or replicating the appearance of the tonal gradations as they appear in the artist's visual field; but this would not in fact have been the case.[51]

Willats point here seems rather obtuse: the fact that the cross-hatching (which can't be seen from a distance or in a poor reproduction eg Fig#) in a picture is composed of parallel lines does not disqualify it from being in the class of pictures which, like photographs, rely on shape-from-shading effects that we naturally recognize. If we can't see the optical primitives how are they relevant to our perception of the picture?[52] He borrows the term 'optic array' from Gibson and uses it refer to the light which arrives at the retina.

Thus even apparently realistic, optical paintings can be thought of not so much as direct facsimiles of the optic array as constructions on the picture surface made up of specific and largely independent pictorial devices: devices which, like line drawing, are related to features of the optic array but not necessarily derived directly from them. For example, artists and illustrators sometimes use thicker lines in the foreground of a drawing and thinner lines in the background in order to give an impression of depth. This device is clearly related to the effects of atmospheric perspective, but it is equally clear that it is not derived directly from any identifiable feature in the array of light from a scene.[53]

I don't think it comes as any surprise to anyone that the techniques of artists are devices. A camera is also a device which employs particular techniques to render an image. Paul C. Vitz and Arnold B. Glimcher in their book *Modern art and modern science : the parallel analysis of vision*[54] provide a detailed account of how Modernist artists including Seurat, Monet and Picasso used developments in vision science to develop depiction techniques based on principles of perception, including those of the Gestalists. I suspect one could trace this tendency among artist right back to the ancient Greeks.

---

[51] {Willats, 1997 #36 36 @145}
[52] Think of all those pictures made up from vast number of other pictures. They still look like photographs from a distance.
[53] {Willats, 1997 #36 36 @133}
[54] {Vitz, 1984 #35}

One reason Willats has trouble making techniques of drawing and painting fit his denotational scheme is because, as I pointed out in the last chapter, artists don't follow strict denotational guidelines. Kennedy presented a taxonomy of lines and their denotations and I struggled to stick to it in my attempts to draw dogs and pubs. I also failed to find any examples, outside Kennedy's book, which held to the scheme.

Vincent van Gogh, Jardin de Fleurs, 1988, Ink over pencil on paper, 61 x 49 cm. Collection of Mrs E Sigrist-Nathan
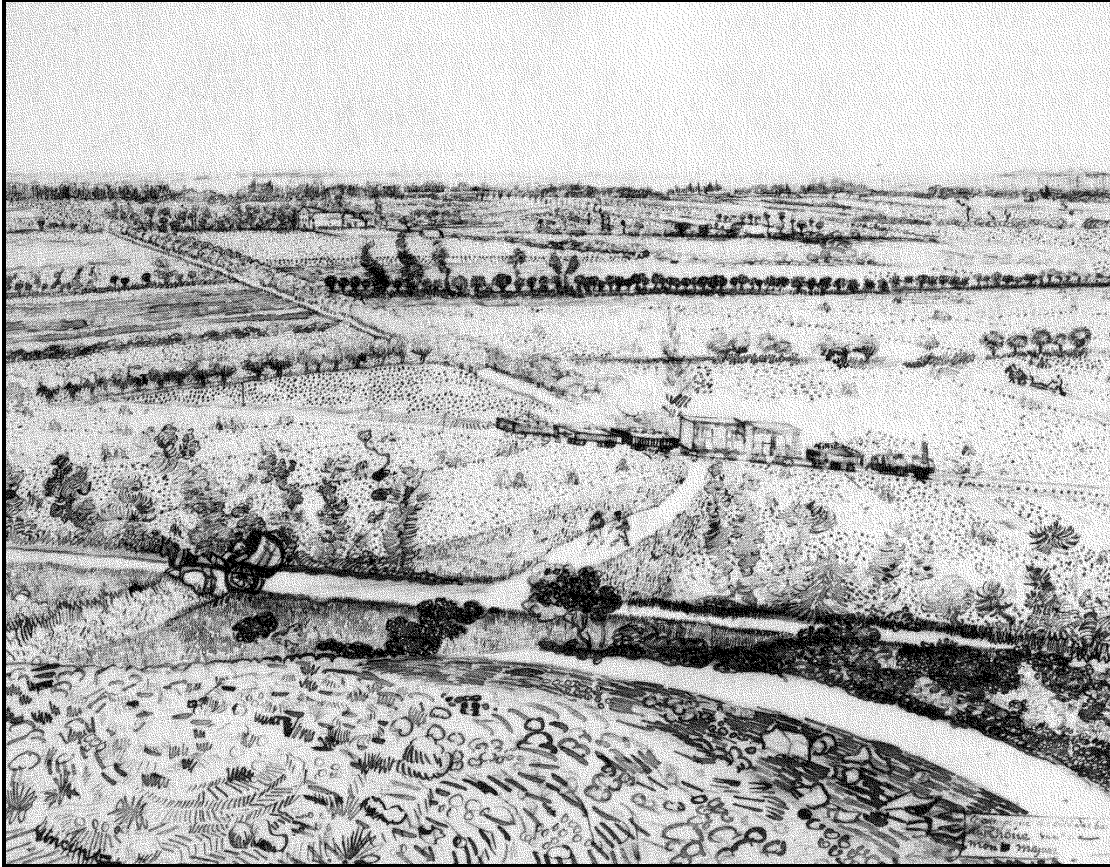
Fig # Vincent van Gogh Landscape

Willats states that the use of thick lines in the foreground and thinner lines in the background is an atmospheric perspective device. However, a simpler explanation is that it could be capitalizing on a fundamental feature of vision, exemplified in the Van Gogh drawings *Jardin de Fleurs* and *Landscape*, that we can see more detail close up than we can in the distance.

The cross-hatching device which Raphael uses adheres to some fairly basic Gestaltian rules about how lines generate perceptions. If the hatching on the Grace's arms, for example, had been at right angles to her outstretched arm instead of parallel to it, her arms would have looked shorter, and therefore less outstretched.

Willats' approach strikes me as rather ingenuous throughout *Art and Representation*. He uses terms such as "optically faithful" in order to draw the reader into a trap where he demonstrates, in true Goodman tradition, that there is no such thing as being 'optically faithful' and that everything boils down to denotation and projection.

Willats' taxonomy is clearly not going to clarify whether there is any way to separate depictive techniques in terms of their underlying process. Look at the variety of kinds of dog picture I have used in this study so far – photographs, silhouettes, line drawings, paintings, even a mosaic. It's hard to believe that they all refer in the same way or that they all cause the same dog-shape selective neurons to fire. This observation, that maybe pictures refer in assorted ways, seems to fuel the denotationist view that depiction is no more than a bag of tricks. On the other hand, we have no evidence that vision itself amounts to one coherent mehod of seeing – our visual system seems to have many strategies for getting information from light and guiding our behaviour. Perhaps the diversity of depictive techniques reflects the diversity of modules and tricks which vision itself employs.

What theoretical ground do vision researchers have to use 2D pictures as substitutes for 3D objects?

Our project here is to understand how the artist "mobilizes the mechanisms of recognition".[55] If techniques of depiction are diverse then we would expect that mechanisms of recognition must be at least equally diverse. In the last section we tried to reason from how the mechanisms of recognition worked with real objects and scenes to some conclusions about how they worked with pictures. We didn't find any evidence that from the perspective of a vision scientist there was anything different about picture recognition and object recognition. So maybe we could take Zeki's advice and reverse the direction of reasoning and look at some of the techniques of depiction and see what they tell us about the mechanisms of recognition. Zeki argues, somewhat rhetorically, in his paper 'Art and the Brain', that "artists are neurologists, studying the brain with techniques that are unique to them and reaching interesting but unspecified conclusions about the organization of the brain."[56]

---

[55] {Podro, 1998 #28@6}
[56] {Zeki, 1999 #324@76–96}