Depiction and Recognition
by
Ron Gallagher

Submitted to
The Department of Philosophy and Bioethics
for the degree of
Doctor of Philosophy
at
Monash University
on
June 16th 2006
# Table of Contents

## Chapter One

### The Recognition Moment

1. **Resemblance vs Convention: What roles might resemblance and convention play in pictorial interpretation?**

2. **Goodman, Resemblance and Selective Commitment**

3. **Goodman’s Five Objections to Pictorial Resemblance**

4. **Resemblance, Realism and the “Innocent Eye”**

5. **Do projection systems affect pictorial recognition?**

6. **Realism and the Western Eye**

7. **The No-learning Debate and the End of Conventionalism**

8. **Summary of Chapter One**

## Chapter Two

### Schier, Natural Recognition Abilities and Common Sense

9. **Convention Maketh Not the Icon**

10. **Natural Generativity**

11. **Schier’s Two-Stage Theory**

12. **Recognition Triggers and Illusions**

13. **Selective Commitment, Frame Knowledge and Communicational Cooperation**

14. **A Viewer Equipped with Conventionalist Interpretative Abilities Looks at the Constable Landscape (or is it a Pink Elephant?)**

15. **A Viewer Equipped with Natural Recognition Abilities Looks at the Constable Landscape**

16. **Summary of Chapter Two**

## Chapter Three

### Recognition-based Theories vs Resemblance Theories

17. **Is Schier’s account vulnerable to alternate theories of depiction?**

18. **Resemblance Theories**

19. **Hopkins and Resemblance Theory**

20. **Outline Shape Resemblance**

21. **Silhouettes**

22. **Some Interim Conclusions on Silhouettes and Recognition**

23. **Are nested outlines any easier to disentangle and recognise than overall outlines?**

24. **Context and Configuration**

25. **Stick figures and Seeing-in**

26. **The End of Resemblance?**


<table>
<thead>
<tr>
<th>Chapter Four</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mechanisms of Recognition</td>
<td>127</td>
</tr>
<tr>
<td>Composition, Change Blindness and Visual Metacognition</td>
<td>127</td>
</tr>
<tr>
<td>Configuration, Orientation, Detail and Information</td>
<td>134</td>
</tr>
<tr>
<td>Camouflage, Evolutionary Psychology and Gestalt</td>
<td>142</td>
</tr>
<tr>
<td>The Assumptions of Vision Researchers</td>
<td>145</td>
</tr>
<tr>
<td>Ramachandran and the “Bag of Tricks”</td>
<td>149</td>
</tr>
<tr>
<td>Shape from Shading</td>
<td>151</td>
</tr>
<tr>
<td>The Smell of an Oily Rag</td>
<td>159</td>
</tr>
<tr>
<td>A Brain Scan Experiment: Tanaka’s Differential Amplifier</td>
<td>166</td>
</tr>
<tr>
<td>Context, Constancy and Gist</td>
<td>177</td>
</tr>
<tr>
<td>Change Blindness, the Grand Illusion and the Sensorimotor System</td>
<td>193</td>
</tr>
<tr>
<td>Sensorimotor Cues</td>
<td>197</td>
</tr>
<tr>
<td>Summary of Chapter Four</td>
<td>202</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter Five</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conclusion</td>
<td>204</td>
</tr>
<tr>
<td>The Halle Berry Neuron</td>
<td>206</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter Six</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>212</td>
</tr>
</tbody>
</table>

| Appendix: The Complete Pictures                                              | 216  |

| Bibliography                                                                | 218  |
Summary

In this thesis I begin by evaluating whether conventional and resemblance theories of depiction can explain how we recognise the content in pictures in the first half-a-second of looking. I conclude that no version of either theory can explain how a pictorial interpretation gets started, and both require an appeal to natural recognition abilities to explain how an interpretation gets a foothold. Using Schier’s recognition-based theory, I analyse the differences between a resemblance account and a recognition account of depiction. I conclude that resemblance cannot work either as a mechanism for interpreting pictures or for recognising objects in the world. The reasons why resemblance is unworkable are analysed in the context of recent work in visual cognition. Research on change blindness, visual metacognition and gist views indicates that we recognise objects, and by extension pictures, on the basis of very little information. I go on to explore the neurological mechanisms which are responsible for pictorial recognition and propose a methodology for research into the theory of depiction based on research on visual cognition.
Statement

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program. I gratefully acknowledge the assistance of my supervisors, Ian Gold, John Armstrong and Dirk Baltzly. I also gratefully acknowledge the editorial assistance of Alison Goodman.

Ron Gallagher
June 2006
Chapter One

The Recognition Moment

Figure 1  John Constable, _Salisbury Cathedral from the Bishop’s Grounds_, 1820. Oil on Canvas, National Gallery of Canada, Ottawa

When we see objects, animals and people in pictures, are we interpreting a symbol system, seeing resemblances or merely using our natural perceptual abilities? Is it possible that we are using all three of these methods? It is hard to argue that the brown smudges in Constable’s painting of Salisbury Cathedral resemble cows. In fact, it’s hard to be sure they are cows. Observations such as this seem to damage the theory of pictorial resemblance and give credence to the conventionalist view that when we look at a picture we are interpreting a symbol system. The conventionalist maintains that the brown smudges are symbols for cows, and a symbol for a cow no more needs to look like what it represents than the word "cow" needs to sound like what it refers to. In response to this claim you might be inclined to ask “In what symbol system do brown smudges stand for cows?” It is hard to believe that there is a convention about brown smudges and cows. We might be inclined to think that the painting triggers some of the perceptual
abilities which we use to recognise cows, trees and buildings in real life. Of course, the problem with this account is that recognising an enormous three dimensional Cathedral whilst standing in a field surrounded by trees and cows on a windy day in Wiltshire will obviously engage a very different range of perceptual abilities from those required to look at a rather sketchy painting of the scene.

None of these candidates for a theory of depiction look very promising when we consider the problem of the cows. Resemblance would seem to require a higher degree of resemblance than exhibited in this painting – the cows don’t even have legs! Conventionalism seems to require that somehow a symbol system exists which we have absorbed, that features a convention that brown smudges can stand for cows. Clearly the convention would need to be more specific about the kinds of smudges that generate cow interpretations for this to be a feasible convention. Recognition theory would have us believe that looking at the brown smudges triggers perceptual processes much like those which are triggered when looking at two tons of beef on the hoof standing in a paddock.

Certainly it seems possible that we use all three methods – perhaps visual cognition is a complex mix of convention, resemblance and recognition. A cognitive account of perception should be able to eliminate, or dissolve, the problems which these three accounts raise. After all, a cognitive theory of perception must include an account of how we interpret symbols as well as how we perceive the real world. Such a cognitive account should also satisfy our intuition about the experience of resemblance we feel when we look at Constable’s pastoral scene.

One thing that we can agree on is that anyone with normal human perceptual abilities will see a building framed by trees in the Constable picture. Moreover, that recognition will occur in less than half-a-second of sighting the picture. Psychophysical experiments on picture recognition indicate that novel images flashed rapidly onto a screen can be categorised by human subjects into animal, forest, building, fruit etc in around one-fifth of a second.¹ It is this primary recognition moment I would like to focus on in this study.

A picture as sophisticated as Constable’s painting of Salisbury Cathedral speaks to us in many ways – expression and symbolism are certainly there – but these interpretations depend on our ability to recognise what is depicted.

![Figure 2](image)

Figure 2  Piero della Francesca, *The Baptism*, 1442. Tempera on panel, 167 x 116 cm, National Gallery, London

The dove in Piero della Francesca’s *The Baptism*, for example, symbolises the Holy Ghost. Clearly one needs prior knowledge of the symbology of Christian religious paintings to interpret the dove as the Holy Ghost, but any child can identify the dove as a bird. The point is that the viewer needs to be able to recognise the “raw” content of the picture – a landscape with people under a tree – before going on to talk about expression and symbolism.

It is this ability to recognise objects, animals and people, and their disposition in a scene in those few first moments of looking at a picture that interests me. The speed of recognition of familiar objects does not seem to be affected by style or

---

technique of depiction. I doubt that you could say (or prove) that it was easier or harder to recognise depicted content in this photograph of Salisbury Cathedral (fig. 3) than it is in Constable’s painting. Perhaps some psychophysical experiment might establish some differences in response time or a brain scan might indicate that different perceptual processes are being triggered. I would be fascinated to see the results of such experiments, and in the latter part of this study I look at what psychophysics and neurological scanning can tell us about these perceptual processes.

![Salisbury Cathedral from the Bishop's Grounds](http://www.salisburycathedral.org.uk/gallery.php)

Before I explore the significance of this work in vision research for a theory of depiction, I would like to clarify which issues are relevant to the pictorial “recognition moment”. There is considerable interest in the intersection of aesthetics and cognitive science, but both are enormous fields with diverse, if not
The Recognition Moment

divergent, interests. The issue of pictorial representation, as Dominic Lopes points out\(^3\), can be split between three related questions.

1. What is a representational picture?
2. What is the relationship between real-object perceptual experiences and picture-object perceptual experiences?
3. How does a picture represent what it does?

My focus is on the last question. And an answer to the last question has implications for the second and first questions. However, the search for an answer to the question of how we see depicted content in pictures based on current theories of visual cognition poses an immediate problem – there is no single theory of visual cognition. Over the last few decades all large theories of vision and recognition have had to be re-written because of new discoveries in psychology and neuroscience about how the brain actually works. Thousands of papers are published every month reporting the results of experiments and developing new theories on workings of the visual system. In this jostling field of theories and counter-theories it’s hard to back a winner. The visual system (in fact, the brain in general) works in radically counter intuitive ways and many a theory based on sound logical principles has been invalidated by the evidence from experimental work. But that’s science! The challenge is to develop theories of depiction that can be tested. Of course, it is setting the criteria for a successful test that is the difficult part of any experiment. A theory of depiction should explain how human beings can see depicted content in pictures, and how they can see the depicted content that the artist or photographer intended. In particular, it should be able to indicate what kinds of knowledge and perceptual abilities are required to kick-start a pictorial interpretation.

Initially I will try to clarify the question: Do pictorial conventions, or resemblance between the depicted object and its depictum, generate recognition?

**Resemblance vs Convention: What roles might resemblance and convention play in pictorial interpretation?**

Before we can ascertain whether resemblance plays a role in the initial pictorial interpretation we need to specify what kind of resemblance is a candidate? For

example, is it a congruence of shapes as they present to the retina, or does resemblance happen elsewhere? There are a number of competing accounts of the role of resemblance in pictorial interpretation and I propose to measure these against current theories of visual cognition.

There are also a number of competing views on the role of convention in pictorial interpretation. Conventionalist theories suggest that a viewer needs knowledge of the relevant pictorial symbol system to interpret a picture, but these theories are generally not very specific about the perceptual abilities involved in applying this knowledge. For example, when we look at Constable’s painting, is it possible, theoretically or practically, to isolate knowledge of pictorial conventions from knowledge of a more general sort? We know that cows are often found in fields and surrounded by trees. The conventionalist claim would be that we also know that landscape pictures often feature cows and this knowledge may help us identify the brown smudges. Of course this doesn’t tell us that knowledge of pictorial conventions of landscape actually contributes to the initial recognition of the picture as a landscape. The conventionalist position isn’t as intuitive as the resemblance position, but it does appeal to the fact that pictures are artefacts that are constructed using certain techniques and styles. It is hard to argue against the claim that pictures are artefacts and we use knowledge of their artifice to interpret them.

Once I have clarified the kind of role resemblance and convention might play in the recognition moment, and flushed out the relevant issues, I will go on to establish some ground-rules for pictorial recognition using the work of Flint Schier.

**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.4 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.

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

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

If Goodman’s objections to resemblance have any weight, they would seem to strengthen the case for the idea that a cognitive account of pictorial interpretation also requires an account of symbol interpretation. His arguments would also weaken the case for any theory which appeals to property matching as the primary mechanism in the pictorial interpretation.

Maybe issues such as the selective commitment problem are a false dilemma encouraged by the opposition of the conventionalist and resemblance accounts. If this is the case then an analysis of the conventional and resemblance positions, particularly with regard to how an interpretation gets started, should clarify which issues remain on the agenda in the development of a cognitive account of pictorial representation.

**Goodman’s Five Objections to Pictorial Resemblance**

Goodman’s arguments purport to show that resemblance cannot play an explanatory role *vis-à-vis* how a picture refers to its subject. He argues that a picture does not refer to its subject because it resembles it. On the contrary, pictorial reference is achieved by virtue of a denotational system much like natural language, and the experience we call “pictorial resemblance” is the result of that process.

His analysis turns on the logic of resemblance relations, and the conditions, such as context, which enable these relationships to hold.

- **Objection 1:** Resemblance is reflexive: reference is not.
- **Objection 2:** A picture resembles any other picture more than it resembles its subject.
- **Objection 3:** Depiction cannot be imitation: a picture always involves selection.
- **Objection 4:** Perspective is an arbitrary convention.
- **Objection 5:** Realism is culturally determined and so is resemblance.

The first two objections are to resemblance as a **sufficient** condition for a picture to refer to its subject – the remainder are objections to resemblance as a **necessary** condition. The first three objections, taken as statements about

---

7 Goodman comments, “B is as much like A as A is like B, but while a painting may represent the Duke of Wellington, the Duke doesn’t represent the painting.” Nelson Goodman, *Languages of Art: An Approach to a Theory of Symbols* (Indianapolis: Bobbs-Merrill, 1968), 4.

8 Goodman argues, “A Constable painting of Malborough Castle is more like any other picture than it is like the Castle, yet it represents the Castle and not another picture – not even the closest copy.” Ibid., 5.
resemblance and depiction, are largely true. The last two objections are debateable and hinge on Goodman’s idiosyncratic concept of realism.

The first two observations on the reflexivity of resemblance relations do not illuminate the question of how we see trees and cows in the Constable painting. It is true, as Goodman points out, that one painting is more like another painting than what it represents, but that does not preclude the possibility that the painting also resembles what it represents. The painting of Salisbury Cathedral from the Frick Collection (fig. 5) obviously resembles the Ottawa study at the beginning of this chapter more than it does the actual Cathedral, but that does not necessarily mean there isn’t some resemblance to the Cathedral. Goodman is claiming here that “no degree of resemblance is sufficient to establish the requisite relationship of reference” and I agree with him. Degree of resemblance isn’t sufficient to establish reference, other conditions need to apply (e.g. context) before we can say that the picture of Salisbury Cathedral refers to Salisbury Cathedral.

Figure 5 John Constable, Salisbury Cathedral from the Bishop’s Grounds, 1826. Oil on Canvas, 88.9 x 112.4cm, Frick Collection, New York

9 Ibid.
He argues, in his next three objections, that resemblance is actually a *product* of context, and that degree of resemblance is dependent on matters of personal choice and circumstance. Consequently, resemblance is not even a necessary condition for reference, because it *arises* from the other conditions rather than works together with them. Thus we see that Goodman ascribes a dominant, even decisive, role to context and association.
An artist cannot copy the full range of visual properties an object may have when looked at in various ways, in different moods, at different times, from different angles etc. The artist must choose which angle and in what light and from what distance to represent an object. Goodman argues that a “faithful picture” cannot be a copy or imitation of an object because the person that creates the picture “selects, rejects, organizes, discriminates, associates, classifies, analyses, constructs”\textsuperscript{10} and so does the eye that looks. What Goodman is saying is that there is an array of aspects of an object from which an artist must select in order to represent the object, and these choices are likely to reflect the artist’s mood, disposition and personal tastes. This, he insists, is as true for photographers as it is for artists. The photographer must choose whether the lens is zoomed in on its subject and thereby flattens the picture, or whether the lens is wide and emphasises shapes. The photographer must also decide whether the picture will be black and white or colour, whether the colour should emphasise flesh tone or not, whether the image will be grainy or fine, etc. There are a whole host of settings which the modern photographer can choose from to change the quality of the image. The photographer does not have the range of techniques open to her that the painter does, but we can see that any of these decisions will affect the way the object looks in the photograph and, presumably, how the picture resembles its subject.

Goodman’s larger point, following Gombrich, is that there is no “innocent eye”; an artist isn’t neutral in the way that he or she looks at an object. So, far from being a copy of an object, a picture is the selection of various aspects of an object according to the disposition of the artist/photographer.

\textsuperscript{10} Ibid., 9.
If, for example, some kind of tree-shape resemblance triggers recognition both in Constable’s painting and the photograph of the Cathedral, one would expect some similarity between the kinds of tree-shapes. In fact the tree shapes are entirely different. When you look at the painting you can see how Constable has emphasised the sinuous organic aspect of the trees – they are almost alive! The trees in the photograph are like cardboard cut-outs.

Goodman’s objective is to undermine our concept of objective representation and thereby undermine any absolute sense of resemblance. Goodman, quite rightly, emphasises the tricks and devices which an artist must use to make the marks on canvas evoke an object or scene in a particular way. He argues that any resemblance between the picture and what it depicts is as likely to be in the eye of the creator as in the eye of the beholder. The phenomenon of resemblance, he says, is therefore variable and relative.11

Goodman constructs his argument around two assumptions:

1. He equates ‘realist’ depiction with trying to make a copy of reality.
2. He assumes that the ideal of realist depiction is the representation of all the visual aspects of the object.

The problem with his analysis is:

1. No-one, not even Gombrich, and certainly no twentieth-century artist, thinks that a realist depiction is a copy of reality.
2. No-one believes that one can depict all the visual aspects of an object at the same time in the same picture.

So, his observation that realist depiction is a style – a convention that involves making choices about how and what to depict – does nothing other than state the obvious. An artist must choose which aspects of a scene to emphasise, just as we must choose what to focus on when look at the real world.

His main objective is to link the comparative degree of realism with degree of resemblance, and thereby undermine resemblance by linking it to conventions of realism. He argues in “Seven Strictures on Similarity”,

---

11 Goodman makes a similar point about relativity of fidelity in sculpture. “The distant or colossal sculpture has also to be shaped very differently from what it depicts in order to be realistic, in order to ‘look right’. And the ways of making it ‘look right’ are not reducible to fixed and universal rules; for how an object looks depends not only upon its orientation, distance, and lighting, but upon all we know of it and upon our training, habits, and concerns.” Ibid., 20.
The most we can say is that among pictures that represent actual objects, degree of realism correlates to some extent with degree of similarity of picture to object. But we must beware of supposing that similarity constitutes any firm, invariant criterion of realism; for similarity is relative, variable, culture-dependent.12

Goodman here tries to tie similarity to realism and thereby undermine resemblance by ascribing to it a following role in the depiction relation. However, a comparison of realist depictions demonstrates that the criteria Goodman criticises are in fact not the criteria we use to judge realism. The photograph of Salisbury Cathedral (fig. 3 above) may be more “technically” realist than the Constable painting at the beginning of this chapter, but that is not to say that it has the greater resemblance to Salisbury Cathedral. The painting of Salisbury Cathedral in fig. 5 is in a more realist style than the one at the beginning of the chapter. It is less impressionistic, and if you squint you might mistake it for a photograph. The cows definitely look more like cows when you examine them closely, and the trees are not so wild. But can we say this more realist painting resembles the scene more than the earlier one at the beginning of this chapter? I’m not sure that I could unequivocally state that, for example, the trees in one resemble trees more or less than in the other. I certainly wouldn’t say that the differences in style influence my ability to recognise the depicted content.

Goodman’s attempt to convince us that resemblance emerges from a culturally dependent criterion of realism fails because we do not equate degree of realism with degree of resemblance. His argument is that we judge that a picture resembles its subject more when it is in a realist style because the conventions of realism lead us to that conclusion. He claims that a high degree of resemblance doesn’t result in realism – realism results in the impression of resemblance. His ploy is to persuade us to disavow the intuition that because a photograph is a faithful record of how light is reflected from a subject that it therefore resembles the subject. He wants to convince us that it is because the lens and camera are constructed according to standards of Western realist depiction, not because of some causal relation between the subject and the camera’s film, that we see the resemblance. If his idea of a high degree of realism is something like photographic detail, he is clearly wrong. We can sometimes remark stronger

resemblances in sketchy drawings, e.g. caricatures, and cartoon characters than in full-colour photographs. Goodman’s last two objections are also designed to show how resemblance proceeds from the conventions of realism, and not vice versa. These objections hinge on his conception of realism and thus provide some insight into how Goodman conceives realism.
The Recognition Moment

There is more to vision than meets the eye. Nelson Goodman

Do projection systems affect pictorial recognition?

Earlier I noted that Goodman is rather vague about what a “system of symbolisation” is and how it works. Goodman’s arguments against perspective give us some insight into which elements of a picture conventionalists construe as the “system of symbolization”. Goodman encourages us to think of pictures as a language-like symbol system but, as Flint Schier points out, “with pictures you have nothing that plays the role of a word.” Schier goes on to point out that there is nothing that can be construed as grammar or syntax either. It may seem curious, at this stage, to be struggling to identify what a “system of symbolization” is, but neither Goodman nor decades of conventionalist theory (including the ‘science’ of semiology) throw any light on which elements of a picture combine in what ways to refer to objects in the world. The best the conventionalist can do is appeal to the idea that during our upbringing we have absorbed the rules of pictorial symbol systems in much the way that infants can learn rudimentary language skills without tuition. This is the strategy Goodman adopts when he tries to convince his reader that perspective is an arbitrary convention which we have learned to associate with pictures that are more realist.

The main elements of a picture are the medium, the style and the projection system. For the purposes of this investigation the question is “How do these elements affect the ability to recognise a picture?” The issue which Goodman’s objection to perspective raises is “Do projection systems affect pictorial recognition?”

Goodman’s argument is designed to undermine the idea that the techniques of perspective enable the artist to represent a true version of how we see the world. He argues that there can be no resemblance between how we see spatial relations in the world and the way spatial relations are depicted using perspective. His argument is two-tiered.

First, he argues that although perspective purports to be a scientific discovery, and an objective means of rendering objects in space on a two dimensional

canvas, the conditions under which we need to view a perspective picture for it to “deliver to the eye a bundle of light rays matching that delivered by the object itself” are so unnatural that it is a misnomer to describe it as “correct pictorial representation”.

The putative aim of a picture painted in perspective, Goodman claims, is to recreate an image on the retina that matches what the artist saw when he painted it. This is almost impossible, he claims, because no matter how we stand in relation to the picture it cannot deliver a “bundle of light” that matches what we get when we confront the scene in real life. When we approach a picture in a gallery we almost always approach it from an angle and pictures in magazines are very rarely viewed straight on. Goodman suggests that conditions of observation rarely, if ever, achieve the objective that is claimed for perspective depiction – correct pictorial representation. An illusion of reality.

Second, he asks, if a perspective drawing is designed to deliver the right bundle of light rays to the eye such that we were fooled into thinking we were seeing the real object, why do most realist pictures omit vertical perspective? He observes that, under “standard pictorial rules”, “railroad tracks running outward from the eye are drawn converging, but telephone poles (or the edges of a façade) running upward from the eye are drawn parallel. By the “laws of geometry” the poles should also be drawn converging. But so drawn, they look as wrong as railroad tracks drawn parallel.” He argues that the omission of vertical perspective from most realist pictures shows that perspective is merely a realist stylistic convention and not an objective means of rendering three dimensional space in two dimensions.

These two related criticisms of perspective build on Goodman’s observation that realism is nothing more than a style of depiction. Both are designed to show that one of the principle techniques of realism – perspective – is not objective and scientific, but an arbitrary convention. A picture of an object in correct perspective...
perspective, in Goodman’s view, cannot re-create the experience of seeing the object in real life because the normal conditions of viewing a picture do not enable anything like the same optical conditions.

Tantalising though it may be to get into the perspective debate, we need to analyse what Goodman thinks this has to do with resemblance. What he actually says is:

Identity in pattern of light rays, like resemblance of other kinds, is clearly no sufficient condition for representation. The claim is rather that such identity is a criterion of fidelity, of correct pictorial representation, where denotation is otherwise established.

Goodman is arguing that advocates of “scientific perspective” are claiming that the image which a picture in perspective casts onto the retina resembles the retinal image from the scene in real life. His counter-claim is that there can be no resemblance between the retinal image and what we see depicted because the conditions of viewing are unnatural and couldn’t be a match. We will see later that Goodman is right, but for the wrong reasons. It makes very little sense to talk about our visual system matching images on the retina. We don’t recognise objects by comparing our current retinal image with a similar one we have had previously. We are not going to be able to explain the experience of resemblance through congruence of retinal images. There are a number of reasons, some physiological and some psychological, why such a scenario is unworkable, and we will explore these later when we analyse how recognition works. The issue at stake here is whether perspective is a technique which enhances the resemblance relation between a picture and a scene, and thereby facilitates recognition of that scene.

Goodman’s second argument against perspective suggests that the elimination of vertical perspective from most realist pictures shows that the Western eye is selective about the properties a realist picture requires to resemble reality. Vertical perspective is corrected for by certain cameras and lenses. He argues that if perspective is a scientifically correct way of rendering three-dimensional objects in two-dimensions we should see “standard” realist pictures which omit vertical

---

21 There is an interesting debate on the subject in Heiko Hecht, Robert Schwartz, and Margaret Atherton, eds., *Looking into Pictures: An Interdisciplinary Approach to Pictorial Space* (Cambridge, Massachusetts: The MIT Press, 2003).

perspective as distorted, but in fact, he claims, we don’t. Rather, he says, we tend to see a picture which features vertical perspective as distorted or “not quite right”. He writes, “The artist who wants to produce a spatial representation that the present-day Western eye will accept as faithful must defy the ‘laws of geometry’.” Goodman concludes that this selectivity about the kind of picture we consider distorted, or otherwise, supports his claim that the rules of perspective are “whimsical” and other realist picturing techniques are arbitrary conventions not scientific discoveries.

Goodman is right – pictures with vertical perspective can look distorted. So can pictures taken from extreme close-up and from unusual angles. But he overstates

---

23 Ibid., 16.
24 Ibid., 19.
The Recognition Moment 20

the case. The picture of Salisbury Cathedral from the east in figure 6 (above) features vertical perspective, and it doesn’t look strange at all. Most photography of tall buildings features vertical perspective. If it is true that many realist paintings omit vertical perspective there are two very plausible, and related, reasons:

- The painting of the building may be from such a distance that the effect of vertical perspective is negligible.
- Painting a picture in vertical perspective is quite tricky. An artist may omit it because it’s hard and usually doesn’t make that much difference.

It’s a lot easier to take a photograph using vertical perspective than it is to paint a picture.

There may be other reasons for “realist” pictures favouring a view from the straight and level. Human beings the world over seem to favour pictures of objects where the object is pretty much straight-on – there is some evidence from work on canonical views\(^2^7\) that this is even the case for objects which are usually seen from above or below e.g. shoes and chimneys. We tend to depict them straight on. The vertical view, the up and down, may not be as important to human beings as the side to side view. Perhaps that’s because human beings don’t have many predators which come from above, or maybe it is because our visual field subtends a small angle, and that angle is usually straight ahead. The omission of vertical perspective from much realist painting may merely be a concession to what we take as being a normal human point of view – the one we usually have when we are standing and confronting the world at eye level. Thus, realism may simply be a bunch of techniques which suit how we normally see the world. Goodman’s claim is that we cannot know what a normal view of the world is – but that is where he’s wrong. Research on perception indicates that there are identifiable norms in the ways that human beings see the world. These norms are not cultural; they are universal features of human cognition. Realism may be a mode of depiction particularly well tailored to the disposition of the human visual system.

---

\(^{26}\) There was a fashion for wide-angle photography in the 1960s, which highlights the vertical foreshortening, and some of those pictures still look a bit strange.

Much work on perception takes an evolutionary perspective and analyses the role of the visual system in terms of how its functions are tailored to the organism’s evolutionary niche.

Figure 7  Richard Estes, *Double Self Portrait*, 1976. Oil on canvas, 60.96 x 91.44cm, MOMA, New York

Thus our feeling that a picture is realist is not determined by the level of detail with which the artist has copied reality, but by features of the picture which we might encounter in everyday life. The criteria for “natural realism” are all concerned with the way human beings engage the world in their normal environment. The paintings of Richard Estes and other super-realistic artists are often disturbing because they fail to meet some of the criteria for a naturally realist picture. In his *Double Self Portrait* (fig.7) for example everything in the scene is far too cleanly delineated. We can make out every detail in the reflection. If we were to encounter this scene in real life there would be glare, ambiguous reflection, lack of detail of distant objects and signs, atmospheric haze – Estes has left all this “noise” out of his painting and achieved a kind of sterile pseudo-realism.

The painting is unlike a photograph because in a photograph it would be impossible to maintain focus throughout the depth of the picture. It would also be hard to maintain details in the dark interior areas whilst preserving the detail in the brightly lit exterior.
In a later chapter we will look at the techniques artists use to exploit our sense of what is natural in a scene to enhance our feeling of realism.

Goodman’s attempt to convince his reader that the techniques of realism, such as perspective, are arbitrary conventions is not very successful. It is fairly clear that a picture of an object or scene from a human viewpoint is more recognisable than one from above or below. The human visual system predisposes us to recognise the world from this human point of view and pictures which use rudimentary perspective capitalise on this pre-disposition. However, we can equally recognise pictures which are not in perspective (eg children’s drawings) and pictures which use vertical perspective. For Goodman’s argument to have any weight he needs to convince his reader that only pictures which are in some kind of perspective are seen as resembling what they depict because during our pictorial upbringing we have absorbed the conventions of perspective. This he fails to do.

**Realism and the Western Eye**

In his fifth objection, Goodman argues that realism is totally conventional and culturally determined, and what we experience as depictive resemblance must also be a product of culture and convention. To some extent the previous two objections are designed to support this major plank of Goodman’s aesthetic theory. He states:

> Realism is relative, determined by the system of representation standard for a given culture or person at a given time. Newer or older or alien systems are accounted artificial or unskilled. For a Fifth-Dynasty Egyptian the straightforward way of representing something is not the same as for an eighteenth-century Japanese; and neither way is the same for an early twentieth-century Englishman.28

Goodman claims that to the Western eye, Western realist pictures are truer to reality than, for example, eighteenth-century Japanese or Fifth-Dynasty Egyptian, and that this is a result of habituation to the ambient Western style. He argues that resemblance cannot be said to establish the relation of reference of pictures because what we call “resemblance” is a product of the prior processes of denotational reference which are characteristic of depiction. He writes:

> Our addiction, in the face of overwhelming counter evidence, to thinking of resemblance as the measure of realism is easily understood in these terms.

---

Representational customs, which govern realism, also tend to generate resemblance. That a picture looks like nature often means that it looks the way nature is usually painted. Again, what will deceive me into supposing that an object of a given kind is before me depends upon what I have noticed about such objects, and this in turn is affected by the way I am used to seeing them depicted. Resemblance and deceptiveness, far from being constant and independent sources and criteria of representational practice are in some degree products of it.29

Goodman is here, once again, trying to invert the intuition that a greater degree of resemblance between a picture and its subject makes the picture more realist. How does this objection show that resemblance plays no part in depiction and what does it say about how we recognise pictorial content?

It suggests that there is a mechanism of learning to see resemblances between pictures and their depictum which is particular to each culture and which is determined by the representational norms of each culture. Goodman mentions that there is “overwhelming evidence” for this “conventionalist” view that each culture develops its own techniques of depiction, which then become the “realist” or “straightforward way of representing something” in that culture.30 He argues that these techniques inevitably seem strange to those in another time or culture and that this is proof that depictive resemblance is not a cultural universal.

In fact, the evidence points the other way. There is ample evidence to suggest that objects in depictions (drawings, paintings, photographs) from just about any era and culture can be easily recognised by people from other eras and cultures.

29 Ibid., 39.
30 Ibid., 37.
The birds in this 15th century Japanese drawing by Orugi Soritsu (fig. 8) are recognisable to anyone from any culture. The power of the drawing to depict recognisable birds is not affected by time or culture.
This eighteenth-century Japanese landscape (fig. 9, above) in the Chinese style by Nakabayashi Chikuto is unmistakably a building surrounded by trees. Our cultural remoteness from the styles of Japan in that period does not prevent us from seeing what is depicted.

The wall painting below (fig.10) has been preserved under the layers of pumice which buried Pompeii following the eruption of Mount Vesuvius in 79AD. It is another example of the stability of canons of pictorial representation. Not only is the body of the main character well-proportioned and realistically modelled but there is a hint of perspective in the depiction of the stairs and the archway to his right. Walking around Pompeii today you will find many examples of what are considered to be some of the earliest uses of perspective.
Guard dog mosaics, like the one above (fig. 11), are common in Pompeii. They guard the entrance to the villa and, if you approach it from the correct angle, it looks as if the dog is sitting up in three-dimensions. This optical trick is thousands of years old and it is just as effective now as it was then.
Finally, this Palaeolithic drawing of a horse from the caves at Lascaux (fig. 12), speaks to us across tens of thousands of years of human culture. It features outline drawing, profile depiction and the suggestion of movement – this is one of the earliest examples of these techniques being used in depiction. The fact that anyone can recognise what is depicted in the all foregoing pictures – and recognise these things despite the fact that the pictures might not have originated in our time and culture – shows that the power of depiction to represent is not culture-relative.

Where is Goodman’s overwhelming evidence that “representational customs, which govern realism, also tend to generate resemblance”? More to the point, where is his evidence that these representational customs are era and culture specific? In fact, there is none. All the evidence points to the fact that techniques of depiction throughout the history of humankind are more similar than they are different. The evidence shows that these techniques have been around for thousands of years and the ability to interpret them is a universal human trait.

Goodman’s claim that techniques of depiction are arbitrary conventions is further damaged by the fact that there are, in fact, many techniques of projection and depiction and many styles which we do not count as realist. These techniques have come about for a variety of reasons. For example, many engineers use

---

31 Ibid., 39.
oblique projection, and many artists (and whole art movements such as Cubism, Fauvism and Futurism) deliberately disrupt our expectation of normal colour, proportion and perspective. These techniques are not arbitrary and, despite the fact that they are commonly used techniques, they have not become part of the realist canon simply through familiarity.

Figure 13  Pablo Picasso, Gertrude Stein, 1906. Oil on Canvas, 100 x 81.3cm, The Metropolitan Museum of Art, New York

Goodman makes a revealing comment in his section on “Invention”. He quotes Picasso, who apparently responded to the complaint that his portrait of Gertrude Stein (fig. 13) did not look like her, “No matter; it will.”33 Goodman is here inviting us to imagine that Picasso’s style will eventually become so familiar that Cubist pictures will seem realist. The “Invention” section in Languages of Art emphasises how artists shape how we see the world by developing new techniques of depiction. Once we have learnt how to interpret these new techniques of depiction and have become habituated to them, Goodman suggests, they become part of our perceived version of realism. It is one hundred years since the first Cubist painting was painted, and although Cubism no longer shocks, it is still not considered a realist mode of depiction. We do not have Cubist photographs in our passports, and no one thinks that Gertrude Stein had a face like an African mask. Robert Schwartz comments in his essay “Representation and Resemblance” that

33 Goodman, Languages of Art, 33.
no matter how second nature interpreting cubist pictures becomes “such pictures will not seem realistic (or at least nowhere near as realistic as Rembrandt). Our judgements of realism are just not as flexible as the familiarity view would appear to require.”  

Our underlying perceptual skills, and the experiences and knowledge we have built up in our engagements with the world, constrain what we consider realist. Our capacity to recognise objects is a crucial component of those skills.

Figure 14  (above left) Pablo Picasso. Les Demoiselles d’Avignon. 1907. Oil on canvas, 243.9 x 233.7cm, The Museum of Modern Arts, New York
Figure 15  (above right) Edmond Fortier, Femme Malinké, 1906. Postcard, Musée Picasso, Paris

It is worth noting that even in the archetypical Cubist painting Les Demoiselles d’Avignon we are still able to recognise human figures with faces. We can conclude that Picasso had an intention to depict and for the viewer to see what is depicted. This is an important point. The Cubist style does not necessarily fail to depict. It is, however, selective about how it depicts spatial relations and proportions. The issue which Goodman’s objections highlight is the question of how the viewer knows which aspects of the picture are not intended by Picasso to be taken as depictive. For example, the woman at the bottom right appears to have her head on the wrong way. The table with fruit on in the foreground seems to be vertical.

For the sake of argument I’m going to observe that we see a resemblance, no matter how tenuous, between the faces in the picture and human faces. Picasso has succeeded in depicting faces and that is how we interpret them. But how does the viewer discount his depiction of spatial relations and the strange body contortions which follow from a literal interpretation? Goodman claims that the viewer must have prior knowledge of depictive techniques and symbolic schema in order to be selective about which aspects of the picture to interpret as depicting.

If we compare the poses of the women in the painting with a postcard that Picasso used \(^{36}\) (figs. 14 & 15) we can see that in some ways Picasso has been quite faithful to the pose of the woman in the postcard. We might even say there was a resemblance in the pose of the central women in the painting to the pose in the postcard. We might also hazard that the woman in the postcard resembles a real woman rather more than those in the painting. Although that does seem like a strange use of the term “resemble”.

Picasso’s commitment to depiction is even clearer when you compare the mask like face of the top-right woman with the mask he used as the model (fig 16).

\[\text{Figure 16} \quad \text{Detail from } \textit{Les Demoiselles d'Avignon} \ (\text{left}) \text{ and African mask from Picasso’s collection (right) }^{37}\]


\(^{37}\) \url{http://www.marin.cc.ca.us/art107/PicassoMaskDemoiselles.jpg}
Picasso has given us a fairly faithful depiction of the mask. Even in a picture as non-realist as *Les Demoiselles d'Avignon* there are robust examples of what we might call resemblance in the depiction.

One of the questions I asked at the beginning of this chapter was: “When we see an object in a painting, why does it trigger the feeling that there is a resemblance between the marks on the canvas and the object in real life? One way to answer this question would be to isolate the mechanism whereby we recognise depicted content in the first half-a-second of looking at a picture. Goodman has an explanation for the experience of resemblance but it does not fit the facts and he fails to tie degree of realism to degree of resemblance.

![Figure 17: Head of a Dog](image)

**Figure 17. Head of a Dog**

**The No-learning Debate and the End of Conventionalism**

Even if one doesn’t accept Goodman’s definition of realism, his observation that the viewer needs to have some understanding of depictive techniques to know which properties of the object are being represented in the picture is in need of an explanation. 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. There are two major problems with this claim. Firstly, it is not clear exactly what one would need to learn about symbol systems in order to arrive at a pictorial interpretation. Goodman’s account of art in Languages of Art is more remarkable for what it fails to address than for what it shows. Goodman fails to define at least two central terms, “denotation”38 and “symbol”,39 and says that the word “language” in the title should be replaced by the term “symbol systems”.40 He does not address how symbol systems come into being,41 and says very little about what they are. The consequence of this omission is that it is not clear what kind of visual education one requires to see dogs in line drawings.

39 Goodman says in his introduction that ‘symbol’ is used in Languages of Art as a “very general and colorless term.” He comes under some criticism for this particularly from Paul Ziff, "Review of Goodman's Languages of Art," review of Languages of Art, Philosophical Review (1971).
40 Goodman, Languages of Art, xi-xii.
Secondly; there is a kind of universal pictorial reference which we noted in connection with the horses in the cave paintings and the birds in the Japanese hanging scroll. Any child, who has seen birds and horses, can recognise what is in those pictures. The twenty thousand year old depiction of a hand above (fig. 18) is recognisable by anyone with human visual abilities. Furthermore, there is strong evidence that no learning is required to see familiar objects in photographs or line drawings even in a first-time encounter with pictures. Our ability to see things in pictures co-varies with our ability to see things in the world. Somehow we see the hand, the horse and the dog in spite of the lines and the paint. This ability appears to be innate and not a product of upbringing or convention. Our explanation of the selective commitment issue needs to explain how this innate ability to see objects in pictures is unfettered by the artifice of the picture.

In his book *Deeper into Pictures*, Schier develops the theory of “natural generativity”. His theory accounts for our capacity to see what is depicted in a picture in terms of an innate capacity to recognise objects. He argues, if you can see mountains, churches, cars, and planes in real life, then the chances are you can see them in pictures. Of course there has to be a “first time” experience of recognising what is depicted in a picture. For most of us it would have been when

we were one or two years old. On the other hand, one doesn’t come to recognise pictures bit by bit. There isn’t a stage where you see lines but not shading, or eyes but not ears, or cats but not dogs. It isn’t like learning a language where you gradually build up a vocabulary. Schier comments:

Although in some cases something like learning how to interpret icons occurs, once the coin has dropped and someone has succeeded in a first pictorial interpretation, they will then be able to go on to pictures of other things and, without further ado, say what the icon depicts, provided that they are able to recognise the depicted object.

The trick of interpreting a picture relies on a person having at some stage seen a picture of something, and that something needs to be something they have encountered in real life – once they have interpreted their first picture they can see content in all pictures thereafter.

Schier is saying that we don’t have to go to school to learn how to look at pictures. The evidence from child psychology and anthropology overwhelmingly supports this view.

In his book *A Psychology of Pictures*, John M. Kennedy cites an interesting 1960’s study by Hochberg and Brooks.

Hochberg and his wife raised their child with restricted exposure to any kind of picture. As far as possible, pictures were removed from the child’s vicinity. His parents even removed labels from cans and bottles. Sadly, there were no picture books for him to leaf through. A few decals and the occasional advertising billboard were the only pictorial displays the child encountered.

The child was never instructed in associations between words and pictures, never told that pictures represented anything, and was never read a story with illustrations in attendance.

Just before the child was two years old, at the time when he had a reasonably large vocabulary, a test was given. Line drawings....and black and white photographs were set in view, and the child was asked what they were. No photograph was shown before a line drawing of the object was offered, and the child’s responses were not corrected. The child labeled almost all the pictures correctly, whether they were photographs, complex line drawings with interior detail (like a doll), or simple outline drawings with minimal interior detail.

43 There will obviously be a stage where an infant or child can see pictures but not put a name to what is in the picture due to lack of vocabulary or lack of experience of seeing those kinds of things. Animals are a good example. It takes a few months to learn the names of major zoo animals, but even without the vocabulary to name the animal the child will still see an animal of some sort in the picture.


This rather extreme experiment demonstrates that, not only is it not necessary to train infants in the perception of pictures, it is not even necessary to expose them to pictures during their upbringing in order for them to be able to recognise what is depicted.

![Figure 8. An outline drawing that was correctly recognized by an infant who had never been taught the meaning of pictures.](image)

This is not to say that babies emerge from the womb with the capacity to recognise depictions – there is a developmental cycle which begins with the capacity to register moving objects, then static objects and then, when the infant is able to recognise the same object from different angles and in different guises, the infant can begin to recognise objects in depictions. As the child matures, the kinds of discriminations the child can make become more sophisticated particularly where it applies to judgements of depth.46

But what of the natives in remote parts of Africa, cited by Goodman 47, who failed to recognise the subjects of photographs and drawings? Aren’t they proof that the acquisition of pictorial recognition abilities is dependent on acculturation? There is some consensus amongst commentators that many of these studies were conducted “under less than ideal experimental conditions.”48 Much of the work

---


47 Goodman, Languages of Art, 15n.

done with in the 1940’s and 50’s on African subjects in remote areas is methodologically suspect. At best it can be said that the studies highlight cultural differences between peoples. Certainly they cannot be used to argue that training is required for people to see depicted content in photographs or drawings. More recent studies, such as those by Deregowski, indicate that pictorially innocent adult viewers, those who have never seen a drawing or photograph, can very quickly (in a matter of minutes), given a few clues, recognise what is depicted.

Kennedy cites a 1966 paper, “The Influence of Culture on Visual Perception” by Segall, Campbell and Herskovits, which concludes “one can regard the photograph as we use it as an arbitrary linguistic convention not shared by all peoples.” In fact, as we shall see, Segall, Campbell and Herskovits have merely interpreted the evidence to support their conventionalist view of how human beings come to recognise depictions. Kennedy notes that Herskovits’, in his 1948 study wrote, “A bush Negro woman turned a photograph this way and that, in an attempt to make sense out of the shadings of grey on the piece of paper she held.” But in a footnote Herskovits reveals “when the details of the photograph were pointed out (to the interested Bush woman) she was able to perceive the subject.” This supports Schier’s view that the uninitiated initially need some hints to recognise their first depictions, but once the first picture has been recognised, all subsequent pictures which the person see become intelligible.

Deregowski’s later work on the perception of depth cues in line drawings and photographs does indicate that cultural differences (particularly differences between the experience of city-dwellers and bush-dwellers) can affect the overall interpretation of a picture – but that is a far cry from failing to see people, animals, objects and buildings. Jones and Hagen in their review of cross-cultural studies of picture perception conclude that the “apprehension of pictures of familiar objects seems to be a universal ability even among people who have never before viewed a picture.” In fact, most of these studies found that the only

51 Kennedy, A Psychology of Picture Perception, 65.
52 Ibid., 69. What is revealing here is that Herskovits puts the uncomfortable truth (uncomfortable for Hersovits theory) that the woman had no trouble seeing the picture after a few hints, into a footnote.
53 Jones and Hagen, "A Perspective on Cross-Cultural Picture' Perception," 220.
problems which naïve viewers had with pictures concerned issues of spatial layout as opposed to object identification. Jones and Hagen add that pictures which use perspective are the most successful in specifying spatial layout.

The conventionalist account of depiction does not emerge well from these studies. All the evidence suggests that the ability to see familiar objects in pictures is an innate ability. It is an ability that transcends cultural differences and there is no evidence that depiction is based on a language-like symbol system.

In “What Goodman Should Have Said about Representation”, Douglas Arrell summarises the core problem with Goodman’s theory of pictorial representation:

Nelson Goodman’s theory of pictorial representation is the best known and most widely rejected feature of his aesthetics. His contention that representation is a form of denotation has achieved notoriety rather than acceptance. A survey of some forty of the articles and reviews that appeared in the wake of Languages of Art reveals that in about three-quarters of them his theory was a major topic of concern, and that overwhelmingly the concern was to refute it; indeed it is hard to find a clear-cut case of someone agreeing with it. There is a danger, I think, that Goodman may go down in the history of aesthetics, like many aestheticians before him, primarily as the inventor of a discredited theory – in this case the extreme “conventionalist” theory of representation – to be ceremonially refuted for the benefit of successive generations of aesthetics students.54

In Arrell’s view, it is Goodman’s allegiance to nominalism which forces him into the extreme view that pictorial representation is purely denotational. The essence of denotational representation is some sort of agreement or rule concerning how a symbol or symbols refer. Clearly such rules or conventions would need to be learned and, in the case of seeing depicted content, we have found no evidence that any such rules are required. Rob van Gerwen in his book *Art and Experience* comments:

Goodman argues for a conventionalist account of representation, while disregarding the fact that persons endowed with natural faculties and using their common sense hardly ever encounter problems with understanding an object as a depiction and recognizing what it is supposed to be a picture of.55

---

The conventionalist assertion that the capacity to interpret a picture is culture specific ignores the fact that, as van Gerwen points out, hardly anyone ever has any trouble seeing what pictures depict.

**Summary of Chapter One**

Our discussion of conventionalism and resemblance has generated four main conclusions:

- Degree of realism is not an important criterion for resemblance or recognition – we can recognise a child’s drawing or a cave painting as easily as a photograph.
- The style and projection system of a picture are largely irrelevant to what happens in the recognition moment.
- Human beings from any era and any culture can recognise all kinds of depictions without any training in pictorial conventions. We do not need to learn how to see depicted content in pictures in order to see pictorial content in that first moment of looking.
- The key problem for resemblance theory is explaining how a black and white sketch which bears very little resemblance to a three dimensional object can generate recognition through resemblance. The subsidiary issue is: how does the viewer discount accidents of style such as thick black outlines without knowledge of depictive conventions?

The ability of naïve viewers, such as Brooks’ child and African natives, to recognise line drawing with no previous exposure to pictures indicates that pictorial interpretation is an innate ability. It seems extremely unlikely that we used knowledge of a symbol system or depictive techniques to recognise trees and a building in Constable’s Ottawa study. It is more likely that, in that first recognition moment, at least, we used a native ability to see objects and scenes in pictures. It seems natural to appeal to the intuition that the feeling of resemblance that we feel when we look at Constable’s landscape is part of the recognition process. Unfortunately, the resemblance account is unable to explain how we can see resemblances in degraded sketches and smudgy cows that bear very little resemblance to what they depict. The conventionalist account easily explains this aspect of pictorial interpretation, but cannot explain how an interpretation can get started where the viewer has had no prior pictorial experience. We are looking for
a mechanism of recognition which does not involve having to learn about depictive techniques, and which can account for selective commitment and the ability to quickly recognise rough sketches.

If infants and naïve viewers can see familiar objects in line drawings and photographs without some kind of training about the symbology of the lines, tones or shapes, we can be sure that, as Rob van Gerwen notes, they see such things because they are endowed with natural recognition faculties and common sense, not because of some prior agreement on symbology. Flint Schier’s theory of “natural generativity” explains our capacity to see what is depicted in a picture in terms of an innate capacity to recognise things in the world. His theory addresses the shortcomings of conventionalist theories, and accounts for the universality of pictorial competence by locating it in basic recognitional abilities. Schier’s theory also resolves the central problems which we have encountered with resemblance theory such as the selective commitment problem.
Chapter Two

Schier, Natural Recognition Abilities and Common Sense

*The respect in which S resembles its depictum O is this: there is an overlap between the recognitional abilities triggered by S and O.* Flint Schier

In his book *Deeper into Pictures*, Flint Schier offers an account of the pictorial experience which highlights what is distinctive about depiction and therefore makes it different from other modes of representation. In particular, he rejects the paradigm of linguistics and focuses on identifying what is distinctive about how pictures refer to their subjects. In his view, genuine pictorial interpretation relies on natural recognition abilities alone.

He begins by “rusticating” the theories of depiction offered by, among others, Gombrich, Sartre, Wollheim, Walton and Goodman. These theories, he observes, variously “suggest that pictures look like their depicta, cause illusions of seeing their depicta, are modes of attending to their depicta, are seen as their depicta, and are make-believedly identical with their depicta.” Some of these theories, such as the “seeing-as” and “illusionist” theories, are often classified as resemblance theories because they ascribe a crucial role to resemblance. In fact, all theories of depiction, even the conventionalist theories, make some attempt to mollify our intuition that the resemblance between picture and object is crucial to the experience of looking at pictures. Not all theories, however, allocate an explanatory role to resemblance. That is, it is not always argued that resemblance has a functional role in the interpretation of the picture.

We saw that for the conventionalist, resemblance is a *by-product* of the interpretation of pictorial symbols. Pictures, for the conventionalist, only *seem* to resemble their depicta because of habit and the inculcation of pictorial techniques. In the case of the “illusionist” and “seeing-as” theories, when we look at a picture and experience pictorial resemblance, we see the object depicted while simultaneously or alternately being aware that we are looking at marks on a

---

57 Ibid., 31-32.
In Schier’s view, these so-called resemblance theories fail to establish an explanatory role for resemblance in the interpretation of a picture. The focus of these theories is on the resemblance experience rather than the role of resemblance in pictorial interpretation. In *Deeper into Pictures*, Schier seeks and develops an account of “how all and only pictorial interpretations are generated.” In his view, resemblance stems from the fact that “S and O trigger some of the same recognitional abilities.” Resemblance is therefore distinctive of the pictorial experience. Schier conjectures that:

there is a peculiar mode of coming to know the meaning or semantic content of pictures in terms of which it will prove possible to define depiction and pictorial experience.

The question his book asks is “What makes a symbol a picture?” or to put it in Schier’s terminology “What makes a symbol iconic?

**Convention Maketh Not the Icon**

Schier’s first point is that “convention alone doesn’t make something iconic”. For example, if an artist stipulates that when he paints a square it actually denotes a sphere, we still do not see the square as a sphere. It requires more than the stipulation of a denotational convention for a viewer to see a sphere depicted in a picture. It might be possible, in some symbol systems, such as Fauvism and Cubism, to stipulate that red denotes green, or that squares denote wheels but, in Schier’s view, these stipulations are unnatural. The techniques of Fauvism and Cubism are not playing by the natural rules of depiction. In this sense, Schier’s intuitions coincide with my observations, in the previous chapter, on the failure of Cubist techniques to enter the mainstream of realist depictive techniques. Schier comments that “convention maketh not the icon”,

---

58 Christopher Peacocke’s ‘perceptual theory’ of depiction explicitly rejects the idea that pictures actually resemble the objects they depict, but holds that our visual impressions of picture and object are similar. See Christopher Peacocke, "Depiction," *The Philosophical Review* 96, no. 3 (1987): 383-410.
59 Schier, *Deeper into Pictures: An Essay on Pictorial Representation*, 188.
60 Goodman sets out to do this in *Languages of Art*, but fails to make the case.
62 Ibid., 31.
63 Ibid., 41.
Manifesting the intention that a symbol shall signify O makes it signify O (in your idiolect at least). By contrast, manifesting the intention that it shall depict O doesn’t necessarily do the trick.  

What makes a symbol iconic, in Schier’s view, is precisely the fact that human beings arrive at an interpretation of what the symbol depicts without any prior agreement about what it is supposed to depict. We arrive at the interpretation “naturally” – not by virtue of conventions which stipulate the content of the icon. These observations on the “naturalness” of successful depiction, and the distinction between iconic and non-iconic representation, are the key to understanding Schier’s concept of “natural generativity”. He writes, “natural generativity is what makes a symbol (system) iconic” and argues that this is a crucial property of iconic modes of representation.

**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.

---

64 Ibid.
65 Ibid., 43.
66 Ibid., 49.
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.\(^{67}\) 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:

\[
\text{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?}\(^{68}\)
\]

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.\(^{69}\)

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:

\[^{67}\text{Pictorially identifying the Constable painting as a picture of Salisbury Cathedral, and not some other cathedral, can only happen if you either know what Salisbury Cathedral looks like in real life, or have seen other pictures of it. Natural generativity is therefore not necessarily the key to identifying it as Salisbury Cathedral. On the other hand, you don’t need prior knowledge to see it as a picture of a building surrounded by trees; that interpretation is caused through natural generativity.}\]

\[^{68}\text{Schier, Deeper into Pictures: An Essay on Pictorial Representation, 46.}\]

\[^{69}\text{Ibid., 186-87.}\]
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:

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?\(^{70}\)

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

\(^{70}\) Gombrich claims we are aware of both at the same time.
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?

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).

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 nowhere 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. 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.

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 required for S’s depicting O”. 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.

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.

---

71 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.
73 Ibid., 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.
74 Ibid., 137.
75 Ibid., 132.
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.

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.\(^76\)

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

\[^{76}\text{Ibid., 135.}\]
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**

It is notable that when Schier talks about “overlapping recognition abilities” he doesn’t specify exactly how many, or which, of our O-recognising abilities are engaged when we see S as a picture of O. This is not surprising considering that it is quite hard to specify which overlapping recognition abilities are triggered when, for example, we see an actual object in different lighting conditions or from different angles. For example, a cat presents different shapes to the eye as it moves and ducks in out of the shadows. What common properties does a cat coming toward you in the daylight and a cat moving away from you at night present to the eye? It is hard, without understanding how our visual system tracks objects, to be specific about how it might track the perception of objects in pictures. Schier maintains that there are strong parallels between a causal theory of perception and a causal theory of iconic reference:

A percept is about a particular object if it tracks that object, so that the percept would have been otherwise had the object been otherwise. Likewise, a depiction S of O tracks O in the sense that there are some features of O such that had they been otherwise, S would have been otherwise. Moreover, the function of a given percept is to track an object; of course the percept itself is not designed to track its object, but it is the product of a visual system that has the evolutionary function of producing percepts which track certain features of the environment. This is analogous to a camera. The mechanism has been designed to produce symbols (photographs) that track the properties of objects which cause or trigger the production of photographs. A photograph is of O because there are features of O such that had these features been different, that photograph would have been different. Finally, an icon not only tracks what it represents, but iconically tracks what it represents. In other words, an icon is something the function of which is to be such that the ability to
recognise the object which the icon tracks should yield the ability to
determine that it is that object and not some other that the icon tracks. Of
course, an icon will serve this particular function only if the function is
made manifest. It will not be made manifest if the symbol is too much like
what it symbolises, or too prone to induce an illusion of seeing what it
symbolises. So the interpreter must cotton on to the fact that S is a symbol.
A picture S of O must both trigger O-recognising abilities and give off
cues as of being something other than O. S must provoke contrary
recognition abilities.77

This long passage on how a percept tracks an object shows that what Schier has
in mind when he refers to recognition triggers is not something as naïve as a
particular colour tracking a particular colour and a particular shape tracking a
particular shape. He is aware that our visual system has evolved to track objects,
not particular colours or shapes. If we follow a white dog from a sunny position in
the garden into a room lit only by a light bulb, not only will we see the dog
constantly change shape as it presents different aspects of itself to us, the colour
of the dog also changes from a brightly lit bluish shade of white to a dimly lit
orange shade of white. A visual system that evolved to track a colour and shape,
as opposed to an object, would be confused by these changes and fail to track the
dog. What is clear from this is that the illusion problem, as it stands, betrays a
naïve idea of how recognition works. If recognition involved one-to-one matching
of properties from one view of an object to another, and the goal of a depiction (as
assumed by Goodman) was to match the properties of the picture as closely as
possible to the properties of the object depicted, perhaps there would be an
illusion problem – people would constantly mistake pictures for reality. In fact,
the human visual system is not a property matching system and the primary goal
of depiction is not illusion.

Schier not only maintains that a partial overlap in triggered recognition abilities
is required for interpretation; he also maintains that a partial as opposed to
complete, overlap is desirable. It is crucial to depiction that a picture should not
resemble the depictum enough to be mistaken for it. He comments:

The object which is made with the intention that it will deceive people into
thinking that it is O when it is not O is not an icon of O. Being a pear
decoy is just not the same sort of thing as being a representation of a pear.

77 Ibid., 194-95.
The deceptive function and the representational function are quite distinct.\(^78\)

In the case of the pear decoy, we can see that although there is a *prima facie* resemblance between the decoy and a real pear, when we apply Convention C we get a negative confirmation that the pear decoy is intended to be a depiction. In this case Schier’s theory agrees with Goodman’s that resemblance isn’t a sufficient condition for the depiction relation, but not for Goodman’s reasons.

By shifting resemblance into an almost pre-interpretational stage, Schier’s account segregates resemblance from the issues which seem to beset the illusionist and seeing- as theories, and makes a virtue of Goodmanesque objections. In Schier’s view, a purely pictorial interpretation of a picture is enabled by, and gets its power from, our natural ability to see things in the world. A picture by definition capitalises on those natural visual recognition abilities. There may be other aspects of the picture which affect our interpretation – a caption, or symbology – but these aspects are taken care of in Schier’s second stage. Seeing an object in a picture in Schier’s first stage is only interpretation in the sense that looking at an object in the world and identifying it as a tree or a church is interpretation. Schier’s second stage of interpretation involves the conventions that surround the activity of depiction. It is when theorists consider what explanatory role resemblance plays in this second stage of interpretation that they get in a tangle with selective commitment. However, in Schier’s scheme, resemblance (the triggering of our visual recognition capacities) has already happened by then and doesn’t require an explanation in terms of the conventions of depiction. We should remember here that it is not resemblance that causes the initial interpretation; it is the triggering of our natural recognition capacities which causes us to see a tree or a building. The experience of resemblance comes when we identify features of the tree such as leafiness or gnarliness.

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

\(^78\) Ibid., 130.
study (fig. 1) 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.

**Selective Commitment, Frame Knowledge and Communicational Cooperation**

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.79

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.

---

79 Ibid., 168.
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 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.80

This observation, and Schier’s emphasis on the term “pictorial”, highlights the chief problem with the conventionalist argument and semiological theories of pictorial representation. Given what we know about the universality of pictorial interpretation abilities and the fact that no learning is required to interpret a photograph or a line drawing, it is simply impossible that one could arrive at a pictorial interpretation through the acquisition of language-like rules of interpretation. In the case of natural language, Chomsky’s “poverty of stimulus” argument hinges on the observation that children do not get enough feedback about their language skills to develop an understanding of the symbolic aspect of language as quickly as they do. The “poverty of stimulus” argument is far stronger in the case of depiction than it is for spoken language. Schier’s theory, on the other hand, accounts for universality and no-learning, and suggests a mechanism whereby visual recognition abilities combine with the communicative conventions of depiction to generate a pictorial interpretation of the content of a depiction. Schier’s theory does exactly the kind of work that is needed to build a comprehensive theory of depiction based on recognition abilities.

80 Ibid., 171-72.
The bankruptcy of the conventionalist account can be seen if we imagine what a viewer might be able to see in Constable’s Ottawa study, equipped only with the kinds of abilities and knowledge which Goodman allows. If we then compare it to a scenario where the viewer is equipped with “natural recognition abilities” we can see the advantages of Schier’s approach.

**A Viewer Equipped with Conventionalist Interpretative Abilities Looks at the Constable Landscape (or is it a Pink Elephant?)**

Naïve viewers, such as our 18-month old child and Kalahari bushman, equipped with the knowledge and abilities allowed by conventionalists will see nothing but a disorganised mess of colour in the Constable painting. Such naïve viewers have never seen a picture before and therefore will have to undergo training in picture interpretation to see what it depicts. Similarly, viewers from remote areas in China who have never seen a Western painting or photograph will also have difficulty interpreting the image, because it is constructed using Western perspective rules, featuring curiously shaped grey lines (possibly a structure) and green and brown wiry shapes (possibly trees). However, these are not the only “conventionalist viewers” who will have difficulty interpreting the picture. Goodman himself claims that with “suitable principles of correlation, Constable’s landscape painting could provide an enormous amount of information about a pink elephant.”

How then, can a conventionalist viewer see the trees in the painting? All one needs to do, according to the conventionalist, is hit on the right schema by which to interpret the painting. This is complicated by the fact that, according to Goodman, there are an infinite number of schemata which might be used by the artist.

Almost any picture may represent almost anything; that is, given picture and object there is usually a system of representation, a plan of correlation, under which the picture represents the object. How correct the picture is under that system depends on how accurate is the information about the object that is obtained by reading the picture according to that system. But how literal or realistic the picture is depends on how standard the system is. If representation is a matter of choice and correctness a matter of information, realism is a matter of habit.

---

82 This is a bit like buying a DVD and knowing that in order to watch it one needs a DVD player to decode it and not a CD or video player.
So how does the conventionalist viewer know whether a picture is an abstract expressionist painting, a pink elephant or a landscape? How can he/she discover, just by looking at the picture, which is the correct plan of correlation between the shapes and colours in the picture and objects in the world? There must be a clue in the painting which will enable the conventionalist viewer to kick-start the interpretation. Perhaps one can assume that the colours in the painting correspond to the colours of objects depicted. After all it seems logical that green paint depicts green things. Goodman says “No”:

Consider a realistic picture, painted in ordinary perspective and normal color, and a second picture just like the first except that the perspective is reversed and each color is replaced by its complementary. The second, appropriately interpreted, yields exactly the same information as the first.84

There is some sleight-of-hand in this comment. What does it mean to say that two sets of data “appropriately interpreted” yield the same information?

Consider these numbers:

6 and 7
21 and 2

Both sets yield the same number if “appropriately interpreted”.

Consider these words:

Cerrado
Chiuso

“Appropriately interpreted” they both mean the same thing.

Consider these two pictures: one is a picture of a Constable painting the other is its negative.

Figure 21  Constable’s painting of Salisbury Cathedral rendered in negative (left) and positive (right).

---

84 Ibid., 35.
What does it mean to say that “appropriately interpreted” they yield the same information about Salisbury Cathedral? Goodman is arguing that realism cannot be defined as yielding relatively more information about a subject than other depictive techniques because the negative has as much information about the cathedral as the positive. He claims that “appropriately interpreted” they yield exactly the same information as each other. Goodman is saying that all we have to do is reverse the colours of the trees in the negative picture and we will get a correct interpretation. Try it now. Look at the negative Constable in figure 21 and interpret the correct colours for the trees, building and sky. You’re not allowed to just imagine what leaves and skies look like – you have to get it from the negative picture. It is clearly impossible, and Schier is right. It doesn’t matter how much you know about the techniques of depiction you cannot see the correct interpretation through a mental decoding operation. Schier’s important point is that if you can’t see the correct interpretation it isn’t a pictorial interpretation.

The conventionalist comeback to this argument is that if our society only used negative images we wouldn’t need positive ones. Negative images would be considered the realist convention. It should be noted here that a camera is a colour-reversal upside-down image generator. If colour-reversed images were just as natural as colour-normal images, photographers wouldn’t need to print contact sheets to check their photographs – they could tell which were the good shots from the negatives. Printing the positive photograph from the negative is the most tedious and time-consuming part of photography. Most photographers would eliminate this step if they could. (In fact, digital photography eliminates the negative processing stage.) It doesn’t matter how long you have been a photographer you will never be able to judge the quality of a colour photograph from a negative.

Goodman appeals to the fact that our eyes can adapt to all kinds of transforming glasses to support his view that colour-inversion is something our

---

85 These are actually photographs of Constable’s painting of Salisbury Cathedral but the way that photographs relate to paintings is another issue. Furthermore, the negative image is not of the same painting as the positive image. The positive is from the Frick Collection the negative from the Met. The point of the juxtaposition is that we have difficulty even recognising which image the negative has reversed – let alone how its palette varies.

86 Unless you are Ansell Adams

87 Obviously you can ascertain whether it is in focus and you can make some judgements about composition, but to choose the good shots from a roll you need a contact print. As you can see from the above example Constable’s careful composition is unbalanced in the negative.
visual systems can easily accomplish for us. He comments, “Upside-down pictures, if we are given the clue, will provide the same information as right-side up ones; and, indeed, a man can adjust rather quickly to inverting spectacles.”

This comment reveals how thoroughly Goodman misconstrues the problem. It is one thing to invert one’s vision, it is quite another to invert an individual picture one is looking at. If you turn your television upside down and watch it that way for a year you will never adjust to it because when you’re not actually watching the television, for example when you are walking or cooking, you need to look at the world the right way up. It is one thing to put devices over a person’s eyes and thereby manipulate the reflexes of their visual system, and quite another to assume that a person can intellectually adjust to upside down photographs and still be able to make the kind of discriminations that they make with photographs that are right-side-up. Studies of the kinds of discriminations one can make when presented with inverted images indicate that the visual system is very sensitive to the orientation of features of an image and that inversion disturbs our ability to discern shape, balance and other important aspects of an image.

These speculations have got us no closer to discovering how the conventionalist viewer can kick-start the interpretation process. There appear to be no criteria for determining which is the correct “plan of correlation”. Moreover if our conventionalist canvasses a few interpretational techniques to see how they pan out, they have no criteria for determining when they have hit on the correct schema.

In the case of the numbers, you no doubt found that you had to multiply to get the correct interpretation. Probably you canvassed adding, subtracting and finally hit upon multiplication, and it was the fact that both sets of figures generated 42 that convinced you that you were right. Now maybe I had something else in mind when I suggested that you interpret them as yielding the same number. There are many ways to generate the same number from the two sets; multiplying them together is the most obvious option. You don’t know that 42 is the correct interpretation, but the fact that both sets of numbers yield 42 from a simple

multiplication makes it a very attractive solution – but you don’t know if that’s the correct interpretation because you can’t read my mind.

In the case of the words, you need to know a little Spanish and Italian. And if you do, you know that they both mean closed. That is the correct interpretation, and you don’t need to read my mind to figure that out, you just need to know the languages.

In both cases, you had to consider what operation on the data was required to get the correct information. You couldn’t have done this without knowledge of arithmetic and languages. In the case of the numbers, multiplication was only one method amongst many of getting the same result – and only I know if you are right. In Goodman’s scenario there are also many ways of getting an interpretation. What kind of knowledge do we need in the case of the positive and negative pictures? How do we know we have arrived at the correct interpretation in the case of the pictures? If we cast around with various operations and interpretations looking for the correct one, how do we know when we have hit on the correct one? If our conventionalist viewer, following Goodman, sees a pink elephant in the Constable, how does he/she know this is right? It would seem that only the artist can know which is the correct interpretation and the artist is not always available to consult. The conventionalist account lacks standards of correctness. Even subsequent conventional interpretations (which actually involve pictorial symbology such as in the Piero della Francesca (Fig. 2)) need pre-communicated standards of correctness to facilitate validation. Schier maintains that our natural recognition abilities enable us to determine the correct interpretation and argues that, even if conventionalism was true, it would also rely on natural recognition abilities. He therefore concludes that the conventional interpretation is parasitic on the natural one. Fauvist paintings are good example of depictions with features which require both natural interpretation and interpretation according to technical conventions.

Schier’s analysis of colour in Fauvist painting shows that the conventionalist can never know whether they have a correct interpretation unless they have a colour correct, naturally generated interpretation of the picture to kick-start their interpretative process.

In Fauvist paintings red may well stand for green, and objects in a room may be in an unidentifiable perspective projection. Twentieth-century art is replete
with such examples where artists have pushed the bounds of depiction through innovative techniques. In such cases we cannot fall back on the assumption that green is green and that normal perspective rules are in play. Does this give support to Goodman’s argument that we cannot do so in the case of this Constable painting? The conventionalist account gives us no grounds for preferring the realist interpretation, and maintains that by interpreting the painting as a building surrounded by trees we are merely decoding according to the dominant Western convention. Notice that the conventionalist isn’t saying that it is impossible to interpret the painting this way, just that it is one of many possible ways of interpreting the painting. The claim is that it could have been painted by a Fauvist who was actually depicting red-leaved trees, and that the grey shapes in the centre of the painting are a schematised version of a pink elephant. If we interpret paintings according to the representational customs of realism we will not see these other interpretations.

None of this tells us how the viewer knows which is the correct interpretation. Neither does it tell us why we intuitively prefer the interpretation where the green paint smudges depict green leaves. In Wollheim’s view, a theory of depiction needs to account for the fact that a colour-reversed system requires more rules for decoding than a colour-normal system. Schier agrees, and asks how it is possible for someone to ascertain which is the correct interpretation when “an icon of a tree with red leaves, one which depicts a tree as having red leaves, may be perceptually identically with a colour-reversed icon which represents the tree as having green leaves by using the inverted convention that red pigment means ‘green’”?90 The capacity to solve this puzzle is, Schier believes, the acid-test of a theory of depiction because a successful explanation will enable us to discriminate between iconic systems and systems that are merely based on codes. An account that fails to do this fails to be an account of depiction. He notes that manifesting the intention that red denotes green does not necessarily mean that red depicts green. Depiction involves more than stipulation and Schier’s account of colour reversal highlights why this is the case.

Goodman suggests that the reason we interpret realist pictures the way we do is entrenchment. The convention that green splodges denotes green leaves is an

90 Schier, Deeper into Pictures: An Essay on Pictorial Representation, 128.
entrenched convention and no more natural than the convention that red splodges denote green leaves. If, for example, green paint was so expensive, or socially repugnant that it wasn’t used to depict green trees, we could imagine a situation where perhaps red paint splodges became entrenched as the way to depict green leaves in realist landscapes. How, in this society, does one determine that an artist is depicting a tree which actually has red leaves? How does the viewer know to omit the red-means-green step? In Schier’s view, the viewer can only be alerted if there is a specific indication, either via a caption or some other method, whereby the artist communicates that the red-means-green convention doesn’t apply. This alert effectively says “use your natural colour vision to interpret the colour of the leaves”. Because the viewer naturally sees red paint as depicting red leaves the red-means-green rule needs to be revoked in order for the person to see the depiction as normal. Thus there is a colour-normal interpretation, and the colour-reversal rule adds a step in a natural interpretation of the painting. This extra-step needs to be invoked even in the case of the complete symbol colour-invert. The person for whom green-symbols mean red and red-symbols mean green. This is because the symbol colour-invert only applies the colour-inversion when he/she looks at a picture. When the symbol colour-invert sees a tree with green leaves in real life, the leaves are green not red. So in both cases Wollheim’s view that symbol colour-inversion is unnatural is vindicated. In each case the natural interpretation of the colour of the icon needs to be inverted later to get the correct interpretation. Hence such systems are less natural than a system which relies on natural recognition abilities where green smudges depict green leaves because green smudges trigger green leaf recognition abilities. Inversion pre-supposes natural generativity and can’t be used to cast doubt on the role of natural generativity in kick-starting a pictorial interpretation.

In short, the Constable painting resembles what it depicts in point of colour and we have no reason to believe that Constable intended otherwise. Schier comments:

The colour-reversed system is parasitic on an iconic system; it is just a bit of conventional vocabulary added to an otherwise iconic system. The point remains that to interpret a colour-reversed icon one must remember, as it were, an item of vocabulary – that red pigment means ‘green’ and so on. 91

91 Ibid., 132.
But before one applies the inversion rule one needs to have already naturally generated a colour-normal interpretation.

If colour-reversed icons did not to this extent rely on the naturally generated interpretation, if only to negate it, the conventions for interpreting them would be unmanageably complex. Thus, Goodman is quite wrong to suppose that a reversed system is just as ‘natural’ as the normal one once you have got the hang of it. It demonstrably involves more labour, in that one has to decode it; moreover......the colour-reversal conventions are parasitic on normal ones.92

The conventionalist explanation for how the viewer kick-starts the interpretation of the Constable is to assume that the viewer adopts the realist interpretation because it is the most common interpretation, not because we just can’t help seeing the picture the way we do – colour-normal. In the case of the Fauvist picture this is not clearly the case. But the above argument show that if our conventionalist begins the interpretation of the Constable assuming a Fauvist palette, rather than a natural palette, the reference point for correctness will always be natural recognition abilities with respect to colour. The viewer interprets the picture in a realist manner because the viewer recognises the colours in the painting in exactly the same way as he/she recognises colours in the real world.

The conventionalist cannot explain the initial recognition moment without natural recognition abilities and cannot explain how a viewer knows which is the correct interpretation without the viewer having acquired some standards of correctness about the particular style of the depiction. This point is significant not just because a conventionalist account cannot explain how a pictorial interpretation gets started, but highlights the fact that conventions of style, technique, projection and symbology are post-interpretation issues.

**A Viewer Equipped with Natural Recognition Abilities Looks at the Constable Landscape**

When we imagine a viewer, equipped with the kind of abilities which Schier allows, looking at the Constable we run into none of the above problems. Equipped with natural recognition abilities that person will see a building surrounded by trees and asks themself whether Convention C is in place. The viewer asks:

92 Ibid., 133.
Did the painter intend me to see this picture as signifying the naturally generated content “building surrounded by trees”?

If there is no reason to conclude that the painting, or parts of it, aren’t signifying non-naturally generated content, the answer is “yes” and Convention C is in place. If the trees had bright-red leaves the viewer might conclude that the answer was “No” for the red leaves, but “Yes” for the other parts of the picture.

Equipped with the Schier-theory, a viewer sees smudges in the foreground that look a bit like cows and concludes that they were probably intended by the artist to be interpreted as cows. Our frame knowledge tells us that cathedrals aren’t usually built with sloping walls so we know that what is depicted is a cathedral with vertical walls.

Thus our Schier-equipped viewer is able to generate an interpretation using natural recognition abilities and discriminate deviant or Fauvist cases without any need to read the mind of the artist. The viewer assumes that if the picture triggers natural recognition abilities then it was intended to trigger those abilities and generate a natural interpretation.

Thus, to review our two Constable painting scenarios – the conventional and the Schier.

In the conventional scenario we found that there was a narrow class of persons able to interpret the painting as depicting a building and trees, and these people must have had a pictorial education in the schemata of Western painting and photography. These people have become habituated to seeing buildings and trees depicted like this, and for them the painting is more-or-less realist. They report that they see a resemblance between Constable’s trees and trees in real life. When pressed on the issue, they admit that they can’t see many actual leaves depicted in the painting and that the branches are just brown smudges. They do, however, admit that if you squint a little the painting seems to have a lot of green and brown tree-like shapes. But, they add, this is merely a realist interpretation, and that if you interpret it according to a non-realist schema a pink elephant can be seen grazing amongst the trees.

With the Schier scenario we found that any viewer equipped with Schier’s theory – infants, Chinese, Westerners – in fact, anyone equipped with primate recognition abilities, could see a building surrounded by trees. When pressed, they
said that the picture resembled a country scene in colour and line, and they also suggested that squinting at the picture made it seem more realistic.

It is clear that Schier’s explanation is not only simpler than the conventional explanation, but the conventionalist explanation is unworkable. Conventionalist theory requires natural recognition abilities in the viewer in order to kick-start the interpretation process and provide criteria of correctness. Thus, the conventionalist superstructure is superfluous to a theory of how we see depicted content.

**Summary of Chapter Two**

Schier’s recognition-based theory of depiction provides a broad theoretical framework which can explain how we recognise depicted content in pictures. In particular, Schier establishes that a conventional interpretation is always parasitic on natural recognition abilities. That is, conventions play no role in the initial “raw interpretation” of a picture. Schier’s theory fits the facts of depiction more closely than other theories and satisfies our intuitions about resemblance and colour reversal. Furthermore it:

- accounts for how everyone, including bushmen, infants and conventionalists, see trees and not pink elephants because they have natural recognition abilities;
- accounts for how, by applying Convention C, we ascertain whether depictive conventions are in play, and by using our frame knowledge determine which aspects of the picture are committed to pictorial representation;
- accounts for how we don’t need to know the artist’s actual intention, just the artist’s general intention to depict something whose content will be amenable to people with natural recognition abilities.

Thus, Schier’s theory accounts for:

- the no-learning issue;
- the universality of depiction issue;
- the selective commitment of depiction techniques;

and

- the riddle of how we ascertain the artist’s intention.

The lack of any evidence that anyone anywhere has to be taught to see depicted content is enough on its own to sink full-blooded conventionalist theory. In
addition, by attending to many of the facts about interpreting pictures which con
tventionalist have ignored or misrepresented, “natural generativity” provides exactly the kinds of explanations required for a robust recognition-based theory of depiction.

But what of the other theories of depiction which Schier “rusticates” in the introduction to his book? What does “natural generativity” offer that cannot be obtained from resemblance theories such as, for example, Wollheim’s seeing-in theory? What is the difference between Schier’s recognition-based theory and simpler resemblance theories? One of the curiosities about Deeper into Pictures is that it is largely devoted to rebuttals of conventionalist attacks on resemblance theories.93 However, “natural generativity” is not a resemblance theory. On the other hand, given that Schier has fought off the objections to resemblance theory, it would seem that he has strengthened its case. Thus resemblance theorists use many of Schier’s arguments without buying into his cognitively-based premise. One such theorist is Robert Hopkins, who adopts many of Schier’s arguments whilst eschewing any allegiance to cognitive theory. In the next section we will see how Schier’s theory fares against modern resemblance theory, and simultaneously tease out the difference between a recognition-based theory and resemblance theories.

---

93 It is also interesting that conventionalists used these attacks on resemblance theory as a kind of smokescreen. They probably need to do this because in the absence of any evidence that we need to learn a visual language to interpret pictures, conventionalist arguments look rather empty. Indeed, despite the fact that humanities department have been dominated by various shades of conventionalism in the form of semiology and structuralist theory for the last 30 years, it hard to find a full-blooded conventionalist.
Chapter Three

Recognition-based Theories vs Resemblance Theories

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.

Flint Schier

Robert Hopkins, in his essay “El Greco’s Eyesight”, argues that Schier’s account does not sufficiently make the case that, given his conditions for depiction, the best explanation is that “the very recognitional capacities engaged in perceiving things are also involved in interpreting pictures of those things.” In short, Hopkins does not believe that Schier has shown that “pictures engage our capacities to recognize the items depicted”.

He summarizes Schier’s position thus:

Schier attempts to understand pictorial representation through the way we interpret it. His main claim is that grasping what a picture represents requires distinctive resources. More precisely, he thinks that the following conditionals are definitive of pictorial interpretation:

SC. If a subject S has general competence in the pictorial system of which a picture p is a member, and S has the ability to recognize visually an object o, then S is able to interpret p as of o.

NC. If S is able to interpret p as of o, then S has general competence in the pictorial system of which p is a member, and S has the ability to recognize visually o.

In other words, given basic competence in the pictorial system, the ability to recognize something is both necessary, by (NC), and sufficient, by (SC), for the ability to interpret a picture of that thing. In contrast, understanding convention-governed representations, such as expressions in a natural language, may often not require the ability to recognize the thing described, and will always require something more, viz., knowledge of the relevant content assigning conventions. Thus, Schier suggests, we can understand the peculiarly pictorial form of representation as precisely

96 Ibid.
that form which can be interpreted pictorially, that is, which can be interpreted given just the resources (NC) and (SC) describe.97

The core of Hopkins’ objection is that the ability to visually recognise an object is also required for other forms of representation. He uses the example of a “describing game” in which you try to guess the identity of a household object or a person from verbal descriptions of their appearance. The winner of this game will succeed only if he/she can recognise the object in question. Crucially, he points out, one does not assume that “representation and represented thing engage significantly the same processing operations”98. He continues:

That we might use one such feature to identify, in visual recognition, a particular object does not show that we in fact do so. No more, then, does the fact that we use one such way to identify something in a representation show that we use the same means to identify it in the flesh.99

This objection challenges Schier’s hypothesis that pictorial interpretation invokes overlapping natural recognition abilities. Hopkins is arguing that we do not assume that the verbal visual descriptions in the description game require significant overlapping recognition abilities to facilitate an interpretation. He therefore asks why we should do so in the case of a pictorial representation. Hopkins concedes that cognitive processing might be matched at a high level, for example, the processing involved in final stage of recognizing an object might be matched by the processing at the moment when a representation of that object is identified, but is sceptical that this is the case, or could be shown to be the case.100

Hopkins’ argument suggests that even if we could show, through an appeal to cognitive science or some other discipline, that overlapping recognition capacities are engaged by picture and depictum, then the evidence would never be more that circumstantial. In a discussion with Jeffrey T. Dean on the subject of visualizing, Hopkins comments that:

appeal to the empirical disciplines, in the form in which we now know them, seems to me to be of necessarily limited use. Not that such disciplines can never illuminate phenomenology. On the contrary, there are concrete instances of their having done so. But of any discovery at the level of processing, there is always a farther question: what, if any, phenomenological dimension there is to the facts thus uncovered. This

97 Ibid.: 446.
98 Ibid.: 449.
100 Curiously, neurological studies confirm that there are neurons which fire when a particular object, picture or letter string are presented to view. We will examine this issue in a later chapter.
farther question needs answering by attention to the phenomenon – here visualizing – “as experienced”. So, as far as phenomenology goes, the contribution of the empirical disciplines, at least in their current state of development, is heuristic. They provide clues as to where to look for interesting aspects of phenomenology.....Perhaps when we visualize we deploy some of the same cognitive resources as when we see, but is overlap in that respect, or even just overlap to a similar degree, necessary if any mental state, of any creature, however structured, is to count as visual imagining?\textsuperscript{101}

Hopkins is arguing that the best cognitive science can do is to develop working hypotheses for philosophers. It can suggest explanations but not provide them. I’m skeptical that twenty-first century philosophy which concerns itself with perception can draw such a rigid line between philosophical enquiry and cognitive science. It will be clear, from the following analysis of his position, that he fails to deliver the goods at crucial points because he is unprepared to step over his imagined dividing line between the disciplines.

Nevertheless, Hopkins criticisms have set us two challenges:

1. To show that Schier’s account is not vulnerable to alternative theories of depiction.
2. To establish, through empirical, psychological or philosophical means, that pictures provide alternative causes for something like the same visual effects as those caused by what they represent.

\textbf{Is Schier’s account vulnerable to alternate theories of depiction?}

There are many theories of depiction which might present a challenge to the Schier hypothesis, including illusionism, seeing-as, mimesis, visualizing, and make-believe. However, it is possible to group theories of depiction into three main approaches:

- **Conventionalism** – interpreting a picture is reading a code;
- **Resemblance theories** – there is a similarity between picture and object;
- **Recognition theories** – pictures provide alternate causes for triggering natural recognition abilities.

These theories are available in various flavours, strengths and mixtures. When Hopkins says that Schier is vulnerable to alternate theories of depiction, he is really claiming that Schier is vulnerable to Hopkins’ own account of depiction. Hopkins attacks most of these alternate accounts for, in some way, failing to explain crucial features of depiction. We have established that the conventionalist case flies in the face of the evidence, and Schier’s hypothesis is one of the most complete philosophical accounts of recognition theory as it applies to depiction. Resemblance theories, therefore, remain the most significant alternate theories.

Resemblance Theories

Some flavours of resemblance theory, particularly experience resemblance, appear to get a boost from Schier in that he argues that the experience of resemblance is the result of overlapping recognition cues. Thus if a given resemblance theorist is agnostic about the underlying causes of the experience of resemblance, whether they be psychological, physiological or neurological, it could be argued that “natural generativity” is compatible with that particular flavour of resemblance theory. However, a theory which merely shows that we experience resemblance when we interpret a picture, does not explain how we interpret pictures. Gavin McIntosh makes this point in his essay “Depiction Unexplained”103. He argues that both Hopkins and Peacocke assume the role of resemblance in pictorial realism, rather than explain it. He comments:

What we want from a theory of depiction is an explanation of the fact that the particular method we employ for projecting the three-dimensional world onto a two-dimensional plane results in pictures we find convincing.104

He concludes that, not only are their accounts fatally flawed, but their whole approach is mistaken.

Earlier we saw that several of Goodman’s objections to the role of resemblance in pictorial interpretation failed because he assumed that resemblance was a sufficient condition. Of course, nobody regards resemblance as a sufficient condition for the depiction relation. But, presumably, all theories of resemblance

---

102 See Mark Rollins, "Pictorial Representation: When Cognitive Science Meets Aesthetics," *Philosophical Psychology* 12, no. 4 (1999), for an interesting taxonomy of theories of depiction, especially as they relate to theories of visual cognition.
104 Ibid.: 287.
Recognition-based Theories vs Resemblance Theories

must include resemblance of some kind as a necessary condition. Not all theories ascribe the same role to resemblance. In some theories the experience of resemblance between picture and object is crucial (experience resemblance). Other theories stress the actual sharing of physical properties between picture and object (resemblance realism).

Schier’s hypothesis may support experience resemblance theory because advocates of this theory maintain that the crucial feature of depiction is that the experience of looking at the picture in some way replicates or evokes the experience of looking at the object. Walton’s make-believe theory, Sartre’s visualization theory, Wollheim’s seeing-in, Gombrich’s illusionism and Peacocke’s visual field theory, all rely on the notion that some kind of experienced resemblance is involved in interpreting a picture.

One of these experience resemblists, Christopher Peacocke, argues that pictures don’t actually resemble what they depict, but argues that the visual impressions we have of a picture are similar to the visual impressions we have of the object it depicts. This notion of a “visual impression” requires some unpacking, but what it boils down to is that when we look at a picture we have something like the same visual experience as when we see the object it depicts. This form of resemblance theory seems largely compatible with Schier’s theory given that neither specifies the underlying causes of the experience of resemblance. Schier simply says “The respect in which S resembles its depictum O is this: there is an overlap between the recognitional abilities triggered by S and O.” Schier is reticent about committing himself on exactly what these overlapping recognition abilities might be. We will analyse the reasons for Schier’s reticence in a later section. Peacocke’s reasons for failing to specify underlying causes are clear. He states:

A philosophical paper on depiction is no place for speculation on the empirical issue of the nature of information-processing underlying the perception of pictures. But it would certainly be a criticism of the philosophical account if it left it hard to see how empirical conditions relevant to the psychology of picture perception could possibly be relevant: the philosophical account must dovetail with the psychological. It can hardly be irrelevant to the perception of a relatively realist picture that the optical array reaching the eye of one who perceives it bears certain

105 I must thank Gavin McIntosh for this paraphrase of Peacocke’s position. Ibid.: 280.
systematic spatial similarities to the array reaching the eye of someone really perceiving a scene of the sort depicted. ¹⁰⁷

Peacocke refers to the visual experience of shape in the optical array as “visual-field shapé”. He uses the notation “shapé” to distinguish this property of the visual field from two-dimensional shape. Peacocke’s is a relatively common strategy for resemblance theorists who don’t want to get bogged down in the physiology and neurology of vision. The argument runs something like:

- Stimulation of the retina by the picture is equivalent in some ways to the stimulation of the retina by the object – hence resemblance.

This is not a bad strategy. After all, a good photograph under controlled conditions can emulate the experience of seeing a real object. It seems quite plausible that this could be explained by the fact that the retina is stimulated in ways similar to when stimulated by the actual object. In this case, one doesn’t need to claim that the picture and object actually share properties; they merely present themselves to the retina in such a way that one could say that there must be a similar pattern of stimulation. Of course, in order for this strategy to work it needs to be shown that 1) this is the case, and 2) that the similarity is registered by the visual system. Peacocke fails to show either of these things.

Thus, on the face of it, there appears to be some broad compatibility in the approach of Schier and the experience resemblists. Schier suspects, however, that resemblance realists would not be happy with his account and would demand evidence of a more fundamental similarity between S and O. This would need to be a similarity that is independent of recognitional abilities. He imagines the resemblance realist arguing thus:

It simply must be a fact that S and O are alike: and we must be able to state the likeness in a way which is independent of the perceiver’s recognitional capacities in such a way that these likenesses can be invoked to explain the perceiver’s recognitional propensities. Once this project is carried out, we will see likeness as a deep fact in pictorial representation, natural generativity being simply the consequence of this deep fact. ¹⁰⁸

Schier argues that even if we could find an independently specifiable similarity such as colour “this does not mean that recognitional similarities can always be

¹⁰⁷ Peacocke, "Depiction," 404.
reduced to or eliminated in favour of any deeper similarities.” This argument is reminiscent of Hopkins’ own argument against Schier:

There are many features of the things around us which individually suffice to differentiate them from their neighbours. That we might use one such feature to identify, in visual recognition, a particular object does not show that we in fact do so.

It is worth remembering here that we are talking about visual resemblance – if something looks like something else it resembles it. Using Peacocke’s terminology, we might say that the visual impressions of the outline shapes of a beach ball and a picture of a beach ball look circular as opposed to spherical. Schier is asking what difference it would make if the beach ball and its picture actually were circular. Clearly the beach ball is spherical and a picture is not, but if this is not observable what difference can it make to resemblance? Why should object and picture need to share properties they actually possess, not the properties which it looks like they both possess?

Once we begin to speculate on the properties object and picture might share that might be considered suitable candidates for realist resemblance, it is clear that the main impediment to realist resemblance is that objects are three-dimensional and pictures of them are two-dimensional. It is hard to argue for shape resemblance if one is looking for actual property sharing. For example, the property of sphericity had by the actual beach ball is not available in a two-dimensional depiction of a beach ball, and so one cannot claim that ball and picture share the property of sphericity.

The ball and picture present to the eye as circular, but this does not amount to sharing non-dispositional properties. They need to at least share the circularity property, or something like it, to satisfy realist resemblance relation. This is indeed the ploy which Hopkins uses. Hopkins argues, against Peacocke’s concept of “visual field shape”, that what Hopkins calls “outline shape resemblance” is a geometrical reality, not something we merely experience:

Visual field shape is the notion of a property of visual experience rather than of a property of objects which may itself be seen. It cannot therefore be identified with outline shape. Nonetheless, both notions yield characterizations of an aspect of our visual experience. For although

109 Ibid., 188.
outline shape is a property of objects, not of experiences, an experience may be one of perception of outline shape.\footnote{Hopkins is not happy for resemblance to be merely experiential – it has to be caused by something like shared properties – in Hopkins’ case geometrical relations between object and picture. Outline shape is a property shared by picture and object \textit{relative to a point} and not just relative to the eye of the viewer. Despite the fact that he denigrates resemblance theories, such as Peacocke’s which invoke mere “visual experience”, Hopkins characterizes his theory as a development of Wollheim’s seeing-in theory which, as we will see, is essentially an experience resemblance theory.}

Hopkins describes his approach to the problem of depiction in these terms:

\begin{quote}
We are attempting to construct a feasible example of the experiential approach to the topic. That approach takes a certain experience to be essential to picturing. Our view must provide characterization of the crucial experience, seeing-in.\footnote{Hopkins argues that his concept of outline shape goes some way towards this characterization, but adds that seeing the shape of something in a surface is not definitive of depiction. Seeing-in requires the concept of a “standard of correctness” in order to be a complete theory of depiction. If you see the Virgin Mary in a piece of cheese-toast\footnote{see \url{http://news.bbc.co.uk/2/hi/americas/4034787.stm} “’Virgin Mary’ toast fetches $28,000”} that does not mean she is depicted there. Strictly speaking, someone should have shaped the cheese for it to be a depiction.}
\end{quote}

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure22.png}
\caption{‘Virgin Mary’ Toast}
\end{figure}

\footnotetext[2]{Ibid., 71.}
\footnotetext[3]{see \url{http://news.bbc.co.uk/2/hi/americas/4034787.stm} “’Virgin Mary’ toast fetches $28,000”}
The point is that both actual shared properties and experienced resemblance are not enough to explain the depiction relation. If resemblists, such as Hopkins could, for example, ascribe an interpretive role to resemblance based on shared properties, that would seem to make the theory more attractive than Schier’s for two reasons:

- It would be simpler, intuitive, and would avoid the appeal to cognitive science.
- It would seem to ground resemblance, and the depiction relation, in something more than accidents of our perceptual apparatus.

On the other hand, if we can eliminate the appeal of this type of resemblance realism we will have established that there are no alternative theories of depiction which seriously challenge the recognition based account.

One problem with Schier’s account is that it is incomplete in crucial ways – in particular, Schier is evasive on the nature of recognition abilities. Hopkins is quick to point out these shortcomings. Thus an analysis of Hopkins’ position and his arguments against Schier will serve to highlight areas in which Schier’s theory requires further work and sharpen the case for a recognition-based recognition account against a pure resemblance account.

Let us first see if Hopkins’ account of resemblance stands scrutiny, we will then go on to examine the perceived shortcomings of Schier’s theory.

**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.

Hopkins, on the other hand takes a different tack, and writes:

It is hard to know how to make sense of resemblance between a particular thing and some, but no particular, item of a certain sort – a horse say. For resemblance is presumably a matter of shared properties – A resembles B provided both have a certain property F. Yet only what exists has properties, and thus can share them with anything else. Since there are no horses which are not particular horses, there are no properties enjoyed by some, but no particular, horse.114

He explains:

Much depiction is not of particulars, but is rather of some, but no particular, item with various properties. Resemblance, in contrast, is a relation which only holds between one particular and another. So depiction cannot be a matter of resemblance.115

But Hopkins has a way out; his strategy here is to argue that experienced resemblance unlike resemblance simpliciter “does not require two particulars, one resembling, the other resembled”116 because if a person experiences a picture, or part of a picture, as resembling something which they recall, or imagine, that is enough for the experience of resemblance. Thus, the first problem poses no problem for the experience resemblist but makes problems 2 and 3 more acute.

114 Hopkins, Picture, Image and Experience: A Philosophical Inquiry, 10-11.- We could recast his objection thus. “It is hard to know how to make sense of resemblance between Phar Lap and some, but no particular horse.” In fact, it’s not difficult at all for us to say that Phar Lap resembles a horse, or a number of good looking Arabian horses, or even other horses sired by Night Raid. So it is not clear where Hopkins’ difficulty is here. It seems to be part of the function of the term ‘resemblance’ that one can use it to mark similarities at any level of generality. The mistake, I suppose, is to assume that the ideal of resemblance is identity.
115 Ibid., 50.
116 Ibid.
This is because it is precisely the fact that when we reflect on the resemblance between, for example, a stick figure drawing of a horse and an actual horse, it is hard to say that we are experiencing a resemblance at all. The usual strategy here is to exclude various types of depiction from the class of those motivated by resemblance, and usually this means that we are left with a theory which explains realist depiction and little else. Hopkins believes that he has found an antidote to the fragmentation and watering down of resemblance theory by identifying shape as the “respect in which picture and object must be experienced as resembling.”

The problem for shape resemblance has always been that shape is usually three-dimensional. We have no problem with the concept that the two rectangular two-dimensional shapes on the left in figure 23 resemble each other – they are both two-dimensional rectangles.

![Figure 23  Squares and a box](image)

The problem arises when we consider how the two-dimensional drawing on the right resembles the shape of a three-dimensional box. Hopkins comments:

> Shape is the key, but not 3-D shape. Rather, we seem to need some way to extract, from the 3-D shape of objects, a more two-dimensional form of shape which the pictures can share. But what could this form of shape be?118

His answer is “outline shape”. But what Hopkins means by “outline shape” requires some elucidation and, given that his resurrection of resemblance theory hinges on it, it is worth spelling out his theory in detail.

### Outline Shape Resemblance

Hopkins sets up his outline shape theory in the following terms:

> Suppose I am looking at one of the pyramids through a misty window, and that on the glass I trace the monument’s contours with my finger. The result is a picture, of sorts. How does it resemble the pyramid?119

---

117 Ibid., 52.
118 Ibid., 52-53.
Thus far all we have is a property tracing and picture might share – ordinary two-dimensional shape. What we need is one the object might also enjoy. It is tempting to think that the relevant property is the objects outline shape, but what sense can we make of this idea? ……To see what property tracing and pyramid might share, we need to explore the geometrical relationship that holds between them.120

Hopkins uses the following drawing (Fig 24. Hopkins’ Pyramid and Tracing) to demonstrate this relationship. He continues:

at the point from which it is made, the tracing subtends, in every plane, the same angle as the pyramid. More accurately, parts of the tracing subtend the same angles as parts of the pyramid they trace. The line tracing the base subtends the same large angle as the base itself, the point representing the apex subtends the same very small one. And so on for every plane……Thus taken together, the marks on the glass will subtend the same solid angle as the face of the pyramid they trace.121

![Figure 24 Hopkins’ Pyramid and Tracing](image)

Figure 24 Hopkins’ Pyramid and Tracing. The angles APB and aPb in the drawing are equal. The same is true for CPD and ePd and any other pair of angles one might select from corresponding points on pyramid and tracing

120 Hopkins, Picture, Image and Experience: A Philosophical Inquiry, 53.
121 Ibid., 55.
The term “solid angle” is rather confusing, but crucial to the argument, because it is precisely this geometrical property of a shared solid angle which Hopkins claims is the basis for outline shape resemblance. The angles APB and aPb in the drawing are equal. The same is true for CPD and cPd and any other pair of angles one might select from corresponding points on pyramid and tracing. Considered together, all these corresponding angles comprise a three-dimensional angle subtended at P, the point from which both tracing and pyramid are viewed. This three-dimensional angle is Hopkins’ “solid angle”:

Two items will resemble in outline shape to the extent that, at some point, one subtends a solid angle similar to that subtended, at the same point, by the other……The proposal is clearly going to be that outline shape provides the respect in which pictures look like what they depict.

Hopkins goes on to qualify this notion by specifying that outline shape is not just another term for silhouette. In particular, “the outline shape of an object may include the nested outline shapes of its parts.” In his pyramid drawing we can see that two planes of the pyramid are visible, thus it has two nested planes. Outline shape can also include other nested features such as doors and other features which are delineated by a visible boundary. Outline shape is not confined to things with determinate edges: clouds of gas and mist can have a fuzzy outline shape: “In short, if something has a 3-D shape it has an outline shape.”

Despite these caveats there is still the worry that, even if this esoteric geometrical relation does exist between a tracing and the object traced, it is not something that we obviously perceive. Hopkins’ claim is that we do perceive this property but don’t call it “outline shape”. He presents the famous example of how a round wagon wheel might look from various angles. We might say that it looks elliptical if viewed obliquely from, say, 45°, yet we do not assume that the wheel has become elliptical. It just looks that way. And if we are disposed to make a drawing of it from that angle the lines will trace an ellipse. Hopkins is at pains to point out that although the drawing (or tracing) is elliptical:

The outline shape of the wheel is not elliptical, for that is a term applying to two-dimensional shapes. Outline shapes are not straightforwardly two-dimensional, for each outline shape is relative to a point – the point at

122 Gavin Mcintosh dismisses the concept outright in his criticism of Hopkins in McIntosh, "Depiction Unexplained: Peacocke and Hopkins on Pictorial Representation," 287.

123 Hopkins, Picture, Image and Experience: A Philosophical Inquiry, 56.

124 Ibid., 57.
which that particular solid angle is subtended. Since 2-D shape is not so relative, 2-D shape terms cannot directly apply to outline shapes. Nonetheless, we can understand why we might want to describe the outline shape of the wheel in these terms. For that outline shape would be shared by a tracing of the wheel, at least at a point directly over the surface of the tracing. And that tracing would really be elliptical. Thus the full situation is this. There is a feature of the wheel which our experience represents, and which we find natural to describe by a term for a 2-D shape – ‘elliptical’. That term cannot apply directly to any represented feature of the object, since our experience does not misrepresent the wheel’s shape, and that shape is not in any way elliptical. But that term can be seen to apply to one of the wheel’s features by a less direct route. It describes the 2-D shape of an item which, from the appropriate angle, share’s the wheel’s outline shape. I suggest that this provides the best construal we can put on this talk of ellipses. It is an attempt to capture the outline shape we see the wheel as having.125

Hopkins argues that we are aware of outline shape, for example when we see the railway tracks converging into the distance, but do not construe that shape as the shape of the actual set of tracks. When we see objects foreshortened we do not take them as being distorted and we are aware of the shape they are presenting to our eye. This, Hopkins construes, is an awareness of the solid angle which determines the outline shape from that point and would be shared by a drawing or tracing executed on a perpendicular plane at any point between point and object.

This argument is designed to head off the usual criticisms of the naive argument that shape resemblance is a phenomenon related to the resemblance of a silhouette to an object. Christopher Peacocke navigates these objections to silhouette resemblance in his essay “Depiction”:

Suppose you see hanging on the wall a black silhouette of Salisbury Cathedral …..if the silhouette is successful, the following is true of the perceiver’s experience: the silhouette is presented in an area of the perceiver’s visual field which is experienced as similar in shape to the region of the visual field in which Salisbury Cathedral itself is presented when seen from a certain angle. The point is not just that the area of the visual field is thus similar in shape, but that it is experienced as being so.126

Peacocke inserts a “visual field” between the silhouette and the object to act as a go-between for the relation and avoid the usual criticism that a two-dimensional drawing does not share many shape properties with a three-dimensional object. Hopkins has inserted “outline shape” between silhouette and object. In fact, what

125 Ibid., 59-60.
126 Peacocke, "Depiction."
Hopkins and Peacocke have identified is one of the most puzzling phenomena in visual perception – it is known as “object constancy”. How does the visual system track an object despite the fact that it presents a different shape to the eye as it moves and as the person looking at it moves? There are many theories as to how this phenomenon of “object constancy” works. In a sense, both Hopkins and Peacocke are trying to explain the constancy phenomenon without the appeal to cognitive science. This, we will see, is fatal for both of their accounts.

If Hopkins’ manoeuvre is to succeed, it must avoid the difficulties which the more traditional experience resemblance account faces and show that outline shape is not only a genuine geometrical relation between object and tracing, but that we really do notice it. Of course his problems don’t end there. Even if we grant that such a geometrical relation as outline shape is shared by picture and an object, Hopkins has not shown that, even if we do perceive it or anything like it, we use it to recognise objects in real life or to interpret drawings.

Artists use the device of a real or imaginary “picture plane” between viewer and object to get proportions and scaling right in a drawing. This is an old technique and one which is also used in photography (although the method is slightly different). It would seem that the fact that the concept of outline shape is based on artist’s technique is a mark in its favour. It should be noted, however, that if you stand at a window and use a marker to trace an outline on the window of the objects in view (let’s say some beach balls and some assorted cubes) you will find that the shapes you actually get on the window are not much like the objects you are looking at. You will find that anything outside 30 degrees either side of dead-ahead is distorted on the glass. The beach balls will be elongated ellipsoids and the cubes will be rectangular – the artist (and cameras) using this technique must correct for this distortion.

It must also be noted that this is not the only technique available to artists, and as Hopkins himself notes, depiction didn’t begin with tracing outlines on picture planes. That is, there are many kinds of depiction, e.g. children’s drawings, which do not rely on the geometrical relation which Hopkins proposes to explain depiction. Thus we have to allow that this geometrical relationship is either specific to particular kinds of depiction (e.g. perspective projections) or has some room for manoeuvring in the latitude of congruence. When one considers that it is
rare to view a picture front-on, the possibility that this geometric relationship holds starts to look bleak.

It is clear that however one construes it, both Peacocke and Hopkins are appealing to the intuition that a silhouette traces the outline of an object from a certain point of view. A silhouette outline can be generated by a causal mechanism and can be readily encountered in our everyday life – for example, the object might be silhouetted against a bright light, or its silhouette projected on the wall. This fact seems to ground resemblance theory because it is difficult to argue against the fact that photographs and perspective drawings and paintings use a picture plane device to capture the shapes which the objects present to the eye. There is a causal relation between the object’s silhouette shape and the shape on the picture plane.

This raises two questions:

How fundamental is a silhouette or an outline to our ability to recognise an object in real life?

And, by extension,

How important is an outline in the interpretation of a drawing?

If we can establish that the outline shape an object presents to the eye is a good basis for recognizing that object, then we would have a case for shape-resemblance. What I am talking about here is not just the overall outline – internal details have outlines as well. If shape-resemblance is the primary way that we recognise objects in the world, it is likely that it is the way we recognise drawings and photographs as well. After all, a line drawing has little other than shape to offer and we have seen that even young children readily recognise rudimentary line drawings. I am not saying that Hopkins’ notion of outline shape boils down to nested silhouettes. Hopkins conceives it as a geometrical relation involving three-dimensional angles subtended at the viewpoint. Thus, he admits, it is a rather esoteric property of objects. Silhouettes are not at all esoteric, and most people would agree that they are very recognizable. It is hard to believe that the more complex relation, which Hopkins proposes, is easier to see and more effective at generating the experience of resemblance than a simple silhouette. If I can show

127 Of course there may be many esoteric relationships between object which our visual system uses to track them and recognise them, but it is clear that they are not geometric relationships between eye and object. We may not be aware of how our visual system builds these relationships
that we don’t use simple outlines of the silhouette kind to recognise objects or interpret pictures, Hopkins’ argument collapses and we need another account of shape recognition.

**Silhouettes**

An initial problem for shape resemblance is that the activity of matching visual outlines of rigid geometric shapes is a very different activity from matching outlines of animals and plants. It is easy to appeal to our intuitions that it is easy to recognise the outlines of simple geometrical objects, and familiar household items, but plants and animals present a different problem: not only do they present themselves from different angles, but they also change shape.

On the next page you will see the outlines of four objects. On the following page you will find their silhouettes, and on the page after that the photographs from which they were traced. Can you guess the object? Is the silhouette more suggestive than the outline?
Recognition-based Theories vs Resemblance Theories

Figure 26
Figure 27
These particular pictures were chosen because they presented the objects from an angle at which we would usually view them. The reasoning was that shoes and children are usually viewed from above, cars and people at eye level.

I traced the outlines onto transparent film from the photographs, scanned them, transferred them to a photoshop program and cleaned them up. I have shown them to a number of people with some interesting results:

- Nobody had any trouble identifying the Drew Barrymore silhouette and her outline as a girl with a handbag (in fact it was the handbag that attracted most comment as a recognisable feature).
- Everybody got the Volkswagen – surprisingly, everybody identified the make of the car!
- Only one person identified the shoe and the child. Many thought the child was a penguin.

I have come to no conclusions as to whether the outline drawing is more or less easy to recognise than the silhouette – some say the silhouette is easier to identify, some disagree. Consequently, in the following examples I will randomly alternate between silhouette and outline. Have a guess at these two shapes (fig. 28) before you turn the page.

---

Figure 28
Once again, the angle I have chosen is quite normal, but a dog can take many shapes and that makes the dog’s outline very hard to identify. These two below are easier (fig.30), but are they the same dogs as above (fig. 29)?

The pose here is probably more commonly encountered than the lying down pose. The picture, as you will see on the next page, is taken from a lower angle. The view does not look down on the dogs as much as the previous example.
It should be noted here that I could find no automatic way of de-articulating the dogs from their background using a photoshop application. You will see from the photoshop rendering below (fig. 32) that in many instances the program cannot distinguish dog from background.
My own efforts to trace dog outlines from photographs were similarly fraught.

In order to obtain the above outlines I first had to de-articulate the dog from its scene. Even in the carefully posed photographs there were parts of the dog obscured by furniture and foliage (figs. 34 & 35). This is not a minor problem for our intuitions about outline resemblance. We constantly de-articulate objects from their environment as we move around in the world, because in the real world one object is usually obscured by another.
It is interesting to try to generate a line drawing from a photograph using Photoshop. Even with a relatively straightforward photograph, it cannot de-articulate the subject from the background and consequently the lines of object and background often merge (fig. 36).

Figure 36  A line drawing of Spike generated by a photoshop application.

The de-articulation issue may prove to be more damaging to the shape resemblance argument than any other issue. It assumes that in recognizing an object one knows which parts of a scene are foreground and which are background. Clearly this is a subjective choice. One motivated by interest and attention.
Recognition-based Theories vs Resemblance Theories

In the photograph of Salisbury Cathedral (fig. 37, left) several trees obscure crucial parts of the structure; the image generated from the photograph by the computer (fig. 37, right) confuses figure and ground. The clouds blend into the shape of the building.

In Constable’s painting (fig. 38), on the other hand, the trees neatly frame (rather than obscure) the building. Constable has arranged the Cathedral and trees so that the subject is clearly picked out. I was similarly motivated when I chose photographs which were easy to trace and in which it was easy to discern a figure.
In the next batch of pictures I would like you to consider which outline is easiest to recognise.

Which is most recognisable?

Figure 39
Most people judge number 4, the Stubbs dog, to be the most recognizable from its outline. Note that Stubbs has posed the dog in such a way that its tail and all its legs are visible. It also has a well-defined muzzle. Stubbs has ensured that the dog strikes a very informative profile. My efforts with the photography and drawing of dogs didn’t ever afford me a profile as complete as that of the Stubbs. The dogs never hold a pose for very long, unless they are lying down, sleeping or waiting for food. The number 3 photograph is the nearest I have got to a posed photograph with all the dog’s limbs in sight. Getting a profile shot in the manner of Stubbs is virtually impossible. Thus, in all my photographs the foreshortening makes the outline tracing more difficult to interpret.

Even this cursory look at the role of silhouette outline indicates that the shape of overall outlines which we are afforded on an everyday basis is rather uninformative. It is often the shape of a detail – a handbag, a tail or the buckle on a shoe – that makes the silhouette identifiable. Furthermore, the whole exercise begs the question of what we mean by one outline being “more recognisable” than another.

On the next page are more silhouette outlines. Which ones are most recognizable?
Figure 42
You probably recognised all the outlines in figure 41 as people, you may even have guessed which were male and which female, but I suspect that you failed to guess the correct pose or action for all of them.

Commentary on the figures from top left to bottom right:

1. In the Demi Moore photograph it is not immediately apparent that her suit is painted on. The lines of the lapels, collar and cuffs draw our attention away from her outline. It is only when we notice her breast and nipples that we realize she is unclothed. On another note: I was very unhappy about the outline which I traced until I realized that the photograph itself was doctored in some way before it was published. Her whole outline has been photoshopped. Consequently when you look at my tracing it looks featureless and distorted.

2. Most people guessed, wrongly, that this guy was throwing something, rather than kneeling down. Obviously, no one guessed he had his back to the camera. A silhouette outline is entirely reversible back-to-front.

3. Most people thought that this was someone jumping and that it was unfair of me to omit the chair.

4. The girl surrounded by gas is a cartoon drawing, and as such has all her limbs and very wavy girl-type hair. I presume that comic artists draw either from the imagination or from a standard set of posed pictures. It is certainly true that the outline tracings of comic characters are much easier to recognise, in terms of their general dynamic – walking, standing, and running – than tracings of unposed photographs.

5. I imagined that everybody would guess that this was someone sitting with a hat on. In fact, the outline of me on the beach caused much debate. I conclude that the level of detail in the outline is too crude and some of my decisions about what parts to trace, especially in terms of towels and foreground objects, were hasty.

6. This cartoon character (Iron Man) also has all his limbs. Consequently, his outline is easily recognizable. I traced him to see if the strange shape of his iron suit and his outsized hands would be highlighted by the outline. I don’t believe they are.
Recognition-based Theories vs Resemblance Theories

In general there were quite a few surprises as you moved from outline to full-colour pictures. One thing which was highlighted quite strongly in a group session, which I conducted with students and staff at Monash University, was that there was quite a lot of guesswork involved. People were speculating as to what the outlines were. Was it a boy or a girl? Were they sitting, standing or throwing? Was it a penguin? In the group, these speculations fed off each other, and people would remark “Oh, yes I see it.” as they became convinced by someone’s construal of the outline. And, at some point guesswork would turn into certainty. This is reminiscent of Minsky’s “pandemonium” model of how the mind works, and Gregory’s theory about how vision works – there are hypotheses, and lots of different speculations about what is before their eyes, and then one interpretation wins out.

Some Interim Conclusions on Silhouettes and Recognition

This was a fascinating exercise, but these tests weren’t conducted under supervised conditions, and my method for choosing the target images and generating the outlines was very subjective. These conclusions should therefore be considered as provisional.

1. The outline shape of an object is not a primary shape recognition cue.
Those readers, who at the beginning of this section, thought that silhouette outline would prove to be a primary shape cue, would now agree that it is not. My drawing and tracing experiments have convinced me that neither artist nor viewer use the overall outline of an object as a primary shape cue. It is clear that pictorial shape resemblance, as espoused by modern resemblance theorists such as Hopkins and Peacocke, does not hinge on overall outline. The problem with shape resemblance theory is that the outline shape is not as informative as you might expect.

2. Outline shapes are generally uninformative.
The only thing an outline drawing does well is distinguish an object from its background. The Monash subjects found it easy to recognise all the people-outlines as people, but were in general mistaken about what the person was doing – whether they were sitting, jumping, throwing etc. The outline does not communicate the dynamics of the figure, that can only be communicated when we know exactly how the limbs are configured, and even then it can be hard. It is
worth noting that recognising movement in still photograph is not as straightforward as one might think.

3. **Objects rarely present an informative and unentangled outline**
Disentangling an object from its background is a primary visual function. That is, our visual system is dedicated to disentangling objects from their backgrounds. When we look at the work of V. S. Ramachandran on visual recognition, in a later chapter, we will see that he maintains that the visual system has largely evolved to see through camouflage and pick out objects and animals occluded by foliage or other obstructions. In a drawing a lot of this disentangling work has already been done.

4. **Disentangling objects from their backgrounds is a subjective process.**
When I made the outline tracings, I did the work for you. At every stage where one object occluded another I had to make a decision as to whether to continue a line, and effectively invent the path of the outline, or go around the occluding object(s).

![Figure 43](image)

If you compare the automatically generated line drawing of Spike (fig. 43.1) and the outline which I traced (fig. 43.2) you will notice that I have cheated. Spike’s tail and paws are buried in the grass in the photograph. The photoshop line drawing doesn’t pick them out – I have indicated them using my knowledge of Spike. You will also notice that some lines in the photoshop version are broken where Spike blends into the background – once again I have added lines where I believed I could guess the shape. In each case where I used a photographic original such as (fig. 43.3) for my tracing, I had to make these kinds of decisions.
This is clearly a very subjective procedure. It was relatively easy with an outline tracing, but is much more difficult when I include internal details.

**Are nested outlines any easier to disentangle and recognise than overall outlines?**

It is worth considering in detail how this tracing process works because it has a direct bearing on how one construes Hopkins’ qualification that his “outline shape resemblance” theory also applies to “nested” outlines. In such a line drawing, internal detail is depicted in outline, but without shading or texture.

The following drawing of Spike, which I have drawn from life, is a good example. I have avoided indicating colour or texture boundaries. The lines of the drawing delineate boundaries between three-dimensional features which stand out against the blank background or occlude each other. For example, his leg partially occludes another leg.

![Figure 44 Spike](image.png)

It is clear that once internal occlusion lines are included, not only does it become easy for anyone to recognise that Spike is a dog, but we no longer have the feeling that we are guessing. The depicted object begins to have distinct properties. The lines describe the shape of his head, legs and tail, and one ear. His paws are roughly outlined and we can see that he is lying down. The internal lines also give the viewer a feeling for the three-dimensional shape of the dog – another
dimension has been added – literally. All this, and more, is evoked by a dozen or so lines. It is clear a line drawing which includes internal lines, and does not restrict itself to silhouette outlines, can trigger powerful recognition cues. We must assume that these are shape recognition cues because it is hard to say what else a line drawing could be describing but shape. Thus Hopkins’ qualification that “outline shape resemblance” must take into account “nested outlines” turns out to be crucial to his argument. We found that without these internal lines, recognizing the object depicted from silhouette outline alone is often guesswork.

It is hard to see why internal outlines should be any more recognizable than overall outlines. Yet when we add these internal (or as Hopkins calls them “nested”) outlines the figure is positively identifiable. What happens if we isolate some of these internal details?

![Figure 45 Line drawing of Spike’s ear (from fig.44) and photo](image)

In the drawing of Spike we can see an ear, however, if we isolate the area of the drawing featuring Spike’s ear (fig. 45 left) it is not immediately apparent that it is an ear – it could be a spinning top or a cup or a heart or any number of other things. Even when we come to look at a full-colour picture of Spike’s ear (fig. 45 right) it is conceivable that someone could fail to recognise the shape as that of a dog’s ear. If we go on to imagine what Spike’s ear actually looks like in real life – a three-dimensional furry object which has a different shape according to which angle you are looking at it from and what mood he is in – it is clear that shape recognition is something more than simply matching the two-dimensional boundaries which the lines delineate against the boundaries which three-dimensional objects create when edges are outlined against a background or another object.

If we take a random selection of details from other pictures it is clear that as soon as a shape is isolated from the other shapes in a drawing, painting or photograph recognition is reduced to guesswork.
Recognition-based Theories vs Resemblance Theories

Figure 46
Recognition-based Theories vs Resemblance Theories

Figure 48
Even if we add colour to the same selection of details we are still guessing.\textsuperscript{128}

\textsuperscript{128} See Appendix for the full pictures from which each of these fragments is extracted from.
It is probably true that the degree of guesswork is reduced as the pictures change from lines, to grayscale, to colour. But that is not surprising. The grayscale contains tonal shape cues, texture and some indication of depth, slant and tilt. The colour doesn’t seem to add much in terms of shape but it does indicate flesh tones and that is a very important cue.

**Context and Configuration**

It is fairly clear that the reason we fail to recognise isolated elements of a line drawing is because context is removed. Context seems to be unconsciously factored into interpretations of drawings and their parts. When I removed parts of drawings from their contexts, or omitted something seemingly important, such as the chair in fig. 42 no 3, my test subjects complained that I had cheated. In fact, I have found that the viewer feels the test is unfair if a section of a line drawing is chosen which is not in some way “naturally” recognizable. However, the main context of an element in a line drawing is the other lines. It is clear that it is the overall configuration of lines and their relation to each other which determines how the lines depict. So it is not surprising that we fail to recognise the triangle in fig. 45 as a dog’s ear. It is only an ear because of its relationship to all the other lines. If we moved it to a position half-way down Spike’s back we would not leap to the conclusion that Spike was the victim of a medical experiment, we would probably read it as a leaf.

Schier points out that it is not just the context of the marks set against other recognizable parts of the drawing such as a muzzle or legs, which enable the marks to depict an ear; it is the overall spatial configuration of the marks on the paper which determines what is depicted. He comments that “a dot or a brush stroke can make a difference to what a symbol depicts even though that dot or brush stroke *by itself* would not depict anything.” Schier emphasizes his point using the example of a Paul Klee painting:

In the centre of the painting we see a primitive depiction of a face, like that a child might produce: two dots for the eyes, a line for the nose, a curve for the mouth, a circle for the head, and a few squiggles of hair. Surrounding this face are all the dots and circles but strewn about randomly. The point is clear: it is the relative position of the marks which depicts a face; and it is only in the context of such a configuration that given marks iconify given features......But the dots-in-the-context-of-

certain-other-marks not only make a difference to what the whole iconifies – a face with eyes – they themselves play a semantic role: the dots represent eyes. I call them sub-iconic because the dots cannot iconify eyes except in the context of the whole. Just as they make a contribution to the whole, so the whole makes a contribution to them. 130

Schier is aware of the objection that this all seems to be rather circular. A line or dot depicts something because it is near another line or dot which only depicts something because it is near the other line or dot etc etc. It is hard to understand how the whole interpretation process can get a foothold. Schier’s explanation is simple:

The key seems to be that we comprehend these marks all at once and not seriatim. If we are given the marks in a temporal sequence, one bit at a time, obviously the sequence will not be an icon of a face.131

This observation of Schier’s that we interpret the marks “all at once” is immensely significant. We see the whole of the picture in one short moment. We don’t read it like we read a sentence. It is the overall configuration of marks and parts that kick-starts our initial interpretation. This observation of Schier’s, not only signifies a clear break with the language analogy, but raises one of the most interesting questions in recognition theory:

Do we recognise objects and scenes by building up a percept from their parts or do we recognise whole scenes and objects all at once?

There is a lively debate in vision research on this issue. On the one hand there is recognition by components theory, most famously pioneered by Biederman and his concept of geons. On the other hand are the view-based theories of recognition. 132 The main challenge for both these theoretical positions is to account for how we recognise objects from unfamiliar angles. If vision research eventually shows that we recognise whole scenes before we are aware of their parts, this may indicate that theories of depiction which attempt to reason from the parts to the whole are mistaken. Schier’s idea is that with certain pictures (e.g. a child’s drawing of a face) no single mark in the drawing can stand alone as

130 Ibid.
131 Ibid., 70.
signifying an eye or a mouth. Whereas if we look at, for example, Leonardo’s
career Head of a Girl (fig. 49) one eye can stand alone as an eye.

Figure 49 (left) Leonardo Da Vinci, Head of a Girl, c. 1483, Silverpoint and white highlights on prepared paper, 181 x 159 mm, Biblioteca Reale, Turin
Figure 50 (right) Paul Klee, Face of a Face,

In Schier’s terminology, the eye-marks in the Leonardo drawing “iconify” an eye. None of the eye-marks in the Paul Klee painting, Face of a Face (fig. 50), iconify an eye. Schier calls these dot and circle marks “sub-iconic”, although he stresses that they are sub-iconic only in the configuration of this particular painting. They are not intrinsically sub-iconic. Of course, any drawing, painting or photograph viewed at a certain scale will be seen to be composed of sub-iconic marks. If we were able to look closely at one of the girl’s eyes in Leonardo’s drawing we would just see lines.

If we apply Hopkins’ terminology to the same issue, we find that pictures are composed of an ever descending “nesting of outline shapes”. It is clear with the Klee painting that this nesting ends rather abruptly with lines, circles and dots, whereas with the Leonardo drawing one might be tempted to zoom in until the image pixelates.

Hopkins’ argument, when reduced to its basics, is that “seeing-in is experiencing resemblance in outline shape.” If we follow this line of argument we might agree that seeing an eye in the Leonardo is to experience resemblance in

133 Hopkins, Pictures and Film; Philosophy and the Empirical Disciplines, a Reply to Dean ([cited]).
outline shape. But the argument that seeing an eye in the Klee picture is to experience resemblance in outline shape is less easy to accept. Hopkins’ theory rests on an appeal to the geometrical congruence of the angles which the shape of the head of a girl might subtend at the viewer’s eye and the angles which the drawing of a girl’s head subtends. This geometrical congruence ensures that some of the proportions of the girl’s head in the drawing can also be found when looking at the head of a real girl, thus we experience resemblance. There can be no such proportional congruence in the Klee painting; these shapes do not share any proportions with real heads – so where is the experience of resemblance which Hopkins’ claims to be the key to seeing-in? Hopkins addresses this objection using the example of a stick figure.

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. 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.

---

134 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.
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 precisely the outline that the stick figure omits. Hopkins’ strategy is to invoke the indeterminacy principle — “pictorial content can in various ways be indeterminate.” 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.

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.

---

136 Ibid., 124.
Recognition-based Theories vs Resemblance Theories

– 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 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.137

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.138 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.139

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

137 Ibid., 124-25.
138 Ibid., 125.
139 Ibid.
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 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\(^{140}\) 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.\(^{141}\)

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

---

\(^{140}\) 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.

\(^{141}\) 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.
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.142

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 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?\(^\text{143}\)

His answer is that she must appeal to “widespread knowledge of a very general nature.”\(^\text{144}\) 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 negatived 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.\(^\text{145}\)

---

\(^{143}\) Ibid., 137.

\(^{144}\) Ibid.

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 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.\(^{146}\)

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:

\[
C: \quad \text{Given that } S \text{ is of } O, \text{ it is } \text{intended} \text{ that those who are able to recognise } O \text{ should be able, on that basis, to interpret } S.\text{[my italics]}\]

\(^{147}\)

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.

---

\(^{146}\) Ibid., 169.

\(^{147}\) Ibid., 137.
Recognition-based Theories vs Resemblance Theories

The End of Resemblance?
The significance of the artist’s intention and standards of correctness for a theory of depiction is highlighted by Leonardo’s example of seeing “fanciful” figures in the stains on a wall. A theory of depiction needs to establish how seeing faces in stains and clouds is functionally different from seeing faces in depictions. The conventionalist would like to claim that when we see a face in Salvador Dali’s Head of a Woman in the Form of a Battle (fig. 52) that our brain is processing symbols and when we see the face in a stain it is more like a hallucination. We know why the conventionalist account doesn’t work, but the point of the example is to show that a theory of depiction needs to indicate the mechanism whereby we are able to see the face. The resemblance theorist would need to admit that in both cases that our visual system was seeing the face because of shape resemblance. Thus the resemblance theorist cannot claim (as the conventionalist does) that the mechanism whereby we arrive at the pictorial interpretation is different in each case. In both cases, the stain and the Dali painting, the resemblance to a face is fairly loose and the features of the face are fairly indeterminate. The problem for resemblance is that resemblance is a method of seeing similarities and as such is a
rather blunt instrument – we can see similarities in all kinds of ways and in all kinds of things. What could make seeing shape resemblance in pictures invoke different brain processes from seeing shape resemblance in stains and clouds? It looks as if the only thing that separates the stain from the picture is signs of artifice in the latter. That is, the marks in Dali’s picture do not look as if it they happened by accident. However, noticing that the marks couldn’t have come about by accident is something that happens after one has seen the horses and figures that compose the face. Noticing Dali’s artisanship comes after the recognition moment. If resemblance is responsible for the recognition, what are we using to determine the standard of correctness? Degree of resemblance cannot be the key to making this determination because a child’s drawing of a house is as much a picture of a house as a photograph. If degree of resemblance is invoked, one is left with the Hopkins dilemma about the very indeterminate resemblance one finds in children’s drawings and Dali’s drawings of horses. There may be stains on walls that look a lot more like faces than any child’s drawing, thus the resemblance mechanism cannot discriminate the cases of accidental images and purposefully created images without appeal to artist’s intention. Therefore, if resemblance is the method we use to interpret images, resemblance by itself cannot establish standards of correctness. This can only be done by the appeal to artist’s intention and this has nothing directly to do with resemblance. One could of course argue that the artist intended that the picture resemble the object depicted and that this opens the door to resemblance theory – but of course this isn’t an explanation of the mechanism whereby we interpret pictures – it’s an explanation of how artists create pictures.

Hopkins suggests that sometimes you should use the marks as a guide and other times not. That seems particularly unhelpful given that what he claims to have developed is an explanation for how we determine what the marks represent. Schier’s explanation is that it’s not the nature of the marks or their style which generates the initial interpretation – so there is no point in analysing the marks to discover what they depict – we already know what they are depicting. The key to understanding how we see content in the marks is understanding how our visual recognition system works. The more we understand about visual cognition the closer we get to understanding what it is about depictions that can so powerfully trigger recognition. The underlying problem with both conventionalist and
resemblance theories of depiction is that they assume that we notice the nature of
the marks before we recognise what the marks depict. That is, they assume that
style and technique are relevant to recognising familiar objects in pictures. The
problems Hopkins encounters trying to make resemblance the functional
mechanism for pictorial interpretation are entirely a result of focusing on the
marks and not our recognition abilities.

To the extent that Hopkins’ theory is a version of seeing-as, his theory should
be able to explain the difference between fanciful seeing and seeing what the artist
intended. Hopkins’ theory, and resemblance theories in general, are unable to
apply their standards of correctness in exactly the cases where they are needed
most; stick-figures, caricature, stains on walls, etc. In this sense, resemblance
theory fails to establish a functional role for resemblance when put to the simplest
of tests.

The truth is that our visual system interprets the stains and the Dali painting in
exactly the same way. In that first recognition moment, the picture and stains do
not act on our visual system in a way that is functionally different. Our visual
system sees the “fanciful” horses and the face in a moment of recognition. It is
only after we have recognised the image that we ask ourselves “Did someone
intend me to see that image?” Even if at first you don’t notice the face in the Dali
painting, it is almost as if, when you finally do see it, that you succeed in spite of
the marks. Thus the only difference between the cases of the stains and the picture
is that when we ask ourselves “Did someone intend me to see that image?” we
notice signs of artifice (e.g. brushstrokes) in the picture and answer is “yes”. In
these cases we can easily find other content in the picture, which maybe we hadn’t
noticed at first, because the artist intended the viewer to find such content.

Notice that the crucial question here invokes the intention of the artist. You
cannot establish standards of correctness unless you can appeal to the intention of
the artist or photographer. With the stain on the wall you have no court of appeal.
This is where Hopkins makes his final mistake when he appeals to “widespread
knowledge of a very general nature.” His appeal should be to knowledge about
human visual systems and human powers of recognition and discrimination which

148 Artist’s intent is also relevant but only to the extent that the artist wants to depict in a way that
other creatures with similar visual systems can see the marks as what he/she intends them to be.
149 Hopkins, Picture, Image and Experience: A Philosophical Inquiry, 137.
constrain the kinds of marks which an artist might use when wishing to make a picture that other human beings can see depicted content in. As Schier puts it, the content of the picture 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.”

A resemblance theory which purports to be a theory of how we interpret the marks shouldn’t need the appeal to frame-knowledge to arbitrate standards of correctness. Such a theory purports to be a theory of how the marks come to mean what they do. If Hopkins needs to invoke frame knowledge to resolve the standard of correctness issue, he is effectively saying that Schier’s “two-stage” theory is correct and we already know what the depicted content of the picture is before we become aware of the marks. That is, his appeal to frame knowledge is a check to see whether his interpretation is the correct one. If it was the case that we arrived at interpretations through resemblance then we would have to assume that the face in the Dali picture (fig. 52) is seen as a face because it shares a property, such as shape, with a face. However, the right eye of the face also appears to be a person on horseback. The resemblance relation in this case must be very stretched. What does that part of the picture resemble? An eye or a person on horseback? How can resemblance establish a standard of correctness in this case? Clearly, it is only an eye when seen in the context of all the other horses and people which make up the face. Context and configuration are the key elements which trigger recognition in Dali’s picture. Once these visual mechanisms have done their work and we have recognised a face, then we might say we see a resemblance to a face. But it is clear that it is not shape-resemblance that is the mechanism whereby we see the face. The mechanism whereby we see the face can only be understood when we understand how the human recognition system works. One thing we can be sure of is that it doesn’t work through shape matching.

The Artist’s Intention
The immediate problem which recognition theory seems to present is that it doesn’t account for the skill of the artist. Surely a drawing by Rembrandt is more easily recognizable than a stain on a wall? Of course it is more likely that we can recognise something that is designed to be interpreted by the human visual system
as depicting an object or scene. But the problem for most people who object to recognition theory is that it is pretty clear that a 4-year-old child can draw recognizable houses and people without any particular skill or training. So, if all there is to depiction is recognition abilities, what makes Rembrandt’s drawing better than the four-year-old’s drawing? This question takes us back to the question I asked earlier about what it means to recognise a drawing? In the picture below (fig.53), are you recognising a person, a man, a man hitting a ball with a bat, a Yankee hitting a home run, the Yankee Clipper hitting a home run at League Park in 1941?  

![The Yankee Clipper hitting a home run at League Park in 1941](http://images.encarta.msn.com/xrefmedia/sharemed/targets/images/pho/t305/t305137a.jpg)

Figure 53: The Yankee Clipper hitting a home run at League Park in 1941

When we talk of the recognition moment, which of these moments are we talking about and where does the skill come in? One of the skills of depiction is to be able to depict the dynamic of the scene – the fact that DiMaggio is hitting a home run – and as we will see, for a depiction to be able to do that it must trigger very specific cognitive abilities. A child may be able to draw a man with a bat but it takes a specific skill to be able to depict the dynamic which is the difference between us recognizing a man and baseball player in action. The study of visual

cognition can help us understand how the old masters evoke such movement with just a few lines. Studies of how human beings perceive motion can give us an insight into what mechanisms are in play in such depictions, and that means trying to understand what happens in the brain in the recognition moment.

The problem with most theories of depiction is that they have paid very little attention to how we recognise objects in the world. Schier’s insight is that seeing a picture of X exploits similar capacities to seeing X. Unfortunately, Schier doesn’t say very much about how we actually see objects in the world. What he does say is that the key intent in depiction is the intention of the artist to make a picture whose content can be recognised by human beings with similar recognition capacities to those of the artist. If an artist just makes a few squiggles and calls it a face, that is not depiction. If a photographer makes a blurred, out-of-focus, overexposed picture of a featureless wall, that is not depiction. The artist must have the intention of engaging object recognition abilities. The proposition is that the abilities the artist is attempting to trigger are of the same character as those engaged in real life encounters with objects. That is, the proposition is that pictorial interpretation abilities co-vary with ordinary object recognition abilities. In Schier’s terms, the very fact that you are holding a camera and focusing it satisfies the condition of Convention C that you intend that the image you create is recognisable. 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.\(^{152}\)

A camera is designed to generate images that are recognisable by human beings, just as a representational painting is painted with human beings in mind. In this sense, Schier’s theory doesn’t seem to discriminate between methods of depiction. His theory suggests that depictions work by engaging natural recognition abilities, but it is difficult to believe that a line drawing of Salisbury Cathedral is triggering a similar set of overlapping abilities as the painting or the photograph or the actual cathedral. But that is not our only problem.

It is easy to nominate visual processes which are engaged when we look at the Constable painting – colour vision, the ability to focus on fine detail, the ability to judge point of view, among other things. Which abilities and how many of them

---

\(^{152}\) Schier, *Deeper into Pictures: An Essay on Pictorial Representation*, 137.
would need to be engaged before we could say that Constable’s painting has provided alternate causes for something like the visual effect of seeing Salisbury Cathedral from the Bishop’s Ground? This is the second of Hopkins’ challenges: to demonstrate that pictures trigger similar recognition mechanisms to the things they depict, not only in terms of psychological responses but in terms of underlying neurological processes. Schier doesn’t provide us with a list of which natural recognition capacities are being triggered by a depiction. He says things like “A photograph is of O because there are features of O such that had these features been different, that photograph would have been different.” But this doesn’t get us any closer to how he conceives the character of these natural recognition abilities. In fact, Schier effectively warns his reader not to worry about specifying which recognition abilities overlap:

it is obviously not necessary to put any number on the capacities in the overlap set (the set of capacities engaged by S and O). We can say that S is similar enough to O if it is possible for someone to generate naturally an interpretation of S as being of O.

Furthermore, he is at pains to point out that if S is too similar to O then S may no longer operate as an icon – “dissimilarity is as crucial to natural generativity as similarity.” That is, if all these unspecified recognition abilities overlap we won’t be able to tell the depiction from the object and the depiction will effectively be a decoy and not operate as a picture.

Schier may not have been specific about what natural recognition capacities are, but he is quite specific, at one point, about what they are not. He argues that the triggering of recognition capacities is not necessarily a conscious process. He states: “the claim that my ability to recognise O comes into play in my interpreting S is not an introspectively verifiable claim.” Thus the similarity that obtains between picture and depictum is the fact that they provoke overlapping recognition abilities, but this does not mean, according to Schier that when we reflect on how S resembles O we should be able to say precisely in which respects this is the case. He argues:

When one says that X recognises a bit of Jane in John, that means that X is conscious of some of the properties that Jane and John have in common.

---

153 Ibid., 194-95.
154 Ibid., 189.
155 Ibid.
156 Ibid.
But……we need not be conscious of any specifiable properties which picture and depictum have in common.\textsuperscript{157}

This comment suggests that Schier’s overlapping recognition triggers operate below the level of consciousness and are not amenable to introspection. This view is confirmed in a section of \textit{Deeper into Pictures} that addresses Gombrich’s illusionist theory of depiction. In this section Schier forwards his “two hypotheses” account of pictorial recognition.

\textbf{The Two Hypotheses Account}

Schier’s concern in the “Illusion and the Cognitive Theory of Pictorial Perception”\textsuperscript{158} section is to clarify how one can see marks on a flat surface as an object. The apparently twofold nature of pictorial recognition has been a sticking point for most resemblance theories. In particular, resemblance theories such as Gombrich’s and Wollheim’s fail to make the connection between seeing the marks and seeing what they depict. Schier’s analysis of Gombrich’s illusionist theory demonstrates that there is a way of making this connection. However, Schier’s analysis reverses the usual direction of interpretation. All other theories attempt to explain how we get from seeing the marks to seeing the object. Schier’s “two-hypotheses” theory shows how we get from seeing the object to seeing the object as marks on a flat surface.

Gombrich uses Jastrow’s duck-rabbit drawing as an analogy for how we perceive the relation of a picture to its subject.\textsuperscript{159} When we look at the duck-rabbit drawing, our perception alternates between seeing a duck and a rabbit. When we look at a picture, Gombrich maintains, our perception alternates between the marks on the surface and the object depicted. Gombrich’s point is that just as we cannot entertain the rabbit interpretation at the same time as the duck interpretation, so too we do not see the subject of a painting and the marks on the surface simultaneously. In his view, the illusion of the object and the perception of the marks are mutually exclusive. Schier calls this the “alternating illusion theory” and Michael Podro calls it the “reciprocal neglect thesis”. Podro comments:

\textsuperscript{157} Ibid., 188.
\textsuperscript{158} Ibid., 188-95.
Gombrich argued that our ordinary nonpictorial recognition of objects relies on mechanisms that we do not control, so that effective depiction involves finding the configurations that activate those mechanisms despite the limiting conditions of painting – for instance, paintings being flat, still and having a very small range of contrast from light to dark compared with the normal illumination of our environment; for Gombrich, to understand how depiction works we must examine how the painter mobilizes the mechanisms of recognition. Painters, so he argued, learn by trial and error to construct forms that yield experiences convergent with – but always distant from – those of the prepictorial subject.

It is clear from Podro’s account, that there are similarities between Gombrich’s and Schier’s positions. Both see natural recognition abilities as the key to understanding depiction. Podro continues:

What led Gombrich to the reciprocal neglect thesis in the first place was the search for a rational explanation of the mental mechanisms underlying depiction by assimilating them to other models of psychological effect; and the kind of effect he chose was of the triggered reactions isolated by perceptual psychologists; they studied the effects of minimal configurations, for example patterns of lines that looked longer or shorter in spite of our knowing them to be the same size, and he thought of pictorial illusion as pivoted on such triggered responses. What marked these phenomena was that the effects were experienced as occurring in our perceiving and not in the object perceived. But even though we think of depiction as eliciting an effect on us, which in some sense we must, it does not follow that it is experienced as discontinuous with what we know is there – the material surface.

What Podro is alluding to here is Gombrich’s study of optical illusions. Gombrich believed that simple illusions such as the Muller-Lyer illusion isolated the kinds of visual effects which were operating in a more complex manner in many works of art. The main difference is that artists attempt to create a seamless (clandestine) illusion which evokes objects and scenes from the real world, whereas the visual psychologist designs illusions that highlight how our visual system can be manipulated by lines and colours. In this sense, Gombrich was right. The visual illusions created by psychologists and visual researchers isolate visual processes – line perception, for example – and provide insights into the way visual system deals with basic visual representations. These visual experiments, because they often use two-dimensional drawings as “targets”, tell us a lot about how our visual system deals with pictures. The assumption of the visual

---

161 Ibid., 7.
researchers is that they also tell us a lot about basic visual processes. However, it is interesting to speculate why Gombrich should maintain that once a picture triggers recognition of an object, our perception should alternate between the object and a flat surface with marks. The twofold nature of pictures and the question of how we move from seeing marks to seeing objects are the central issues in “resemblance” theories of depiction. Earlier in this chapter we analysed how the brown smudges in Constable’s painting were recognised as cows. Is Gombrich saying that at some level our visual system actually sees the smudges as real cows? Or do we always see picture-cows? Gombrich says we alternate between raw marks and object. Podro and Schier take issue with Gombrich’s reciprocal neglect thesis and argue that we are simultaneously aware of the marks on the surface and the object depicted. In fact, they claim that this simultaneity of perception is the defining characteristic of depiction. We see the brown smudges as picture-cows.

In fact, Schier argues that the duck-rabbit is a bad analogy because there is nothing problematic about the marks being interpreted on the one hand as a duck-picture and on the other as a rabbit-picture. Whereas there does appear to be a problem with interpreting marks on a flat surface as being of an actual three-dimensional object. Furthermore, the duck-picture and rabbit-picture are visually incompatible but logically compatible. Seeing a flat canvas as a three-dimensional object is visually compatible (and true to our experience) but logically incompatible. Thus Schier rejects the “alternating view” and instead suggests what I will call the “two hypotheses” account. Schier comments:

On this view, when S depicts a peach, the visual system processes cues from S, some of the cues being evidence for the presence of a round, ripe peach, and some of the cues intimating the presence of a flat, rectangular piece of canvas. For some reason the visual system prefers the evidence of S’s being a flat, rectangular object to the evidence of its being a round peach. Nonetheless, the fact that one’s visual system entertained the peach hypothesis affects one’s perception of S: one sees S as a picture of a peach. So the difference between the duck-rabbit and picture-peach cases is this: in the duck-rabbit case you consciously entertain two different but equally good interpretations of the figure and (since the interpretations are compatible and equally good) you alternate between seeing a duck-drawing and seeing a rabbit-drawing; in the picture-peach case your visual system (or your visual homunculus if you like) entertains the conflicting ‘peach’ and ‘flat’ hypotheses; they are incompatible so it must reach a
verdict and it decides that the ‘flat’ hypothesis is better grounded; the result is that you see a flat picture of a peach.\footnote{Schier, Deeper into Pictures: An Essay on Pictorial Representation, 192.}

Schier predicts that talk of the visual system entertaining hypotheses invites the objection that he is putting forward a homuncular model of cognitive activity. He does not deny this. In fact, he explicitly dismisses the objection that postulating a little homunculus which processes visual cues merely shifts the act of seeing onto the homunculus and thereby fails to explain how we see. He argues that it isn’t necessary that this homunculus has a visual experience, merely that the homunculus processes the visual cues:

It is only necessary to suppose that [the homunculus] is a mechanism which tracks certain visual cues (input from the eyes) and who processes these cues using a certain ‘programme’. Of course, how a mechanism with this programme may have evolved is another question, though one which must eventually be answered.\footnote{Ibid., 193.}

In short, Schier’s theory suggests that S triggers O-recognising capacities even though that does not amount to recognising S as O. He explains:

Surely the ‘cash value’ of this claim must be equivalent to the claim that S induces me to entertain (albeit unconsciously) the O-hypothesis with respect to S.\footnote{Ibid., 194.}

He continues:

No doubt it contravenes the empiricist assumption that intelligence begins with experience, a homuncular or mentalist account suggesting instead that experience is a product or ‘output’ of intelligent activity. But what of that? A theory cannot be dismissed simply on account of its treading on the empiricist’s corns.\footnote{Ibid., 193.}

This comment throws down the gauntlet and challenges the empiricist view that the recognition triggers must be amenable to third-person observation or first-person introspection.\footnote{In fact, Hopkins criticizes Schier on precisely this point. Hopkins argues that Schier cannot argue that it is ‘conceptual truth’ that recognition capacities are engaged in clusters, (that there could not be bare depiction). Because he maintains that Schier is committed to the notion that ‘recognition, whatever else it involves, has to amount to some grasp, at the level of consciousness, of how the environment is” Hopkins, Picture, Image and Experience: A Philosophical Inquiry,} To the extent that we buy into Schier’s theory, we might
be inclined to say we were off the hook here – his view seems to suggest that it is fruitless to look for these triggers. But this is not the case. By saying that these triggers are not available to inspection or introspection, Schier is not saying that there isn’t an explanation. Rather he’s saying that in order to explain how we can see a cow in a painting, we should first be able to say something about how we see a cow in real life. He also maintains that once we have asked the question “How do we recognise a cow?” we should be prepared to look for answers based on “some computational or cognitive theory of mental activity.”

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.

So, rather curiously, in order to explain how our brown smudge can trigger enough recognition abilities for us to recognise it as cow, we must first explain how we might recognise a real cow in the first place.

**Summary of Chapter Three**

Modern resemblance theory as epitomised by Peacock, Hopkins and Wollheim continues to struggle with the same problems which have dogged resemblance theory from its inception. The chief problem is that if resemblance is defined as property sharing, it is hard to claim that a two-dimensional picture and a three-dimensional object share any real world shape properties. The only shape property which objects and pictures can be said to share is the outline which they present to the eye. My preliminary experiments with outline shape recognition of familiar objects and parts indicated that people do not use the outline of objects as the primary means of identification. Even when there is an exact match between an outline on paper and the outline which it would project on a picture plane situated

---

45. Hopkins argues that Schier must be so committed because his theory ties recognition so intimately to interpretation. “Schier is obliged to restrict recognition in this way because he is only interested in it thanks to its ties to interpretation, and the latter is clearly a matter for consciousness. Unless I form a conscious belief about what a picture before me represents, or at least see it in a way which reflects its having that content, I cannot be said to have interpreted it all.” Hopkins, *Picture, Image and Experience: A Philosophical Inquiry*, 45.


168 Ibid.
between the object and the eye, people simply do not notice the similarity. My thesis is that this is also the case when we look at objects in the world – we don’t notice this outline shape similarity. A mechanism of our visual system known as “object constancy” enables us to discount the fact that objects present different shapes to our eye from different angles and enables us to track them nonetheless. Context and configuration play a key role in “object constancy” and enable us, when we look at a picture, to recognise objects in spite of the flat shapes which the marks make on the canvas.

The other issue which resemblance theory struggles with is the stick figure/caricature problem. It is impossible to adapt any resemblance account to cover degraded sketches and stick figure drawings without either fatally watering the theory down or admitting that resemblance is a concept which only applies to realism. It is clear from looking at how line drawings and stick-figures work that the configuration of the parts and their context is crucial to recognition. Somehow our visual system takes in a picture, or a real world scene, all at once, and recognition can be triggered on the basis of very little information.

It is impossible to develop an account of how we see what we see in Constable’s Ottawa painting without an understanding of how the visual system works. It is clear that Hopkins and modern resemblance theorists have not taken into account the significance of recent discoveries about the visual system. Schier makes it clear, in the quotation which I used to open this chapter, that the experience of resemblance is caused by the triggering of our recognition abilities. Resemblance isn’t the cause of recognition. Resemblance theorists are trying to define how we see pictures in terms of a superficial symptom of visual cognition. We will see in the next chapter that the resemblance account fails because it focuses on the phenomenology of the visual experience and not the underlying cognitive mechanisms which trigger recognition.
Chapter Four

Mechanisms of Recognition

*We see with the brain, not with the eyes.*  Paul Bach-y-Rita

Figure 54  Constable cows from the Ottawa study and from Wivenhoe Park

We are faced with the issue, which perplexed us in the opening chapter, of relating how we see the brown smudges in Constable’s Ottawa study of Salisbury Cathedral (fig. 54) to how we see cows in real life. If Schier is to be believed, looking at that brown smudge is triggering something in us that is triggered by an encounter with the real thing. This seems unfeasible. Even when we look at the Wivenhoe Park cows (Friesians and Herefords, I think) it is still hard to say that this is much like seeing real cows. However, our working hypothesis is that the techniques of artists and photographers mobilize the same mechanisms of recognition which we use to see objects and scenes in the world. Schier has proved a reliable guide so far, and his “two-hypotheses” thesis requires that we entertain the possibility, however remote, that our visual system at some basic level takes these paint marks to be cows. In short, I need to show that seeing a picture of a cow is far more like seeing a cow in real life than you might suppose. I propose to demonstrate this by relating aspects of how we see pictures to aspects of how we see things in the world. Following Podro, I will show how artists “sustain recognition” by mobilizing the mechanisms of visual cognition. My first example shows how pictorial composition mobilizes recognitional abilities.

**Composition, Change Blindness and Visual Metacognition**

David Wooding’s experiment at the National Gallery’s “Telling Time” exhibition in 2000-2001 tracked the eye movements of over five thousand gallery attendees as they looked at paintings. The results show that in the majority of cases we only
look carefully at parts of the picture towards which the artist’s composition directs our eye. One of the paintings in the study was Veronese’s *Christ addressing a Kneeling Woman* (fig. 55) and we can see that viewers’ eyes have been strongly drawn to the upper torso of the two main figures, to the exclusion of the other parts of the painting.

Veronese has directed the viewers’ attention through colour and brightness contrast, and by connecting the gazes of the main characters. This compositional technique is augmented by the alignment of heads and the gaze of the other characters who are looking towards Christ and the kneeling woman.

What Wooding’s experiment reveals is that contrary to Gombrich, the symmetry of composition does not work against representation of the subject, but enhances it by capitalising on mechanisms of visual cognition. We are drawn to look at the two main characters because Veronese is manipulating our visual mechanism using gaze, colour light and composition. Podro takes issue with Gombrich treating “convincing representation and the pursuit of pictorial order as reciprocally curtailing” and argues:
these demands may sustain each other: symmetries or correspondences are ways in which we see one form on analogy with another and project one aspect onto another; this is integral to how recognition is sustained and developed.169

Podro examines a similar technique of alignment, gaze and symmetry being used in Ghiberti's panel of the Adoration and Leonardo's Adoration. He further argues that, not only do deliberate techniques of the artist sustain recognition of the subject, but seemingly accidental features of a drawing or paintings, such as unfinished edges, roughly sketched lines and the texture of the paint, are recruited to the depiction of the subject. If we apply this line of thinking to the Leonardo drawing Head of a Girl, which we discussed in the last chapter (fig. 49), we can imagine that if we used Wooding's eye-tracking technique to reveal which parts of the picture we were most fixated-on, it would almost certainly be the girl's eyes. The rest of the picture would hardly get a glance. Thus, one reason Leonardo hasn’t rendered the rest of the picture is because he knows that, given his composition, the viewer simply wouldn’t attend to the hair and the shoulders even if he had drawn them.

This explanation seems at odds with Schier’s speculation that it is “frame knowledge” that enables us to discount these accidents of technique. In fact, the two explanations are compatible – both claim that our visual system effectively ignores or throws away a good deal of the visual information that presents to our field of view. This is confirmed by the research on change blindness and inattention blindness that has been conducted over the last few years. All the evidence from these studies indicates that we notice and remember very little from a visual scene. Furthermore, we vastly overestimate our own and other people’s abilities to attend to and recall what we see. Rather curiously, this phenomenon was predicted by Daniel Dennett in 1991.170 In a review of change blindness research, Laura Spinney summarises its impact on vision research:

Until the last decade, vision researchers thought that seeing really meant making pictures in the brain. By building detailed internal representations of the world, and comparing them over time, we would be able to pick out anything that changed. Then in 1991, in his book Consciousness Explained, the philosopher Daniel Dennett made the then controversial

169 Podro, Depiction, 8-9.
claim that our brains hold only a few salient details about the world—and that this is the reason we are able to function at all. We don't store elaborate pictures in short-term memory, Dennett said, because it isn't necessary and would take up valuable computing power. Rather, we log what has changed and assume the rest has stayed the same. Of course, this is bound to mean that we miss a few details. Experimenters had already shown that we may ignore items in the visual field if they appear not to be significant—a repeated word or line on a page of text, for instance. But nobody, not even Dennett, realised quite how little we really do "see".

Just a year later, at a conference on perception in Vancouver, British Columbia, John Grimes of the University of Illinois caused a stir when he described how people shown computer-generated pictures of natural scenes were blind to changes that were made during an eye movement. Dennett was delighted. "I wish in retrospect that I'd been more daring, since the effects are stronger than I claimed," he says.¹⁷¹

Grimes’ experiments tested whether subjects noticed changes in a scene introduced between saccades. In his 1996 paper, “On the Failure to Detect Changes in Scenes across Saccades” Grimes concludes that very little information is carried across from saccade to saccade.¹⁷² Grimes’ work prompted a renewed interest in the human ability to detect change and has inspired much speculation and experimentation on the change blindness phenomenon. It raises the question “Is the visual world a grand illusion?”¹⁷³.

In Dennett’s view, the whole phenomenology of perception is largely an illusion. The common conception is that there is a seamless stream of consciousness; Dennett’s view is that consciousness is a fragmented, disjointed thing and that we piece it all together in retrospect. Vision researchers who set out to explain how we arrive at the seamless detailed, integrated, deep and colourful reality which we experience are actually falling at the first hurdle. Dennett maintains that the little we do see is largely in black and white, patchy and superficial. That is, we are not registering the world in anything like the detail that we think we are and we do not recall from moment to moment most of what presents to our retinas. Dennett’s experiment with the playing card demonstrates this nicely. He comments:

> The visual field seems to naive reflection to be uniformly detailed and focused from the center out to the boundaries, but a simple experiment

shows that this is not so. Take a deck of playing cards and remove a card face down, so that you do not yet know which it is. Hold it out at the left or right periphery of your visual field and turn its face to you, being careful to keep looking straight ahead (pick a target spot and keep looking at it). You will find that you cannot tell even if it is red or black or a face card.\footnote{Dennett, \textit{Consciousness Explained}, 53-54.}

You will be aware of motion at the periphery of your visual field but will not be able to identify the card until it is almost right in front of your eyes. The reason that we cannot see colour or identify objects outside 5 degrees of dead centre is because only the receptors (cones) of the fovea around the centre of the retina generate a colour signal. The rest of the receptors (rods) cannot generate a colour signal. These facts are not remarkable – the remarkable thing is that we don’t notice this staggering deficiency in our vision. Given these, and other facts about the limitations of the human visual system (we have two upside down distorted retinal images and our eyes constantly dart around in saccade movements during which the retinal image “greys out”) it is also remarkable that vision research didn’t predict change blindness and inattention blindness earlier.

In fact, the results of the hundreds of change-blindness experiments that have been conducted since 1991 are so surprising that they have generated a new interest in the role of attention in both perception and consciousness. They have also given birth to a new discipline – visual metacognition – the study of people’s beliefs about vision. Research into visual metacognition is exploring how people not only fail to notice or attend to most of a scene, but always overestimate their ability to remember what they have seen. That is, we massively overestimate our powers of perception, attention and memory, and we similarly overestimate other people’s powers of visual cognition.

In his paper entitled “Is the Visual World a Grand Illusion?” Alva Noë summarises the significance of the change blindness work as follows:

\begin{quote}
The fact of change blindness is widely thought to have several important consequences. First, perception is, in an important sense, attention-dependent. You only see that to which you attend. If something occurs outside the scope of attention, even if it's perfectly visible, you won't see it. In one study, perceivers are asked to watch a video tape of a basketball game and they are asked to count the number of times one team takes possession of the ball (Neisser, 1976; Simons & Chabris, 1999). During the film clip, which lasts a few minutes, a person in a gorilla suit strolls
\end{quote}
onto the centre of the court, turns and faces the audience and does a little jig. The gorilla then slowly walks off the court. The remarkable fact is that perceivers (including this author) do not notice the gorilla. This is an example of inattentional blindness. Second, perception is gist-dependent. Some changes, for example, in the features that affect the gist of the scene, are more likely to be noticed (Simons & Levin, 1997). Third, it seems that the brain does not build up detailed internal models of the scene; that is, it doesn't perform the integration of information across successive fixations, contrary to the assumption of traditional orthodoxy (Blackmore et al., 1995; Rensink et al., 1997; O'Regan et al., 1999; Rensink et al., 2000; Noë et al., 2000). Or if it does, we have little easy access to this detail. If we did, then presumably we'd keep track of change better than we do.\(^{175}\)

The “basketball and gorilla” experiment to which Noë refers has now become part of the popular folklore of change blindness, and has been repeated on network television to the general dismay of unprimed viewers. Daniel Levin and Daniel Simons embarked on a series of increasingly surprising change blindness experiments in which subjects failed not only to notice changes of clothes and furniture in a scene but even failed to notice if one actor was switched for another. There was, at the time, some scepticism as to whether these change blindness results could be achieved using subjects in real world situations (as opposed to on video). In response to this scepticism, Levin and Simon developed a variety of situations where subjects were exposed to a quick substitution of their conversation partner (see fig. 56). Levin comments:

We still found that about half failed to detect this change. In one case, an initial experimenter approached a subject on a college campus and asked for directions to a building on campus. Midconversation, two other experimenters carrying a door walked between the subject and the first experimenter. While the subject's view was briefly blocked, one of the experimenters carrying the door traded places with the first experimenter, who walked off behind the door. Thus the subject's conversation partner suddenly changed from one person into another. Yet, even though the change occurred right in front of them, about half of the subjects missed it, continued the conversation as if nothing had happened, and were later quite surprised to find that the person who finished the conversation with them was not the same person who had started it (Simons and Levin, 1998). In replicating this effect using a number of different scenarios, we have found that subjects also missed substitutions both when they were photographing the experimenters and when they were receiving consent forms from them (Levin et al., 2002).\(^{176}\)

---


\(^{176}\) Daniel T. Levin and Melissa R. Beck, "Thinking About Seeing: Spanning the Difference between Metacognitive Failure and Success," in *Thinking and Seeing: Visual Metacognition in
The results of these trials amazed even the inventors of the experiment. It is hard to believe that one could fail to notice if one’s conversational partner was suddenly switched. It further confirms that we overestimate the level of visual detail which we notice and retain when we look at a scene. If indeed our visual system is working with such impoverished stimuli, are we justified in concluding that the visual world is a grand illusion? Noë says “no”. He cites the fact that despite the results of change blindness experiments we are still able to navigate the world successfully, and concludes that at the sensorimotor level at least, our visual system is giving us an adequate account of the world. I will return to the significance of the sensorimotor aspect of the visual system when I have analysed the significance of this change blindness work for a theory of depiction.

![Figure 56](image)

Figure 56  An experimenter asks a subject for directions (a), the experimenter switches places with another experimenter under cover of a passing door (b), the subject fails to notice (c)

There are at least two major implications in the change blindness studies for a theory of depiction:

---

We seem to perceive a far less densely populated and detailed image than we imagine. It is therefore possible for an artist to trigger recognition with a picture which is far less detailed than a full-colour photograph or realist painting or drawing.

If we are radically mistaken about our visual phenomenology, intuition is not serving us well in our quest to understand how we see pictures, and the technique of introspection is also suspect. It is clear that only rigorous experimentation using the methodologies of psychophysics can illuminate how we see pictures.

It should, however, be remembered that we are not attempting to solve all the mysteries of visual perception; we are merely trying to ascertain how an artist exploits the abilities (and disabilities) of the human visual system to evoke objects and scenes using pigment. Wooding’s studies of how artists use composition to direct our attention tell us something about how the configuration of a whole scene can lead the eye to significant areas. What we now need to discover is how our visual system recognises the subjects once the eye has been drawn to the relevant parts of the picture. That is, we saw in the Veronese that light, colour, gaze and presumably faces were the main attention grabbers. I’m interested in how our attention can be grabbed so quickly by key features of a scene such that recognition is triggered almost instantly. Podro, and to some extent Wooding, are interested in continuing recognition. I am primarily interested in how the interpretation gets a foothold. I suspect we need to understand the mechanism whereby we make that instantaneous rough judgement about what we are looking at, before we can analyse how recognition is sustained. I originally became interested in the change-blindness experiments because they seemed to confirm that we visually recognise things on the basis of very little or very poor visual stimulus. It seemed that this research might throw some light on how we can so easily recognise very sketchy line drawings or smudgy partial depictions such as Constable’s cows.

**Configuration, Orientation, Detail and Information**

We saw in chapter one that our raw ability to see what is depicted is not affected by style or adherence to a perspective scheme.
For example, I don’t believe that the landscape with trees and animals in the child’s version of Constable’s *Wivenhoe Park* (fig. 57, right) is any harder to recognise than the original. The omission of a sophisticated perspective system and a fairly diagrammatic approach to trees and swans does not radically hinder our ability for instant recognition. It is not a commitment to detailed realism that makes a picture recognisable, it is something far less easy to define. Obviously a realist picture, such as Constable’s painting, will reward detailed examination in a way that a child’s drawing may not. The child’s drawing, however, gives us the important essentials and it is our recognition abilities that are the key to unravelling exactly what these essentials are.

Here are two line drawings (figs. 58 & 59) one is by a blind girl called Gaia, the other by Matisse. Once again there is no radical difference in the speed of recognition of the subject in each picture but there is a sophistication in the Matisse that is hard to place. Matisse is attempting a far more difficult task than Gaia. The point of view he has chosen and the pose of the subject make this quite tricky to draw. In the drawing below (fig. 60) I have erased the eyes, nose and mouth from the Matisse drawing and, as you can see, the girl’s forelimbs and head are now a jumble of lines. Matisse’s careful orientation of the girls face sets up the pose and orientation of the reclining figure. Furthermore, the face is crucial to speed of recognition.

![Matisse drawing with face erased](image)

The point of this comparison is to demonstrate how the configuration of the image – particularly orientation – appears to be crucial to recognition. It also shows how a relatively small part of a picture, such as a face, can bring structure to the whole picture.

When we start thinking about the power of a line drawing to depict, we should remember Schier’s observation that we “comprehend the marks all at once and not seriatim.”¹⁷⁸ That is, we need to consider how the overall configuration of the marks helps start the initial interpretation. What kind of visual stage-setting is required before the marks can be recognised? Perhaps that question will need to

be recast in terms of recognition abilities when we consider how experiments in visual cognition can illuminate pictorial interpretation. We might ask: What does the visual system look for first? Does it first see an overall blurry configuration of uprights and horizontals, dark patches and light patches? How does the visual system zero-in on the salient details so quickly? The line drawings above show that the visual system doesn’t need much to trigger recognition, however, altering one feature of a drawing or photograph, can make it hard to recognise.

Figure 61
Can you recognise the figures in these pictures?
In the photograph of the ballet dancer (fig. 62, left) the head was masked\textsuperscript{179} and I inverted the drawing of the sleeping cowgirl (fig. 62, right). The curious thing about the inverted drawing is that the trickiest thing to decipher is the cowboy hat, whereas that is the most striking feature of the correctly oriented drawing.

The picture below was correctly identified by the majority of students I showed it to, provided they were looking at it from distance of about three metres. It’s difficult to guess close-up, but if you stand back a few metres you may be able to see...

\textsuperscript{179} From John P. Frisby, \textit{Seeing: Illusion, Brain, and Mind} (Oxford: Oxford University Press, 1979). Incidentally most people think that the masked picture is of a rose.
...a reclining girl. Most of the students who identified a girl reclining in the pixelated version also noticed that she had her arm up. The surprising thing about the recognisability of the pixelated version is that the original is a very rough drawing made with an eraser.

![Bather Girl](image)

Figure 64  Bather Girl  This drawing was drawn from life using an eraser.

There is very little information in the original drawing (fig. 64) and hardly any information in the pixelated version (fig. 63). We can safely conclude that recognising pictorial content has very little to do with amount of information in the picture. The richness of the image does not greatly enhance its identifiability.\(^{180}\)

Our visual system needs very little information to trigger recognition. The previous examples show that in some pictures overall configuration of parts is the key; in others isolated details seem to be crucial. There seems to be a tension between our ability to recognise small details and our ability to see the whole configuration. On the next page you will find four images that have been masked with a white mask. Maybe you can guess what they are. If you then turn another page you will see the same images with the mask coloured grey.

Figure 65 Camouflage pictures
Most people report that they cannot see the figures, or see the figures with difficulty, in the original fragmented pictures, but clearly see an elephant, a cheetah and a box in the images with the grey plant-mask. The fourth image is a Salvador Dali drawing from his Don Quixote series – it’s quite hard to get. Why is it easier to see the figure when we can see what is masking it from view? There is no more information, in terms of lines, in the latter set of images, however, it is much easier to see the figure, and our identification of it, and its pose, is suddenly very definite. I suspect that although the plant leaves don’t add explicit information they enable us to work out which bits of the figure are hidden. The leaves enable us to work out which elements are connected and thereby give context to small details.
For example, our visual system automatically adds rough contours of the cheetah which are hidden behind leaves (fig.67).

**Camouflage, Evolutionary Psychology and Gestalt**

This “find the elephant in the bushes” trick was used by the Gestalts to demonstrate how our visual system groups disparate element of a scene. The Gestalts, unfortunately, had some bizarre ideas about how our visual brains achieved this grouping, so their account of the physiological mechanism which enables us to do this is of no use. However, evolutionary psychology may help explain many of these strategies of our visual system. V.S. Ramachandran goes so far as to suggest that “vision evolved mainly to defeat camouflage and to detect

---

objects in cluttered scenes".\textsuperscript{182} The visual system automatically tries to “glue together” elements in a scene with similar colours or textures because it has evolved to spot things hidden behind foliage. Ramachandran speculates that such principles of visual brain organization might explain why many of the Gestaltist observations about visual organization have been used so successfully by artist and designers. That is, the Gestalts were right about the principle of grouping but wrong about the physiological processes underlying our grouping ability. He goes on to relate this grouping principle, which enabled our “arboreal ancestors”\textsuperscript{183} to spot a lion hidden in a bush, to how artist and designers use colour and shape repetition to achieve their effects:

Little does the salesgirl at Nordstrom realize that when she picks the “matching” red scarf for your red skirt, she is tapping into a deep principle underlying brain organization, and that she’s taking advantage of the fact that your brain evolved to detect lions seen behind foliage — so grouping “feels good”. Of course the red scarf and red skirt are not one object, so logically they shouldn’t be grouped, but that doesn’t stop her from exploiting the “grouping law” anyway, to create an attractive combination. The same holds for paintings and mattes, or even blobs of similar color on different objects within a painting. The point is the rule was statistically valid in the treetops in which our brains evolved. It was valid often enough that incorporating it into the visual brain centers as a law helped your ancestors leave more babies behind and that’s all that matters in evolution; the fact that an artist can misapply the rule in an individual painting, making you group splotches from different objects, is irrelevant, because the brain is “fooled” and it enjoys the grouping anyway.\textsuperscript{184}

Matisse is clearly using this principle in his drawing of the girl reclining next to the plant. If you look at the shapes which denote her elbows you can see that they mimic the shape of the leaves. Even her face is leaf-shaped. Veronese has also used the same blue on the robes of Christ and the dress of the kneeling woman to concentrate attention on these figures. If Ramachandran is right and this grouping principle is a “deep principle underlying brain organisation”, it could help explain many features of pictorial composition. This particular principle not only explains how we can recognise objects based on very little information but explains how our attention can be directed using similarities in shape, texture and colour.

\textsuperscript{183} Ibid.
\textsuperscript{184} Ibid.
This discussion is moving very freely between how we see objects in the world and how we see objects in pictures. The discussion presupposes that the abilities which enable us to see things in pictures are the very same abilities which enabled our “arboreal ancestors” to see things behind bushes. In order to head off the accusation of circularity, I would like to resolve some methodological issues before I further analyse what vision research can tell us about how we recognise objects in pictures.

The first methodological issue was raised by Hopkins, who claims that Schier’s account fails to adequately show that pictures “engage the same processing operations as what they depict.” It would seem that before we can enlist vision science in our quest to understand pictorial recognition, we need to show that the visual processes being triggered by pictures of objects or scenes are the same or similar to those triggered by the objects and scenes depicted. Just as the above observations about how we see pictures do not necessarily tell us anything about how we see the world, the reverse is also true. It is not necessarily the case that because we use a particular perceptual mechanism when recognising things in the world, we use the same perceptual mechanism to recognise pictures. If I merely present evidence from vision research which reveals how visual cognition works – by what methodological principle can I relate that to how we recognise pictures? Wooding’s work tells us something about how we look at pictures, and how an artist can manipulate our visual processes, but there is no more than circumstantial evidence that the process which the artist is manipulating is one which we use when we look at real world objects. One might object, for example, that the world is not a composed place, and despite what Ramachandran says about the evolutionary advantages of visual grouping, it is not self-evident that the process at work in the painting is one we use in real-life. A similar objection could be aimed at my camouflage examples. In Ramachandran’s view, when we recognise the objects in the camouflage pictures we are using the same processes our simian ancestors used to spot predators and prey in the jungle. This leaves Ramachandran open to a further objection that his observations are no more than conjectures – or “just so” stories as they have been called. In fact, this is a generic objection to evolutionary psychology which I believe can be answered by an appeal to the

methodological principles of vision science and evolutionary psychology. However, before I invoke these methodological principles, I need to show that the same recognition abilities are in play when we look at pictures as are in play when we look at objects. This will establish that Schier’s hypothesis is correct in principle that research into human visual cognition contributes directly to a theory of depiction, and *vice versa*.

**The Assumptions of Vision Researchers**

In “El Greco’s Eyesight” Hopkins attacks Schier’s recognition-based account of depiction on the grounds that it does not sufficiently support the assumption that:

> Pictures act as visual substitutes for what they represent, in the sense that they provide alternative causes for (something like) the same visual effects.\(^{186}\)

Hopkins claims that Schier’s account fails to support the assumption because it fails to adequately show that pictures “engage the same processing operations as what they depict.”\(^{187}\) Hopkins’ attack on what he calls “the processing assumption” is aimed at vision researchers who use pictures as substitutes for real objects when conducting experiments. His project in “El Greco’s Eyesight” is to test whether Schier’s theory might serve to support the assumption. He observes that:

> a good deal of empirical research into vision…..uses pictures of things in investigating the perceptual processes at work in our cognizing the things themselves. But if pictures do not induce the same effects as what they represent, we can hardly study the processes at work in cognizing the latter by showing subjects the former.\(^{188}\)

I would go further than Hopkins and say that the *majority* of research on vision over the last forty years has used pictures, as opposed to three dimensional objects, as its “target” material. It seems for the purposes of vision research, pictures are an adequate substitute for a real three dimensional object. In nearly all recognition experiments, subjects are shown computer generated images or photographs or drawings. This classic experiment by Cutzu and Tarr\(^{189}\) on canonical views and saliency features is typical. The experiment required subjects

---

\(^{186}\) Ibid.: 441.  
\(^{187}\) Ibid.: 455.  
\(^{188}\) Ibid.: 442.  
\(^{189}\) Florin Cutzu and Michael J. Tarr, "Inferring Perceptual Saliency Fields from Viewpoint-Dependent Recognition Data," *Neural Computation* 11, no. 6 (1999).
to compare views of this computer generated cat (fig. 68) for “goodness-of-view” (which was defined as being “more informative” or “more representative”). They were then asked to rate views for similarity. In fact the experimenters were testing which parts of the image of the cat (ears, paws etc) were being used to make the comparisons.

The features which turn out to most significant for identifying cats are the light areas shown in the illustration below (fig. 69). It may be possible to surmise that features such as the muzzle, legs and tail are the most invariant from different angles. It would be interesting to find out what the results would have been with a real cat. Unfortunately, there appear to be very little research on how experimental subjects recognise real three dimensional objects in normal environments except those which are aimed at discovering how our haptic system recognises shapes. It seems that the bulk of object recognition research is actually research into recognition of depictions of objects. Thus, one could say that much vision research is actually work on how we perceive pictures. This work extrapolates results based on how subjects respond to pictures to how human beings see the world. This is good news for a theory of depiction, but probably not so good for a generalized theory of vision and recognition. If I was to characterize how I would recognise a real three dimensional cat in the real world – I would expect fur, I would expect it to be less than half-a metre long (if you include the tail), be sitting on a mat, in a doorway, on a wall or in a garden, it would have little pointy ears
Mechanisms of Recognition

and round eyes – in short, I’m not sure that many of the Cutzu and Tarr saliency fields really address how cats are encountered in the real world.

Figure 69  The top cat is derived from goodness of view data and the bottom cat from view similarity data. The white areas are areas of maximum saliency.190

A major objector to the assumption that pictures are adequate substitutes for objects was J.J. Gibson, who maintained that “The perplexities connected with the making and seeing of pictures are problems in their own right, independent of the problems of direct visual perception.”191 He believed that it was harder to get information from the static display of a picture than it was from the ambient array of light which we are presented with in real life. As his wife Eleanor Gibson, in a foreword to his collected essays, explains:

It became clear to him, especially during his World War II experience, that a psychology of perception based on drawings (as so much of it had been) simply would not do. In a real world, things are solid and extended; we

190 Ibid.
move around in it, and often things move around us. It turned out that what seemed at first to be complications produced by all these changes are really advantages, and that static perception, on the other hand, is drastically impoverished.\textsuperscript{192}

The “World War II experience” Eleanor Gibson is referring to, is Gibson’s work with the Army Air Force, on the practical problems of the takeoff, flying and landing of aircraft. These could not be studied in a laboratory because the problems all arise when the aircraft is in motion. Gibson established that patterns of movement, flowing textures and gradients over the ground, all furnished information for depth, distance and height. This work changed the course of his research and thereafter he began to study how the reciprocity of a creature and its environment afford the information a creature needs to survive in the natural world. His study of vision emphasised the capacity to avail oneself of the information in ambient light by moving one’s head, and moving around. In his opinion, showing static pictures to a subject with her head in a clamp was not a proper study of vision. Gibson was sceptical that testing subjects with pictures is functionally the same as testing them in real life situations. The reasons which Gibson enumerates relate to his concept of active vision. However, there is another reason to be sceptical of what these experiments can tell us about real world visual cognition – if you look at the drawings, photographs and CAD models which experimenters actually use, one is struck by how little they look like the real life object or animal. The Snodgrass and Vanderwart’s object pictorial set,\textsuperscript{193} for example, which is so often used in these experiments, consists of some very rudimentary line drawings. The assumption of these vision researchers, that a picture is the functional equivalent of an object for recognition purposes, seems to further include the subordinate assumption that a photograph, a drawing and a CAD drawing, of whatever quality, are also equivalent.

These facts seem quite damning for vision research into how we see the world, but looks promising for research into depiction; the last forty years of vision research has effectively been research into how we see pictures!


Hopkins complains that vision researchers are silent about the grounds for their assumption that pictures are adequate substitutes for objects. In fact, vision researchers are not entirely silent on the assumption; one of the chief justifications for using two-dimensional targets, instead of solid objects, is that the image on the retina from which we generate our perception of depth and solid shape, is two-dimensional.

**Ramachandran and the “Bag of Tricks”**

Ramachandran begins his 1988 paper “Perceiving Shape from Shading” thus:

> Our visual experience of the world is based on two-dimensional images: flat patterns of varying light intensity and color falling on a single plane of cells in the retina. Yet we come to perceive solidity and depth. We can do this because a number of cues about depth are available in the retinal image: shading, perspective, occlusion of one object by another and stereoscopic disparity. In some mysterious way the brain is able to exploit these cues to recover the three-dimensional shapes of objects.\(^{194}\)

The importance of the two-dimensionality of the image cast on the retina is also emphasized in Marr’s computational approach to vision and in some of the empiricist work. The assumption which Marr makes, for example, is that the human visual system and the computer visual system are dealing with the same high-level computational problem – how to determine “what” and “where” from two-dimensional data. Ramachandran summarizes Marr’s position thus:

> According to Marr, since any given computational problem can be tackled by many different algorithms and any given algorithm can be implemented in different kinds of hardware, these different levels of description must be kept quite separate and we must be very careful not to get confused between them.\(^{195}\)

This is sound advice for an engineer building a visual system from scratch. But our task is to determine how evolution engineered the human visual system, and what Ramachandran suggests is that evolution hasn’t engineered the human visual system using the kind of sound engineering principles which Marr advocates. The human visual system evolved over hundreds of thousand of years through a series of adaptations. Ramachandran advocates that we deliberately confuse Marr’s three


levels of description and assume that the “computational problem” is constrained by the hardware which evolution had to work with at each stage in this long adaptive process.\textsuperscript{196} Thus, even if the high-level computational problem is the same (which is open to dispute) the hardware implementation which evolution has handed us changes the way that the solution can be implemented and therefore, to some extent, changes the high level problem.

The danger of Marr’s approach is that if we work on the assumption that evolution has developed elegant engineering solutions to the problem of vision, we may end up building what Daniel Dennett calls “cognitive wheels”\textsuperscript{197} – a great solution to a problem which the human brain faces, but not the solution which evolution has actually handed us. And, more importantly, not a solution that is possible given the constraints of our visual hardware. In Ramachandran’s view, the mechanisms of recognition, which neurologists and psychologists are searching for, are a “bag of tricks”:

According to this view perception does not involve intelligent reasoning as implied by some psychologists; does not involve resonance with the world as suggested by Gibsonians; and does not require creating elaborate internal representations or solving equations as implied by AI researchers. One could argue, instead, that perception is essentially a ‘bag of tricks’; that through millions of years of trial and error the visual system has evolved numerous short-cuts, rules-of-thumb and heuristics which were adopted not for their aesthetic appeal or mathematical elegance but simply because they worked (hence the ‘utilitarian’ theory).

This is a familiar idea in biology but for some reason it seems to have escaped the notice of psychologists who seem to forget that the brain is a biological organ just like the pancreas, the liver or any other specialized organ……

It may not be too far-fetched to suggest that the visual system also uses [a] bewildering array of special-purpose tailor-made tricks and heuristics to solve its problems. If this pessimistic view of perception is correct, then the task of vision researchers ought to be to uncover these rules rather than attribute to the system a greater degree of sophistication that it simply does not possess. Seeking overarching principles may be an exercise in futility.\textsuperscript{198}

Ramachandran’s approach recommends itself to developing a theory of depiction because it allows us to be relatively agnostic about the “overarching

\textsuperscript{196} Ibid.
\textsuperscript{198} Ramachandran, "Interactions between Motion, Depth, Color and Form: The Utilitarian Theory of Perception."
principles” which govern perception. There are a plethora of ‘large’ theories of perception and each is freighted with a heavy burden of philosophical, psychological and scientific issues which do not necessarily concern us in our pursuit of an explanation of how depiction works. To the extent that we are prosecuting a recognition based theory of depiction, it would seem inevitable that if a theory of depiction emerges in these pages that it must attach to an existing theory or show allegiances to an emerging theory. Ramachandran’s approach may serve our purposes for now, given that it offers the prospect of comparing artists’ “bags-of-tricks” with the visual system’s “bag-of-tricks” and getting a purchase on the slippery issue of how we can relate the two. His general principle that the brain, and in particular the visual brain, is a specialized, but jerry-built, piece of hardware which cannot be understood by applying sensible engineering principles seems like a good rule-of-thumb. For example, Ramachandran’s argument that the way that human beings derive shape-from-shading “is unlikely to be based on a single algorithm or single constraint”¹⁹⁹ seems to be borne out by the evidence.

![Figure 70](image.png)

**Figure 70** An illusory circle was superimposed on a simple one-dimensional luminance-ramp. This creates the impression of an illusory ‘sphere’ even though there is no change in shading across the border of the sphere. (Ramachandran, 1998)

**Shape from Shading**

The example of the “illusory sphere” (fig. 70) is used by Ramachandran to illustrate that when we look at a ping-pong ball (or any spherical Lambertian surface), our visual system isn’t calculating the curve of the ‘sphere’ using an algorithm based on the cosine function which calculates how the light falls off from the illuminated pole towards the equator²⁰⁰.

In figure 70 it is possible to see a sphere in the drawing on the far right, despite the fact that the picture is composed of four half-circles (centre) superimposed on

¹⁹⁹ Ibid.
²⁰⁰ Ibid.
the luminance ramp (left). What Ramachandran finds interesting is that we see the
sphere despite the fact that the shading which defines its shape is actually the one-
dimensional luminance-ramp which is also its background, and there is no
boundary separating the sphere from this background at its equator and poles:

The implication is that the segmentation boundary that delineates an object
from its background can powerfully influence the processing of shading
information. If the visual system were making detailed measurements of
shading to recover surface orientation there would be no basis for seeing a
‘sphere’ in this image since the shading does not change abruptly across
the border of the sphere.201

In fact, the longer you look at the image the more powerful the “cue ball”
effect gets. Ramachandran’s larger point is that the visual system has a variety of
ways of recovering shape from shading, but solving equations isn’t one of them.
In his paper “Perceiving Shape from Shading”, he demonstrates that it is not the
shading alone which determines the shape we recover, but the outline, elementary
features such as oriented edges and the visual system’s knowledge of objects such
as spheres. He also concludes that the human visual system makes at least two
major assumptions when looking at an object:

The extraction of shape from shading information incorporates at least two
"assumptions" or constraints - first, that there is a single light source
illuminating the whole scene, and second, that the light is shining from
"above" in relation to retinal coordinates.202

Ramachandran has discovered what every photographer and cinematographer
learns on the first day of their training:

A photograph will look natural if lit from a point source positioned at least
45° above the horizon and the modelling on the shaded side is best filled
by a broad reflected illumination.

This rule-of-thumb, adopted by painters and photographers, is crucial not just
for recovering the shape of the depicted object, but for the object to look
“natural”. The lighting scheme emulates how an object looks when the sun is in
the sky and there is some ambient light around. In order to discover which cues
trigger recognition in pictures, we need to correlate the kinds of “gimmicks”
which Ramachandran and other researchers have identified in the visual systems’
“bag of tricks”, with the techniques which artists use to “mobilize” these

201 Ibid.
mechanisms. In their 1992 paper, Kleffner and Ramachandran speculate on the significance of their “shape-from-shading” findings for cognitive studies:

Taken collectively, these findings imply that the extraction of shape from shading is an “early” visual process that occurs prior to perceptual grouping, motion perception, and vestibular (as well as “cognitive”) correction for head tilt. Hence, there may be neural elements very early in visual processing that are specialized for the extraction of shape from shading.203

Subsequent studies have confirmed their suspicions that shape-from-shading is a very “primitive” visual mechanism. For example, in their paper “Differences in perceived shape from shading correlate with activity in early visual areas”, Humphrey, Goodale et al. confirm that the early visual areas are activated when determining depth and shading. They further speculate that there could be a “specialized low-level network for computing shape from shading”.204

Ramachandran’s justification for using pictures rather than real objects assumes that recovering depth from a retinal image is functionally equivalent to recovering depth from a picture. I have no idea if this is the general assumption that vision researchers using pictures as targets assume, but it may be that in individual instances such an assumption is justified. For example, it is hard to see how he could have conducted his “shape from shading” experiments using three-dimensional bumps and dimples, because the point of his study was to highlight the assumption subjects make about the lighting source. If you look at the targets below (fig. 71) you will note that these lighting conditions would be almost impossible to rig with real life dimples and bumps.

![Figure 71](image)

Figure 71 The apparently conflicting light-sources shining on the dimples and bumps influence our perception of the depth of the centre shape. (Kleffner & Ramachandran, 1992)

203 Ibid.
Nevertheless, suppose that somehow Ramachandran conducted a series of parallel shape-from-shading experiments with real bumps and dimples and picture bumps and dimples and achieved the same responses from the subjects. That is, suppose the parallel test established that, for these particular shape-from-shading experiments, pictures and objects were functionally identical. Could we then accept his assumption that we recover depth from pictures and objects in the same way? This kind of parallel test would seem to be the acid test of the Schier position. That is, if we could conduct some kind of brain imaging experiment which showed that some of the underlying brain processes were the same when a subject looked at an object and a picture of that object, we would have a confirmation of his “overlapping recognition abilities”. In fact, as I will show later, this is not as difficult to show as one might imagine. The problem with such an experiment is that it is very easy to find “overlapping” brain processes triggered by object and picture – the problem is identifying which ones are the crucial overlapping functions. Hopkins defines this problem as defining “which effects pictures are taken to share with their objects, that is, at which stage in the visual processing chain the match in processing supposedly occurs.”

He pursues the notion that it may be possible to establish at which stage in the processing chain that the stimulus and response match and at which they differ. However, he points out that the assumption:

\[
\text{cannot claim of any stage in the processing chain, that from there onwards the match in processing between the picture and the object cases is complete. For if pictures had precisely the same effects, either at the retina or anywhere further down the processing chain, as the objects they represent, it would be impossible to account for our ability, which survives almost every circumstance, to distinguish pictures from what they represent.}\]

Of course Hopkins is correct that we are usually aware that we are seeing a picture and not a real object, but he is mistaken to assume that visual cognition is a serial processor. One thing vision researchers seem to agree on is that the visual system is in fact several systems all working in parallel. It may be the case that pictures trigger the same mechanisms of visual cognition as objects \textit{as well as} a mechanism that tells us we are looking at a flat picture. In Schier’s terms – the

---

206 Ibid.
processes which tell us that we are looking at a peach happen simultaneously with the processes which tell us it is a picture-peach. Picture perception, by this account, is something like ordinary perception with a side-order of flatness awareness.


Despite the assertion by the majority of depiction theorists that the ability to make a clear distinction between object and depiction is crucial to the depiction experience, it is not always the case that we do make that discrimination. For example, at the theatre we may be fooled into believing that objects which are painted are three dimensional. When we look towards the stage, we may believe we are having the experience of seeing dining room doors opening onto a patio garden when we may actually be looking at paint on canvas. If we conducted psychophysical and neurological studies of subjects having both experiences – the view of the real patio garden and the painted version – how likely is it that all, or any, of the visual processes would turn out to be identical? The fact that we don’t consciously notice the difference between the experiences doesn’t mean that they don’t trigger fundamentally different brain processes and abilities. Schier’s hypothesis, however, suggests that depictions work, not necessarily by triggering identical abilities, but by triggering a subset of the same abilities. Even in cases where we might not think there is enough in common between a picture and an object to trigger similar recognition abilities (eg a rough sketch or Constable’s cow) Schier’s account suggests that there must be some underlying triggers which are the same. So we must beware of assuming that our awareness of the marks is an indication that the underlying abilities being triggered are different from those triggered by full-colour theatrical illusions. We must also beware of assuming that because we understand the trickery being used to produce the picture, that this also indicates that the recognition abilities being triggered are not natural.

We know that when we watch a film our visual system is looking at 24 still photographs per second flashed on the screen in quick succession. In between each frame is darkness.\(^{207}\) This doesn’t seem a very natural way to see a scene. Nevertheless if we stand in a room with a large curved screen and see a film from

---

on board a roller coaster, we will have trouble standing still. Our visual system processes the movement in the film as if it is real movement, and our body responds accordingly. There is a similar effect when you are sitting on a train in a railway station and the train next to you moves away — for a moment your body has the experience that the train you are on is moving. This is a powerful visceral effect which can only be offset by looking at some point of reference which confirms that the train you are on is stationary. Film technology can only work because it exploits such fundamental perceptual process concerning motion and object recognition.

It is clear that the justification for using pictures and not real objects is:

1. Pictures engender roughly the same psychological responses — for example, a baby reacts in the same way to a real smiling face and a picture of a smiling face.

2. Pictures are as easily recognisable as objects — if I show you a trumpet and a picture of a trumpet your response time will be roughly the same.

Thus for psychophysical experiments which gauge levels of response by response speed, pictures are the perfect “target”. It is clear that a picture can engender all or some of the psychological effects of looking at a real object. Some aspects which are common are that we can often identify properties such as shape or more abstract properties such as weight, or in the case of a picture of a face, we might identify a mood. Thus at this commonsense level we can agree that there is an overlap. We have discounted accounts of depiction based on conventionalism and resemblance and are thus committed to a cognitive account. When we ask what evidence for overlap there is at a cognitive level, we are therefore bound by the evidential rules that govern psychophysical and neurological studies. In the case of Ramachandran’s shape-from-shading experiments, for example, we appear to have evidence that the way people obtain shape information from two-dimensional targets has similarities with the way we obtain shape information

208 There is such a room at the amusement park in Stockholm – but I’m sure there are many more at other amusement parks.
209 Johansson work with ‘point-light figures’ and such work on perception of biological motion, indicates how little information is required for our visual system to recognise complex aspects of human motion and bodily disposition. Subjects can ascertain how balanced a person is, and even guess what kind of weight they are lifting, just by the movements of 16 point source lights at major joints. It seems likely that in figure drawing artists exploit similar perceptual sensibilities about the human form in order to evoke the mood and balance of a figure.
from three-dimensional objects. That is, his experiments confirm certain assumption that people make about light sources. The important point is that even if Ramachandran’s theoretical assumption equating the two-dimensional image on the retina with two-dimensional drawings turns out to be false, that does not undermine his experimental methodology. We have to assume that Ramachandran and other vision researchers adhere to sound psychophysical principles. That is, the objective of the experiment is clear; the subjects have been prepared in such a way that no unfair bias has been introduced, the methodology is transparent and the whole experiment supervised responsibly. One does not need to agree with the theoretical position of a researcher to benefit from the result of his experiment. The assumptions of vision researchers are irrelevant if they have adhered to psychophysical methodology. If the theoretical position of researchers (whether they be vision researchers or any other) undermined the value of their experiments there would be no point in conducting scientific experiments. The reason experimental methodology exists is so that competing theoretical views can be tested against agreed and verifiable standards. In theory, we can learn as much from an experiment which purports to show that “Outline-shape information may be particularly important in the recognition of an object” as one which demonstrates the opposite. In short, the evidence from the last forty years of vision research indicates that vision researchers of all persuasions, excepting advocates of direct perception, have found very little evidence that recognising pictures and recognising real objects involves different visual processes. In principle we have identified a sound methodology for making Schier’s case.

The second methodological issue which I raise involves the accusation that the claims of evolutionary psychology, such as that various aspects of visual cognition evolved to spot predators behind bushes, are no more than “just so” stories. That is, they do no more than suggest possible reasons how certain human abilities and traits might have arisen through adaptation. Once again, scientific methodology takes care of this objection. Ramachandran calls it the “what, why and how” of the biological perspective. He makes this point in his interview with Anthony Freeman on his controversial eight laws of art:

In order to understand any complex mental attribute in humans—be it humour, art, dance, or sex, one needs to have in place three cornerstones: First, the underlying functional logic (e.g. what I call ‘laws’). Second, the evolutionary rationale, i.e., speaking teleologically, why do the laws have the ‘form’ that they do? (e.g. evolution has wired into your brain the ‘rule’ that grouping is pleasing and attention grabbing). Third, an understanding of what is the neural hardware in the brain that mediates the law in question. As a specific example of these three ‘corners’ of an argument consider the following.

In the ‘law of grouping’, the functional logic is to link scattered fragments into a whole. It’s evolutionary rationale is to help defeat camouflage and find objects in noisy environments: vision evolved in our primate ancestors mainly to find objects quickly and efficiently but not infallibly. And third, we suggest that as soon as the fragments are bound, there is a synchronization of neuronal spikes of those neurons (Singer and Grey, 1995) that fire for different parts of an object and it is this synchrony that causes an ‘AHA’ reward signal to be sent to the limbic system (Ramachandran and Hirstein, 1999). So while most artists, fashion designers and art historians may be aware of the grouping law they may not be aware of the evolutionary rationale nor of the neural mechanism (synchrony of spikes causing a reward signal to be sent to the limbic system). And we attempt to do this not just for grouping but for all eight of our laws.

Finally, as Richard Gregory points out in his commentary, we make novel predictions and propose specific experimental tests (e.g. our physiological experiment on Picasso). Without such empirical tests, ‘theories’ of art are merely intellectual exercises of the kind philosophers engage in.212

In this interview, Ramachandran effectively throws down the gauntlet to critics of his “science of art” approach and makes the claim that he can test his theory of why Cubism works using neurological scans. His point is that he has a theory based on evolutionary psychology and it is, in his view, provable or disprovable using the technique of cognitive science. We await the results.

Clearly we cannot assume that everything about vision cognition can be traced back to some evolutionary adaptation. Dennett notes that the rules-of-thumb for reverse engineering evolutionary adaptations were proposed back in 1966 by George Williams. These rules include:-

1. Don't invoke adaptation when other, lower-level, explanations are available (such as physics).
2. Don't invoke adaptation when a feature is the outcome of some general

developmental requirement.

3. Don't invoke adaptation when a feature is a by-product of another adaptation.213

We can be absolutely sure that our ability to see content in pictures is not a specific evolutionary adaptation. Just as human beings cannot have evolved a specific adaptation for reading words, natural selection could not have evolved a module for such a recent activity as seeing depictions. In order to understand how we see pictures we simply need to understand how we see. These kinds of methodological constraints ensure that when we speculate that the reason we can see the elephant hiding amongst the bushes in my camouflage pictures, we invoke principles which have scientific credentials. Crucially for this enquiry, we can now be confident that if we can observe a phenomenon of visual cognition in pictures, it most likely has it roots in how our visual system evolved to recognise things in the world.

Perhaps the most daunting prospect for developing a theory of depiction based on a theory of visual cognition is that there are so many theories and so much experimental work. Luckily we have restricted ourselves to explaining what happens in the first half-a-second of looking at a picture. We can afford to be entirely agnostic about the allegiances of various theories because if it turns out that the mechanism whereby we see objects in pictures is, for example, that object-specific neurons are just waiting to fire even at the hint of their object (unlikely as that seems), then that is the winning theory. A theory of depiction needs to integrate with theories of visual cognition that identify brain processes; this is key to understanding how we see pictures.

The Smell of an Oily Rag

The “bag of tricks” principle suggests that there are very few overarching principles which can tell us what makes one picture more recognisable than another. The examples of “Bather Girl” (figs. 63 & 64) and the ballerina (figs. 61 & 62) show that sometimes the whole configuration is the key and other times a small detail. How does an artist know exactly how much detail and which details are enough to trigger recognition? In their 2003 paper “Line and Borders of Surfaces: Grouping and Foreshortening”, Kennedy, Juricevic and Bai test the

intuition that the salient parts of a figure will be at junctions.\textsuperscript{214} In theory, the most “information bearing” parts of the cheetah would be where the leg meets the body or the tails meets its rump. Their experiments completely confound this intuition.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{cups.png}
\caption{Cups with partial deletion of outline. Omitting sections between vertices (\textit{left}) creates the same problem for perception of the cups as omitting sections with vertices (\textit{right}).\textsuperscript{215}}
\end{figure}

They conclude from examples such as the above (fig. 72) that omitting the section of a figure at junctions (3 cups on right), makes it no harder to identify than omitting the section between junctions (3 cups on the left). Thus, it is not easy to predict what will be the most salient feature of a figure. I found this to be the case when I was preparing the examples of the hidden animals. As I moved the plant mask around the image, various parts of it would be hidden or revealed. I was convinced that revealing the toe-nails and tail of the elephant, combined with the junction of the legs with its body would be a “dead giveaway”\textsuperscript{216}.

\begin{flushright}
\begin{footnotesize}
\textsuperscript{215} Ibid., 338.
\textsuperscript{216} Hence the joke: “Why did the elephant paint its toenails red? Answer: So it could hide in a cherry tree.”
\end{footnotesize}
\end{flushright}
In fact, I don’t believe that these details are as informative as I had thought (see fig. 74). It’s probably the tusk, trunk and ears that do the work.

A picture can be badly drawn and contain very little information but still trigger recognition. It would seem that visual cognition runs on “the smell of an oily rag”. I would like to explore three related pieces of research on visual perception which indicate that:

1. Our visual system prefers the kind of sparse abstract stimuli which a line drawings and rough sketches provide;
2. The reason we can easily see the camouflaged figures is that our visual system is ignoring the marks and tracking the objects;
3. Our visual system is primarily motivated by sensorimotor contingencies not by the need to perceive and store representations of scenes.
Taken together, these three approaches to visual recognition can indicate the character of the mechanisms of recognition which are most significant in that first half-a-second of encountering a scene.

The first piece of research I would like to look at involves the use of brain scans. In fact, it involves precisely the experiment that I imagined earlier where the subject looks at an object and then looks at a picture of that object. I also mentioned earlier that there were some questions concerning the concept of “overlapping brain functions”.

The first question is:

*Is it possible that a brain scan could confirm that looking at an object and a suitable depiction of that object engaged overlapping brain functions?*

The answer is definitely “yes”.

The second question is:

*Does the fact that overlapping brain functions are triggered by an object and a picture of that object indicate that the underlying recognition processes are the same?*

The answer is almost certainly “no”.

I put this scenario to the neurologist Semir Zeki who pointed out that it would be more useful to identify which brain activity *didn’t* correlate. His reasoning goes something like this:-

1. There would be so many overlapping cues of colour, lines and textures for, let’s say a picture of a fire engine and a toy fire engine, that it would be surprising if the same modules in the brain didn’t light up.
2. It would probably be the case that largely the same areas would light up for a red fire engine and a red car (so we are looking for differences).
3. We can easily tell the difference between the picture and the object, so areas would “light up” that are not common to both visual experiences; these may indicate areas which a) don’t respond to pictures, 2) respond to pictures but not objects.
4. It is the connectivity of entire systems and pathways in any given function, rather than single areas acting in isolation, that reveals most about what the brain is doing.

Zeki also commented that there were practical problems associated with such an experiment, not least of all the reliability and resolution of positron emission
tomography (PET) – brain scans based on increases in cerebral blood flow. Of course there are other kinds of brain scanning methods which measure brain activity in more precise ways than PET, but each method has its own advantages and drawbacks.

There is also a further complication which requires us to be more precise about what we mean by a brain scan. We noted earlier that the visual system is best regarded as several systems working in parallel. If looking at a cat and looking at a picture of a cat triggered exactly the same brain activity in one part of the system for, let’s say the first 200 milliseconds, I think we could reasonably say that Hopkins was wrong and that at this point some underlying visual processes are the same – the same recognition-clusters have been triggered. Schier imagines something like this when he constructs his two-stage theory of picture interpretation. In his example, in the first stage we momentarily entertain the hypothesis that the picture of a peach is a real peach, and then reject it because the cues for a flat picture-peach are more convincing. Establishing that a single overlapping process was identical for the first 200 milliseconds vindicates Schier, but does not fully answer the question about which cues trigger recognition. We need to look at the other processes happening in parallel if we are going to be able to say exactly how the cues in the picture are doing the work. For example, even if the processes in the pathways of the early visual system from the retina to the lateral geniculate nucleus are the same for both a cat and a picture-cat, we need, for completeness, to take into account how the raw data from the lateral geniculate nucleus onward is identified by the system as a peach. Surely some area in the higher brain involving memory must also be involved in the actual recognition of catness?. To put this in neurological perspective, I’d like to consider the case of a patient, cited by Zeki, suffering from a visual agnosia. The patient in question could see objects and people and describe their parts – for example, colour or shape – but could not identify familiar people and familiar objects.
The patient could see and draw complex objects but could not make any sense of the object which he had drawn. The drawing above (fig. 75) is a copy of a picture of St Paul’s cathedral made by such a patient. The question is: “How can the patient copy a picture without having some concept of what the picture depicts?” Zeki comments:

There is good agreement in the literature that the drawing is piecemeal, small segments of the picture or of its outline – segments that the patient can see and understand – being drawn one after another. Once drawn, the patient can still only recognize small segments of the drawing and not its entirety. The patient’s report of the process itself is more or less uniform. One patient stated that when he copied a complex figure, ‘all he saw was a complex pattern of lines, which did not correspond to a particular object.’……the authors of this fascinating report state that the patient ‘has intact registration of form elements (single lines, edges), but…his ability to integrate these elements into “perceptual wholes” is in some way impaired. The intact information about the local form elements enables him to make accurate copies of stimuli he cannot identify’.218

Zeki does not believe that the behaviour of this patient, and others like him, can be explained by a separation of seeing and understanding, or perception and association. He maintains that damage to the integrative pathways in the brain is a more likely explanation. My point is that presumably a brain scan would show that in the first 200 milliseconds of looking at the picture this patient registered something…….I’m really not sure what. My assumption with the Constable

218 Ibid., 314-15.
Mechanisms of Recognition

painting is that we take in the whole picture at once. This patient is able to get the parts but not the whole. Zeki’s assumption is that it is failure to integrate the parts that is the problem with this patient’s vision. And I must confess that is the assumption which I first made. Zeki (and I) have assumed what many a vision researcher and art historian before us has assumed – that we build up our impression of the whole scene from the parts. But what if the part of the brain that was damaged is the recognition mechanism which is prior to or even parallel to recognising the individual parts? Could this patient merely have lost the ability to get the gist of a scene? I am speculating here that one needs to get a sense of the overall configuration of a scene before you can integrate its parts. Without this sense of the configuration of the whole one would be like the apocryphal blind men feeling different parts of an elephant and coming to different conclusions about its shape. I will return later to this idea that there are multiple visual systems for gist and parts.

The St Paul’s example show that a recognition theory of depiction based on current work in visual cognition presents its own, heightened, version of the perennial puzzle about how we simultaneously see the marks on the paper and the object depicted. In the cases which Zeki cites, the patient can see the marks and shapes but is still unable to see what is depicted. It is important to remember that patients who suffer from this kind of agnosia can describe the properties of actual objects and still fail to recognise them. The disability is not restricted to pictures. This is a powerful confirmation that the kinds of cognitive abilities which enable us to see and recognise objects in pictures overlap with the abilities to see and recognise objects in real life. In a sense, what we have here is an overlap of natural recognition inabilities.

If we conducted brain scans on the patient who drew the St Pauls Cathedral picture looking for overlap in brain processes when looking at an object and a picture of that object, what kinds of processes would it find in common? Are we looking for areas that are triggered by the object and a drawing of the object or are we looking for areas that are trigged by parts? I suspect that when we scan subjects without brain lesions, we are not just looking for overlaps in primitives such as lines and colours, we are looking for overlaps in recognition centres. This requirement to map the entire path of the visual brain seems to take Schier’s challenge to another level. Brain processes don’t follow a well-ordered
hierarchical progression, there is feedback at all levels and context informs the firing of even the most lowly neuron. Brain processes may be to some extent modular, but they are not encapsulated. Consequently, for completeness, we cannot restrict ourselves to looking for correspondences in only one level of the brain.

**A Brain Scan Experiment: Tanaka’s Differential Amplifier**

A brain scan experiment aimed at discovering which object views and which features of an object most powerfully trigger recognition has been conducted on monkeys by Keiji Tanaka at the RIKEN Brain research Institute in Saitama. Tanaka’s method is to monitor “spike” activities in cells in the inferotemporal cortex (TE) in response to looking at familiar objects and then images of familiar objects. He then reduces the complexity of the images using a specially designed image reduction technique:

After spike activities from a single cell were isolated, many three-dimensional (3D) animal and plant models were first presented by hand within the animal’s visual field to find the effective stimuli for the cell. Different aspects of the objects were presented in different orientations. Second, images of several most effective stimuli were taken with a video camera and displayed on a television monitor by a computer to determine the stimulus that evoked the maximal response. Finally, the image of the most effective stimulus was simplified step by step to determine which feature or combination of features contained in the image was essential for maximal activation. The minimal feature required for maximal activation was determined to be the critical feature for the cell. The magnitude of responses often increased as the complexity of an image was reduced. This may be due to the adjustment of size, orientation and shape, as well as the removal of other features, which may suppress the activation by the critical feature.219

Here we have a neurological confirmation of why some sparse or “refined” line drawings and sketches are often more effective in triggering recognition than full-colour photographs. Tanaka found that the “magnitude of responses often increased as the complexity of an image was reduced”. Tanaka’s work indicates that there are columns of cells that are selective for abstract but rather complex shapes and sets of deformations of those shapes.

---

The kinds of deformation that the cells tolerated seems to reflect the way that shapes deform when they are rotated, especially in the depth plane.

Figure 76  Examples of reductive determination of optimal features for 12 inferotemporal cortex (TE) cells. The images to the left of the arrows represent the original images of the most effective object stimulus and those to the right of the arrows, the critical features determined by the reduction.

Above (fig. 76) is the test set that Tanaka used. The cell that originally responded well to the laboratory cat (image 12) also responded well, or better, to the two circles with horizontal and vertical bars. The column of cells for which these two circles are the critical stimulating feature was tested for its response to the image at different sizes; reversed in contrast polarity; with reverse texture direction; with an altered aspect ratio of size along one axis to the orthogonal axis. Tanaka comments that the “invariances of TE cells suggest that they are actually more sensitive to certain types of deformations than others. The types of deformations that often occur when an object moves around appear to be more tolerated.”

The suggestion here is that the columnar organization of cells in TE is specifically structured to enable primates to recognise an object when looking at it from different angles and in different lighting conditions. The cell responds to the cat shape even when the shape is not presented head-on or when the lighting

220 Ibid.: 93.
changes. This response invariance seems to suggest massive latitude in the object properties required for a positive identification. Even if there are thousands of cell columns with thousands of shapes to which they respond, it seems to take us back to the problem which we had with resemblance – how do we make the fine discriminations between similar shapes that we need to make when faced with a world of infinite and infinitely changing objects? We noted earlier that it is not finding similarities between shapes of objects that is hard – there are ovoids, triangles and rectangles everywhere – it is discerning the subtle differences between shapes that is the key to recognition.

Tanaka suggests a mechanism whereby the columnar structure enables us to achieve these fine discriminations. In observations of areas of TE involving groups of cells (as opposed to single cell columns) using optical imaging, Tanaka noticed that “the presentation of a single feature activated multiple spots.” In figure 77 we can see the shapes which Tanaka used to test this overlap of firing numbered 1 to 8 and a map of the overlaps as they appear on the cortical surface. He explains:

the spots activated by eight moderately complex features are indicated by different kinds of lines and superimposed, i.e. spots activated by one set of four features are shown in the upper half and those by another set of four features in the lower half. For example, feature 1 evoked six spots and feature 2 evoked two spots. This example demonstrates that a single feature is processed in multiple columns in TE.

Another interesting observation here is the partial overlaps between the activation spots evoked by different features. Some of the overlapping regions, which were activated by many stimuli, likely represent columns of non-selective cells. However, others that were activated by only two of the stimuli may represent specific overlaps. For many of these overlaps, we can find similarity between the two features, although the judgment of similarity is only subjective.

---

221 Ibid.: 94.
222 Ibid.: 94-95.
Thus we can see that when shape number 2 was presented to view, the columns which spiked had overlaps with a column which also spiked for 1 and 3. This indicates that the same shape causes reactions in multiple columns. Which is not surprising given the latitude of distortions which a given column can tolerate. One could imagine that the invariant set for the column of which shape 2 is from might include the following:

To arrive at this set of shapes I have imagined what kind of 3D object the first figure might be – a sphere with small cylindrical neck or spout – I have then imagined how it would look rotated in the horizontal plane. Tanaka speculates that columns that are selective for heads and faces might be invariant to views around the horizontal plane (around a vertical axis) but doesn’t believe that this
Mechanisms of Recognition

continuous horizontal mapping can be generalized to “non-face objects”. Thus, the set which I have generated is probably too regular for the kinds of deformations which a column might allow a non-face object such as the spherely-bottle thing. There would no doubt be other deformations in the set which conform to the “types of deformations that often occur when an object moves around”\(^{223}\) or indeed when we move around in relation to the object. To some extent we are assuming that the kinds of deformation that the set includes are those which occur when we shift point of view on a three-dimensional object. It is possible that an object that is not spherical could have generated the first shape. It could be generated by a grape shape with a bottleneck-like protrusion – something like this:

![Diagram of a grape shape with a bottleneck-like protrusion.]

If I rotate this grape shape so that the protrusion is south pointing and reverse its contrast polarity we get:

![Diagram of a transformed grape shape.]

This is remarkably similar to shape 1. It is even easier to transform shape 3 into shape 2 via such deformations. Thus, it seems that multiple columns are activated by one shape because that shape is one of the deformations of another shape that the column responds to.

This column structure may help explain what Gombrich describes as the “constancy” phenomenon which we get when we change our viewpoint with regard to a painting.\(^{224}\) It may also relate the phenomenon which Hopkins notes that when we see a wheel from an angle we don’t mistake it for an ellipse. We can see this illustrated with the picture of a peach. If we approach the image from an

---

\(^{223}\) Ibid.: 93.

angle, the peach becomes an ellipse. You may note that the centre peach painting (fig.78), despite being elliptical, is still very peachy. The constancy mechanism which Tanaka has identified enables us to continue to see it as a spherical peach despite the distortions which the angle of the picture plane introduces.

The column structures that Tanaka has identified also act as what he calls a "differential amplifier". The vertical column structure enables us to track objects despite the different faces they present to us, because the column will respond more or less invariantly to standard deformations. The horizontal relationships between the columns (indicated in fig. 77) enable the brain to make fine discriminations about which object has been presented to view. Multiple firings of different cell columns could be caused by a spherical shape as well as a grape shape. Tanaka’s experiments seem to indicate that adjacent cell columns are selective for similar but related features:

The borders between neighboring columns are not necessarily distinct. Instead, multiple columns that represent different but related features partially overlap with one another and as a whole compose a larger-scale unit. At least in some cases, some parameter of the features is continuously mapped along the cortical surface. The systematic arrangement of related columns could be used for various computations necessary for object recognition. For example, object generalization might be mediated by horizontal excitatory connections between nearby columns representing related features, to achieve a selective blurring of activation. In addition, object discrimination might be achieved through mutual inhibition among nearby columns for winner-take-all selection. 225

Tanaka represents this columnar organization using figure 79.

Figure 79 Tanaka’s diagram of columnar organization in the inferotemporal cortex (TE).

Tanaka suggests that if neighbouring cells arranged horizontally are selective for related features that they can discriminate finely between objects. He is therefore suggesting that the columnar organization serves as a differential amplifier, which simultaneously attends to the gross properties of an object and enables us to track an object as it presents different views, but also enables us to discriminate fine differences between related or similar objects:

Representation by multiple cells in a columnar module, in which the precise selectivity varies from cell to cell while selectivities for most effective stimuli largely overlap, can satisfy two apparently conflicting requirements in visual recognition: disregarding subtle changes in input images under different viewing conditions; and achieving a preciseness of representation in discrimination of objects in subordinate or individual levels. Clusters of cells having overlapping and slightly differing selectivities may work together to confer object recognition abilities that are invariant to viewing conditions. Although single cells in TE tolerate some changes in size, contrast polarity and aspect ratio, these invariant properties at the single-cell level are not sufficient to explain the entire range of flexibility of object recognition. In particular, the responses of TE cells are generally selective for the orientation of the shape in the frontoparallel plane. Cells preferring different orientations and other parameters of the same 3D shape may be combined in a column to provide invariant outputs.²²⁶

²²⁶ Ibid.
It is unlikely, given the relative complexity of the shapes for which the cells are selective, and the variability and complexity of objects in the world, that one object would cause only one cell-column in TE to fire. The only object that is invariant from all angles is a sphere, and even sphere can be variable in illumination. Thus is would appear that an object, or picture of an object, will trigger multiple shape recognition columns and the brain makes a calculation, depending on the overlap of firings, as to what object is in view. The work of Yamane et al suggests that shape recognition cells are not just linked to related neighbouring cells but are linked to cells that respond to the global arrangement features of an object.\textsuperscript{227} This would suggest that the overlapping selectivity for parts of objects is augmented with input from cells which take into account the configuration of parts. This sensitivity to configuration of parts may explain why Shape 4 (the square with a black rectangle protruding from its south face), which looks nothing like the other three shapes, has an overlap with Shape 1 (see fig. 80).

![Figure 80 Shapes 1 and 4](image1)

Figure 80  Shapes 1 and 4

The configurational similarity is clearer if each is a silhouette (fig. 81).

![Figure 81 Shapes 1 and 4 as silhouettes](image2)

Figure 81  Shapes 1 and 4 as silhouettes

It is remarkable to think that the visual brain registers a similarity between these two shapes. This capacity of the brain to register the overall configuration of parts of an object may explain how we recognise stick figures so easily.

If we look at Hopkins’ stick figure (fig. 82) in conjunction with Leonardo’s Vitruvian Man (fig. 83) and imagine how the cells in our inferotemporal cortex are being stimulated by the shapes and their configuration in the overall figures, we are no longer surprised at the feeling of resemblance which Hopkins reports. But of course, we now know that it is not resemblance that is the mechanism behind recognition, it is the organisation of shape-selective columns of neurons and their connection to the rest of the visual system that drives recognition.

Yamane’s work and other work in this area confirms that neurons in the inferotemporal cortex have receptive fields (the area of the retina which stimulates the neuron) of varying size. Thomas P. Trappenberg, Edmund T. Rolls and Simon M. Stringer report:

An analysis of IT neurons that responded to the target stimulus showed that the average size of the receptive fields shrinks from approximately 56 degrees in a blank background to approximately 12 degrees with a complex scene.\textsuperscript{228}

Thus, some neurons are able to process a large percentage of an image and adjust their scope down according to some kind of “top-down” stimulus about the nature of the scene that is being scanned and the task that has been set. One method of testing for “top-down” bias is to prime the subject with a picture of the object which is contained in the scene. These kinds of studies suggest that context, in the broadest sense of the term, plays an important role in the recognition process.

In terms of the methodology which Ramachandran summarized earlier – the what?, why? and how? – Tanaka’s work provides us with a good illustration of how research on depiction can be illuminated by work on vision. We have been trying to understand how the constancy phenomenon works, and in particular how it is that we easily recognise objects and pictures from unusual angles.

- **What? - The underlying functional logic (or law):** Objects can be recognised despite their appearances (let’s call this the law of “object constancy”).
- **Why? - The evolutionary rationale:** Evolution has wired our brain so that we can track objects despite the fact that they look different from different angles.
- **How? - The neural hardware in the brain that mediates the law in question:** Tanaka’s “differential amplifier”

We now have a candidate mechanism which not only helps to explain why very abstract drawings can be very evocative, but we also have a possible mechanism for the constancy phenomenon. It is now possible to imagine why our visual system is largely indifferent to the angle that we approach a picture from. Of course, Tanaka’s explanation that the brain assesses how the image conforms to a set of thousands of bizarre shapes such as this , or this , is probably not going to satisfy critics of the neurological approach, such as Purves and Lotto. We will see later how the “new empiricists” explain constancy.
Clearly there is more work to be done here, and to some extent Tanaka’s experiment throws up more questions than it answers. It is interesting that Tanaka does not speculate as to whether the set of shapes which stimulate particular columns is acquired or innate. Given the heuristic nature of the brain and its plasticity, it would not be surprising if we found that our sensitivity to shapes co-varies with our upbringing.

For our purposes the most significant aspects of Tanaka’s work are:

- The brain scans seem to indicate that, rather surprisingly, the visual system is more finely tuned to abstractions of the shapes of objects than to the detailed object.
- The organisation of Tanaka’s “differential amplifier” columns provides a mechanism whereby our visual system can track objects as opposed to the appearances of objects. His differential amplifier confers “object recognition abilities that are invariant to viewing conditions.”

This latter point provides a neurological explanation for some of the more puzzling illusions that we come across in the vision literature (eg the Ponzo illusion), and explains how it is that the visual system tracks objects not mere appearance. What Tanaka’s work doesn’t tell us is how the context of a scene influences our ability to recognise the object (or its parts). The “differential amplifier” is not going to help us identify a brown smudge of paint as a cow.

---

Context, Constancy and Gist
The ‘frieze’ above consists of squares taken from Constable’s Ottawa study. The cows are, in Schier’s terms, almost sub-iconic; they don’t mean much on their own. The building and the trees are not as recognisable as you might imagine when taken out of context. Yet when we look at the picture as a whole it is vividly a cathedral in a landscape framed by trees. The proposal is, in line with Schier’s theory, that when we first look at Constable’s painting our visual system momentarily entertains the possibility that we are looking at real trees and a real cathedral. Our landscape recognition abilities are thus triggered before our visual system begins to consider the flatness, the paint etc. As we continue to look at the painting and we begin to notice the smudgy cows, the painted surface begins to assume more importance. The interesting phenomenon here is that, even when paying attention to the paint, one is still seeing a landscape with trees. It is almost impossible for our visual system to let go of the “worldly” interpretation. This is because our visual system is always on the look-out for the objects in a scene and not the mere appearance of the objects. In a sense we look past the paint, just as we look past the oddities of inverted distorted image that is on our retina. This is the essence of object constancy and it is a measure of the powerful drive of our visual system to construe everything we see as if it exists in the world. Steven Pinker makes this point in his 2002 defence of the biological basis of human nature The Blank Slate. He explains:

Our perceptual systems are designed to register aspects of the external world that were important to our survival, like the sizes, shapes, and materials of objects. They need a complex design to accomplish this feat because the retinal image is not a replica of the world. The projection of an object on the retina grows, shrinks, and warps as the object moves around; color and brightness fluctuate as the lighting changes from sun to clouds or from indoor to outdoor light. But somehow the brain solves these maddening problems. It works as if it were reasoning backwards from the retinal image to hypotheses about reality, using geometry, optics, probability theory, and assumptions about the world. Most of the time the system works; people don’t usually bump into trees or bite into rocks.230

Pinker notes that our perceptual systems are most probably best suited to a hunter-gatherer lifestyle. Evolution hasn’t “designed” our visual system to expect the kinds of objects and situations which we encounter in the modern world or the kinds of visual illusions designed by artists, psychologists and philosophers. In fact, he argues that when we are “fooled” by such illusions, what it tells us is that our visual system is working just as nature intended. Pinker illustrates with a couple of his favourite illusions by Roger Shepard and Edward Adelson – “Turning the Tables” and the “Checker Shadow Illusion”.

Figure 84 “Turning the Tables” by Roger Shepard. The two parallelograms which form the tabletops are identical in size and shape. The right-hand tabletop can be rotated 90 degrees anti-clockwise to match the left-hand tabletop.

Figure 85 The “Checker Shadow Illusion” by Edward Adelson. The light square, B, is the same shade as the dark square, A.
In the former (fig. 84) the two parallelograms which form the tabletops are identical in size and shape. The right-hand tabletop can be rotated 90 degrees anti-clockwise to match the left-hand tabletop. In the latter, the “Checker Shadow Illusion” (fig. 85), the square marked “A” is the same shade as the square marked “B”.

Such visual illusions are generally used to promote the view that the visual system is easily fooled and that our higher cognitive centres need to be constantly on the lookout for false information. In fact, as Pinker points out, these depictions are carefully designed to fool our visual system. He comments:

In each case we may see the lines and patches on the page incorrectly, but that is only because our visual systems are working very hard to see them as coming from a real world. Like a policeman framing a suspect, Shepard and Adelson have planted evidence that would lead a rational but unsuspecting observer to an incorrect conclusion. If we were in a world of ordinary 3D objects that had projected these images onto our retinas, our perceptual experience would be accurate. Adelson explains “As with many so-called illusions, this effect really demonstrates the success rather than the failure of the visual system. The visual system is not very good at being a physical light meter, but that is not its purpose. The important task is to break down the image information into meaningful components, and thereby perceive the nature of the objects in view.”

If we saw the tables in real life we would easily be able to judge whether the tabletops were the same size and shape – if there was any ambiguity about the relative shapes we would simply move closer or move our head sideways to resolve the issue. If the legs are removed (see inset fig. 84) the illusion is almost non-existent, thus confirming Pinker’s claim that it is the drive of our visual system to construe them as existing in the real world that is the key to the illusion.

Adelson’s comment about the visual system not being a very good “light meter” and his reference to breaking the image down into “components” highlights crucial aspects of the visual system which in some ways seem counter intuitive. The visual system does not register absolute wavelengths of light and levels of brightness, but contrasts signals from receptors in the retina in a process

---

231 Ibid., 200.
232 Of course, the tables aren’t depicted in regular perspective - if they were they would be very strangely shaped. I confess that I did build the tables, and they were indeed bizarre constructions. I ascertained that one could photograph them from such an angle that the tables looked perfectly normal.
called “opponency”. Effectively, it computes the brightness of objects relative to their surroundings, and estimates the colour of objects regardless of the colour of the ambient light – this is the phenomenon of “color constancy” and Purves and Lotto define it thus:

**Color constancy** The similar appearance of two or more surfaces despite different spectral returns from them; usually applied to the approximate maintenance of object appearances in different illuminants.\(^{233}\)

It is this form of constancy which explains the checkerboard illusion. We are surprised that squares marked “A” and “B” in the picture are the same tone because in the real world the squares on a checkerboard alternate between black and white. Our visual system knows that a light square in shadow is still a light square. Once again our visual brain is attending to the enduring real world properties of the squares – the fact that they are light squares – rather than the transient properties lent to the squares by being in shadow. Our visual system has a powerful tendency to discount transient illumination phenomena.\(^{234}\)

The constancy process at work in the checkerboard illusion highlights an important aspect of how we track objects in the world. It is also intimately related to the grouping phenomenon which we noted with the camouflage pictures and the Veronese painting. We do not just group elements together because they are the same colour or texture, we group things together because they are the same shape – or more precisely because our visual system takes them to be the same kind of object.


\(^{234}\) Incidentally, the secret of the success of this particular illusion is the fact that the squares aren’t black and white, but shades of grey. In real life the brightness range between the lightest and darkest square would be hundreds of times greater than the range which can be rendered on a printed page. It is almost impossible to get the white square which is in shadow to be the same as a black square in the light on a checkerboard composed of black and white squares - unless one cheats and reflects a strong light off one of the black squares. That is, it doesn’t work in real life and it is extremely hard to make it work in a photograph.
Imagine that there are some pieces of white paper lying scattered in the street – some of them are on the road and some are under cars. Even though the pieces of paper are in shadow under the cars, we still know they are the same white pieces of paper as the ones on the road. We relate them because of their rectangular shape, and presumably some kind of shape constancy (perhaps enabled by Tanaka’s columns) is in play. In the above photograph (fig. 86) we assume that the piece of paper under the car is the same tone as the one in the foreground. Of course it isn’t the same shade because it is in shadow. But our visual system registers that they are the same shape and most likely the same colour – white. It notices the enduring tonal properties of the pieces of ‘white’ paper, not the transient properties given them by sun or shadow. In fact, if you look at the inset of the paper under the car superimposed on the door of the car, you can see that it has roughly the same tone as the bodywork. The visual system discounts this information because the most important function of the visual system is to track objects, even if they move in and out of the sun, and not measure the amount of light reflected from them.
The constancy phenomenon permeates the whole of vision. In the early 1960’s, a series of experiments by David Hubel and Torsten Weisel indicated that cells in the primary visual cortex deconstructed the signals from the retina into a range of abstractions corresponding to line orientation and line segmentation. Some of these cells are very selective about the visual stimulus they will respond to. For example, some respond to just a particular line orientation. (fig. 87).

However, despite the remarkable selectivity of these cells for lines and edges at a particular orientation, it would appear that our visual system is not taking note of the angle which the stimulus subtends on the two-dimensional plane of the retina, but judges the angle according to how the stimulus fits into its real world context.

Below (Fig. 88) is a drawing developed by Purves and Lotto to demonstrate how our visual system insists on seeing the angles of the figures as real world angles, not angles that exist on paper (or on a picture plane, or even on the retina). They comment:

---

235 In Eric R. Kandel, James H Schwartz, and Thomas M. Jessell, Essentials of Neural Science and Behavior (London: Prentice Hall International, 1995), 437. the authors comment on such cells which operate on the principle of opponency, “Simple and complex cells in the primary visual cortex are thus sensitive to the outline of an object, to its contours, boundaries, and contrasts, but not to the interior or background of objects. These monotonous interior or background surfaces contain no visual information!” They go on to suggest that these particular neurons that only care about edges are responsible for our ability to see line drawings: “It is the information carried by edges that allows us to readily recognize objects in a picture even when the objects are sketched only in rough outline.”
The perception of the angles ……is biased by what each of the identical angles would have typically turned out to be during interactions with the sources of such stimuli in three-dimensional space.\footnote{Dennis Meredith, \textit{Tricking the Eye or Trapping a Reflex} (Duke University Alumni Magazine, 2006 [cited 27th January 2006]); available from http://www.dukemagazine.duke.edu/alumni/dm29/purves.html.}

That is, despite the fact that all the angles can be measured as right angles on paper (and presumably on the retina) we see them as differing by at least 60 degrees between one another (see fig. 89)

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure88}
\caption{If you measure the angles subtended by each of these figures on the paper with a protractor you will find that they are all 90 degrees.(also see fig. 89)}\footnote{Purves and Lotto, \textit{Why We See What We Do: An Empirical Theory of Vision}, 153.}
\end{figure}

Our visual system instinctively sees the objects they would be in real life, and not in terms of the geometry which they form on the plane of the paper. In order for an artist to draw these objects, she must fight against the constancy phenomenon which compels her to see them as real three-dimensional objects in a scene. If you were to draw the picture above (don’t trace it) you will find that you will make the leftmost angle two obtuse and the rightmost angle too acute. It’s hard to believe that in order to draw the red and green prongs on the left correctly they both must subtend 90 degrees on the page.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure89}
\caption{Purves and Lotto’s figure with right angles superimposed.}
\end{figure}
It is the tendency of our visual systems to spatialise things, and see them as
they are in the world, not how they translate into a two-dimensional image, that
makes drawing in perspective so difficult.

This spatialisation phenomenon is not restricted to the visual channel.
Kennedy’s work on drawing and the blind, and Paul Bach-y-Rita’s work on
tactile-vision substitute systems (TVSS) confirm that the ability to recognize
depictions is not restricted to the vision. Kennedy’s blind subjects, even those that
have been blind from birth such as Gaia (see fig. 58), are able to interpret the
raised lines on a drawing tablet as edges and shapes and after some training can
even “see” the depth in perspective drawings. Bach-y-Rita’s work, although not
specifically directed at understanding depiction, reveals that people who have
been blind from birth or early infancy, can quickly learn to “see” shapes and
navigate using a system which transfers camera images to the skin by an
ingenious tactile matrix. Bach-y-Rita writes:

In our first sensory substitution project, we developed tactile vision substitution systems (TVSS) to deliver visual information to the brain via arrays of stimulators in contact with the skin of one of several parts of the body (abdomen, back, thigh). Optical images picked up by a TV camera were transduced into a form of energy (vibratory or direct electrical stimulation) that could be mediated by the skin receptors. In these sensory substitute systems, the visual information reaches the perceptual levels for analysis and interpretation via somatosensory pathways and structures. After sufficient training with the TVSS, our subjects reported experiencing the image in space, instead of on the skin. They learn to make perceptual judgments using visual means of analysis, such as perspective, parallax, looming and zooming, and depth judgments.

Recent experiments by Segond, Weiss and Sampaio on tactile vision confirm
that even sighted, but blindfolded, subjects very quickly spatialise the “tactile
visual stimuli” that is generated by TVSS. The task which the subjects were set
was to navigate a maze and carry out certain tasks using triangular cues as a
guide:

---


240 Bach-y-Rita, Tyler, and Kaczmarek, "Seeing with the Brain."
Blindfolded sighted subjects had to remotely control the movements of a robot on which the TVSS camera was mounted. Once familiarised with the cues in the maze, the subjects were given two exploration sessions. Performance was analysed according to an objective point of view (exploration time, discrimination capacity), as well as a subjective one (speech). The task was successfully carried out from the very first session. As the subjects took a different path during each navigation, a gradual improvement in performance (discrimination and exploration time) was noted, generating a phenomenon of learning. Moreover, subjective analysis revealed an evolution of the spatialisation process towards distal attribution.\(^{241}\)

The blindfolded subjects received their “visual” information about the maze environment from a matrix of 96 electrically activated touch receptors mounted on a 7.5 x 9.5 cm plate attached just above the navel. It is surprising that with such a low resolution matrix that the subjects could navigate at all. However, not only did the subjects navigate easily, they were able to locate targets and get involved in the navigation game. The surprising result is that even with such degraded stimuli they spatialised the maze and very quickly forgot that they were seeing with their skin! The experimenters comment:

> Subjects interpreted decreasing tactile images on the matrix as moving-away cues - and, conversely, increasing tactile images as moving-closer cues to be a consequence of their actions; they did not merely perceive the decrease or increase of shapes.\(^{242}\)

It is probably an understatement to say that they “spatialised” the maze; according to the subjects reports they felt that they were present in the maze. As they were navigating they would exclaim things like: "I'm going backwards", “I'm going to turn around”, “I'm reaching a cue”, “I'm going forwards”, ”I'm stuck” \(^{243}\). The subjects identified completely with the robot even with only 8 x 12 pixels of image to navigate by. The kinds of environment which they were looking at would be something like figure 90, but a 96 pixel image is more like figure 91.

---


\(^{242}\) Ibid.: 1247.

\(^{243}\) Ibid.
It was not the subjects’ ability to navigate the maze which surprised the experimenters; it was their involvement in the task. The experimenters comment:

Subjects had a particular attachment to the upwardly opening triangular cues. When recognising them, subjects explicitly expressed their jubilation: “That's it, I've made it!” “It's just in front of me!” “It's the right door!” This cue category represented the item target they had to find. It was the one that was required to be correctly identified in order to progress in the navigation.²⁴⁴

For the subjects of this experiment, the powerful drive to spatialise tactile visual cues was accompanied by the emotional thrill of achieving an objective, such as going through a door, or finding the target. It is the same drive to make everything real that causes us to see the faces in the stains on walls, teddy bears in the clouds, and elephants behind bushes. It is also what makes us see the trees and

²⁴⁴ Ibid.
grass which give context to the brown smudges which turn out to be cows. Even the most degraded stimulus is construed as something in the world. Evidence from vision research indicates that in the first 100 milliseconds we categorise the scene we are looking at as a landscape, a street scene, an interior etc. That is, the first thing our visual system does is set up the context for the objects in the scene. This initial categorisation is called the “gist” view.

Figure 92  Constable’s Ottawa study in black and white applying Gaussian blur.

I have blurred the Ottawa painting in figure 92 to emulate the kind of resolution your visual system is getting when it takes in the gist of the scene. It is black and white because 95% of our visual field is monochrome. It is blurred because the resolution of 90% of our visual field is very coarse. In order for us to see colour and detail and actual objects, our eyes must scan the scene in a series of saccades. This takes time. Our eyes saccade around a scene at the rate of about 10 times a second, each time fixating on a point. In figure 93 I have emulated the kind of resolution our visual system is getting as we scan a scene.

Figure 93 Salisbury Cathedral with foveal views indicated in colour
From an evolutionary perspective being able to see highly detailed colour image in the “recognition moment” is not a priority. Being able to make extremely fast judgements about what is facing us is a high priority. Thus our visual system has developed a way to roughly categorise a whole scene based on very general, and very coarse, features in less than 100 milliseconds. In her paper “Gist of a Scene”, Aude Oliva writes:

Behavioral studies have shown that observers can recognize the basic-level category of the scene (e.g., a street; Potter, 1976), its spatial layout (e.g., a street with tall vertical blocks on both sides (Schyns and Oliva, 1994), as well as other global structural information (e.g., a large volume in perspective) in less than 100 msec. Observers may also remember a few objects (e.g., a red car and green car), the context in which they appear (e.g., parked on the side) and other low-level characteristics of regions that are particularly salient. 245

Other studies confirm that the gist of a scene is available for purposes of categorisation of a scene on or before the first saccade – sometimes less than 50 milliseconds. 246 Oliva makes a distinction between an object-centred categorisation and a scene-centred categorisation. The gist of a scene is developed before objects are identified; consequently the “spatial envelope” needs to be defined before the object can be recognised. The categories which comprise a scene-centred description are more abstract than an object-centred description. Oliva and Torralba 247 provide examples (fig. 94) of these two types of categorisation of the same scene.

![Fig 94 Oliva and Torralba’s object-centred and scene-centred descriptions](http://cvcl.mit.edu/Papers/OlivaTorralbaPBR2006.pdf)

---

In their view, we initially use such categories as natural/man-made, large space/small space, open/enclosed. After tests with observers, who were asked to develop a vocabulary for scene-description, they further refined these categories into measures of the volume of the space and the scene properties such as, depth range, openness, expansion, ruggedness, verticalness, naturalness, busyness and roughness. The theory is that we categorise the “spatial envelope” using something like these properties before we go on to spot a definite object. This basic level categorisation enables us to recognise a street-scene, a forest, a highway, a panorama or a living room in our first glance. Throughout this study I have been using an object-centred description of Constable’s painting. In fact, I should have described it as a large space, natural scene, semi-enclosed, maybe even as forest-like.

![Figure 95 Examples of space-centred descriptions from Oliva and Torralba (2006)](image)

It is clear from this recent work on gist that before we recognise objects we identify the spatial envelope. The spatial envelope provides the context for object identification. The phenomenon of constancy ensures that we construe that space as real and populate it with appropriate objects. The constancy phenomenon recruits everything to the illusion – the trees, the building, the pond and finally the cows.

The neurological mechanisms for gist recognition are the subject of some debate. There has been very little work done in this area; vision research has tended to concentrate on object recognition.\(^{248}\) However, there is general

\(^{248}\) Some earlier work on scene perception can be found in Stephen Kaplan and Rachel Kaplan, *Cognition and Environment* (New York: Praeger, 1982). and M.C. Potter, “Short-Term Conceptual
agreement that we do not build up the gist of the scene from the parts. We take in the whole scene at once. Consequently, many of the hierarchical models of object perception, and indeed their neural models, do not work for gist recognition. Oliva observes:

speed and accuracy in scene recognition are not affected by the quantity of objects in a scene (for a review, see Biederman, 1995), and recognition can be achieved equally well even when object information is degraded so much that objects cannot be locally recovered.249

One of the impediments to developing neurological account of gist recognition has been the assumption that there is either a local-to-global or global-to-local feedback mechanism which drives the process. However, any cortical feedback mechanisms are likely to be too slow for gist recognition.250 For example, in the famous Fuchs saccade study of how our eyes read a face, we can see that the gaze moves around and fixates on key areas in the matter of a second or two. The problem, as Fuchs points out, is that there isn’t enough time for given fixation to get feedback from the higher areas to tell the eye where to go next.

Saccades are so rapid that they are over before visual feedback can help guide them to the target. Nevertheless, they are very accurate. Therefore, the neural command that drives the eye muscles must be programmed very precisely in advance.251

Fuchs project is to discover “how a sensory stimulus elicits an appropriate eye movement response.”252 One of the puzzles is how the eye can move so accurately to its target when the gist view is so rough. One suggestion is that our eyes are pre-programmed to read a face (see fig. 96). However, there must be some kind of early feedback to the visual system that tells it that a face, or even landscape, is in view.

---

249 Oliva, "Gist of a Scene."
251 Albert Fuchs, Fuchs@U.Washington.Edu Community of Science Page (University of Washington, 2006 [cited 15th May 2006]).
252 Ibid.([cited].)
The question is which area of the brain is giving feedback to which area? Rasche and Koch tentatively suggest that there are a number of local feedback mechanisms routed through the lateral geniculate nucleus (LGN). Rasche and Koch, Recognizing the Gist of a Visual Scene: Possible Perceptual and Neural Mechanisms ([cited]). Epstein and Kanwisher, "A Cortical Representation of the Local Visual Environment," Nature 392 (1998). Rasche and Koch, Recognizing the Gist of a Visual Scene: Possible Perceptual and Neural Mechanisms ([cited]).

Epstein and Kanwisher254 found a region of cortex referred as the parahippocampal place area (PPA) which responds more strongly to pictures of what they call “intact scenes” (these correspond roughly to Oliva and Torralba’s classifications into indoors, outdoors etc.) than to objects.255 This work on scene recognition and gist promises to become immensely important in the study of how we see paintings in terms of their whole composition.

The gist view is not just coarse but it is black and white. So now we have two reasons why it’s so easy to recognise crude black and white sketches. Our visual system abstracts objects into the basic shapes that the columns in TE are sensitive
to. Our early visual processing registers the whole picture coarsely in black and white. It looks as if we are able to take in a whole scene with our whole retinal array in around 100 milliseconds – in this first glance we are not attending to anything in particular and somehow on the basis of that first glance we are able to direct our fovea towards salient objects in the scene. In the real world, the first objects we notice will be moving objects. Our visual system is particularly good at picking out movement at the periphery of our visual system. Such a movement will trigger an eye saccade in that direction or a head turn. Our eyes then saccade around the scene picking out individual details. Until the early change blindness experiments were conducted, it was thought that we stitch all these successive saccades together and build an internal representation of the scene in memory. On this view, our visual world seems rich, colourful and detailed because our visual system is constantly piecing together a jigsaw puzzle of the scene before us. This jigsaw puzzle is our mental representation of the world. The change blindness experiments taken together with what has been discovered about the neurological architecture of the visual system indicate that we do not have rich internal representations that we use as a guide for action or as a memory store. We don’t piece together the jigsaw, we just make a note of the important pieces.

Earlier, Noë noted three issues which summarise the significance of the change blindness work for a theory of perception:-.

1. Perception is attention-dependent.
2. Perception is gist-dependent.
3. It seems that the brain does not build up detailed internal models of the scene.

However, Noë denies this is evidence that the visual world is a grand illusion. Instead, Noë argues what the “grand illusion” debate reveals is a widespread and mistaken assumption amongst philosophers and psychologists about the phenomenology of perceptual experience.
The central problem of visual theory is not: how do we see so much on the basis of so little? It is, rather, why does it seem to us as if we see so much when in fact we see so little?

Alva Noë

Change Blindness, the Grand Illusion and the Sensorimotor System

Noë argues that, contrary to the claims of Dennett, Blackmore and the “new sceptics”, people in general do not believe they see a complete, universally detailed world all the time. He argues that to “normal perceivers” “it does seem to us as if we have perceptual access to a world that is richly detailed, complete and gap-free.” However, it is not our internal representation that is highly detailed, it is the external world that contains the detail:

We take ourselves to be confronted with and embedded in a high-resolution environment. We take ourselves to have access to that detail, not all at once, but thanks to movements of our eyes and head and shifts of attention.

That is, they/we accept that we need to move our eyes and our heads and our bodies in order to be able to see things more clearly and make out detail in things that are on the periphery of our vision or obscured by other objects. Noë argues that those who see the recent change blindness and attention blindness experiments as demonstrating that we, in fact, do not register rich details of a scene, are probably right. On the other hand, Noë argues that those who see this result as significant have assumed that the general population and many vision researchers have a “snapshot” conception of the visual system. In fact, he argues, the snapshot view was never tenable, and all the change blindness experiments demonstrate is that we retain less information about a scene than we previously thought.

The crucial point is that the visual system is not entirely dedicated to constructing an internal representation of what we see. A large part of the visual brain is dedicated to sensorimotor contingencies which enable us to navigate the world and examine its rich detail. The sensorimotor view of the visual system, which Noë supports, holds that we register where and what things are to the extent

257 Ibid.
that we can reach for them, move closer to inspect, etc. The sensorimotor aspect of the visual system has gained ground in the neurological literature since Ungerleider and Mishkin’s 1982 paper “Two Cortical Visual Systems”. In their paper they argued that there were two visual processing streams:

- the ventral (what) stream which routes to the inferior temporal cortex (IT).
- the dorsal (where) stream which routes to the posterior parietal cortex (PP).

The idea that there is a simple division of labour between the what and the where areas of the brain has now been replaced with an account which identifies “complex patchwork of visual areas occupying the posterior 50% or so of the cerebral cortex” Milner and Goodale comment:

> Although the evidence available at the time fitted well with Ungerleider and Mishkin’s proposal, recent findings from a broad range of studies in both humans and monkeys are more consistent with a distinction not between subdomains of perception, but between perception on the one hand and the guidance of action on the other.

Milner and Goodale’s work suggests that although the ventral stream is largely given over to colour and form perception, the dorsal stream is not properly characterised as the where channel because, in their view, its role is to provide the immediate visual control required to guide actions. They identify a series of visuomotor modules in the posterior parietal cortex which respond to different kinds of stimuli:

- some cells respond when the stimulus is the target of an arm reach; others when it is the object of a grasp response; others when it is the target of a saccadic eye movement; others when the stimulus is moving and is followed by a slow pursuit eye movement; and still others when the stimulus is stationary and the object of an ocular fixation.

An important aspect of these stimuli is that in order to elicit the appropriate motor response – reaching, grasping, looking – the stimulus does not necessarily present to consciousness – in a sense the stimulus is not perceived. They describe a case of a subject D.F. with impairment to the ventral stream which resulted in

---


261 Ibid.
her having no conscious visual experience of object shape. However, despite misdescribing the sizes and shapes of a series of objects in front of her “showed excellent visual control of anticipatory hand posture when she was asked to reach out to pick up blocks of different sizes that she could not distinguish perceptually.”

Just like normal subjects, D.F. adjusted her finger-thumb separation well in advance of her hand's arrival at the object, and scaled her grip size in a perfectly normal and linear fashion in relation to the target width (Goodale et al., 1991). Yet when she was asked to use her finger and thumb to make a perceptual judgement of the object's width on a separate series of trials, D.F.'s responses were unrelated to the actual stimulus dimensions, and showed high variation from trial to trial.

This connection between the vision and touch is significant when we consider how we come to recognise the shapes of objects. The above example highlights something which we often take for granted; the integration of our eye and hand enables us to pick up objects by precisely preparing our grip appropriately. For example, when D.F. was shown a flashlight, she could describe its colour (red) and material (plastic), but not identify it. On the other hand she oriented her hand to grasp it correctly and once she had picked it up recognised it (haptically) as a torch. In 2004, Goodale and Milner published a case study of D.F. (aka Dee Fletcher) and it is clear that although her case has superficial similarities to the agnosia patient who could draw St Paul’s but not recognise it, her agnosia is functionally different. For example, when shown line drawings, Dee could neither recognise the object depicted nor copy the shapes. The damage to her ventral steam means that she has no conscious visual experiences of the edges and shapes in drawings or things in the world. In fact, she has considerable difficulty separating objects from their background. In her words, they all “run into each other”.

In their study, Goodale and Milner make the case for at least two visual systems; the “ventral perception stream” which is largely concerned with visual representation and bringing perceptions into conscious awareness, and the “dorsal

---

262 Ibid.
263 Ibid.
265 Ibid., 10.
action stream” which guides action and is by definition unconscious. They use the example of playing tennis to highlight the different roles these visual system play:

Dee's case, along with evidence from a broad range of studies from frogs to humans, tells us that visual perception and the visual control of action depend on quite different brain systems. What we have learned from these studies is that conscious visual experience of the world is a product of the ventral not the dorsal stream. You might perceive the tennis ball that has just been lobbed over the net by your opponent, but you can never be conscious of the particular information that your visuomotor system uses to guide your successful return. This visuomotor computation happens entirely unconsciously. You are not aware of the fact that the ball is expanding at a certain rate on your retina and that this is an important cue for knowing exactly when to swing to hit it with the ‘sweet spot’ of the racquet. When you are running around the court chasing the ball, the visual scene is changing on your retina quite dramatically. The shape of the projected image of the net, for example, will be constantly changing - and yet you will continue to see the net as a stable and unchanging object in the scene. It is perhaps a good thing that you are not aware of all of these viewer-dependent changes. If you were, the world would become a bewildering kaleidoscope of unrelated and disconnected experiences in which objects change their sizes and shapes as you move about. What you need are the enduring constancies of perception in order to make sense of the world.266

If your opponent is serving at 120 kph you have roughly half-a-second from when your opponent serves to when the ball arrives at your end of the court. Obviously, if you wait until the ball leaves your opponent’s racket your response time is reduced. There is certainly not enough time to compute the trajectory of the ball – the sensorimotor system somehow positions you and your racket for the return without the need for you to make conscious decisions. If you relied entirely on the ventral perception stream you would probably never even get in position to return the ball.

Goodale and Milner point out that it is a mistake to refer to the visual processing in the ventral stream as ‘unconscious perception’267:

Use of that phrase carries an implication that such visual processing could, in principle, be conscious. The fact is that visual activity in the dorsal stream can never become conscious – so ‘perception’ is the wrong word to use. The dorsal stream is not in the business of providing any kind of a visual representation of the world: it just converts visual information directly into action. The visual processing that it carries out is no more

266 Ibid., 109.
267 Ibid., 114.
accessible to conscious scrutiny than the visual processing that elicits the pupillary light reflex.\textsuperscript{268}

The authors are effectively warning against supposing that the unconscious processing which guides our actions is comparable to the unconscious perception which sometimes characterizes our perceptions of the world. For example, it is well known that we can see things subliminally and that this “unconscious seeing” can affect how we quickly we recognise things later. This priming technique is often used by psychologists to test perceptual abilities. It seems to be the case that we consciously become aware of some things, but not others. Goodale and Milner speculate that the reason some only some ventral stream activity becomes conscious could be to do with limits on the processing power of the system, or it could be to do with attention, but they remain agnostic on the issue.

![Figure 97](image)

**Figure 97** “We might recognize a can of beer on a commercial but we could never pick it up.”

**Sensorimotor Cues**

Could this unconscious processing taking place in the dorsal stream contribute to recognition of depictions? That is, when we look at the Constable painting is there some part of our sensorimotor system that is responding to features of the picture? If so, what kind of features could they be? Dee could discern colour and texture despite the fact that she could not consciously recognise shapes. It is possible that

\textsuperscript{268} Ibid.
our sensorimotor system is responding to pictures and we are unaware of this. Goodale and Milner are sceptical that a television picture, for example, provides our dorsal stream with stimuli for action. They argue “We might recognize a can of beer on a commercial but we could never pick it up.” However, the fact that we can’t reach into the picture and grab the can doesn’t mean that somewhere in our sensorimotor system the reaching and grasping processes are not being primed. When you look at the peach in figure 78, maybe what makes it so peachy is the fact that you unconsciously prepare to shape your hand to grasp it. The sensorimotor system does not only comprise activities such as reaching and grasping; it is involved in every action that we take. It is also implicated with our other senses. It tells us if we are upright or prone, high or low, it positions us in relation to things and gives us the sense of whether they are in reach or far away. It also controls our involuntary eye movements. Thus, if we construe the activity of the dorsal system very broadly, one could imagine that it could play a role in predisposing us to recognising depictions. Mark Rollins in his 2003 essay “Pictorial Strategies and Pictorial Content” argues in favour of what he calls a “hybrid perceptual theory” applied to a theory pictorial representation. Rollins’ “strategic design theory” suggests a way of combining Ramachandran’s “bag of tricks” with sensorimotor contingencies. For example, Rollins suggests that perspective, by placing the viewer relative to a scene, triggers a sensorimotor response. He comments that linear perspective “emphasized the centrality of the picture’s spectator, who could comprehend all with little effort and in a single glance.” And further suggests that the appeal of perspective may be something to do with “mechanism of perceptual control”. In the case of the Constable painting, we have a sense that we are smaller than the trees and the cathedral, and that we could walk into the scene. Constable has positioned us such that our sensorimotor system can tell us where we are in relation to the objects and prepare us for action. Arguably, the designers of beer commercial have done much the same and attempted to trigger primal grasping mechanisms to make us believe the beer is within our reach. If indeed this is the case, artists and designers are tapping

269 Ibid.
271 Ibid., 119.
272 Ibid., 120.
into unconscious motor control mechanisms. It is not enough just to render the object, in this case the beer can, as a three-dimensional object. It must also have qualities which trigger the sensorimotor response of reaching or grasping. It will trigger such a response more readily if it is upright and looks as if it is the kind of size that our hand can wrap around.

We established earlier that our visual system treats depictions as if they were real world objects not flat representations. The visual illusions we looked at capitalize on this fact and demonstrate that we look past the superficial properties of the flat surface. Thus, to enhance visual recognition, an artist should attend not so much to the transient appearance of objects – its visible shape and colour – but to its real world attributes such as weight, flexibility, and stability; properties that affect how we interact with the object or scene. How does an artist represent these interactive object properties, as opposed to transient properties?

Figure 98  *Fern Drawing* stage one  *Fern Drawing* final stage

I decided to try an experiment where I examined my decision making process as I drew a fern tree from life. My purpose was to ascertain whether there was functional split between drawing how the fern looked, in terms of its shape, from where I was sitting, and attending to its embodied properties.
Above (fig. 98) are two stages of my first drawing. In the first stage I have indicated the stems of the fronds in a way which identifies the three-dimensional structure of the fern tree. It is hard to say whether the three-dimensional structure can be readily seen in stage one. However, I do remember that when I was drawing the picture, I felt that getting the curve of the stems correct and in perspective might give the impression of the fronds bending from the weight of leaves. In the final stage, I have tried to indicate depth by making the closer fronds more detailed and by indicating that one frond is in front of another. These are elementary techniques for indicating depth; our visual system assumes that objects which are detailed are closer than objects which are not. It also assumes that if one object occludes another, the occluding object is closer. Thus in my drawing process I began by trying to ensure that the embodied properties such as the weight and flex of the frond were evoked, and I ended by trying to make the fern look three-dimensional.

When I was making the drawing, I would look at the fern before making each mark. I would identify the part I wished to draw and gauge the angle and length of the mark needed to represent that part.

My primary concerns were:-
- The angle of the mark;
- The thickness of the mark;
- The relation of the mark to other marks on the page.

After making each mark, I would assess whether the mark enhanced the impression of the frond which I was trying to achieve. There is a constant feedback process. One doesn’t imagine the drawing on the page and trace around it. The drawing always grows through trial and error. This judgment of the success or failure of each mark was not based on a strict one-to-one correspondence. Remember, I was drawing from life. My head was moving in relation to the plant and the plant was moving because of a slight breeze. My primary considerations in this regard were:-
- To give the feeling that the leaf was bending under its own weight;
- To represent the spikiness of the leaves;
- To indicate the progressive change in angle of the leaves towards the tip of the frond.
Drawing fronds in such detail is probably a mistake unless you are a naturalist and it is important that someone identify the species. There are between 50 and 100 leaves on the frond and each is a different length and joins the stem at a different angle. In my subsequent attempts, I abandoned the “naturalist” approach and concentrated on an overall impression of the plant. The following sketches (fig. 99), of a different fern tree, show what can be achieved with a few strokes.

![Figure 99 Quick sketches of ferns.](image)

Each of these sketches took only a few minutes. The lack of precision does not detract from the ferny quality of the drawing, in fact, it emphasises the different weights of the each of the fronds and the spikiness of the leaves. You couldn’t construct a frond from these sketches. You might, at a pinch, be able to identify the species, but I doubt it. I don’t know enough about trees to be able to say whether the trees in Constable’s Ottawa study are identifiable, but I’m sure the trees in his Victoria and Albert painting are (fig. 100). I am equally sure that any farmer could identify the breed of cow grazing around the pond. When we compare the two paintings, we can see a “life” in the Ottawa study which the V&A painting lacks. I am speculating here that the more impressionistic style of painting, which emphasises the sinuous, windblown nature of the trees, appeals more to the dorsal stream than the rather static finished painting. Perhaps the kind of realism that we find in old master sketches triggers a different aspect of our
visual system from the kind of realism that we find in a photograph. This is a topic which, I believe, deserves further research, and could illuminate concepts of realism from the perspective of recognition theory.

Figure 100 John Constable, *Salisbury Cathedral from the Bishop’s Grounds*, 1823. Oil on canvas. 87.6 x 111.8cm, Victoria and Albert Museum, London.

**Summary of Chapter Four**
Recognising objects and scenes in pictures is more like recognising objects and scenes in the world than we could ever have suspected. We are largely mistaken about the richness of our visual phenomenology and consequently assume that a picture needs to have rich visual content to trigger recognition. In fact, the mechanisms of recognition are radically counter intuitive. Our visual system has evolved hundreds of little tricks which enable us to quickly identify things in the world and their properties. Artists have learnt to tap into these trigger mechanisms, these shortcuts to our evolutionary past, to evoke objects and scenes in pictures.

Conventional and resemblance accounts of depiction speculate on the cognitive processes underlying pictorial interpretation, but cannot produce psychophysical or neurological evidence to back up their claims. A cognitive account, on the other hand, builds on vision research to create an account which explains how we
see pictures in terms of how we see the world. Schier’s big insight was to realise that in order to understand how we recognise scenes in pictures, we must understand how we see scenes in the world. Schier doesn’t say very much about how visual cognition actually works, but his legacy is to highlight the need for a marriage of aesthetics and research into visual cognition.

There are many competing theories of visual cognition. They can’t all be right, but they are all based on research and experimentation with sound methodological practices. If we want to understand the psychology and the neurology of looking at pictures there is a wealth of work being done in visual science, and the bonus for a theory of depiction is that nearly all of these experiments use pictures as their “visual targets”. Vision science is effectively the study of how we see pictures.

Work on object recognition, scene gist, change blindness, and the sensorimotor system has the potential to explain many of the puzzling features about how rough sketches can be so evocative.
Chapter Five

Conclusion

Figure 101 John Constable, *Salisbury Cathedral from the Bishop’s Grounds*, 1820. Oil on Canvas, National Gallery of Canada, Ottawa

In a sense, my task is done. We are beginning to see how in that first moment of looking at Constable’s painting we are so powerfully transported to the scene. Our visual system registers the gist of the scene in the first 100 milliseconds and the spatial envelope is defined. The first saccade of our eye takes us to the cathedral; Constable has structured the painting so that our eye goes there first. Then our eyes begin to saccade around the scene and pick out details. But the context is already established and even in the Ottawa study it is clear that the brown smudges must be cows. Perhaps you might then wish to speculate why there are black clouds over the church in 1823 – but that is where recognition rests and symbolism takes over.

Schier’s hypothesis, that looking at a scene in a picture triggers the same recognition abilities that are triggered when we look at a scene in real life, is vindicated. I have even suggested that this might be true right down to the level of
sensorimotor contingencies. His important insight is that to understand how we see things in pictures, we need to understand more about how we see things in the world. Clearly, we have a lot more to learn about how the visual system enables us to recognise scenes and objects in the world. In 1993 Zeki wrote “the visual areas which are involved in the recognition of even simple objects, as well as the details of the integrative processes, are not known, especially in man.” In the 13 years since Zeki wrote this, there has been much progress in research on the visual brain and the mechanisms of recognition. Brain scan technology has advanced considerably over the last decade or so and has produced some fascinating, if sometimes puzzling, results. We are learning more about the specialized functions of the brain, but there is still much that we do not understand about how all these areas integrate to produce what we call vision. One thing is clear; an account of how we recognise depictions must also involve an account of how the visual brain gives rise to percepts. It may also involve an account of how the visual brain guides action. This kind of account will inevitably involve a description of brain processes. In the light of Zeki’s comment and given the broad scope of theories and approaches to visual cognition which we have touched on in this study alone, it looks like it is going to be a long time before we have a unified account of visual cognition. How can the development of a theory of depiction position itself so that it can benefit from the ongoing research on vision without necessarily attaching itself to a particular theory? After all, a well-designed experiment can make or break a theory.

In fact, as we have seen, work on visual cognition is very largely work on a theory of depiction. There are hundreds of experiments which are going on at this very moment unravelling the mysteries of perception and most of these studies are using pictures! A good example is a recent brain-scan experiment which used pictures of familiar people and objects to test individual neuron response to images. The experiment not only identifies a short-cut mechanism at the neuron level for visual identification, but compares response times to drawings, photographs and letter-strings.

The Halle Berry Neuron

Brain scan work on recognition by Rodrigo Quiroga at the University of Leicester suggests that there are neurons which fire when an iconic object or person is seen in a photograph, drawing or even the name of the object is seen. Quiroga et al analysed responses of neurons primarily from the medial temporal lobe. The dedicated neurons that they have identified are “near the end of the transformation from visual information about object structure to memory-related conceptual information about object identity” comments Charles E. Connor.

Quiroga et al conducted the experiment on eight patients who had 64 electrodes implanted in their brains before epilepsy surgery:

While each participant was shown a large number of images of celebrities, animals, objects and landmark buildings, electrodes recorded the brain cells’ firings. This screening stage determined which images elicited a strong response in at least one neuron. The team then tested the responses to three to eight variations of those images from the narrowed list.

Some of the persons and objects which elicited a strong response from individual neurons were, Jennifer Aniston, Halle Berry, Sydney Opera House and the Eiffel Tower. Quian Quiroga adds:

[In a] patient with a neuron specific for actor Halle Berry; the neuron responded not only to photographs but also to a drawing and an image of her name. What is more, even when Berry was costumed as the masked Catwoman, if the patient knew it was Berry, the neuron still fired. "This neuron is responding to the abstract concept of Halle Berry rather than to any particular visual feature. It's like, 'I won't recall every detail of a conversation, but I'll remember what it was about.' This suggests we store memories as abstract concepts."

If you look at figure 102 you will note that although Quiroga was testing for visual responses to faces and objects, it is actually the letter string “Halle Berry” (number 96) which elicits the strongest response. Each image was flashed on a screen for one second. The vertical blue dashed lines indicate image onset and offset (see fig. 102 for an explanation). Interestingly, in all the experiments the letter strings generally elicit a strong and sustained response but the response time to letter strings is around 100 msecs later than to drawings and photographs.

276 Ibid.
A single unit in the right anterior hippocampus that responds to pictures of the actress Halle Berry. Graph a) Vertical dashed lines indicate image onset and offset (1 second apart). Graph b) indicates the picture number in a which causes the most spikes in the Halle Berry neuron. The response of the Halle Berry neuron to the drawing is weakest. Although it is arguably, not a very good likeness, and it is drawn in black and white (fig. 103).

![Halle Berry drawing used in Quiroga experiment (left) and the spike which it elicited (right)](image)

Figure 104  A single-unit in the right posterior hippocampus that responded to pictures of the actress Pamela Anderson, including a caricature of her and the letter string “Pamela Anderson”, but not to other letter strings. Note that it is difficult to identify common visual features that would explain this invariant response.

Colour drawings of Pamela Anderson and Mother Teresa were also used in the experiment. In the Pamela Anderson case, the drawing elicits a better response than photographs or the letter string (fig. 105). The drawing of Mother Teresa also elicits a very strong response (fig. 105).

Figure 105  Test results for Mother Teresa neuron.
It is interesting to note which pictures of Halle Berry triggered the strongest response. For example, not only did the drawing (fig. 106) score badly, but so did the picture with her wearing sunglasses (25) and a rather generic looking three-quarter view (23). The responses to the Mother Teresa and Pamela Anderson pictures were far more uniform. One could speculate that this may be because Mother Teresa and Pamela Anderson are more iconic. I would speculate that it may be because the pictures used present salient features more consistently. For example Pamela Anderson’s hairline and the line of Mother Teresa’s scarf are absolutely consistent. The Halle Berry hairline is very variable – this makes her harder to identify quickly. The response times to photographs, drawings and letter strings are within the 300-600ms range. There seems to be no significant pattern in response speed or level between the various types of depiction.

The response to the letter strings has led to speculation that although the neurons were tested for a visual response, they may not strictly encode or respond to visual characteristics. Connor comments:

These results may be best understood in a somewhat non-visual context. The brain structures that they studied stand at the far end of the object representation pathway or beyond, and their responses may be more memory-related than strictly visual. In fact, several example cells responded not only to pictures but also to the printed name of a particular person or object. Clearly, this is a kind of invariance based on learned associations, not geometric transformation of visual structure, and these cells encode memory-based concepts rather than visual appearance.279

278 Unfortunately Quiroga ran into copyright problems when he came to publish his findings and appends this note: “Note that owing to insurmountable copyright problems, all original images were replaced in this and all subsequent figures by very similar ones (same subject, animal or building, similar pose, similar colour, line drawing, and so on). Ibid.”
Connor notes that some of these cells seemed to be responding not just to the “target” but to associations with the target. For example, a cell which was selective for Jennifer Aniston also responded very strongly to a picture of Lisa Kudrow: both actresses starred in the sitcom *Friends*. Connor comments “What seems to be a sparse representation in visual space may be a distributed representation in sitcom space!” and concludes:

Thus, Quiroga and colleagues’ findings may say less about visual representation as such than they do about memory representation and how it relates to visual inputs. Quiroga et al. have shown that, at or near the end of the transformation from visual information about object structure to memory-related conceptual information about object identity, the neural representation seems extremely sparse and invariant in the visual domain. As the authors note, these are predictable characteristics of an abstract, memory-based representation. But I doubt that anyone would have predicted such striking confirmation at the level of individual neurons.\(^{280}\)

It is almost certain that the icon specific neurons in these subjects would have fired if they had also been presented with the real Halle Berry or Jennifer Aniston. In our quest to identify underlying processes that are common to seeing drawings, photographs and real objects, we have here found reliable evidence that in the medial temporal lobe, at least, there are neurons which are likely to be triggered equally by an object, a drawing and the letter string. Maybe they would be triggered by simply a memory of one of these icons.

\[\text{Figure 106} \quad \text{A single-unit in the right anterior hippocampus responding preferentially to pictures of Catwoman. Interestingly, this cell also responded to other animal pictures.}\]

In Quiroga’s final experiment he tested neurons selective for the Catwoman picture and found that the cells also responded to pictures of animals (fig. 106). The fact that the most exciting animals were a spider and a rhinoceros – which do not strike me as anything like a cat or a Catwoman – seems to indicate that the way these cells fire has a lot in common with the cells in the inferotemporal cortex which Tanaka tested. Quiroga comments:

\(^{280}\) Ibid.
Conclusion

We do not mean to imply the existence of single neurons coding uniquely for discrete percepts for several reasons: first, some of these units responded to pictures of more than one individual or object; second, given the limited duration of our recording sessions, we can only explore a tiny portion of stimulus space; and third, the fact that we can discover in this short time some images—such as photographs of Jennifer Aniston—that drive the cells suggests that each cell might represent more than one class of images.\(^{281}\)

In fact, if Connor is right and what Quiroga has identified are cells that are encoded with a kind of “invariance based on learned associations, not geometric transformation of visual structure”, then these neurons may work in a similar way to those Tanaka identified, but “encode memory-based concepts” based on stimuli that are not associated with one sense and are very abstract. This population of concept-based neurons explains why it is easier to recognise Lisa Kudrow if she is in a picture with the cast of *Friends*, why we recognise the spoon more easily when it is next to a cup and saucer and why we see a brown smudge as a cow if its next to a tree. It also suggests a neural level explanation for some of the Gestalt rules about grouping, similarity, proximity, good continuation and context. The visual system, like the rest of the brain, is a learning system. It is a plastic, heuristic system that specializes in association, quick rough-and-ready learning and integration. We might speculate that if something occurs frequently, the recognition system makes a note by dedicating a neuron. If things occur together it makes a note of that by connecting the neurons. If things which occur frequently also occur together it associates the neurons. These associations are cross-sensory. If the sight of a cup of coffee and the smell of a cup of coffee are associated, it only needs the sight or the smell alone to trigger recognition. The first thing our visual system does with any scene or any object is check whether it is familiar. It does not try to rebuild scenes or objects – it has a shortcut. In fact the brain appears to have thousands of shortcuts. It would seem that the visual brain is more of a “bag of tricks” than even Ramachandran envisaged. Visual recognition is dependent on the accumulated associations of all the senses. What Quiroga’s experiment tells us is that within 600 milliseconds of seeing, smelling, touching, tasting or hearing something our brain has associated the stimulus with a lifetime of learning.

Chapter Six

Summary
Depiction is an invention of human beings, like language and mathematics. However, there is much less learning involved in learning how to see pictures than in understanding a sentence or multiplying by three. In fact, vision research on human beings, monkeys and chimpanzees indicates that no learning is required at all. Given these facts, what kind of role could convention play in our ability to see a landscape with trees, and possibly some cows, in Constable’s painting? The techniques of depiction such as foreshortening, shading and perspective are discoveries rather than inventions. But, as Flint Schier points out, one cannot just invent a rule of depiction. It is useless just to stipulate a rule that a particular kind of mark represents a particular thing (for example, that a dot is an eye) – the mark has to be seen to be what it represents. The marks which represent trees in the Constable picture can be seen as trees by anyone from any culture and probably by chimpanzees and macaque monkeys. The styles and techniques of depiction through the ages involve discovery and invention, but the artifice of depiction is constrained by the fact that by definition it must be possible for anyone with primate visual abilities to naturally recognize what is depicted. The reason that we can see a bull in this twenty-thousand year old cave-painting from Lascaux as easily as we can see a cow in a Constable painting or in a photograph is that the ability to see depicted content is universal.

Figure 107  From Lascaux
There is no evidence that we need to absorb a symbol system to see depicted content in pictures, and all the evidence indicates that if there was a symbology of depiction it would need to be universal. A universal and unlearnt symbol system is by definition innate and therefore non-conventional. If convention, habit and custom have a role in how we recognize content in pictures, it is exactly the same role that convention, habit and custom have in how we see things in the world. An important factor in how we recognize cows in real life is that they are in field and surrounded by trees – this is as true in the Constable painting as it is in real life. So “yes”, convention, in the sense of context and familiarity, does have a role in recognizing content in pictures, but the context and familiarity which we are using to recognize the cows in the painting are the same kind as we use in real life.

Constable’s cow looks far less like a cow than the twenty-thousand year old cow in fig. 107, but is nevertheless recognizable as a cow because Constable is capitalizing on the fact that the grass and the trees will lead us to expect a cow in this field. The difference in style between the pictures does not significantly affect our ability to see the animals. We can recognise things depicted in sketches and paintings despite the fact that they are smudgy and blurred and seem to look nothing like what they depict. It is this inescapable fact that presents such a problem for the resemblance account. Constable’s cow is at the very limits of cow resemblance – yet we recognized it within a second of looking at the painting. It is not the sharing of shape properties between the cow marks and real cows, or even the congruence of a cow shape in the visual field which sparks our recognition. We saw with the outline experiments that, even with silhouettes that faithfully replicate the shape that an animal presents to our visual field, we often failed to recognize what is represented. Modern resemblance theory relies on the appeal to shape congruence in the visual field to justify its claim that the resemblance experience is part of the mechanism of recognition. Against this we found that people didn’t recognize things even when shape congruence could be shown to be present and did recognize things such as a stick figure and smudgy cows, when it clearly was not present. These facts about how we see depictions, especially degraded or sketchy pictures, indicate an important role for the context and configuration of marks. We see a depiction all at once, not by virtue of the reference of each of its parts. This raised the possibility of a role for resemblance in seeing the overall configuration of a scene or an object. For example, maybe it
Summary

is the overall configuration of the stick figure’s limbs which make it a human, not
the straggly lines of which it is composed. The problem with this strategy is that it
stretches the concept of resemblance too far. If we can see a resemblance between
a straggly stick figure and a man as easily as we can between a photograph of a
man and a man, how can shape resemblance be the mechanism whereby we
recognize what is depicted? After all, we have to admit that the shape of the stick
figures legs could not be said to resemble legs. Thus, resemblance theory lacks a
plausible mechanism whereby the actual shape that the drawing presents to the
eye triggers recognition. Something else other than the actual shape must be doing
the work.

In the experiments with the camouflage pictures (figs. 65 & 66) and the
pixelated Bather Girl drawing (fig 63), I showed that our visual system was able
to see what was depicted despite the object being obscured and reduced to a few
grey patches. In the camouflage pictures it was the obscuring foliage-mask which
triggered our recognition ability. In the Bather Girl case it is seeing the picture
from a distance that enables the girl to be recognized. The visual mechanisms that
these examples trigger indicate that to the human visual system a picture is more
than the sum of its parts. It is not the just the information that is contained in the
actual marks that is doing the work, it is the visual system’s “bag of tricks” which
enables it to make quick judgments about the kinds of things which are in view
based on very little information. A depiction works in concert with the visual
system’s bag of tricks and triggers basic recognition mechanisms such as gist
recognition, visual grouping, shape-from-shading etc. These mechanisms are
radically counter-intuitive, very rough-and-ready and are often in conflict with
each other. That is, these rule-of-thumb recognition systems have built-in
redundancy – we can simultaneously see a picture as a round peach and a flat
picture of a peach without any recognitional conflict.

The neurological mechanisms underlying these recognitional abilities are only
now beginning to be understood. It is clear that our visual system comprises many
systems working in parallel and in concert with our other senses to enable us to
navigate the world and perceive objects. Although there is a hierarchy of visual
processing in the brain, there are feedback mechanisms between all levels.
Consequently, even in the first few hundred milliseconds of looking at something
our eyes are directed in their saccades by a combination of instinct, learning and
feedback. When we look at the Constable painting, it may seem that in the first half-a-second we see all the rich, coloured detail of the trees, the clouds, the cathedral, the pasture and the cows – in fact we don’t. We recognize a landscape all at once using the gist mechanism and probably use our knowledge of what to expect in a landscape to populate that landscape with grass, trees and clouds. How long is it before you notice the cows? How long is it before you notice the people? With all our visual systems working together and Constable triggering our visual system’s bag of tricks using his own artist’s bag of tricks the scene bootstraps itself into view.

The conventional and resemblance accounts speculate that our ability to see a landscape with trees in the Constable painting involves cognitive mechanisms of, on the one hand symbol decoding and on the other of property matching. Neither theory proposes a plausible mechanism for this speculation and in neither case have we found any evidence from cognitive science to support this speculation. We have established that resemblance is a very blunt instrument and it is impossible to identify a mechanism whereby shape-matching could be the basis of recognition. In fact, the constancy phenomenon suggests that resemblance, in terms of property matching, does not even play a functional role in how we recognize objects and scenes in real life.

Research into visual cognition is beginning to identify the neurological mechanisms which enable us to perceive the world and navigate through it. The bonus for a theory of depiction is that these studies invariably use pictures to test their theories. From a methodological perspective no one has yet identified significant differences in the ability of people to recognize pictures and real objects. All available evidence suggests that the neurological mechanisms are the same for recognizing pictures as they are for objects and scenes in the real world. Vision research is research into pictures and vice versa. It will probably be decades before we can identify all the cognitive processes which enable us to see Constable’s landscape with cows, but we know that they will be intimately linked to our ability to see real landscapes and real cows. And it is within the realms of possibility that one day we will find a neuron in our brain labelled “cow”.
Appendix: The Complete Pictures

These are the pictures from which the details in figures 46-48 were extracted.
Bibliography


Levin, Daniel T., and Melissa R. Beck. "Thinking About Seeing: Spanning the Difference between Metacognitive Failure and Success." In *Thinking and


Martindale, D. "One Face, One Neuron - Storing Halle Berry in a Single Brain Cell." Scientific American 293, no. 4 (2005): 22-+


