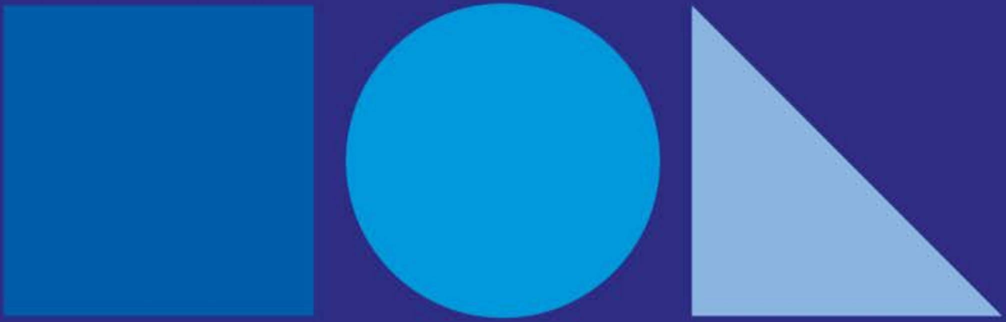
IOP Concise Physics | A Morgan & Claypool Publication

The Physics and Art of Photography

Detectors and the meaning of digital

John Beaver

VOLUME
THREE



IOP | ebooks

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver
University of Wisconsin Fox Valley, Menasha, WI, USA

Morgan & Claypool Publishers

Copyright © 2018 Morgan & Claypool Publishers

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher, or as expressly permitted by law or under terms agreed with the appropriate rights organization. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency, the Copyright Clearance Centre and other reproduction rights organizations.

Certain images in this publication have been obtained by the author from the Wikipedia/Wikimedia website, where they were made available under a Creative Commons licence or stated to be in the public domain. Please see individual figure captions in this publication for details. To the extent that the law allows, IOP Publishing and Morgan & Claypool Publishers disclaim any liability that any person may suffer as a result of accessing, using or forwarding the images. Any reuse rights should be checked and permission should be sought if necessary from Wikipedia/Wikimedia and/or the copyright owner (as appropriate) before using or forwarding the images.

Rights & Permissions
To obtain permission to re-use copyrighted material from Morgan & Claypool Publishers, please contact info@morganclaypool.com.

ISBN 978-1-64327-386-0 (ebook)
ISBN 978-1-64327-383-9 (print)
ISBN 978-1-64327-384-6 (mobi)

DOI 10.1088/2053-2571/aaf0ae

Version: 20181201

IOP Concise Physics
ISSN 2053-2571 (online)
ISSN 2054-7307 (print)

A Morgan & Claypool publication as part of IOP Concise Physics
Published by Morgan & Claypool Publishers, 1210 Fifth Avenue, Suite 250, San Rafael, CA, 94901, USA

IOP Publishing, Temple Circus, Temple Way, Bristol BS1 6HG, UK

For Judith

Contents

Preface **xi**

Acknowledgements **xiv**

Author biography **xv**

Part I The physics of light detectors

1 Detectors and the characteristic curve **1-1**

1.1 The physics of photons 1-2
1.2 Photoelectric detectors 1-3
1.3 Photochemical detectors 1-4
  1.3.1 Negative and positive 1-4
1.4 Basic photochemistry 1-6
1.5 The eye as a detector 1-8
1.6 Exposure, density, and the characteristic curve 1-12
  1.6.1 The characteristic curve and photoelectric detectors 1-21
  References 1-22

2 Silver gelatin photochemical detectors **2-1**

2.1 Black-and-white silver gelatin emulsions 2-1
2.2 Chromogenic color emulsions 2-4
2.3 Reversal-processed silver gelatin emulsion 2-4
2.4 Lumen process 2-6
2.5 Ephemeral process (EP) 2-8
  2.5.1 How does it work? 2-8
2.6 Instant film 2-10
  2.6.1 Peel-apart instant films 2-11
  2.6.2 Integral films 2-14
  References 2-14

3 Other photochemical detectors **3-1**

3.1 Daguerreotype 3-1
3.2 Wet collodion, ambrotype, and tintype 3-2
3.3 Cyanotype and Van Dyke processes 3-3
  3.3.1 New cyanotype 3-4
  3.3.2 Van Dyke brown process 3-5

3.4 Platinum and palladium processes 3-5
3.5 Gum bichromate 3-5
3.6 Anthotypes and chlorophyll prints 3-6
  3.6.1 Anthotypes 3-7
  3.6.2 Chlorophyll prints 3-8
  References 3-8

4 Some interesting technical details 4-1

4.1 Reciprocity failure 4-1
  4.1.1 Reciprocity failure and VLS photography 4-2
4.2 Solarization 4-3
  4.2.1 Sabatier effect 4-5
  4.2.2 True solarization 4-7
  4.2.3 Mackie lines 4-7
  4.2.4 Negative or positive? 4-9
  References 4-11

5 A brief diversion into the weird world of the photon 5-1

5.1 Young's double-slit experiment and the wave model of light 5-2
5.2 The photoelectric effect and the particle model of light 5-3
5.3 Young's experiment reconsidered 5-5
  References 5-6

6 Digital photoelectric detectors 6-1

6.1 CCD and CMOS array detectors 6-2
6.2 The physics of CCD arrays 6-3
6.3 Color digital detectors 6-5

7 Unusual detectors and 3D photography 7-1

7.1 Stereo photography 7-1
7.2 Light-field photography 7-3
7.3 Autochrome Lumière process 7-4
7.4 Holography 7-4
7.5 Lippmann process color photography 7-5
  References 7-7

Part II Photography as an art and the meaning of digital

8 Comparison of digital and film techniques 8-1
8.1 Borders and cropping 8-2
8.2 Brightness and contrast adjustments 8-4
  8.2.1 Digital contrast adjustments 8-6
  8.2.2 Contrast adjustments in the darkroom 8-6
  8.2.3 Levels and curves adjustments 8-9
  8.2.4 Levels and curves in the darkroom: the zone system 8-13
8.3 Dodging and burning 8-15
  8.3.1 Dodging and burning with GIMP 8-15
  8.3.2 Spot healing and retouching 8-15
  8.3.3 Digital retouching 8-16
8.4 Color darkroom vs digital 8-17
  8.4.1 Contrast control 8-17
  8.4.2 Color balance 8-18
  References 8-18

9 The digital and the analog 9-1
9.1 Pixels and granularity 9-1
9.2 Resolution 9-3
9.3 Signal and noise 9-7
  9.3.1 Pennies and Poisson 9-7
  9.3.2 Photons, signal and noise 9-11
  9.3.3 Signal-to-noise ratio 9-12
9.4 Digital photography and the data revolution in astronomy 9-14
  9.4.1 Digital detectors are reusable 9-14
  9.4.2 Linear response 9-15
  9.4.3 Dynamic range 9-16
  9.4.4 Quantum efficiency 9-20
  9.4.5 Image calibration 9-20
  References 9-23

10 Is digital manipulation cheating? 10-1
10.1 Paying one's dues 10-1
10.2 Honesty 10-2
10.3 Retouching 10-4

10.4 Digital filters and cliché 10-5
References 10-6

11 The image, the object, and the process 11-1

11.1 Some preliminary ideas 11-1
  11.1.1 Photographic and representational content 11-2
  11.1.2 The picture plane 11-4
  11.1.3 Control and happy accidents 11-5
  11.1.4 Negative versus positive 11-6
  11.1.5 Order, complexity, and randomness 11-7
  11.1.6 The new antiquarian movement 11-9
  11.1.7 The archival ethos 11-9
11.2 Four photographers and a musician 11-10
  11.2.1 Almudena Romero 11-11
  11.2.2 Caitlin Noll 11-11
  11.2.3 Diane Fenster 11-13
  11.2.4 Chrystal Lea Nause 11-14
  11.2.5 Hal Rammel 11-15
11.3 Examples from lumen and ephemeral process photography 11-17
  11.3.1 EP pictures from pictures 11-17
  11.3.2 Limited edition prints from EP negatives 11-20
  11.3.3 Ephemeral prints 11-22
  11.3.4 To …, or not to … 11-25
  11.3.5 EP accelerator transfers 11-26
11.4 Drawing from negatives 11-27
11.5 The *camera stupida* 11-30
References 11-33

12 Towards an art and science of nature 12-1

12.1 A personal note 12-1

Appendices

A Making ephemeral process (EP) negatives from chromogenic prints A-1

B The optics of the camera stupida B-1

C Units, dimensions, and scientific notation C-1

Preface

Early drafts of this book were written for a course I first taught in the Fall of 2013 at the University of Wisconsin–Fox Valley, in Menasha, Wisconsin.

I assume no specific prior knowledge of the reader except for a very basic understanding of physical units, dimensions and scientific notation; a brief review can be found in appendix C. The mathematics presented in the text is rudimentary, with only the most basic of algebra (more detailed derivations, or those that require calculus, are relegated to the appendices). A familiarity with the material in Volumes 1 and 2 is not essential prior to reading this volume, but it is helpful.

If you have little experience with photography, it is my goal that *The Physics and Art of Photography* will help form a useful foundation from which to learn about photography in whatever way that works best for you. If you are a seasoned pro, but looking to set off in a new direction, then I still hope that you will find much here that is fresh and inspiring, and it is my goal that the book will help to open new possibilities. *The Physics and Art of Photography* is in three volumes:

- **Volume 1: Geometry and the nature of light**
    - Part I: Some preliminary ideas
    - Part II: The nature of light
    - Part III: Geometry and two-dimensional design
- **Volume 2: Energy and color**
    - Part I: Energy and photography
    - Part II: The art and science of color
- **Volume 3: Detectors and the meaning of digital**
    - Part I: The physics of light detectors
    - Part II: Photography as an art and the meaning of digital

The Physics and Art of Photography covers some material that is typical of discussions that link physics and photography. But it is also personal; it is very much my own take on the two subjects. I would not say that my personal views regarding science and art are controversial, but they are perhaps somewhat unconventional. There are few details here that other artists and scientists are likely to strongly disagree with. It is, rather, what I have chosen to emphasize, what I have left out all together, and the particular connections I point to, that most shows my own personal likes and dislikes.

Since my formal training is in physics and astronomy, while I am essentially self-trained in art (with informal mentoring from many others), the science part of this book is perhaps more conventional and straightforward than is my portrayal of art. And so my choice of physics-related topics should give one a fairly balanced and conventional taste of that subject as it relates to photography. Regarding photography as an *art*, however, I am surely on shakier ground.

Certainly, I do not pretend to present a comprehensive or balanced overview of art photography; I am unqualified to attempt such a thing. But I do try to make a case that the particular thin slice that I present here has some merit and is worth spending a little time to consider, even if it turns out not to be your particular cup of tea. This book is a bad place to get a sense of what are the hot topics in ArtForum, but I believe that it does at least point to important and interesting questions about art photography in general. And since it is my goal to get you thinking, it doesn't matter much whether you agree with me or not. Thus, it is fitting that my discussion of art is more personal, since my own art is the wee bit for which I really do know what I am talking about.

And so one might complain that *The Physics and Art of Photography* is a very long artist's statement, justifying the value and relevance of my own art. That may be partly true, but I do try to approach it in a way that emphasizes broad *questions*, rather than the particular answers I try to give (tentatively) with my own art. And I hope this book does help a little to make you a better photographer, and as such I do spend time on some of the very basic technical aspects of photography that I find important. But in doing so, I try to use these technical issues as points of departure to consider the status of photography as an art, finally exploring some issues relating to this status in the digital age.

This book may also be read as a manifesto of sorts for the aspects of science that have always moved me the most. I am interested in science not for the technological gizmos it has produced, or for some notion of inevitable human 'progress.' Rather, science is, for me, part of *the study of nature*. My interest in Einstein's General Relativity, for example, is essentially the same as my interest in bird watching. Because I have spent some time to learn a bit about birds, I can now walk through the woods free of binoculars, looking only at the ground at my feet, and a world is open to me just by the sounds I hear. And when I stumble on my way up the stairs, as a physicist I can take comfort in the idea that my shin in contact with the stair prevented me from following my normal straight-line path through four-dimensional spacetime.

You will find throughout the book illustrations from my own photography as examples. This is convenient, since I know my own pictures and the stories behind them, and I don't need permission to use them. But of course I also want you to look at other photography, and so I have included some examples from a few other artists whose work I admire.

A useful companion is *The Photography Book* (Phaidon Press, 2014), which presents hundreds of photographs, spanning the entire history of photography. Each has a short analysis, with cross references to other photographs that are related. The photographs, only one per photographer, are arranged in alphabetical order by photographer's name. Thus, the ordering of the pictures is thematically random, which often results in unusual juxtapositions on facing pages. I sometimes refer to pictures in *The Photography Book* as examples, and so it is useful to have it handy. But all of these pictures are famous and can easily be found online as well.

The reader will also find, scattered throughout the three volumes and their appendixes, details and examples from what I call *ephemeral process (EP) photography*. EP photography is my own invention—sort of—and I spend so much time on it because it is perfect for illustrating many of the concepts in *The Physics and Art of Photography* in a way that I believe goes directly to the heart of the matter. Furthermore, it is *accessible*. The materials and equipment are inexpensive, it requires no specialized facilities (such as a darkroom), and it is surprisingly versatile. But most importantly, it is a lot of fun. Practical details of the technique can be found in appendix A and in volumes 1 and 2 of *The Physics and Art of Photography*.

The larger concerns of *The Physics and Art of Photography* are to give the reader some background that is helpful for asking important questions about the nature of art and science. But the practice of photography is the point of departure for these bigger issues, and as such *The Physics and Art of Photography* does contain a lot of simply practical information as well. And so *The Physics and Art of Photography* has five basic goals:

1. To ask basic questions about how photography fits in as an *art*, and about the nature of art itself.
2. To ask basic questions about the nature of physics *as part of the study of the natural world*, and about the nature of science itself.
3. To gain some practical knowledge that will allow the reader to more easily learn technical aspects of photography, as they are needed.
4. To gain some practical knowledge that will help the reader more easily learn to be a better photographer.
5. To expose the reader to a set of interesting photographic processes and tools that are not usually covered in a beginning photography course.

One of the themes of this book is the meaning of digital technology and what it has to say regarding photography as an art form. This may seem like I am speaking out of turn here, since I have neither formal training in art, nor have I ever been a professional photographer using professional digital equipment. Nevertheless, there is a sense in which I am well-positioned to say something of interest about these issues.

My own photography is almost entirely devoid of the use of a digital camera. I often use equipment and old physical processes that are about as far removed from modern digital photography as one could imagine. But I use these in new ways that depend absolutely on the digital; many of my photographs could not exist without modern digital processing and scanning and printing. This kind of interplay between the old and new is one of the running themes of *The Physics and Art of Photography*.

And despite my collection of old cameras, I am not a knee-jerk hater of digital imaging technology. In fact, I am one of its early practitioners, having used digital cameras and sophisticated digital image processing long before most photographers. My formal training is in astronomy, and I was there (in graduate school) for the digital revolution as it transformed astronomy in the 1980s. The CCD digital detectors used in modern digital cameras were fairly new then, and still too expensive (and with insufficient resolution) to be of much practical use for photographers. I am the last person one would want to ask about the latest multi-thousand-dollar model of DSLR camera. But I do have a decades-long understanding of some of the most basic underlying principles of digital photography.

Acknowledgements

I thank Valeria Sapiain for all-around support and patience over the several years I spent working on this book. I received much valuable feedback, support and mentoring from Laura Andrews, Doug Fowler, Caroline Geary, Anne Haydock, Diana Ludwig, Dawn Patel, Teresa Patrick, Judith Waller, and Frank Zetzman.

The software packages GIMP, Gnuplot, Inkscape, SciDAVis, OpticalRayTracer and IRAF were used for many of the illustrations. All photographs and illustrations are by the author, except as noted below:

Figure 1.1: By Unknown—illustration from the Scientific American Supplement, January 11, 1879, Public Domain.

Figure 1.2: By Fig_retine.png: Cajalderivative work Fig retine bended.png: Anka Friedrich (talk)derivative work: vectorisation by chris—Fig_retine.pngFig retine bended.png, CC BY-SA 3.0.

Figure 1.3: By RicHard-59—Own work, CC BY-SA 3.0.

Figure 4.4, right-hand side: Image used by permission of Teresa Patrick.

Figure 5.1: Public Domain.

Figure 6.3: By en:User:Cburnett—Own work This W3C-unspecified vector image was created with Inkscape., CC BY-SA 3.0.

Figure 7.3, left-hand: Public Domain.

Figure 9.3: Georges Seurat, Art Institute of Chicago, CCO Public Domain Designation.

Figure 10.1: Ralph A. Clevenger, 1999, used by permission of Ralph A. Clevenger.

Figure 11.7: Photograph used by permission of Almudena Romero.

Figure 11.8: Photograph used by permission of Almudena Romero.

Figure 11.9: Image used by permission of Caitlin Noll.

Figure 11.10: Image used by permission of Diane Fenster.

Figure 11.11: Images used courtesy of Chrystal Lea Nause.

Figure 11.12: Images used by permission of Hal Rammel.

A portion of this work was carried out with the support of the University of Wisconsin Colleges sabbatical program, to which I extend my heartfelt thanks.

Author biography

John Beaver

For nearly 20 years, John Beaver has used old processes to make new negatives, often in ways that can only be realized as a print with digital scanning and printing. This includes his development of the cyanonegative process, innovative work (in collaboration with Teresa Patrick) with instant film, and most recently his development of an accelerated, unfixed printing-out process he calls (perhaps annoyingly) 'Ephemeral Process photography.'

He is Professor of Physics and Astronomy at the University of Wisconsin–Fox Valley, where he teaches physics, astronomy, photography, and interdisciplinary courses. He earned his BS in physics and astronomy in 1985 from Youngstown State University, and his PhD in astronomy in 1992 from Ohio State University. His published work in astronomy is on the topics of spectrophotometry of comets and gaseous nebulae, and multi-color photometry of star clusters.

He has exhibited photographs in many juried competitions in Wisconsin, Ohio, New York, Louisiana, Missouri, Oregon, and Colorado, even occasionally winning an award or two (well, two actually). He has had several solo exhibitions, as well as joint shows with artists Judith Waller, Diana Ludwig, Dawn Patel, and Teresa Patrick. Beaver has long been involved in art–science collaborations (many with artist Judith Waller) in the classroom, at academic conferences, and in art galleries and planetaria.

Some of John Beaver's photography can be seen at http://www.JohnEBphotography.com

Part I

The physics of light detectors

The Physics and Art of Photography, Volume 3
Detectors and the meaning of digital
John Beaver

Chapter 1

Detectors and the characteristic curve

Part of the essence of photography is that light falling on a surface *is recorded*. This is the *graphy* in photography. A permanent record is made of the intensity (and maybe color) of light that fell on each point of a given surface. To do this one must have some material that undergoes a physical change when light interacts with it. This material must form a surface for the image to focus on, and there must be some way to record different physical responses at different positions on the surface.

Throughout *The Physics and Art of Photography* I use the word *detector* to represent, in a general sense, any light-sensitive surface used to record an image. It could be the retina of the human eye, traditional photographic film, a digital detector in a digital camera, or a light-sensitive material that is used to make a print. Ultimately, there are three crucial features of every light detector:

1. Light must cause some physical change to occur in the detector.
2. Once the exposure is completed, there must be a way to stop further physical changes from occurring. Or barring that, there must be some way to 'read' the results of the detector, and transfer those results to some permanent, non-light-sensitive storage.
3. There must be some way to tell, after the exposure is complete, which part of the detector underwent what physical change. That is, we need to be able to record *separately* the physical changes that occur at different locations on the detector. Otherwise, we would have not an image, but rather only a single measure of brightness.

Light *interacts* with some physical material. That is, some of the light disappears, and the material changes in some way. The idea that some materials change visibly upon exposure to light is far from modern, since there are many common everyday examples. For example, many dyes and pigments fade with light over time; when a piece of furniture shields part of a painted wall from light, its outline is visible when it is moved years later.

doi:10.1088/2053-2571/aaf0aech1

My usage of the word detector is somewhat unconventional in that I use it to refer to any two-dimensional surface used in photography to make a record of an exposure to light—whether or not its use is for the original capture of an image in a camera, or instead to make a permanent photographic print. We can divide most photographic light detectors into two broad categories—photoelectric and photochemical. The retina of the human eye (section 1.5) is more complex than either, and includes aspects of both.

1.1 The physics of photons

The physics of light is described in some detail in volume 1 of *The Physics and Art of Photography*, but here I recap and expand upon some concepts that are especially relevant to light detectors. Light has a *wave-like* property; it does all of the basic things that waves do. But it also has a *particle-like* nature; in many circumstances it transfers energy in discrete clumps, at particular points in time and space. Such a particle of light is called a *photon*. These aspects of light seem at first glance to contradict each other, but they do not. The two concepts are unified by *quantum physics*, and we take up this subject in more detail in chapter 5. Photon interactions are of particular importance for understanding photographic light detectors, but the wave-like properties of light play a role too. Below I outline some of the most basic principles of the physics of light, as it pertains to photographic light detectors.

- Light is an electromagnetic wave—a changing pattern of the electric and magnetic properties of space, that travels through space at the speed $c = 3.8 \times 10^8\ \mathrm{m\ s^{-1}}$.
- The perceived color of light is related to its *wavelength*—the distance between successive maxima in the electric (or magnetic) fields, at a given point in time. It can also be described by its *frequency*—the number of electric field maxima that pass a given point in space per second. Frequency is measured in *Hertz* (Hz) or repetitions per second. Wavelength is a length, and so it can be measured in meters—but the wavelengths of visible light are sub-microscopic, and so the nanometer (nm), or 10^{-9} m, is more commonly used. The human eye is sensitive over the range approximately 400–700 nm. We perceive 400 nm light as violet and 700 nm light as red.
- There is an inverse relationship between wavelength, λ, and frequency, f, for light. And the speed of light, c, relates the two:

$$c = \lambda f \tag{1.1}$$

$$f = \frac{c}{\lambda} \tag{1.2}$$

$$\lambda = \frac{c}{f}. \tag{1.3}$$

- Light carries both energy and momentum, and when light interacts with matter, one or both must be transferred. It is the transfer of *energy* that is important for understanding light detectors; in order for light to be detected, some energy must be transferred from the light to the detector.

- The smallest unit of energy transferred by light is the photon. A photon interacts with an individual atom or molecule at a particular point in time.
- When light interacts in a wave-like manner, resulting in wave phenomena such as interference and diffraction, we can determine the frequency of the light. If that same light is allowed to interact in a particle-like manner, we can measure the energy of those photon interactions. There is a direct relation between the energy, E, of the *individual* photons and the frequency, f, of the light of which they are a part:

$$E = hf \tag{1.4}$$

$$= \frac{hc}{\lambda} \tag{1.5}$$

  where h is a tiny constant called Planck's constant.
- Photons of higher energy correspond to light of shorter wavelength (and higher frequency). Photons of lower energy correspond to light of longer wavelength (and lower frequency).

1.2 Photoelectric detectors

A *photoelectric detector*, upon exposure to light, directly produces some kind of measurable electronic response. Most digital photographic detectors are of this type, and they are usually an *array* of rows and columns of separate light-sensitive sites, called *pixels*.

There are different ways in which light can interact with conductors or semiconductors and cause a measurable electrical effect. For example, individual photons (particles of light) can remove electrons from their atoms. If this happens in an evacuated glass tube, the electrons can be accelerated through an applied voltage to make a measurable electric current. This procedure is not ordinarily used by photographers to record images, but devices called *photomultipliers* employ this principle for the precise measurement of very faint light (mostly by astronomers and particle physicists). Photomultipliers can easily detect individual photons.

Other devices can be designed for which light alters its electrical resistance (or conductance). A *photoresistor* becomes less resistant to electrical current when light shines on it. This can be used in a circuit with a current meter and a battery to make a light meter—an instrument for measuring the brightness of light. Indeed, many older cameras used photoresistors for their internal light meters. But photoresistors are not particularly useful for recording *images*.

A photovoltaic cell turns light directly into electrical energy. A voltage is produced, and power can be delivered, when it is illuminated by light. Very low power photovoltaics (the element selenium is a good example) were used for some of the first electronic light meters. But like photoresistors, photovoltaics have not been particularly useful for forming images.

Finally, semiconductor *photo-junction devices* use the interaction between light and a boundary between two dissimilar semi-conducting materials. These devices can be manufactured in arrays of microscopic, separately light-sensitive pixels. We will examine two types of photo-junction photographic imaging devices in chapter 6.

1.3 Photochemical detectors

A *photochemical detector* can be thought of as any material for which the exposure to light produces some kind of chemical change. If these changes are visible to the eye (or can be made visible by further chemical processing), then the material can be used as a detector. There are two main categories of photochemical detectors:

1. A *latent image* is formed upon exposure to light. A change occurs at the molecular level, but no visible change is apparent until the detector undergoes a chemical treatment called *development*. In this case we say the detector *develops out*. Ordinary silver-based color or black-and-white film photography is of this type.
2. The detector *prints out*; visible changes occur in the detector as it is exposed to light, with no chemical development needed. Many antiquarian and alternative processes are of this type.

There is an important complication with most photochemical detectors, and this was an important stumbling block in the history of their invention. There must be some way to stop the detector's sensitivity to light after the image is formed. Otherwise, the act of looking at the picture would continue to expose it to light. This process is traditionally called *fixing* the image, and it usually requires some kind of additional chemical treatment of the detector.

Another important aspect of photochemical detectors is the nature of the particular *base* or *substrate* that is used. What type of flat surface is the light-sensitive material attached to? What is that substrate made from, and by what means does the light-sensitive chemical attach to it? Is the base transparent or opaque? And if it is opaque, is it black or white? If the substrate is transparent, one can shine light through it, perhaps to use lenses to enlarge and project the image onto another surface.

Two different approaches are most common for attaching a photochemical detector to its substrate. The simplest method is to allow the light-sensitive chemical to penetrate into the fibers of a semi-porous surface such as paper, in essentially the same way that the dyes soak into the paper in a watercolor painting. The second technique is to allow the light-sensitive chemical to form microscopic crystals, and to then suspend these in some sort of gelatinous material, which is then applied to the top of the substrate as a thin coating. This is called an *emulsion*, and most modern photochemical detectors use this method.

1.3.1 Negative and positive

Since photochemical detectors may result in a *visible* change in the detector upon exposure to light, there is the possibility of a direct physical relation between the picture and the original exposure. That is, a greater exposure to light leads to a greater *visible* effect on the detector. But this leads to an important question: does a greater exposure to light cause the detector to become visibly brighter, or does it instead make the detector visibly darker? A *negative process* results when the detector is rendered darker by greater exposure to light. A *positive process*, is of course the opposite.

If our immediate task is to produce a picture for the wall, then we most likely want bright areas in the picture to represent more light in the world. It would seem, then, that the detector of choice would be one that uses a positive process. It would allow for the picture on the wall to be the actual detector that was exposed in the camera. We call this a *direct positive*, and we will describe a few photochemical detectors that work this way.

But the majority of photochemical processes suitable for photographic detectors work as a negative process. And for those processes we must take some extra steps in our journey from the camera to the gallery wall. We shall see, however, that negative processes have some distinct practical advantages over positive processes. There are two traditional approaches to get from a negative image to a positive image, and we consider them in turn.

Negative-on-negative

The most straightforward way to get a positive image from a negative process is to recognize that a negative of a negative is a positive. If one simply repeats the negative process, *but using the result of the original exposure as the source for a second exposure*, then a positive image results.

For example, one can expose a sheet of 8 × 10 inch light-sensitive *enlarging paper* in a pinhole camera (see volume 1 of *The Physics and Art of Photography*). Upon processing, a negative image results; the brightest features in the scene become the darkest areas on the paper. This sheet of paper with its negative image can then be placed (in a darkroom) image-side-down directly onto another sheet of the same kind of paper, with its light-sensitive side facing upward. One can then simply expose the second sheet by shining light through the back of the negative image. Upon processing the second sheet, a positive image appears because it is exposed the least (and so appears the brightest) wherever the original negative was darkest.

This is called a *contact print*, and it is an easy way to make a positive from a negative. Clearly, the print is the same size as the negative. So if the negative is made in a camera, it must be a very large camera if one wants a large print. If the original negative is made on a transparent film, then one can expose the light-sensitive print paper by projecting light through the negative with a lens—sort of a camera in reverse. Such a device is called an *enlarger*, and it is the most common traditional method for making large prints from small negatives.

The negative-on-negative process has the obvious practical advantage that it separates the process of recording an image in a camera from the process of making a print to hang on the wall. And so it opens the possibility of making a large number of prints from a single image capture. It also allows for adjustment of the image in the step of making the print—and so maybe one can get a good print out of a not-so-good image capture. For these reasons, the negative-on-negative approach has dominated much of the history of photochemical photography.

Reversal processing

There is another traditional way to get a positive image out of a negative photochemical process, and it is called *reversal processing*. It is sometimes possible to process the light-sensitive material in a clever two-step way, first producing a negative, and then reversing the negative to a positive directly in the same piece of light-sensitive material. So we have, in the end, something that is much like a direct positive—the picture on the wall is the same object that was in the camera.

The history of reversal processing is almost exclusively associated with silver gelatin emulsions, and so we take up some of the technical aspects more fully in that context (chapter 2, section 2.3).

Digitally-printed negatives

And finally, there is a not-so-traditional way to get a positive image with a negative process. Any digital image can now be easily made into a transparent negative with image processing software and digital printing. The digital image, inverted to a negative by the software, is printed onto acetate with an inkjet printer, at the same size as the intended print. This negative can then be exposed in contact with a negative-process light-sensitive paper to make a positive print. Some photochemical printing materials are sensitive only to ultraviolet light, and this makes it very difficult to expose them with an enlarged projected image. And so they must be contact printed, and for that a negative is needed that is the same size as the final print. It is increasingly common for photographers to use digital inkjet negatives for this purpose.

1.4 Basic photochemistry

There is a sense in which most photochemical detectors are also photoelectric—a photon is absorbed and some transfer of electric charge results between atoms and molecules. Chemistry is about the electrical attraction between the positively-charged nucleus of an atom and its negatively-charged electrons. The nuclei do not change (if they do, it is called a *nuclear* reaction). But it is possible for the electrons, especially those that are most distant from the nucleus, to transfer from one atom to another, thus taking negative charge away from one and adding it to another.

An atom or molecule that has a net electric charge is called an *ion*[1]; it is an *anion* if it has a net negative charge and a *cation* if it has a net positive charge. The process of an atom or molecule decreasing its number of electrons is called *oxidation*, while the process of increasing its number of electrons is called *reduction*.

---

[1] A *molecule* is often defined to be a bonded-together group of atoms that is electrically neutral. And so, for example, the proper term for a similar structure, but containing a net electric charge, is a *polyatomic ion* or *molecular ion*.

The *oxidation state* is related to the net charge of an ion, and it is denoted by a superscript number with a following plus or minus sign, affixed to the chemical symbol[2]. And so, for example, Ag^+ is a silver (Ag) atom that has lost one electron; taking away -1 is the same as adding $+1$, and the number 1 is assumed, if no other digit follows. Neutral silver, with just the right number of electrons to balance its internal charge, is simply denoted Ag, with no superscript.

Like charges repel and unlike charges attract, and so an isolated atom wants to be neutral, with just the right number of electrons to balance the positive charge of its nucleus (it is the positive charge of the nucleus that gives an atom its identity). Because of the intricate rules of quantum physics, however, there is more to it than that, and when atoms are put together with other atoms, other factors come into play.

Certain molecular combinations of atoms are only stable if they are ions—if they are either missing or have extra electrons. A good example is *nitrate*, which is a nitrogen atom bound to three oxygen atoms. But this only happens if it can find an extra electron, and so form NO_3^-, a nitrate cation. Silver, on the other hand, is happy as just ordinary Ag. But in the right context, it is more quantum-mechanically happy to give up one of its electrons and instead be Ag^+. Put Ag^+ together with NO_3^- and they stick together (unlike charges attract) to form the neutral *compound* silver nitrate ($AgNO_3$).

A compound such as silver nitrate does not mean that one particular Ag^+ is stuck specifically to one particular NO_3^-. Instead, it forms a *crystalline solid*—a regular arrangement of alternating Ag^+ and NO_3^-. It can also be dissolved in water to form an *aqueous solution* in which the ions are separated from each other by the water molecules, constantly associating and dissociating, no ion belonging to any other ion in particular. Compounds that will do this easily are called *soluble*, while those that don't are called *insoluble*.

Some atoms or compounds have electron structures that make it easy for them to lose electrons, donating them to some other chemical species in a chemical reaction. The act of losing electrons is called oxidation, but the atom or compound that does so is called a *reducing agent*. In the opposite sense, an atom or compound that easily accepts electrons—and so is likely to 'steal' them from some other chemical species —is called an *oxidizing agent*. When an oxidizing agent accepts electrons, it is *reduced* in the process.

And so what does all of this have to do with *light*? In certain circumstances the absorption of a photon—a particle of light—can provide the energy needed for the transfer of electrons from one chemical species to another, and so cause chemical changes to occur. When this happens, it does so on a one-on-one basis. A photon (or several photons in a very short period of time) transfers its energy all at once, and at one particular place. If these chemical changes enabled by the absorption of light produce some visible change, then they can be used as the basis for a photochemical detector.

One of the great feats of early 20th-century modern science was the discovery of fundamental physical laws that provide explanations for the mostly already-known laws of basic chemistry. The principles of quantum physics allow one to calculate the

---

[2] It is, strictly speaking, only the net charge if all bonds to the atom are ionic.

electron structures of atoms, and also the rules by which those structures may change when a given atom is combined with other atoms. These rearrangements of atomic electron structure involve transfers of energy, and this is the ultimate source of the energy transfers associated with chemical reactions.

Why, for example, is NO_3^- one of the ways in which nitrogen and oxygen like to arrange themselves? The question can be answered on the one hand, in terms of pre-20th-century concepts of chemistry such as valence, the periodic table of elements, acids, bases, etc, and the complex chemical rules that relate these concepts to each other. The fundamental laws of physics, however, were too incomplete in 1900 to provide an underlying physical explanation for these chemical rules. It was not until the development of quantum mechanics in the first third of the 20th century—a revolution in the fundamental understanding of matter and energy—that physics was up to the task of providing an explanation for many of the facts that chemists already knew (Chomsky 2000, pp 110–1).

1.5 The eye as a detector

They eye looks like a camera. There is a lens, a dark box (albeit nearly-spherical in shape), and a light detector at its back—the retina—upon which an image is focused. The *retina* is covered by a matrix of millions of light-sensitive cells, seemingly similar to the light-sensitive pixels on the detector of a digital detector. And these cells are 'wired' to the brain by the optic nerve. And so it is tempting to think of the eye as simply a biological version of a modern digital camera. But the differences are profound.

The English word 'photography' was apparently coined by John Herschel, astronomer and one of the pioneers of photography[3]. The combination of *photo* (light) and *graphy* (writing) implies that we are 'writing with light,' and indeed that was literally the first photographic method. Centuries before photographic detectors were invented, the *camera obscura*—a dark box with a lens to focus an image—was used to trace images by hand.

An even more convenient arrangement was patented in 1807 by William Hyde Wollaston, three decades before the first successful photographic process. The *camera lucida* used a clever combination of mirrors and prisms that allowed one to see a virtual image of the subject superimposed onto a piece of drawing paper, as in figure 1.1. This allowed the artist to mark key points in the image—corners, ends of lines, locations of important features—in order to draft an exact photo-realistic drawing.

John Herschel was an avid user of the camera lucida (see Evans *et al* 1969, for many examples). Its use influenced both Herschel and the more-famous pioneer of photography, William Henry Fox Talbot. Both were intrigued by the idea of making a permanent photographic record, but somewhat frustrated by the imperfection and tedium of the hand-drawing process. And so the idea of making a direct, permanent recording of the image formed by a lens was already 'in the air' by the late 1830s when the first chemical photographic detectors were invented.

---

[3] The French word *photographie* was coined independently, at about the same time, by Hèrcules Florence.

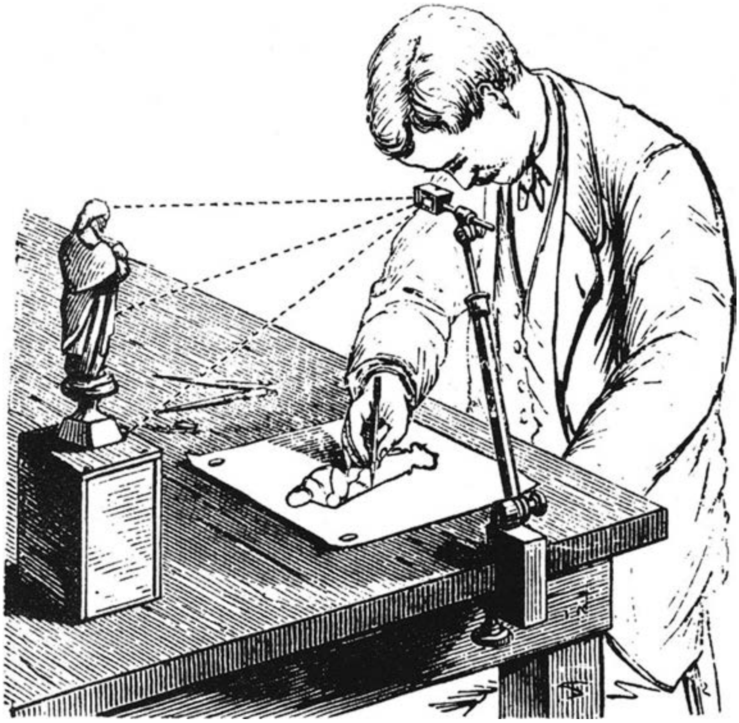


Figure 1.1. A *camera lucida* uses an arrangement of lenses and prisms to superimpose an image onto a piece of drawing paper. The image is then traced by hand. Graphic credit: Public Domain.

The physical process by which the retina detects light is complex and multi-stepped. There are both chemical and electrical changes, as well as changes in the physical structure of individual molecules. The retina is elaborately structured, and it is inside out; the light-sensitive part is on the back surface, not the surface facing the eye lens. Thus, the light focused by the lens must pass through several layers of only semi-transparent cells before being detected. Figure 1.2 illustrates this structure. The two basic types of light-sensitive cells—called *rods* and *cones* because of their shapes—are on the right of the diagram; the light focused by the eye lens enters from the left.

As is the case for most light detectors, we must appeal to the particle-nature of light to understand even the simplest aspects of the process. Within the rod or cone cells, individual photons are absorbed, and this triggers a complex set of chemical pathways that ultimately leads to neurological stimulus. The stimulus is transmitted to other parts of the brain by neurons (seen at the far left edge of figure 1.2) that ultimately bundle together into a nerve cord—the optic nerve. I say 'other parts' of the brain, because the retina and optic nerve are considered to be a part of the brain, not separate organs; the retinal tissue develops directly from part of the embryonic forebrain (Kolb 2003).

The rods and cones are elongated in shape, and they are structured like a stack of thousands of thin disks. Each of these disks is separately light-sensitive. Since the photons arrive grazing along these stacks of disks, there is a high probability that an

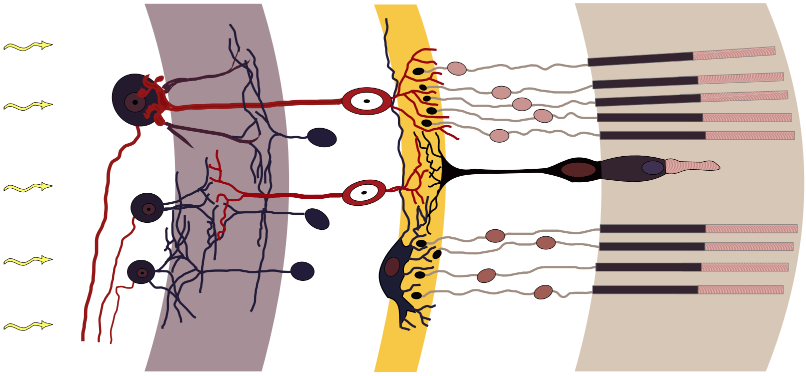


Figure 1.2. A diagram of the structure of the human retina. Photons focused from the eye lens enter from the left. The light-sensitive rod and cone cells are on the far right side of the diagram. Graphic credit: CC BY-SA 3.0.

individual photon will interact with one. And so the rods and cones are highly sensitive to light, apparently able to detect even single photons. The retina overall, however, is much less efficient; the majority of photons are absorbed by the overlying layers of the retina and so never make it to the layer of the rods and cones (Kolb 2003; Holmes 2016).

The key to the light sensitivity of the rods and cones is the molecule *retinal*. It couples to different proteins called *opsins* to form *photo-pigments*—*rhodopsin* (also called *visual purple*) in the rods and three different types of *photopsins* in the cones. The opsins are permanently located in the rods and cones, but the source of the retinal necessary for the light sensitivity is a layer of tissue called the *retinal pigment epithelium*, located just outside the retina (the far right-hand edge in figure 1.2).

The visual process is initiated when an individual photon interacts with the retinal part of a photo-pigment molecule. The photon disappears, but its energy does not; it is used to raise an electron in the retinal to a higher energy level, and this causes the molecule to change shape in a process called *photoisomerization*.

An *isomer* is a particular way in which a given set of atoms can be arranged to form a molecule. As an example, the molecules allene and propyne are both made of three carbon atoms and four hydrogen atoms. But these same atoms can link together in two different ways. And although these two molecules have the same combination of atoms, they have different geometrical shapes, and this gives them somewhat different chemical properties. And so *isomerization* means simply that the atoms of a molecule rearrange themselves in some way; the molecule changes shape but still contains exactly the same atoms as before. Thus, *photo*-isomerization is a change in the shape of a molecule brought about by exposure to light. Figure 1.3 shows the two isomers of retinal, the second being the result of photoisomerization.

Once the retinal changes to a different isomer, a completely new set of chemical pathways lies before it, and this triggers a complex chain of electrochemical events that eventually leads to a nerve impulse traveling along the optic nerve to the brain.

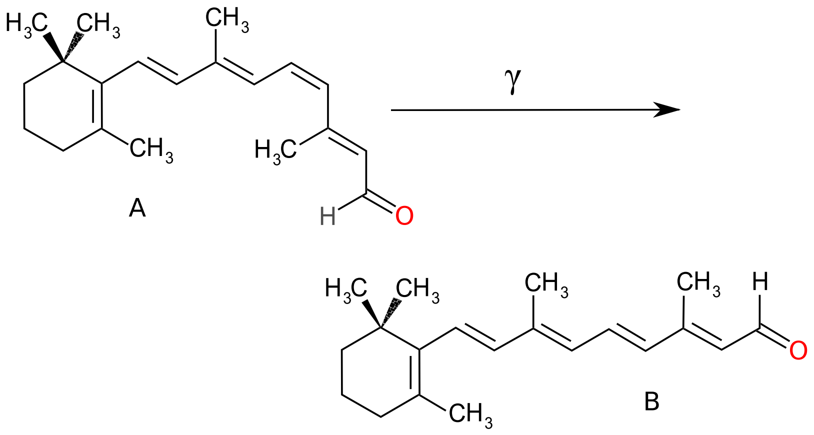


Figure 1.3. Photoisomerization of retinal. Graphic credit: By RicHard-59—Own work, CC BY-SA 3.0.

At a superficial level, the retina of the eye is similar to the film or digital detectors one would use with a camera. In both cases the detector is covered by many individual microscopic detectors, each of which is separately sensitive to light. For the retina of the eye, these are special light-sensitive cells. But for most of the detectors that we will consider, the process of going from absorption of photons to the formation of an image is relatively straightforward. Not so for the human eye! This book is not the place to work through the complex details of human vision, but I list some of the more striking differences. Further details can be found in, for example, Kolb (2003) and Kolb *et al* (accessed 2018), and regarding color vision in Conway *et al* (2010) and Schmidt *et al* (2014).

1. There are two basic kinds of light-sensitive photo-receptor cells in the eye, and they are not distributed uniformly. The cone cells are responsible for color vision, and they are very densely concentrated at the center of the retina (in a region called the *fovea*). For humans, the cone cells come in three types, sensitive to three overall parts of the spectrum—short wavelengths (S), middle wavelengths (M) and long wavelengths (L). The rod cells are more densely arranged at the periphery of the visual field and they are far more sensitive to light than the cones. They play a key role in night vision.
2. The electronic signal from the detector in a digital camera is typically processed by a computer; some of this is performed after the exposure by microprocessors within the camera itself. But the light detector in the eye—the retina—elaborately *pre-processes* the direct signal from the rods and cones *before* signals are transmitted along the optic nerve, in ways that are only partially understood. This is perhaps not too surprising when we recall that the retina *is* part of the brain. This pre-processing is quite sophisticated, and some of it occurs via the very neural wiring in two layers of specialized cells and neurons that lie over the layer of rod and cone photo-receptors (Kolb 2003).
3. Color photographic detectors make use of colored dyes or filters to render sets of light-sensitive elements responsive to different ranges of wavelengths, in order to

ultimately synthesize color. In this general sense, the retina is similar—it too has three types of cone cells, sensitive to different ranges of wavelength. But the similarity ends there. The neural wiring of the retina combines the signals from the different cone cells to make *opponents—direct comparisons* of photoreceptor responses, rather than the responses themselves.

4. Unlike photographic detectors, the retina has two separate visual systems—*photopic vision* for relatively bright light and color vision, and *scotopic vision* for dim light. These two systems involve different pre-processing structures as well as distinct photo-receptors. The high-sensitivity photo-pigment rhodopsin, used by the rod cells, is bleached by light. Upon exposure to bright light, this high-sensitivity pigment is destroyed, leaving the lower-sensitivity cone photo-pigments to take over. Rhodopsin is then regenerated in dim light conditions. And thus the retina can modify its own sensitivity according to the intensity of light. Any given photochemical detector has a far more limited *dynamic range*. Some photoelectric detectors, however, do have a sensitivity that can be adjusted according to the lighting conditions.
5. The retina engages in *parallel processing*. The photo-stimuli from neighboring rod and cone cells are combined with each other in complex ways, in real time, to rearrange and compress the data (in ways that are still only partially understood) so as to maximize the efficiency of many visual tasks, such as detection of edges and the direct perception of contrast.
6. The structure of the retina enables it to process direct and peripheral vision in different ways. A spot of light focused on the retina is processed differently by specialized *ganglion cells*, regarding whether it is focused to the center or periphery of the visual field (see, for example Kolb 2003, figure 1.10).
7. The chemical and neurological processes in the retina constitute a *cycle*. The photoisomerizaton of retinal by photons leads to a chain of chemical and neurological responses that circle back upon themselves to their starting point, allowing for the detection and interpretation of new photons. This cycle takes longer for dim light than for bright light (Kalloniatis and Luu accessed 2018), allowing for better detection of dim light.

I list some of these weird ways in which the retina/brain work (insomuch as they are understood, and I understand them) because I find them fascinating. But also, these facts have implications for photography as an art. Seeing is mostly in the brain. And this means that, as practitioners of two-dimensional visual art, perhaps our task is not really to make a literal visual mapping of some part of *The World* onto a flat surface. Perhaps instead we are simply trying to somehow cause interesting things to happen in the brains of people looking at our pictures. And so, how best to do that?

1.6 Exposure, density, and the characteristic curve

I use the term *density* to mean the measurable physical effect on the detector, caused by the *exposure* to light. Exposure can be related to the total light energy transferred to the detector per surface area. Photographers have their own way of describing this (see volume 2 of *The Physics and Art of Photography*), but it can be expressed in physical units of joules per square meter (J m^{-2}), or the number of photons of light per square meter that strike the detector. Since the illuminance (brightness) of the light falling on the detector is a flow or *flux* of this energy, it is expressed not in J m^{-2}, but rather in J m^{-2} *per second.* And so we have the following basic relation:

$$\text{exposure} = \text{illumance} \times \text{time}. \tag{1.6}$$

For a photoelectric detector, exposure results in an electrical signal of some kind, and so greater density means a larger signal. For a photochemical detector, density represents the visible change in the detector. It becomes darker (for a negative process) or lighter (for a positive process), and the degree to which this happens is the density.

My usage of the word density in this general sense is somewhat unsatisfactory because the word is usually only applied to photochemical detectors; the word *signal* is more commonly used in regard to photoelectric detectors. And so density, in the case of photochemical detectors, usually has a more technical meaning, and is defined according to a particular mathematical relation. Furthermore, most photochemical detectors are developed out, and so most of the density is a direct result not of the action of light during the exposure, but rather the chemical reactions during development. And so, for a typical photochemical detector, the relation between density and exposure is complex and indirect.

A graph that shows, for a given detector, the quantitative relationship between density and exposure is called the *characteristic curve* for the detector. It tells, for any given exposure, what density results. But before we can plot density and exposure together on a graph, we must agree upon their precise mathematical definitions.

Figure 1.4 shows a scan of a particular photochemical detector; black-and-white enlarging paper used in the method of ephemeral process photography, which we will consider in more detail in chapter 2, section 2.5. This is a negative process, so the paper turns darker with greater exposure. I did the experiment twice, each time

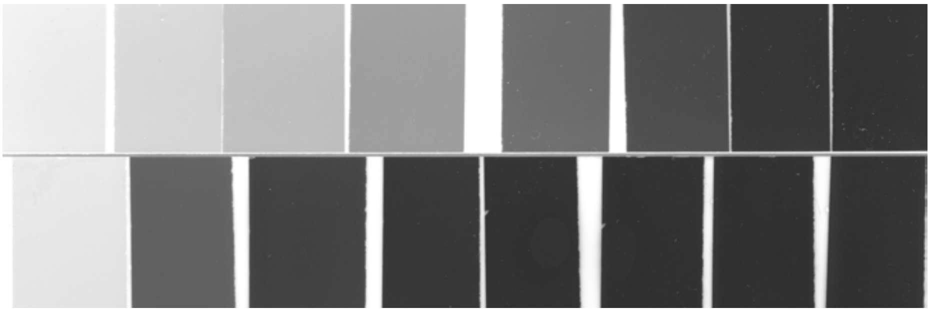


Figure 1.4. The same overall range of exposures applied to silver gelatin enlarging paper, but with two different choices for the exposures steps in between. Although the top set of exposures appears to be in roughly equal steps, it is actually the bottom set for which equally-stepped exposures were applied, with each strip receiving 9.07 s more exposure than the previous. For the top strip, each exposure was a *factor* of two greater than the previous.

giving a range of different exposures, to see the range of densities that result. The illuminance was the same for both, and so I varied the exposure by varying the time, according to equation (1.6). For both of the strips, the lightest portion was exposed for 1/2 s, while the darkest portion was exposed for 64 s.

Even though both examples in figure 1.4 show the same total range in exposure—and the same total range in density—I have broken up this range in two different ways. The example on the top appears to be in roughly equal steps, as if I had added the same exposure each time from one to the next. For the bottom strip on the other hand, it appears as though I made the biggest step in exposure between the first and second, then added progressively *less* exposure each time.

And so what were the actual exposures I used? The answer may be surprising; it is the opposite of how it looks. I added exposure in equal steps for the *bottom* example, not the top. It was made with this list of exposures (in seconds): 0.5, 9.6, 18.6, 27.7, 36.8, 456, 54.9, 64. That is, each exposure was 9.07 s longer than the previous. For the top example, on the other hand, in order to produce a set of densities that *appear* to be in equal steps, I used the following list of exposures (in seconds): 0.5, 1, 2, 4, 8, 16, 32, 64. And so a set of densities that appears to be roughly in equal steps was produced by a set of exposures each of which is *twice as great* as the previous.

The basic result shown in figure 1.4 is typical of photochemical detectors. In order to achieve what appears as the next added step in density, it is necessary to *multiply* the previous exposure by some factor. For this reason, photographers work in terms of what they call *exposure steps*, that are actually successive *factors* of two. The name is somewhat misleading; imagine walking up a staircase for which the distance doubled with each step.

A sequence of numbers such as 1, 2, 4, 8, 16, 32, …, can be written instead as follows: 2^0, 2^1, 2^2, 2^3, 2^4, 2^5, …. This sequence follows the function $y = 2^x$. This is known as an *exponential function*, and its inverse is called a *logarithmic function*. Photographers base their exposure steps on an exponential function with a *base* of 2, but other bases are also commonly used. So, for example, we can set up an exponential function as follows:

$$y = 10^x. \tag{1.7}$$

And so values for x of 1, 2, 3, 4, … simply represent y values of 10, 100, 1000, 10 000, …, or the power of 10. The inverse of this particular exponential function is called a *common logarithm*, and it is defined such that:

$$\log(10^x) = x. \tag{1.8}$$

And so $\log 10 = 1$, $\log 100 = 2$, $\log 1000 = 3$, …. A graph of the common log function can be seen in figure 1.5.

The common logarithm is also called a *base 10 logarithm*, since it is the inverse of the base 10 exponential function. Logarithms have interesting and useful mathematical properties, the most important of which are the following:

$$\log(ab) = \log a + \log b. \tag{1.9}$$

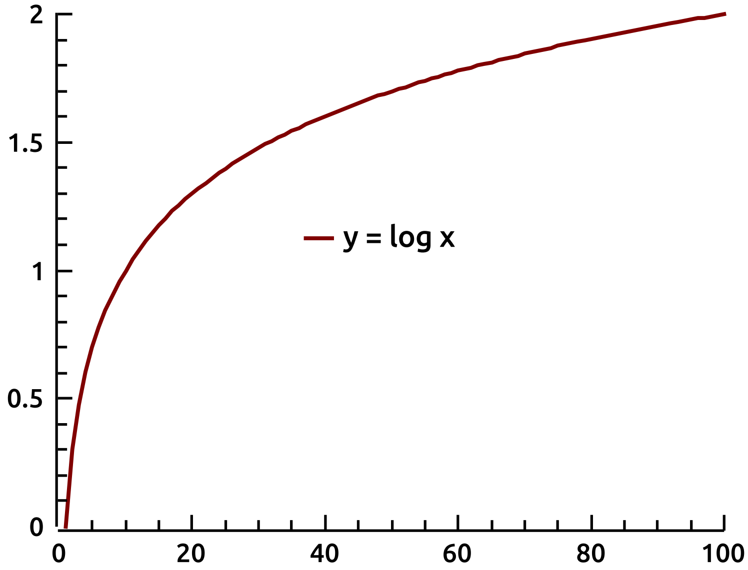


Figure 1.5. A graph of the common logarithm (log x) is defined in terms of powers of 10, and so depicts 10 as 1, and 100 as 2.

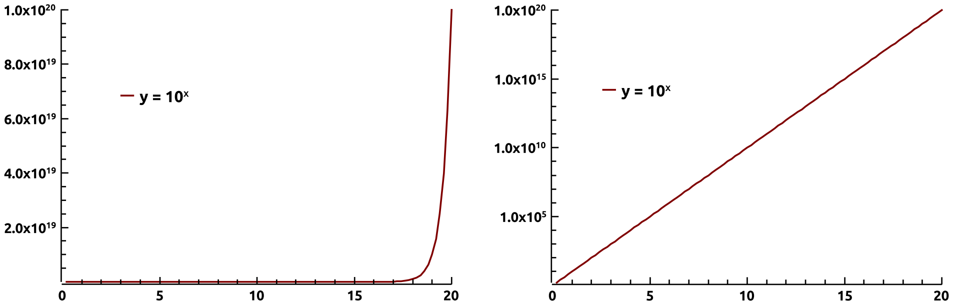


Figure 1.6. The exponential function $y = 10^x$ plotted on an ordinary linear scale (left), and a logarithmic scale (right). The logarithmic scale straightens the exponential function, and makes a large range of values more manageable.

$$\log a^b = b \log a \tag{1.10}$$

Because of equation (1.10), logarithms 'undo' exponents, turning them into simple factors. And equation (1.9) shows that logarithms turn multiplication into addition. For this reason, published tables of logarithms were important in pre-computer times, since addition (and subtraction) is easier to perform by hand than multiplication (or division). This led to the old joke that certain types of snakes only reproduce when placed on a wooden table[4].

The left side of figure 1.6 shows the exponential function $y = 10^x$ plotted versus x, while the right-hand graph shows that same function plotted with a *logarithmic scale*

---

[4] Even adders can multiply on a log table.

for the y-axis. On a logarithmic scale, each tick mark of the graph represents not an amount, but rather a *factor*. For the example shown here, each tick is 10 times greater than the tick below it. Since $\log(10^x) = x$, this is simply a straight line. And so use of a logarithmic scale on the vertical axis of a graph has the effect of turning an exponential function into a straight line. Notice how the logarithmic scale compresses an enormous range in values to a much smaller scale. Most of the direct graph of the exponential function, as shown on the left, is almost useless; at the scale of the graph, 90% of it is either indistinguishable from zero or nearly vertical.

We can also define logarithms with bases other than ten—for example a *base 2 logarithm*, written $\log_2 x$, that is the inverse of the base 2 exponential function that photographers have chosen for their exposure steps. It would then represent not powers of 10, but rather powers of 2. But it is common to use the base 10 'common logarithm' even so. The base 10 logarithm has the convenience that it is directly related to our base 10 numbering system, and there is a very simple relation between logarithms of different bases. Equations (1.9) and (1.10) are written for common logarithms, but they hold true no matter what the base, and so we have:

$$y = 2^x \tag{1.11}$$

$$\log_2 y = x \tag{1.12}$$

$$\log_{10} y = \log_{10} 2^x \tag{1.13}$$

$$= x \log_{10} 2 \tag{1.14}$$

$$\log_{10} y = \log_2 y(\log_{10} 2) \tag{1.15}$$

$$\log_{10} y \approx 0.301 \log_2 y \tag{1.16}$$

where I have made the base 10 of the common logarithm explicit for clarity. Thus, logarithms of different bases are simply proportional to each other, and to convert from base 2 to base 10, a simple factor of roughly 0.3 is all that is required. And so we have the important result that on a common logarithmic scale, the photographer's factor of two steps appear as even increments of approximately 0.3.

Our goal is to make an appropriate graph that relates density to exposure—the characteristic curve of our detector. And so we see from figure 1.4 that—for the photochemical detector shown in the example at least—the appropriate horizontal scale for our graph is not exposure, E, but rather $\log E$. But what is the appropriate measure of density? To answer this question we must consider not only the physical nature of the detector in question, but also the way in which the human eye/brain combination perceives steps in lightness and darkness.

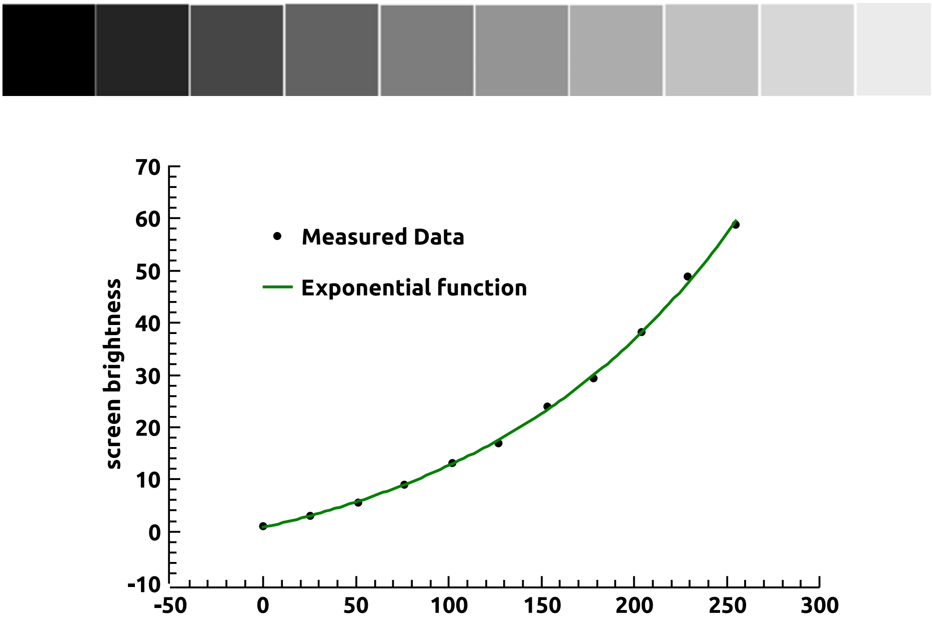


Figure 1.7. **Top:** A digital grayscale image produced so as to appear in roughly equal steps of brightness on a computer monitor. **Bottom:** A graph of the measured brightnesses of the grayscale sections, plotted on an additive scale. An exponential, rather than linear, function fits the data (green curve). This illustrates that we see levels of brightness in multiplicative, rather than additive, steps.

Figure 1.7 shows another set of gray levels that appear in roughly equal steps in brightness. This is an image created digitally, rather than from a negative-process photochemical detector, and so I show the levels arranged from darkest to lightest. When I display this image on my computer monitor, it appears to me as roughly equal steps in brightness. In a darkened room, I used a light meter to measure the brightness of the monitor for each of the gray levels shown in figure 1.7, and the result can be seen graphed at the bottom.

The graph of brightness levels shows an increasing slope, like some type of exponential function—and indeed, the data is well fit by one (the green curve). This crude experiment illustrates the well-known fact that our vision is better described by a logarithmic scale; what appear to us as equal steps in brightness are more accurately described as equal *factors* (Adler *et al* 2014). And so photographic density might also be better defined according to a logarithm of measured levels of brightness. This was first recognized in the early days of silver gelatin photography. In the late 1800s Ferdinand Hunter and Vero Charles Driffield made the first quantitative measurements of the density of exposed photographic emulsions, and we still use their definition today (Ferguson 1920).

One way to quantitatively measure the ‘darkness’ of a piece of exposed and developed photographic film is to shine light through it, and measure by what factor the light has dimmed while passing through the film. A device that does this is called a *densitometer*; it directly measures the *transmittance*, T, of the film.

The transmittance is simply the fraction of light that is transmitted, the rest being absorbed by the film. For a negative-process film, greater exposure means *less* transmittance, and so it makes sense to define density as $1/T$, rather than T. However, this simple measure would have little relation to how we perceive steps in brightness. And so instead, the photographers' definition of density, D, also incorporates the logarithm:

$$D = \log(1/T). \tag{1.17}$$

And so it appears that it is appropriate for both density and exposure to be described in terms of logarithms of the physically measured quantities. For this reason, the *characteristic curve* of a photochemical detector has, since Hunter and Driffield, been portrayed as a graph of D versus $\log E$. Since density is also defined in terms of a logarithm, the characteristic curve is a *double-log* or *log–log* plot, with a logarithmic scale on *both* axes.

But there is another important consideration—the physical process by which exposure causes the light detector to undergo its physical change. For a negative-process photographic film, the exposure to light ultimately causes the transmittance of the film to decrease. And so what *is* the mathematical relation between the exposure on the one hand and the inverse of the film transmittance on the other? And how will that relation appear on the log–log plot of our characteristic curve?

We have already seen that an exponential function appears as a straight line on a graph that has a logarithmic y-axis. But an exponential function is a strongly *curved* line on a *double-log* plot. And so, is there a mathematical function that appears straight on a log–log plot? And does the mathematical relation between $1/T$ and E follow such a relation for a real photochemical detector? The answer is yes to the first question, and sort-of yes to the second.

A double-log plot straightens out not an exponential function, but rather a *power law*. A power law is defined by:

$$y = ax^{\gamma} \tag{1.18}$$

where a and γ (the Greek letter gamma) are constants. Notice, as with an exponential function, that there is an exponent involved—but for a power law it is the *constant* that is the exponent, not the variable x. We can easily see that this function will appear as a straight line on a log–log plot by taking the logarithm of both sides:

$$\log y = \log(ax^{\gamma}) \tag{1.19}$$

$$\log y = \log a + \log x^{\gamma} \tag{1.20}$$

$$\log y = (\log a) + \gamma \log x \tag{1.21}$$

$$y\text{–axis} = \text{intercept} + \text{slope} \times x\text{–axis}. \tag{1.22}$$

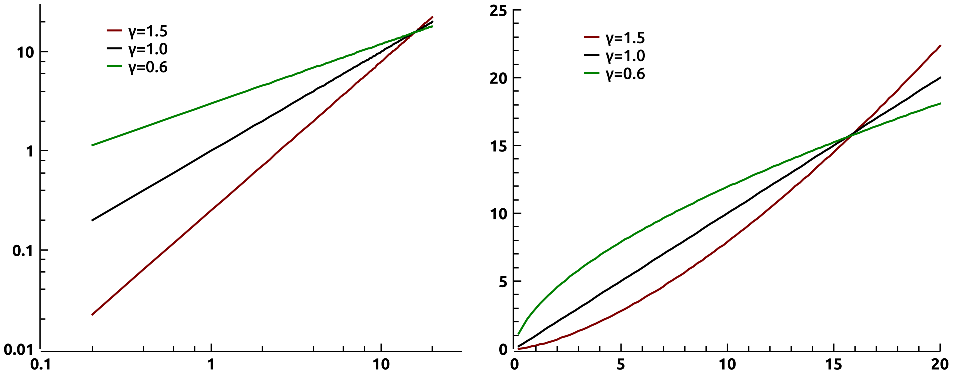


Figure 1.8. **Left:** Three power-law curves plotted on a double-log graph. They all appear as straight lines; the greater slope corresponds to the higher power, gamma (γ). **Right:** The same mathematical functions plotted on an ordinary linear scale for x and y. For $\gamma > 1$ the curve is concave upward, while the opposite is true for $\gamma < 1$. A straight-line, linear curve corresponds to $\gamma = 1$.

And so if we use a log–log plot to graph a power law, with $\log y$ plotted versus $\log x$, then we get a straight line with a slope equal to the exponent, γ, and a y-intercept equal to the logarithm of the scale factor a. The left side of figure 1.8 shows the log–log plots of three different power laws, each with a different γ. They all appear as straight lines on this double-log graph, with the steeper line belonging to the larger value of γ. The right side of figure 1.8 shows these same three functions plotted without the logarithmic axes—and here we see a qualitative difference. A power law with a $\gamma > 1$ curves upward, while $\gamma < 1$ produces a downward-curving graph. A value of $\gamma = 1$ leads to a simple straight line.

It turns out that the relation between E and $1/T$ for developed-out photographic film *is* often related approximately by a power law, but with two important exceptions. First, equation (1.18) has the property that $y = 0$ when $x = 0$. But this is not the case for the relation between film transmittance and exposure; there is a *minimum density* that occurs even with zero exposure. This is called the *fog density*, or D_{min}. And so a more appropriate mathematical relation would be:

$$y = ax^{\gamma} + b \tag{1.23}$$

where b is the value of y for $x = 0$. Taking the logarithm of both sides of equation (1.23) cannot be simplified to a straight line, like equation (1.18). But it is clear that for large enough values of x, the addition of the constant b would make an insignificant difference. And so we expect equation (1.23) to appear very much like a straight line on a log–log plot for large values of x, but to deviate significantly from that at small values.

Figure 1.9 shows log–log plots of power laws with $\gamma = 2$, both with and without an additive constant. And so we would expect our characteristic curve to appear more like the green curve, with a straight-line part, but also a flat *toe*, corresponding to the expected D_{min} of our photographic film.

Figure 1.10 shows a D versus $\log E$ curve for some real densitometer data from Stroebel *et al* (2000, p 96). Notice that the data points are separated on the

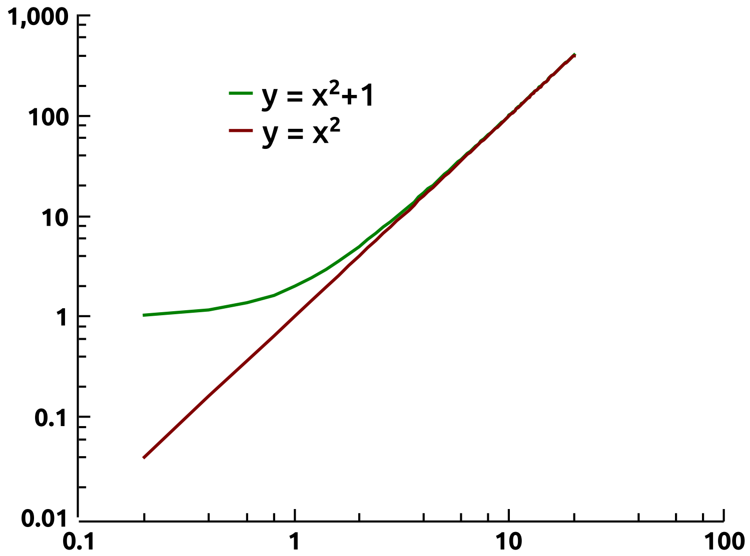


Figure 1.9. A simple power law (red curve) appears as a straight line on a log–log plot, unless it also includes an additive constant (green curve). The characteristic curve of photographic film has a 'toe,' like the green curve, because the film absorbs some light even with zero exposure.

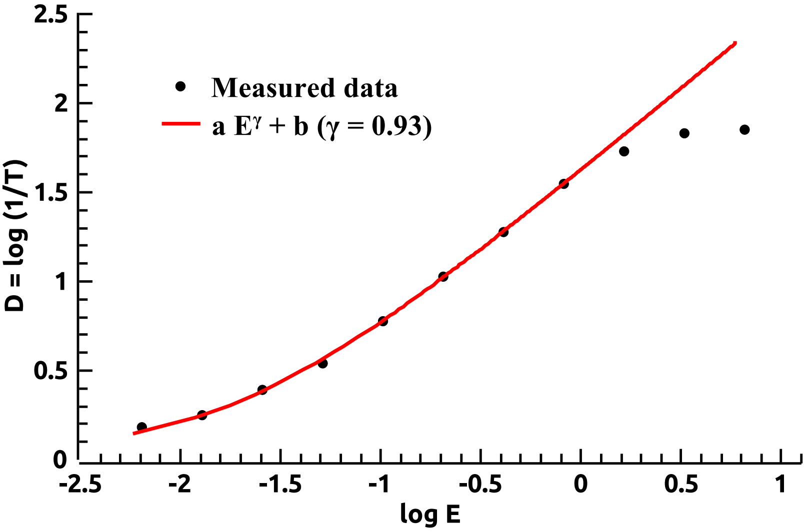


Figure 1.10. The characteristic curve for real densitometer data from Stroebel *et al* (2000, p 96). The red curve is a power law, with $\gamma = 0.93$, fit to the toe and straight-line part of the data. There is also a shoulder because real photographic film has a maximum density that is not exceeded with increased exposure.

horizontal axis by $\log E = 0.3$, corresponding to a difference of 1.0 if we had used a base 2 logarithm instead. Thus, the different exposures were made in the photographer's traditional steps, corresponding to factors of two.

Plotted on top of the measured data in figure 1.10 is the mathematical function of equation (1.23), with values of a, γ, and b chosen so as to best fit the data. It clearly has a *toe* and a *straight-line part*, as we would expect from a power law with a D_{min}. But there is also a *shoulder*—the density increases more slowly than the power law at large exposures, appearing to level off to a maximum density.

It should be unsurprising that the characteristic curve of a real photographic film has a maximum-density shoulder, or D_{max}. There is, after all, only so much of the photochemical reactant to be darkened by light and the chemical development process. Once the film is as dark as it can get, one would expect greater exposure to have no measurable effect.

The true mathematical shape of the characteristic curve for photographic film cannot be summarized by any simple formula. It can only be modeled accurately by a complex statistical analysis that includes all aspects of the process, from exposure to light, through chemical development and fixing, to the final measurement by the densitometer. But the toe and straight-line portions are often well fit by a power law. Some special kinds of technical film have a D_{max} that is too high to measure with a typical densitometer, and so the measurable part of the characteristic curve shows only the toe and straight-line part, with no shoulder (see for example Stroebel *et al* 2000, p 109, figures 4-28 and 4-29). And in those cases the measured part of the characteristic curve appears very similar to a simple power law plus constant, as in the green curve of figure 1.9.

1.6.1 The characteristic curve and photoelectric detectors

Whether one is using a negative or positive process, developed out or printed out, a photochemical detector produces a visible image in itself. Expose a roll of black-and-white silver gelatin negative film in a camera, develop and fix the film, and then hold it up to the light. The image is right there, visible in the very physical object that interacted with the light. Not so for a photoelectric detector—only an electronic signal is produced by exposure to light. The visible image must be produced from that signal by other means, either a lighted screen or some kind of printing process. And this means that we can, somewhat independently of how the detector responds to light, *transform* the electronic signal into something that corresponds to the logarithmic manner in which we perceive levels of brightness. And so for a photoelectric detector, the characteristic curve—a graph that relates exposure (the cause) to density (the effect)—can be connected more directly to the physical relation between exposure and signal. Thus, for a photoelectric detector there is little motivation for the nearly 150-year-old $\log(1/T)$ versus $\log E$ graph of Hunter and Driffield.

For these reasons, photoelectric detectors are more commonly described by the direct relationship between exposure and the resulting electronic signal. For the most important photoelectric detectors we consider in this book, that response is approximately *linear*, corresponding to the following graph:

$$y = ax + b. \tag{1.24}$$

Note that equation (1.24) is the same as equation (1.18), for the special case of $\gamma = 1$.

Photographers often use the term 'linear' in a different sense—to signify the straight-line part of the D versus $\log E$ characteristic curve of photochemical film emulsions. This usage is both wrong and right. First, why is it wrong? We have already seen why; the response is, at its simplest, not linear but rather a power law. Furthermore, the toe of the curve is part of the same power law as the straight-line part, yet photographers call only the straight-line part 'linear.' If γ is not equal to one, then *no* part of the characteristic curve is linear. But ironically, if $\gamma = 1$, then not only the straight-line part, but also the toe of the characteristic curve is linear. And so, from a perspective of the physical relation between exposure and the actual darkening of the film, photographers' common use of the term 'linear' is incorrect.

But as is the case for many commonly-used 'incorrect' terms, there is a method to the madness. If we are using a photochemical detector not for scientific measurements, but rather for making pictures, then we have already seen that it is density—as defined by photographers with a logarithm—that is the point.

References

Adler M, Mayo A and Alon U 2014 Logarithmic and power law input-output relations in sensory systems with fold-change detection *PLOS Comput. Biol.* **10** 1–14

Chomsky N 2000 *New Horizons in the Study of Language and Mind* (Cambridge: Cambridge University Press) https://doi.org/10.1017/CBO9780511811937

Conway B R, Chatterjee S, Field G D, Horwitz G D, Johnson E N, Koida K and Mancuso K 2010 Advances in color science: From retina to behavior *J. Neurosci.* **30** 14955–63

Evans D S, Deeming T J, Evans B H and Goldfarb S (ed) 1969 *Herschel at the Cape: Diaries and Correspondence of Sir John Herschel* (Austin, TX: University of Texas Press)

Ferguson W B (ed) 1920 *The Photographic Researches of Ferdinand Hurter and Vero C. Driffield Bath* (Bath: The Royal Photographic Society of Great Britain)

Holmes R 2016 Seeing single photons *Phys. World* **29** 28–31

Kalloniatis M and Luu C Temporal resolution *Webvision: The Organizaton of the Retina and Visual System* ed H Kolb, R Nelson, E Fernandez and B Jones (Salt Lake City, UT: University of Utah Health Sciences Center) [https://webvision.med.utah.edu/book/part-viii-gabac-receptors/temporal-resolution/] (accessed 2018)

Kolb H 2003 How the retina works *Am. Sci.* **91** 28–35

Kolb H, Nelson R, Fernandez E and Jones B (ed) *Webvision: The Organization of the Retina and Visual System* [https://webvision.med.utah.edu/] (accessed 2018)

Schmidt B P, Neitz M and Neitz J 2014 Neurobiological hypothesis of color appearance and hue perception *J. Opt. Soc. Am.* A **31** 195–207

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

Chapter 2

Silver gelatin photochemical detectors

2.1 Black-and-white silver gelatin emulsions

The most common type of black-and-white photochemical detector is the *silver gelatin emulsion* (also called gelatin silver emulsion). This is the basic photochemical process used both for ordinary black-and-white film and for making black-and-white prints in the darkroom. It is also the foundation for the *chromogenic* color film and printing process (see section 2.2).

Microscopic crystals of a *silver halide salt* are suspended in a thin, transparent layer of gelatin, which is coated onto the surface of the plastic film or print paper. The word 'halide' refers to a column (group 17, second from the right) in the periodic table of elements, called the *halogens*. As with other columns in the periodic table, halogens are grouped together because they would require the same number of electrons to complete their outermost electron shell. For the halogens, the outermost shell is only one electron shy of being filled, and so they easily accept electrons from other chemical species. Since they all have this in common, halogens all have similar chemical properties, and this is why they are grouped together as a column of the periodic table. Examples of halogens include fluorine (F), chlorine (Cl), bromine (Br), and iodine (I).

Halogens form halide salts with elements that easily donate electrons, such as those that have only one electron in their outermost shell. Silver (Ag) is an example of such an atom. Since atoms like to have their outermost shell filled, everyone can be happy. Thus, a silver atom gives one electron to a halogen, and now both have their outermost shell filled (the silver had one extra while the halogen had one too few). In addition, the silver now has a positive electric charge (since it lost an electron), while the halogen now has a negative charge (since it received an extra electron), and so the two stick together due to the electrical attraction of unlike charges.

Examples of such silver halide salts are AgCl, AgBr, AgI, and AgF. Three of these in particular are of interest for photography, as they share similar physical and light-sensitive properties: silver chloride (AgCl), silver bromide (AgBr) and silver

doi:10.1088/2053-2571/aaf0aech2

iodide (AgI). Like many ionic compounds, they may form an ordered and ongoing three-dimensional alternating pattern of silver and halogen ions—a crystal.

It is easy to make a silver halide by mixing the compound silver nitrate ($AgNO_3$) in an aqueous solution with a halide salt such as potassium or sodium chloride (KCl or NaCl), potassium or sodium bromide (KBr or NaBr) or potassium or sodium iodide (KI or NaI). Both silver nitrate and the halide salt are soluble. In solution, a halide salt such as KCl dissociates into the ions K^+ and Cl^-. Similarly, silver nitrate dissociates in solution into Ag^+ and NO_3^-. But when the two compounds are dissolved together into the same solution, Ag^+ has a greater affinity for Cl^- than the NO_3^- from which it came, and K^+ has a greater affinity for NO_3^-.

This would not matter so much if all of the ions stayed in the solution. But the combination of the silver and halide ions—AgCl in this example—is markedly *in*soluble in water. And so solid crystals of AgCl form and sink to the bottom of the container, leaving a solution of, in this example, potassium nitrate. The overall reaction works like this (using, as an example, potassium chloride as the halide salt):

$$AgNO_3 + KCl \longrightarrow KNO_3 + AgCl. \tag{2.1}$$

Silver chloride is sensitive to light, by mechanisms we will discuss later. But producing a fine sludge of AgCl in this simple way is not very useful for photography. And so there is a trick—the reaction must be carried out in a thick solution of gelatin, so that the silver halide crystals remain suspended and separated from each other as they form. The gelatin is mostly transparent, and it allows the crystals to grow separately to a consistent size. This thick liquid gelatin with silver halide crystals suspended in it can then be applied as a thin coating onto film or paper and allowed to dry.

And so a silver gelatin emulsion is a transparent gelatin matrix of tiny crystals of some combination of AgCl, AgBr, and AgI (other, more complex silver salts are sometimes used as well). Silver halide crystals are sensitive to light, by way of the following process (Stroebel *et al*, 2000, pp 198–200):

1. Photons are absorbed at an imperfection in the silver halide crystal. The energy of the photons is transferred to negatively-charged electrons in the crystal. This higher energy level makes the crystal electrically conductive.
2. Electrons migrate to the imperfection site, and neutralize (reduce) Ag^+ to make simple elemental silver, Ag. It is this ordinary silver that makes the dark part of the image (the silver halide crystals are mostly white).
3. When used for developing out, this image of silver is far too faint to be visible; there may be as few as only 4 reduced silver atoms in a given crystal. The image is then amplified by a chemical *developer*, which further reduces Ag^+ to Ag by chemical means—but only in the presence of existing silver. And so a *developed-out* image forms that is made of millions of times more Ag than the original *latent image* originally reduced by the photons. The step of chemical development turns the invisible latent image into a visible ‘blatant’ image.
4. Ordinarily, the remaining silver halide would then be chemically removed (using a *fixer*) in order to render the image no longer sensitive to light.

5. If the image is *printed out*, rather than developed out, then millions of times more photons must be absorbed in order to make a visible image, since there is, by definition, no chemical development to amplify the image. This means that printing-out processes inevitably require far more light to make an image.

In a photographic emulsion these crystals of silver halide are microscopically small. But each is still comprised of an enormous number of individual atoms. When photons strike these atoms, there is a small possibility for any individual photon that it will be absorbed, and in the process a few atoms of metallic silver are freed from their halide companions. These silver atoms would appear dark, but in an ordinary exposure there are far too few of them to see.

The more photons that arrive at a given silver halide crystal, the more likely it is that at least several silver ions have been reduced to metallic silver, producing a tiny metallic silver speck in that crystal. And so in regions of the emulsion that received more light, a greater percentage of silver halide crystals contain metallic silver atoms, and those crystals that received more light are likely to contain a greater number of such atoms. And so we have a *latent image*, invisible at this point. This latent image can persist in the emulsion for years, especially if it is kept at a sufficiently low temperature.

The process of making the latent image visible in an exposed silver gelatin emulsion is called *development*, and a (usually liquid) chemical that does this is called a *developer*. The developer accomplishes this feat by chemically reducing the silver halide to metallic silver, and thus turning the entire silver halide crystal dark. *But the developer will do so only wherever some free silver atoms already exist*. And so only those silver halide crystals that contain the latent silver atoms from the exposure to light are reduced to dark silver metal. Thus, only areas that were exposed to light turn dark, and a visible *negative* image is formed.

The development process is more-or-less rather than either-or. Put the emulsion in the developer solution and, as time passes, silver halide crystals with more latent silver atoms will develop first, while crystals with fewer latent silver atoms will take longer to develop. And thus more and more of an exposed part of the emulsion will be reduced to silver, making that part darker and darker. But those parts that received more light during the exposure become darker faster during development. Thus, an *amplification effect* occurs.

The trick is to develop the exposed emulsion just enough that the dark areas are as dark as needed, while leaving the unexposed areas still white (or clear if it is a transparent piece of film). We will revisit this idea a little more in chapter 8, section 8.2, in the context of controlling the *contrast* of a photograph. But since the emulsion must be developed only for a set length of time, the development must then be *stopped* by placing it in a *stop bath* (usually a weak acid solution) that chemically neutralizes the developer.

The final step is to remove the remaining still-sensitive-to-light silver halide left in the emulsion, otherwise the image would not be permanent; it would gradually turn dark all over if exposed to light. This process is called *fixing* the image, and a chemical that accomplishes this is called a *fixer*; the fixer dissolves the silver halide crystals, allowing them to be washed away. It is sometimes under-appreciated just how important to the history of photography was the discovery of a practical method for fixing silver-based photographic emulsions. The astronomer and photography pioneer John Herschel is usually given credit for first solving this problem, with his discovery that sodium thiosulfate (sometimes called 'hypo') can be used as a fixer for silver halide photography.

Silver gelatin emulsions are made for different purposes; the two most common uses are image capture in a camera and producing a photographic print to hang on a wall. Emulsions for image capture are usually designed with higher sensitivity and a greater range of wavelength response. Black-and-white photographic film, used for in-camera image capture, usually contains a combination of AgCl, AgBr, and AgI, along with organic dyes that extend the sensitivity to longer wavelengths. Black-and-white photographic *enlarging paper* on the other hand, used to make prints in the darkroom from film negatives, is much less sensitive, and typically has a wavelength response that extends only to blue-green wavelengths[1]. These emulsions use only combinations of AgCl and AgBr, with no AgI.

2.2 Chromogenic color emulsions

The typical color film or printing process is called *chromogenic*, or *C-type*. As with black-and-white photography, the chromogenic process uses silver halides as the light-sensitive material. But there are typically three layers of silver halide emulsion, each sensitive, respectively, to red, blue, and green light, the additive primary colors.

When processed, the silver is developed (which makes it black), but it is then removed in a two-step process. First, a colored dye is chemically attached to the developed silver in each layer. This is called *dye coupling*. Each additive primary layer is coupled to its respective subtractive complement (the subtractive primary colors). And so a cyan-colored dye is coupled to the red-sensitive layer, a magenta dye is coupled to the green-sensitive layer, and a yellow dye is coupled to the blue-sensitive layer.

Once the dyes are coupled to their respective silver layers, the black silver is then bleached away, and the remaining undeveloped silver halide is removed with a fixer (these two steps are often done at the same time, combining the bleach and the fixer into a bleach-fix, or *blix*). What remains are the dyes, and they are the negative of the colors that were exposed by the three silver halide layers.

2.3 Reversal-processed silver gelatin emulsion

The ordinary silver gelatin process is a negative process. It is nonetheless straightforward to get a positive image directly from a silver gelatin emulsion through what

---

[1] This has the advantage that one can handle them under an orange 'safelight,' so the photographer can see to work without exposing the light-sensitive paper.

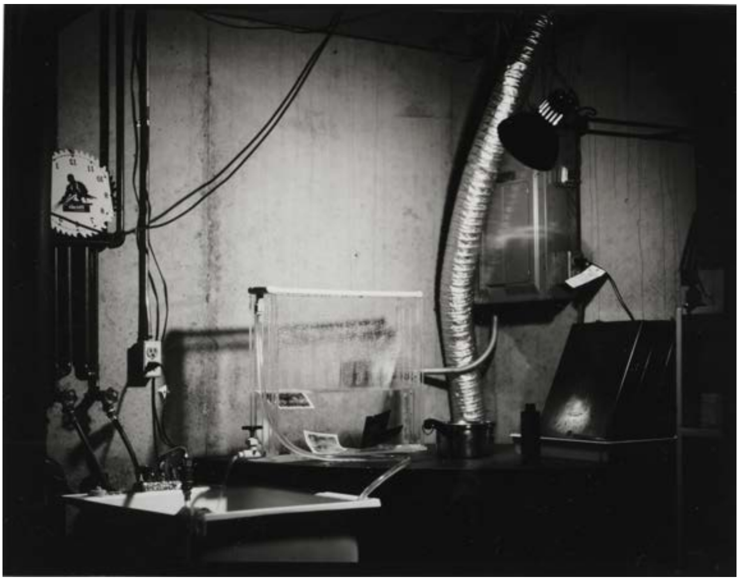


Figure 2.1. A direct and straightforward scan of a 4 × 5 inch positive image on reversal-processed enlarging paper.

is known as *reversal processing*. The emulsion is exposed and developed, which brings out a negative image in the ordinary way. But instead of fixing the image—removing the remaining undeveloped silver halide—the *developed silver* is bleached away. This makes the negative image disappear, and what remains is undeveloped silver halide.

But this remaining silver halide is not uniform. Where the original exposure was large, most of the silver halide was developed to silver and then bleached away, leaving relatively little silver halide remaining. The opposite is true for areas that received very little exposure. For these regions most of the original silver halide remains, since very little was exposed, developed and bleached.

Thus, one can then simply expose the emulsion to uniform light, exposing all of the remaining silver halide. When this is developed, a positive image appears. Figure 2.1 shows an example of a 4 × 5 inch print made with ordinary silver gelatin printing paper exposed in a large format camera and then reversal processed. As is the case with a direct positive, the picture on the wall is the same object that recorded the light in the camera.

Reversal processing can be used to make 35 mm color transparencies for direct projection with a slide projector. The technique can also be used in the darkroom to make an enlarged *negative* transparency from a film negative. Figure 2.2 shows a 24 × 18 inch print made with the Van Dyke printing process (see chapter 3, section 3.3). The print could only be made by exposure to sunlight, in contact with a negative the same size. But the original 35 mm film negative was only $1 \times 1\frac{1}{2}$ inch, and so an enlarged negative was needed.

Figure 2.2. *Goat*. John Beaver, 2004. Van Dyke print on watercolor paper, 24 × 18 inch.

2.4 Lumen process

The most common use of silver-based emulsions is to form a latent image with exposure to light, which is then made visible by a development process, as described in section 2.1. But silver halide emulsions will eventually turn visibly dark on their own, with no development, if exposed to large amounts of light. This is called *printing out*, and the earliest photographic processes were of this type, most notably the early experiments of William Henry Fox Talbot.

Since a much greater exposure to light is needed to reduce enough of the silver halide directly to a visible amount of silver without the amplification effect of the chemical developer, *printing-out papers are of very low sensitivity*. Because of this, they have traditionally been used mostly for the printing stage, where a large negative could be placed in direct contact with the paper and exposed in a bright light box or to direct sunlight. Alternatively, objects can be used in order to make a direct negative *photogram*—a photographic shadow print.

The so-called *lumen process* is a modern take on printing out with silver. But instead of using papers specifically designed for printing out, ordinary black-and-white *enlarging paper* is used. This paper is designed to be developed out, not printed out, and as such many varieties are quite slow indeed when used for this unintended purpose. Direct sunlight is used for the exposure, and photograms are more common than positive prints from negatives. Exposure times can be as much as many hours or even days long, depending on the particular variety of paper used. In most cases the lumen print is chemically fixed in the ordinary way, to make a permanent print.

When enlarging paper is printed out, subtle tones and hues often appear that are not present when the paper is developed out. Furthermore, the printing-out speed depends critically on the level of humidity and moisture, and so botanical photograms are especially popular; the added sensitivity from the moisture of the leaves

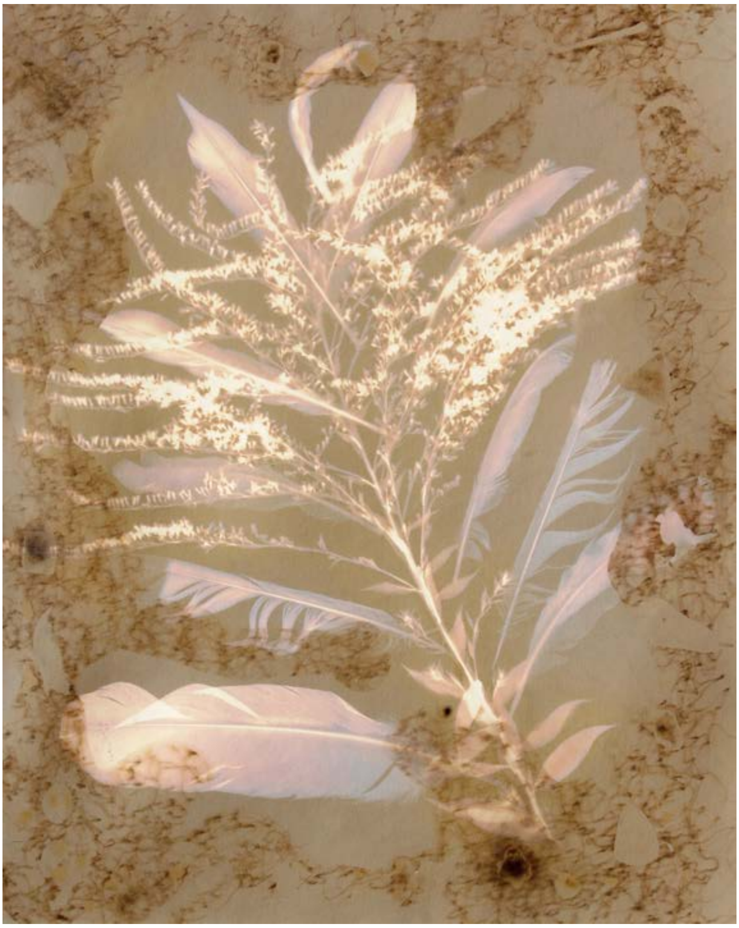


Figure 2.3. John Beaver, 2018. A lumen print made by placing feathers and leaves directly onto a piece of enlarging paper, and exposing to sunlight.

can produce dark outlines and a halo effect. The final step of fixing often greatly reduces the darkness (requiring an initial overexposure), lowers the contrast and mutes the colors, giving lumen prints a very particular and appealing look. Enlarging paper that is well past its expiration date and no longer suitable for developing out may still work perfectly when printed out. Of the many hundreds of varieties of enlarging paper produced over many decades, even those that are nearly identical when developed out may produce strikingly different results when printed out. As such, lumen photographers often use papers that are decades old.

Figure 2.3 shows one of my own lumen prints, made in 2018 with an enlarging paper that expired in 1975. The flowers and feathers were exposed in sunlight, but the meandering dark squiggles were added afterwards, using techniques described in section 2.5. Better examples of the use of lumen process, by photographers Diane Fenster and Caitlin Noll, can be seen in chapter 11, section 11.2.

2.5 Ephemeral process (EP)

What I call ephemeral process (EP) photography uses printed-out silver gelatin enlarging paper to make a picture (Beaver 2017; Fudge and Beaver 2017). In this respect it is much like the lumen process (see section 2.4), but there are many differences. In particular, lumen prints are usually fixed, while EP usually is not, and this is the source of the *ephemeral* in EP. But also, EP uses what I call an *accelerator*, a liquid that is brushed on just prior to exposure, that allows the printing-out process to occur up to many hundreds of times faster[2]. And so only seconds of exposure to direct sunlight may be required for an EP photogram. Because of this much higher sensitivity, EP photograms can be made easily with artificial light, and it is practical for daylight photography as the light detector in a camera (although exposures of many minutes are required).

After exposure, the paper is washed and dried, and it returns to something like its initial very low sensitivity. It can then be scanned to capture the image. The scanning process exposes the paper further, and so damages the image. But if done carefully, the damage can be insignificant. Since the paper is exposed mostly only where the accelerator is applied, one can control what parts of the paper are exposed and what parts are not. The ideal enlarging paper for EP photography prints out very slowly when dry, but is very fast when the accelerator is applied. This provides the greatest contrast between the accelerated and unaccelerated areas.

It is important to note that the accelerator is not a developer. A developer has its own reducing agents that change the silver halide to silver metal, and when developing a silver gelatin emulsion the developer acts in the dark, *after* the emulsion has been exposed to light. The EP accelerator does not work in this way; it has no effect at all if it is applied in the dark, after the paper has been exposed to light.

The use of EP for photograms is further described in volume 1 of *The Physics and Art of Photography*, and volume 2 describes how to use the process as an in-camera light detector. In chapter 11, section 11.3, we consider some features of the ephemeral process as they relate to art photography, and some practical details are outlined in appendix A. Here we describe the physical mechanism behind ephemeral process photography.

2.5.1 How does it work?

The description in section 2.1 of silver gelatin emulsions leaves out one important detail in the chain of events that leads to the formation of an image. When light reduces Ag^+ to ordinary Ag, the electric charge must be conserved. And so the inevitable consequence is that some halide will also lose its excess negative charge, and become a neutral halogen—Cl, Br, or I, depending on which silver salt is involved. If nothing is done with the halogen, then it will have a strong tendency to simply find those silver atoms and re-form new silver halide. This reverse reaction

---

[2] The same term is also used by photographers in a different way; see Stroebel *et al*, (2000, p 217).

happens at a lower rate, and so the overall effect is still to reduce Ag^+ to Ag, but it has the effect of greatly reducing the reaction rate.

In practice, this is unimportant for the extremely faint latent image that is meant to be developed out. There is so little halogen released (because so little silver is reduced) that the gelatin matrix surrounding the crystals can accept the halogen easily. During development millions of times more halogen is released, but the developer solution contains its own ingredients to deal with that. For a printed-out image, however, millions of times more silver is reduced, and there is little means to take-up this halogen—and this is the prime factor that slows down the whole process. Because of this, silver gelatin emulsions for printing out are ordinarily made differently than those intended for developing out.

Recall that silver nitrate is combined with a halide salt to make crystals of silver halide. If the two compounds are not mixed in a balanced proportion, then an excess of one or the other will remain. *Printing-out emulsions are made with an excess of silver nitrate, while developing-out emulsions are made with an excess of halide salt* (Eston, 1957, p 21). When printing out, this excess silver nitrate can then accept the liberated halogen (this also makes new silver halide). Developing-out emulsions are made with an excess of halide salt to *prevent* this from happening. Otherwise, the paper would eventually turn dark on its own, even if kept in the dark. And so printing-out papers often have a rather short shelf life and are usually prepared immediately prior to use (Lawless, 2005).

For printing out, the take-up of halogen by the excess silver nitrate is greatly accelerated by the presence of water. Thus, it is well-known that printing-out emulsions are 'faster' in conditions of high humidity (Lawless, 2005). And so, if one simply wets the printing-out paper it becomes more sensitive. But photographers rarely do this. In most cases of printing out, a photographer is making a contact print with a large negative—and the wet paper would damage their negative.

So how does EP photography fit into all of this? The difference is that we are printing out with papers designed for developing out. These papers are mostly designed to be poor at accepting liberated halogens, and so they print out very, very slowly. The simple addition of water greatly accelerates their printing-out speed, because the halogen can then go into the solution. The reaction (with chlorine as an example) is as follows:

$$4\,\mathrm{Cl} + 2\,\mathrm{H_2O} \longrightarrow 4\,\mathrm{HCl} + \mathrm{O_2}. \tag{2.2}$$

But in order for this to happen, one must get rid of the oxygen (O_2), or this reaction will simply undo itself (albeit at a slower rate). The O_2 can simply bubble out of the solution as a gas, but it can be disposed of far more efficiently by a chemical compound that easily reacts with O_2 to make a different, stable compound.

Sodium sulfite, Na_2SO_3, is such a compound. It reacts with oxygen to make sodium *sulfate*, Na_2SO_4, in the following way:

$$2\,\mathrm{Na_2SO_3} + \mathrm{O_2} \longrightarrow 2\,\mathrm{Na_2SO_4}. \tag{2.3}$$

Indeed, sodium sulfite is one of the ingredients of Dektol, the most commonly used developer for enlarging paper, partly because it is a preservative, but perhaps also for this reason. Ascorbic acid, better known as vitamin C, will also work for this purpose. And so a simple water solution of either sodium sulfite or ascorbic acid makes for an effective accelerator for printing out enlarging paper. The two accelerators are not exactly equivalent in their action with all enlarging papers, as sodium sulfite has other effects as well (Stroebel *et al*, 2000, p 215). More details on the preparation of EP accelerators can be found in appendix A.

Since we want a paper that, when dry, prints out very slowly, it would seem that developing-out papers with a large excess of halide salt might have the right properties. It is difficult to know for sure, because the chemical formulas of commercial enlarging papers are proprietary. But to test this theory, we have made six different silver gelatin papers—three each of AgCl and AgBr (AgI is ordinarily used only for film, and not for enlarging papers). Emulsions of each of the three silver halides were prepared in three different ways—with a balance of salt and silver, with a 100% excess of salt, and with a 100% excess of silver.

Figure 2.4 shows an exposure test of all six papers, both dry and with a sodium sulfite solution (Fudge and Beaver 2017). Almost the entire paper was covered with emulsion, but the accelerator was applied only to a small area. The accelerated areas achieved a similar density whatever the mix of silver and salt. But as expected, the papers with an excess of salt did not darken at all in the amount of time required to darken the accelerated area. The examples with a near balance of salt and silver showed slight, but noticeable darkening in the unaccelerated areas. Not shown in this test is the additional fact that unexposed examples of both papers with an excess of silver nitrate eventually fogged in the dark, over the course of a couple of days. But the papers with an excess of salt resisted fogging.

The formulas of commercial developing-out print papers are far more complex that this test, and so it is unclear how to generalize these simple results. But tests of different commercial papers often show markedly different tones, both wet and dry, when printing out. And some individual papers give very different tones depending on what is applied to them. In figure 2.5 we have compared several different types of papers, exposing them all to sunlight. The papers were all exposed to achieve a range of densities, by varying the exposure time. For each paper, there were four tests: dry, wet with deionized water, wet with a sodium sulfite solution, and wet with a sodium chloride solution.

For all of the papers, the greatest acceleration occurred with the sodium sulfite solution (which aids in the removal of halogen), and the least with the sodium chloride solution, which *adds* chloride (but not bromide) ions.

2.6 Instant film

The Polaroid corporation was the first to market systems of silver gelatin processes for which the light-sensitive material and chemical processing are self-contained in the same package, and so no wet processing in a darkroom is required. These films require special cameras; many of Polaroid's designs were marvels of ingenuity. I will not go into the complex and amazing history of these films and cameras in detail, but they fall into two basic categories: *peel-apart* and *integral films*. In later years, Fuji

Figure 2.4. Exposure test with simple silver halide papers. The left column is silver bromide, while the right column is silver chloride. The top row was prepared with a 100% excess of silver nitrate, while the bottom row was prepared with a 100% excess of halide salt. The middle row was made with a balance of silver and salt. All were exposed simultaneously with a broad-spectrum, high-intensity LED lamp.

manufactured both kinds of instant films as well. As of this writing only integral films are manufactured, but there are old stocks of peel-apart film still in existence (in my refrigerator, for example).

2.6.1 Peel-apart instant films

Peel-apart instant film consists of a silver gelatin negative that is exposed in the camera. It is then, by a clever mechanism inside the camera, matched up face-to-face with a transfer print material. This paper is not light sensitive, but it has a chemical developer incorporated into it. The exposed negative and transfer paper are then pulled together through rollers, which first crush a paper package of gelled processing chemicals. Upon passing through the rollers on the way out of the camera, the processing gel is spread out between the negative and the print transfer paper, both of which have opaque backsides.

Outside of the camera, this negative-gel-print sandwich is allowed to process for a minute or two; a positive print is then peeled off and the negative (with the processing gel stuck to it) is usually discarded. The processing gel contains a silver halide solvent

Figure 2.5. Six different kinds of enlarging paper were exposed to direct sunlight in 2 s intervals. The tests on the left column were brushed with deionized water, the middle column with a solution of sodium sulfite, and the right column with a solution of sodium chloride (table salt).

and an activator that triggers development of the latent image in the negative to reduce it to silver. The silver halide solvent dissolves the remaining silver halides in the negative, which then diffuse through the gel to the print transfer material. The silver halides are then made dark by the developer incorporated into the transfer print material. Since the light areas of the negative are the regions with the least reduced silver, they are the regions with the *most* silver halide. And so in those regions, more silver halide diffuses to the print, and that part of the print is darker. The net result is then a positive print. A variation of this *diffusion transfer process* incorporates a multi-layered emulsion and colored dyes to make a color print (Stroebel *et al*, 2000, p 210).

The color peel-apart films have long been popular among art students because of the possibility of performing the transfer onto ordinary watercolor paper instead to make what is a called an *instant film transfer* (or *Polaroid transfer*). Some versions of peel-apart instant film used a transparent negative that could be saved and used in the darkroom to make multiple prints. No darkroom was needed for processing; they only needed to be cleared of chemicals in a sodium sulfite solution. The highest speed (ISO 3000) versions of black-and-white peel-apart film had an opaque negative, but it could be easily scanned to make an interesting digital image quite unlike the instant film print. See figure 2.6 for two examples. The peel-apart film came in single sheets for large format 4 × 5 and 8 × 10 inch cameras, and also a 3.25 × 4.25 inch medium format film. Special cameras or film holders were needed to process the films.

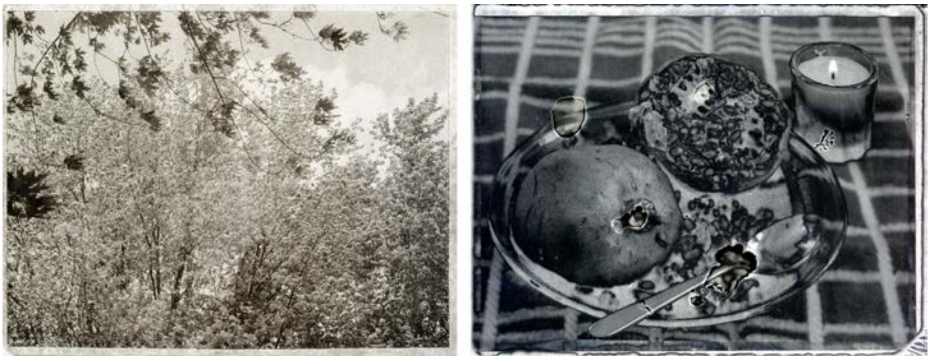


Figure 2.6. **Left:** *I Was Home*. John Beaver, 2008. The gooey throwaway paper negative from high-speed Polaroid peel-apart film was allowed to dry for several days, and then it was scanned and digitally reversed. An interesting diagonal pattern of wrinkles resulted. **Right:** *Pomegranate*, John Beaver 2005. The 3000 speed Polaroid black-and-white peel-apart negative was pulled off of the print early, and then flashed with light to solarize the image (see chapter 4, section 4.2) before scanning to make this digital image.

Figure 2.7. An example of a transfer of an integral film image onto watercolor paper.

2.6.2 Integral films

Integral films are complex multi-layer emulsions with all components of the silver-based photochemical process included in one thin package, with a transparent plastic cover on the top through which the resulting image is visible. After exposure, the chemistry is activated by passing the film package through a set of rollers (built into the specialized cameras) that crush tiny capsules of chemicals. The positive image appears gradually after a few minutes. Polaroid originally made only color versions of this film, in small sizes of a few inches on each side. Both color and black-and-white versions are now available, as of this writing. The black-and-white integral film can be cut apart and the silver gelatin emulsion carefully lifted off in warm water and then transferred to watercolor paper; see figure 2.7 for an example.

References

Beaver J 2017 Using new-antiquarian photographic processes to integrate art and science *AGU Fall Meeting Abstracts*

Eston G T 1957 *Photographic Chemistry in Black-and-White and Color Photography* 2nd edn (Hastings-on-Hudson, NY: Morgan & Morgan)

Fudge H and Beaver J 2017 Ephemeral process photography: Fresh ideas for artistic expression using old techniques *Research in the Rotunda (Madison, WI)*

Lawless L 2005 Making p.o.p. (printing out paper) *Unblinking Eye* [http://unblinkingeye.com/Articles/POP/pop.html]

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

IOP Concise Physics

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Chapter 3

Other photochemical detectors

I describe here several photographic processes that are well out of the mainstream of commercial photography. Many of these are antiquarian—they predate not only digital but also the standardized silver gelatin film and printing processes that immediately predated the advent of digital photography. Some of these processes are finding increasing use in contemporary photography; the reasons have nothing to do with convenience and everything to do with art, issues we will explore in Part II of this book. This list of alternative photographic processes is nowhere near comprehensive; for more details I highly recommend James (2016).

3.1 Daguerreotype

Daguerreotype was the first commercially successful photographic process, introduced by Louis Jacques Mande Daguerre in 1839. It is in some ways one of the most physically complex ways to make a picture; the precise mechanism behind the production of a daguerreotype image is still not fully understood.

A highly polished silver-plated sheet of metal is first exposed, in a special dark box, to a halogen gas (iodine, chlorine, bromine, or some combination of the three). This results in the formation of silver halide directly onto the polished silver plate. The plate is then exposed in a large-format camera and a latent image of metallic silver forms, in a manner somewhat like that of the silver gelatin process. This latent image is then made visible, usually by development with mercury vapor.

But the daguerreotype process differs in a fundamental way from the silver gelatin process, and indeed most other photographic process. In the silver gelatin process, the reduced metallic silver absorbs more—and so reflects less—of the light that falls on it. And so a negative image results when the emulsion is placed on a white or clear background; the reduced silver reflects less light than the white paper base in a silver gelatin print, for example. The daguerreotype process relies instead on a physical change in the *direction* in which light reflects off the developed plate.

doi:10.1088/2053-2571/aaf0aech3

The unexposed parts of the daguerreotype consist of highly polished silver—it is in essence a mirror, and so when illuminated a mirror-like *specular reflection* occurs. But areas of the plate that have been exposed to light and reduced back to silver metal undergo a physical change in structure, and they *scatter* light that reflects off of those areas. And so a somewhat more *diffuse reflection* results from the exposed and developed areas (Stroebel *et al*, 2000, pp 203–4). The daguerreotype is illuminated and viewed such that the reflection off the mirror-like surface of the unexposed regions misses the viewer altogether. But the diffusely-reflecting exposed regions reflect light in all directions, some of which is toward the viewer. Thus, the exposed areas appear bright while the unexposed areas appear dark.

There is still controversy regarding the exact physical process by which this happens. It is often stated that the Mercury mixes with the silver to form an *amalgum*, which is dull instead of shiny (see for example Stroebel *et al*, 2000, p 204). But this has been called into question, most particularly by the fact that there is a variation of the daguerreotype (called the Becquerel process) that uses no mercury at all (James 2016, p 18).

The traditional daguerreotype process is very difficult and quite dangerous, as toxic gases are used. By 1860, other processes were discovered that were more practical, and so its use became increasingly rare. Nonetheless, there is nothing quite like a daguerreotype. The image has a shimmery quality that is captivating. Because of this appeal, in recent years there has been something of a Renaissance in the use of the process, and techniques have been worked out to make the process less dangerous. One of the most successful contemporary daguerreotype photographers is Jerry Spagnoli.

3.2 Wet collodion, ambrotype, and tintype

In the two decades before the invention of the silver gelatin process, silver halide emulsions were made instead with *collodion*—a sticky liquid made of cotton dissolved in ether. The collodion is mixed with halide salts and then poured by hand onto a glass plate that will fit in the camera. If tipped carefully, the collodion flows over the surface of the plate, and then quickly thickens.

Once the collodion is thick enough to no longer flow, the coated plate is dipped—in the dark—into a tank of silver nitrate. The reaction of the silver nitrate and the halide salts then forms, on the surface of the collodion, crystals of light-sensitive silver halide. The now-sensitive plate can then be loaded into a light-tight plate holder, and exposed in the camera.

This is called the *wet collodion,* or *wet-plate* process. The plate must be exposed in the camera, and then developed, before the collodion completely dries. And so to do wet-plate photography in the field, one needs a portable darkroom to prepare the plates immediately before exposure and develop them immediately after.

Wet collodion was the first truly successful photographic negative process. And despite its inconvenience, it is capable of finely detailed negatives that can be used to make many prints. The sensitivity is moderate—an exposure in daylight with a typical lens takes a few seconds. But earlier processes required exposures of several minutes.

The collodion plate can be developed in a different way that leaves the reduced silver metal as a dull whitish color, instead of black. If the back of such a collodion plate is painted black, we then have whitish developed silver on a black background. And thus the collodion process can be also be used to make a direct positive—called an *ambrotype* in the US, and a *collodion positive* in the UK.

A collodion positive can also be made by pouring the collodion onto a pre-blackened metal plate instead of glass, and this is usually called a *tintype*. The name is somewhat misleading, since the plates were almost always made of black-enameled iron, rather than tin. For that reason, some photographers instead use the term *ferrotype*. But that term has its own problems (the same word, as a verb, names a particular method for putting a gloss finish on a silver gelatin print).

A tintype can also be made by hand-coating a silver gelatin emulsion onto the same enameled black plate. It too can be developed so as to make the developed silver whitish instead of black, in order to make a direct-positive image. Using silver gelatin instead of collodion has the advantage that plates could be coated in the darkroom well ahead of time, and developed in the darkroom long after the exposure was taken. And so the photographer did not require a portable darkroom, as in the wet-plate process. Tintypes were relatively easy and inexpensive to make. And although the wet-plate version was less convenient from one point of view, there was a commercial advantage to the fact that they could be exposed, developed, and sold on the spot.

Within two decades of its introduction, most uses of wet-plate photography were eclipsed by the dry, silver gelatin developed-out process. But it is now quite popular among art photographers, attracted by the 'poured' look of the emulsion. The process and equipment required are much simpler than daguerreotype, and it can be used as both a negative and direct-positive process. Probably the most famous contemporary practitioner of wet-plate photography is Sally Mann; her book What Remains is a good example.

3.3 Cyanotype and Van Dyke processes

Cyanotype is one of the oldest of photographic printing processes, invented in 1842 by the English astronomer and intellectual John Herschel. It uses relatively inexpensive iron salts to make a print that is Prussian blue in tone.

In its traditional form, a liquid sensitizer, made by mixing solutions of potassium ferricyanide and ferric ammonium citrate, is brushed onto watercolor paper and allowed to dry. Upon exposure to ultraviolet light, the ferricyanide is reduced by photons to ferrocyanide, and ultimately combines with the ferric ion from the ferric ammonium citrate (the detailed reactions are complex). The result is ferric ferricyanide—Prussian blue. The single letter difference—ferro instead of ferric—means that the former involves Fe^{3+} while the latter is Fe^{2+}. Thus, an electron is gained (reduction), and it is the action of photons that enables this.

The final result, Prussian blue, involves *both* states of iron. Four Fe^{3+} (ferric) ions are bound to three combinations of the Fe^{2+} (ferrous) ion with six CN^- (the *cyanogen* ion, comprised of carbon and nitrogen). The formula can be written $(Fe^{3+})_4[Fe^{2+}(CN^-)_6]_3$.

This mixed state of ionization is unusual, and it is responsible for the deep blue color. It absorbs red light easily, the low energy from the long-wavelength photons causing electrons to skip from the ferrous to the ferric iron (Ware 2016, p 67). The red light is absorbed, but the rest of the spectrum—blue and green—is reflected.

There is a lucky feature that makes all of this practical as a light detector. The potassium ferricyanide and ferric ammonium citrate used for the sensitizer are soluble in water, but the Prussian blue that results from the action of light is not. Thus, after exposure one needs only to wash away the remaining sensitizer, and the Prussian blue stays in the paper, yielding a negative blue-tone image.

Cyanotype is not an emulsion; the sensitizer goes into the fibers of the paper rather than forming a layer that sits on top. It is sensitive to violet and near-ultraviolet light only, with its peak sensitivity near the short-wavelength edge of human vision. Thus, it must be exposed either in a special ultraviolet light box, or to direct sunlight. It is a printing-out process, and so has a very low sensitivity to light compared to developed-out silver gelatin emulsions.

3.3.1 New cyanotype

There are many accounts in print of the cyanotype process, but by far the most comprehensive—covering every imaginable aspect of the process and its history in engaging and erudite detail—is the book *Cyanomicon II* by photographer and chemist Mike Ware (Ware 2016). Ware is, in fact, the originator of what is likely the most significant innovation in the cyanotype process since its discovery—what is called *new cyanotype* (James 2016, pp 212–6). He greatly improved both the light sensitivity and the tonal range of the traditional process by substituting some of the ions used (ammonium and oxalate instead of potassium and citrate). This also has the practical advantage that all of the ingredients of the sensitizer mixed together are stable. The traditional formula requires the potassium ferricyanide and ferric ammonium citrate to be kept separately, mixed only just before use. The left-hand image in figure 3.1 is an 8 × 10 inch new cyanotype print on watercolor paper; it is a contact print with a negative of that same size.

Figure 3.1. The same 8 × 10 format negative was used to make a cyanotype print (left) and a Van Dyke print (right).

3.3.2 Van Dyke brown process

The right-hand image in figure 3.1 shows the same negative used to make a *Van Dyke brown* print. Soon after the discovery of cyanotype, John Herschel realized it could be adapted, using the light sensitivity of the iron salts to make prints from other metals. Many others added to this knowledge over the years, using this foundation of iron-salt sensitivity to make prints with silver, gold, mercury, platinum, and palladium. The Van Dyke process, invented in 1889, couples the sensitivity of iron salts to silver, to produce a brown-tone print (Ware 2016, p 50). Although it is silver that forms the dark areas of the print, it is the iron salts—rather than salts of silver—that provide the light sensitivity. Thus, Van Dyke printing is far more closely related to cyanotype than to the silver processes described in chapter 2.

3.4 Platinum and palladium processes

Platinum printing is the most expensive, and according to many the very best, of all the black-and-white printing processes. Although the final image is made of platinum—one of the most expensive of metals—regarding its light sensitivity it is, like cyanotype, an iron-based process. The chemicals details are different, but in the platinum process, as is the case for cyanotype, ferric salts of iron are reduced by absorption of photons to ferrous salts. But this happens in the presence of salts of platinum, initially producing a very faint negative image. The faint image is then chemically developed to produce a strong image made of platinum (James 2016, chapter 13).

In skilled hands, the platinum printing process can produce the most vibrant range of values, with a lush and complex tone. The main difficulty is twofold: it is difficult to master, and it is very expensive. The element palladium is chemically similar to platinum, but sometimes it is less expensive, and it can be substituted for platinum to make a *palladium print*. Some photographers exert further control over the tonality by using a strategic mix of both metals. Apart from its unparalleled tonal range, the other prized advantage of platinum/palladium printing is that, when done correctly, it is considered to be one of the most permanent or *archival* of all black-and-white printing processes.

3.5 Gum bichromate

Gum bichromate (sometimes called gum dichromate) is, in some ways, the conceptually simplest of all photographic processes. It is also one of the most difficult to master if one wants excellent results. Watercolor pigment is mixed with a gum arabic base, and made light-sensitive with potassium or ammonium dichromate. This sensitizer is then brushed onto watercolor paper and contact printed in the same manner as platinum or cyanotype. Exposure to UV light makes the gum arabic, and thus the pigment suspended in it, insoluble to water. The print is then gently rinsed with warm water, and the unexposed portions wash away. The final print is, in essence, a photographically-printed watercolor.

Figure 3.2. An awful example of a four-color gum bichromate print, using four separate exposures with black, cyan, magenta, and yellow pigments. The four negatives were made by performing a digital color separation of a scan from a color negative, and then printed onto acetate with an inkjet printer.

In order to achieve good results, multiple exposures in layers are needed in order to build up proper contrast. The precise mixture of the gum arabic, dichromate, and pigment must be carefully controlled in conjunction with the exposure and final washing. It has the advantage, however, that virtually any pigment can be used to make the image. In fact, one can even make color prints by using separate exposures with four color-separated black-and-white negatives, each using pigments of black, yellow, cyan, and magenta.

It is very easy to make a gum bichromate print. But it is very difficult to make an excellent one. Figure 3.2 shows a very bad example—my only attempt to make a four-color gum bichromate print (my black-and-white attempts are slightly less awful).

3.6 Anthotypes and chlorophyll prints

Most photochemical light detectors become darker with greater exposure to light, resulting in a negative image. But it is well known that many things respond in the opposite sense, through *bleaching or fading*. Most color dyes fade with light over time. This is, in fact, one of the key difficulties for the archival preservation of color photography. Whether it be a color negative, a color slide or a color print, the dyes that make those colors are susceptible to fading if exposed to light over a long period of time.

Leave a color film negative sit out on a table exposed for weeks to bright afternoon sunlight, and disaster is sure to ensue. The dyes in the negative will fade. This may not be too disastrous if everything fades by the same amount, for one could then correct for this in the process of making a print. But what if, for example, one of the negatives is left half-covered for those weeks by a copy of Noam Chomsky's *The Logical Structure of Linguistic Theory* (which you have been meaning to read)? You will likely find a definite line down your negative; the uncovered part will have faded while the covered part, protected by the opacity of Chomsky's tome, will remain unfaded. It may be possible to scan your damaged negative and fix it with clever use of Photoshop or GIMP, but the task will likely set you to swearing.

A clear lesson to be learned from this thought experiment is that one should always quickly clean up one's mess. But also, this means that color dyes are sensitive to light, and so they could possibly be used as a photographic detector.

3.6.1 Anthotypes

An *anthotype* is a print made by the bleaching action of light on (usually) vegetable dyes on paper. The history of the anthotype goes back to the earliest days of photography, having been first described in detail by John Herschel, the inventor of cyanotype (Ware 2016, p 42). Paper is coated with dye, typically made from flower petals or fruits, and allowed to dry. Objects or a film transparency are placed upon it, and it is then exposed to bright light (usually direct sunlight). After a prolonged exposure of at least several hours, and even as much as several weeks, the uncovered areas fade, while the dye remains unfaded in the covered areas (James 2016, p 42f).

Thus the anthotype is a direct-positive process. But the sensitivity is far less than even the very low sensitivity of a printing-out process such as cyanotype. And so its use as a detector in a camera is impractical. And the anthotype, once completed, is still sensitive to light, and so viewing the image would eventually destroy it. But the extremely low sensitivity of such dyes means that very long exposures in full sunlight are required to make an image. That fact is inconvenient when making the image, but it also means that the resulting photograph can withstand quite a bit of additional exposure without degrading significantly. Even some of John Herschel's anthotypes from the 1840s still survive today (Ware 2016, p 42).

In order for light to have any effect, it must be absorbed. If a particular dye is red, it means that it reflects red light well. Thus, it is unlikely that it will be easily bleached by red light, for the simple fact that little red light is absorbed. The opposite is true for a blue dye, and this leads to the basic notion that dyes are bleached primarily by light that is a complementary color to that of the dye. This is likely true for visible light, but it has no bearing on bleaching by ultraviolet and infrared light, not visible to the human eye and so not related in any direct way to the perceived color of the dye.

Figure 3.3 shows one of my rather-poor attempts at an anthotype. I stained a piece of watercolor paper purple by rubbing it with a wilting purple iris. I then layered objects on it, held in place by a sheet of glass, and exposed it to direct sunlight for about 12 h (over two days). It is still sensitive to light, but the act of scanning it to make the image in this book was insignificant compared to the original exposure.

If I were to frame it and display it in a moderately-dim hallway, with 1000 times less intensity than sunlight (and probably several thousand times less ultraviolet light), it would take many months to fade noticeably due to the additional exposure to light (the dyes may be unstable for other reasons).

Figure 3.3. My own mediocre example of an anthotype photogram, using the juice from a pressed purple iris.

3.6.2 Chlorophyll prints

Chlorophyll is the ultraviolet-sensitive pigment that both allows a plant to produce energy from sunlight and also makes the leaves green. Pluck a green leaf from a plant, and the leaf no longer has access to the plant's system of vascular circulation of water and nutrients. The leaf will dry and become brittle, of course. But also, the green chlorophyll will fade upon prolonged exposure to ultraviolet light. Thus, one can use a chlorophyll-laden green leaf to make a photographic image in the same manner as an anthotype, but with no need to brush on a dye (James 2016, p 56f).

This is called a chlorophyll print, and the artist Binh Danh was the first to produce a successful body of work using this idea. He renders the leaf/images more permanent by casting them in resin to preserve their delicate physical nature, and also perhaps because the acrylic resin blocks a large percentage of ultraviolet light. Some examples of chlorophyll prints by the London-based photographer Almudena Romero can be seen in chapter 11, section 11.2.

References

James C 2016 *The Book of Alternative Photographic Processes* 3rd edn (Boston, MA: Cengage Learning)

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

Ware M 2016 *Cyanomicon II–History, Science and Art of Cyanotype: Photographic Printing in Prussian Blue* [http://www.mikeware.co.uk/mikeware/downloads.html]

Chapter 4

Some interesting technical details

4.1 Reciprocity failure

It is desirable that the density of a photographic emulsion depends only on the exposure, and not the details of how that exposure was made. Exposure is the product of illuminance and time (equation (1.6)). The *law of reciprocity* states that the density achieved depends only on the exposure—and not specifically on either illuminance or time individually. This means one could (for example) quadruple the illuminance while reducing the exposure time by the same factor, and it would still have the same effect on the detector. When the law of reciprocity holds, a photographer's job is easier.

Reciprocity failure means that a different density results when using a long exposure time at low illuminance compared to a short exposure time at high illuminance—even if both combinations are the same exposure. The most commonly-encountered example of reciprocity failure relates to silver gelatin emulsions. It is well known that silver gelatin emulsions are, in effect, *less sensitive at very low levels of illuminance*. Because of this, a greater exposure is needed to achieve the same result (the same density) whenever the illuminance on the detector is very low.

What would cause a very low level of illuminance on the detector? Two things: either a very tiny aperture (for example pinhole photography) or very dim scene lighting (for example night photography). In either of these cases (or a combination of both), a greater exposure is needed than what one would expect from tests of the detector at higher levels of illuminance. And thus one must compensate for the lower sensitivity of the detector by choosing a *longer* exposure time than would be used if reciprocity failure were not a factor. Reciprocity failure greatly complicates the life of the silver gelatin film based astrophotographer or pinhole photographer.

But what is the *physical* cause? Why *is* the sensitivity of a silver gelatin emulsion reduced at very low levels of illuminance? To answer this, we must recall an important detail regarding how a latent image is formed in a silver gelatin emulsion. Recall from chapter 2, section 2.1, that the first step is for a stable speck of silver to

doi:10.1088/2053-2571/aaf0aech4

form, from the action of photons, at a crystal sensitivity sight. It is estimated that at least four Ag^+ ions must be reduced by photons to elemental silver before the speck is stable. Otherwise, the affinity the silver has for the halogen that is also formed in the process causes them to simply re-trade that electron and again form, for example, Ag^+ and Cl^- ions (Stroebel *et al.*, 2000, p 200).

A very low level of illuminance means that very few photons arrive per unit of time at a given sensitivity speck. *And this means that very few silver ions are reduced to silver per unit time*. At extremely low levels of illuminance, it essentially *never* happens that photons succeed in reducing four silver ions to silver in the same place in a short enough time for them to not, in the mean time, convert back to ions. And so a stable speck is never formed.

At high levels of illuminance, on the other hand, none of this is a problem. Photons reduce silver ions to silver at such a high rate, that they essentially never have time to revert back to ions before a stable speck of four silver atoms is created. This is the ordinary use of silver gelatin emulsions where reciprocity failure is not a factor. And so reciprocity failure is the case in between. The illuminance is low enough that the natural 'undoing' of the reduction of silver ions to silver is a factor, but it is still possible to record a latent image.

4.1.1 Reciprocity failure and VLS photography

One can expose a piece of silver gelatin enlarging paper in a pinhole camera on a sunny day and develop the paper negative after the exposure is finished. The exposure time required to get a successful photograph can be many minutes long.

In volume II of *The Physics and Art of Photography* I discussed the case of VLS photography—taking pictures with very low sensitivity detectors such as cyanotype or printed-out silver gelatin emulsions. Ephemeral process (EP) photography is a good example, and it provides a useful comparison because the same type of enlarging paper is often used for pinhole photography. But in the case of EP photography, the paper is not developed; it is left to print out instead, with an accelerator applied to increase the speed. The exposure time on a sunny day for EP photography is comparable to that of pinhole photography.

The successful pinhole photographer knows that reciprocity failure is important. If the light meter says an exposure time of five minutes is needed, it will certainly take much more. Pinhole photography is a textbook case of a situation where reciprocity failure is a significant, or even dominant, effect. If the sky becomes hazy and the light meter now says an exposure time of 10 min is needed, then reciprocity failure will be an even *bigger* effect.

The EP photographer uses the same light detector (in a printing-out, rather than developing-out mode) as the pinhole photographer, and needs similar exposure times. And yet reciprocity failure is of no consequence. Why?

The answer is obvious if one recalls that reciprocity failure is not about exposure time, it is about *image plane illuminance*. Reciprocity failure is a large effect whenever the light falling on the silver gelatin emulsion in the camera is very, very dim. That is true for pinhole photography, because of the tiny aperture. The long exposure time is not the cause of the reciprocity failure, it is the *consequence* of the very low image plane illuminance (which is the cause of reciprocity failure).

In EP photography, on the other hand, the long exposure does not result from low image plane illuminance. Far from it! The EP photographer typically uses a fast lens in broad daylight, and the image plane illuminance is about as high as it gets in photography. Instead, the long exposure times in EP photography result from the *very low sensitivity of the printing-out process*. But the high image plane illuminance means that reciprocity failure is not a factor.

4.2 Solarization

On occasion, the relation between exposure and density—the characteristic curve—can be far stranger than the typical example shown in figure 1.10. Consider instead a characteristic curve that is shaped like a hump. Looking at such a graph from the left to right, increasing exposure would cause greater density, as expected. But after the peak of the hump, density would *decrease* with increasing exposure. Or, what if the characteristic curve were the opposite of that—a ‘U’ shape. In that case, the density would decrease with increasing exposure until a minimum density is reached at some particular mid-point exposure. And then as the curve turns upward, density would increase with exposure. For either a hump-shaped or a U-shaped characteristic curve, there are particular densities that are caused by *both* small and large exposures.

If by density we mean the darkness of a silver gelatin emulsion, the hump-shaped characteristic curve would represent an ordinary negative for the parts of the image that received the least exposures, where the characteristic curve slopes upward. But the downward sloping side of the hump, corresponding to the highest exposures, would be a *positive* image. And so the image would be part positive and part negative. The same would be true for the U-shaped characteristic curve—it would be part positive and part negative. The difference is whether it is the shadows or the highlights that have been reversed to a positive.

A picture that shows one of these relations between density and exposure is usually said to be *solarized*, and there is a very particular look to it. Figure 4.1 shows an example. The image on the left is much as the scene appeared in life, while the image on the right has been solarized. In this case, the solarized effect has been accomplished with digital image processing on an ordinary digital image. The hump-shaped curve superimposed on the solarized right-hand image is something like a characteristic curve. But rather than a relation between the exposure of a detector and the density that results, this is an *adjustment curve*. It represents changes made to the brightness levels in the image on the left, in order to produce the image on the right.

We will consider adjustment curves in more detail in chapter 8, section 8.2, but it has much in common with a characteristic curve. In particular the hump shape of the adjustment curve means that two different brightness levels in the left-hand image can result in the same brightness level in the solarized image. Notice that some of the darker tones in the solarized image are also darker on the original. But some of the darker areas in the solarized version were actually the brightest areas on the original.

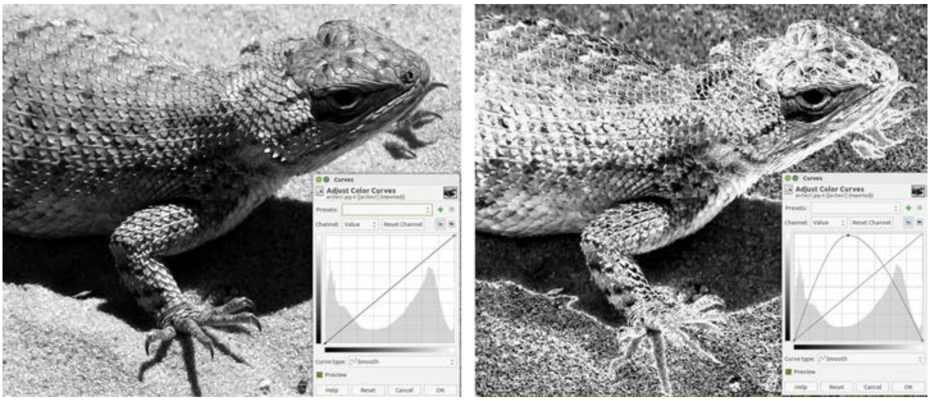


Figure 4.1. The image on the left has been digitally solarized to produce the image on the right. The brightest range of values has reversed from a positive to a negative, while leaving the darker areas as a positive image.

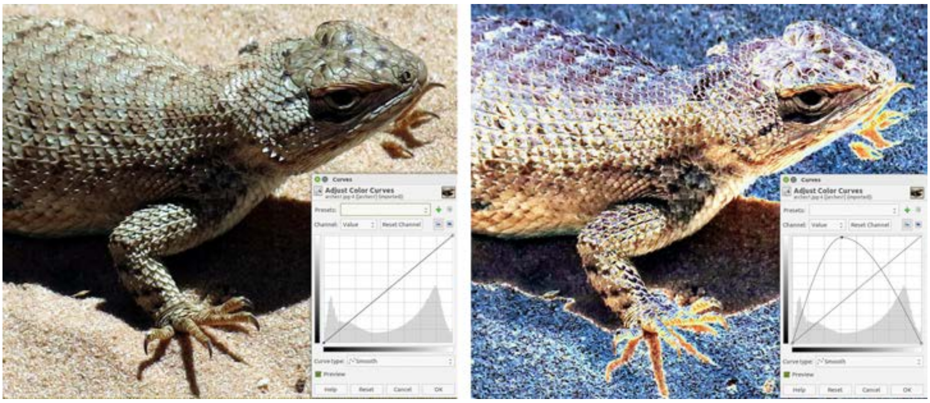


Figure 4.2. A color image can also be solarized. Not only do the values of the highlights reverse to a negative, but also the hues of the highlights switch to their complements.

The adjustment curve works in the opposite sense of the silver gelatin film characteristic curve, which shows an upward slope for a *negative* image. And so an upward sloping adjustment curve represents a *positive* image, while a downward slope results in a negative. For the image in figure 4.1, the shadows are left as a positive, but the highlights have been reversed to a negative.

A color image can be solarized as well. Although the details are more complex, it is usually the case that not only do the values at the high end reverse to a negative, but also the hues associated with those parts switch to the opposite side of the color wheel. See figure 4.2, and note that the reversed parts have turned from a bright yellowish color to a dark bluish color.

There are two ways in which silver gelatin emulsions can be made to respond with an effective characteristic curve that gives a solarized look. The names for these effects, however, are confusing. The preferred technical term for the effect that most commonly produces a solarized look in a silver gelatin emulsion is not ‘solarization.’

Rather, the ‘correct’ term is the *Sabatier effect*. Nonetheless, most people still use the term solarization when they see the Sabatier effect in action. The Sabatier effect results in a U-shaped characteristic curve in the silver gelatin emulsion.

I also describe below another process by which silver gelatin emulsions can result in something like a solarized look, but the characteristic curve is hump shaped instead, the opposite of the Sabatier effect. It is used much more rarely than the Sabatier effect, and it more often than not affects only a small part of a picture. Nonetheless, the preferred technical term for this effect is *true solarization* (Stroebel *et al.*, 2000, pp 124–6, 201–2). The reversal in true solarization is the opposite of that in the Sabatier effect. The Sabatier effect reverses the areas of the emulsion that received the *least* exposure to light. True solarization, on the other hand, reverses only the regions that received a very large amount of *over*exposure.

4.2.1 Sabatier effect

The Sabatier effect produces a solarized look during the development phase of the silver gelatin image. The emulsion is exposed in much the normal manner (but with a somewhat reduced exposure). But while it is in the process of being developed, just as the visible developed image is beginning to appear, it is given a brief exposure to bright, diffuse light. Under the right conditions this second exposure to light will, as the development proceeds, darken the unexposed areas while leaving the previously developed areas relatively unaffected.

The process is difficult to control, and it works best with particular combinations of emulsion, developer, initial exposure to the image, and the details of how it is flashed with light during development. As such, it is most-commonly used to make a silver gelatin positive print from a negative, under controlled conditions in the darkroom. See figure 4.3 for an example. Both silver gelatin prints were made from

Figure 4.3. **Left:** A silver gelatin print made in the darkroom from a film negative. **Right:** The Sabatier effect was used to make a solarized print from the same negative. I used a different developer and enlarging paper, and flashed the print with light while it was only partially developed. Some of the highlights (dark areas on the original negative) reversed to a negative on the solarized print.

Figure 4.4. **Left:** *Emma No. 1*, John Beaver 2006. **Right:** *Yard and House*, Teresa Patrick 2005 (used by permission of Teresa Patrick).

the same negative. The left-hand example was made in the ordinary way, while for the right-hand example I applied the Sabatier effect. Notice that the highlights on the print have reversed to a negative, while the shadow areas have remained a positive. But remember that this is at the stage of a print from a negative. And so the un-reversed shadow areas on the print are those that received the *most* exposure.

It is also possible to solarize an in-camera negative. Figure 4.4 shows two unconventional examples. The image on the right by photographer Teresa Patrick was made by employing the Sabatier effect on silver gelatin print paper in the darkroom, using the ordinary techniques for solarizing a print. But, rather unconventionally, the paper was not exposed by light projected through a negative in order to make a positive print, as in figure 4.3. Instead, Patrick exposed the 8 × 10 inch sheet of enlarging paper directly in a pinhole camera, and then brought it into the darkroom for development.

Ordinary development would have produced an ordinary negative—from which a positive print could have been made by making a contact print with a second sheet of enlarging paper. But instead Patrick solarized the original exposure such that about half of the image reversed from a negative to a positive. She then toned the image, and framed it. So the piece of paper went directly from the camera to the wall—a sort-of direct positive, but with a solarized look.

The Sabatier effect causes the low-exposure areas of the emulsion to reverse and become high-exposure areas. If this is performed at the step of the final print, highlights reverse to become shadows. When the Sabatier effect is instead performed on the negative, the same thing happens. But if a positive print is then made from the solarized negative, the roles are switched. In the final print it is instead the *shadows* that are reversed to highlights.

The left-hand image in figure 4.4 was made by solarizing a negative, and then making a positive print from the solarized negative. The dark-haired model was wearing dark clothing and lying on a black cloth, all of which switched to become bright areas of the picture, while leaving the skin tones unaffected. With a combination of experience and luck, pleasing effects can sometimes result, highlighting important parts of the picture.

Figure 4.5. Two examples of accidental true solarization. **Left:** Testing a new lens with instant film, the intense illuminance from the light bulb, combined with the particular circumstances of the instant film development, reversed both the values and hues. **Right:** This cyanonegative photograph—a direct image of the Sun—resulted in an effect much like true solarization because the focused image of the Sun literally burned a hole in the cyanotype paper detector.

4.2.2 True solarization

In true solarization highly overexposed areas reverse to a negative. This most often happens during the exposure of the film negative in the camera—and the direct image of the Sun focused onto the film is the usual culprit. Sometimes it is accidental, as in the leftmost image in figure 4.5, where the image of the intensely bright surface of the light bulb reversed to a negative in this Fuji instant film photograph.

The image on the right has a similar effect, but it did not involve a silver gelatin emulsion, and so is technically not an example of true solarization. Rather this is one of my early cyanonegative photographs (see Ware 2016, p 105, and volume 2 of *The Physics and Art of Photography*); I exposed cyanotype paper for several minutes in a homemade camera using a fast simple magnifier as a lens. The direct image of the Sun literally burned a hole in the paper during the exposure.

I am somewhat embarrassed by this picture, not because of burning a hole in the negative; I expected that would happen, and wanted to see what the effect would be. Rather it is because of the name I picked for the image; I named it *Black Sun* after the novel by Edward Abbey. It was a year or so later that I discovered Abbey had named his novel after the famous photograph, ‘The Black Sun, Tungsten Hills, Owens Valley, California,’ by Ansel Adams—one of the most famous examples of true solarization in a photograph. And I plead an embarrassing amount of ignorance, rather than unspeakable arrogance, for accidentally naming one of my pictures after a photograph by Ansel Adams.

4.2.3 Mackie lines

If one looks closely at some of the detail in the solarized image in figure 4.1, the border areas between reversal and non-reversal to a negative reveal a strange detail. One part of this area is enlarged as the left-hand image in figure 4.6. This border region seems to be outlined by a bright line, as if someone outlined the edges with

Figure 4.6. The regions on the border between reversal and non-reversal are outlined. These are called Mackie lines, and they give solarized images a graphic arts feel.

white paint. These are called Mackie lines, and they are one of the hallmarks of solarization.

In the right-hand image of figure 4.6, a detail from the left-hand image in figure 4.4, the Mackie lines are dark instead of bright. This is because this image was made by applying the Sabatier effect to solarize a *negative* rather than a positive. The Mackie lines were white on the negative, but now the reversal to a positive has made them black.

Many authors, for example London *et al*., (2005, p 300) and Stroebel *et al*., (2000, p 202) suggest that the Mackie lines result from the retardation of development by the physical migration of chemical byproducts in those areas during the solarization process. These chemical 'edge effects' may be a factor, but there must be more to it than that. For even pure digital processing—with no silver gelatin emulsion involved—can easily produce the same effect.

See, for example, the leftmost image in figure 4.6, which was solarized *digitally* from an ordinary digital photograph; the Mackie lines are very clear even though no chemistry was involved. I performed this solarization effect not by clicking on a 'solarize' filter in Photoshop (which could be specifically coded to simulate the look of Mackie lines), but rather by a simple adjustment of the brightness values of the image in a way that is similar to how the Sabatier effect changes the characteristic curve.

Consider a region, such as the shadow under the lizard, that has a very sharp boundary between light and dark. Upon solarization, the dark area will remain dark, while the light area switches to dark. So it was dark next to light, and now it is dark next to dark.

But there is usually, in any image, a *transition* between a dark area and a neighboring light area—a narrow region where there are shades of gray in between. And what happens to these shades of gray? Some of these shades of gray are dark enough that the solarization process does not reverse them—but they are still light enough to contrast strongly with both the unsolarized dark areas and the solarized

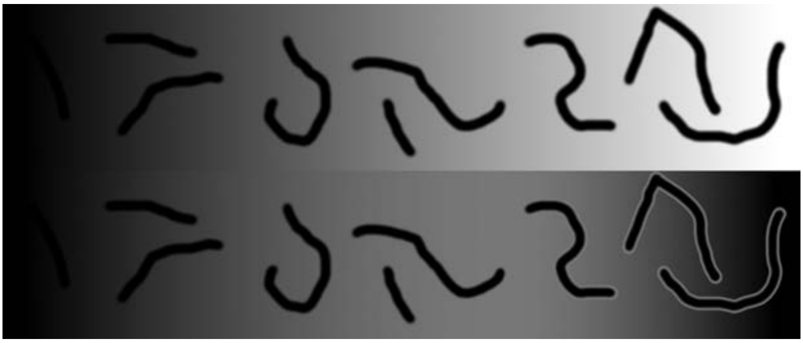


Figure 4.7. The Mackie line phenomenon appears even in this simple digital image. The top image was created with a soft-edge paintbrush tool in GIMP. When the brightness levels are transformed (bottom image) in a way similar to the characteristic curve produced by the Sabatier effect, Mackie lines naturally appear at the soft borders of the marks in the reversed areas.

light areas that have switched to dark. Thus, there is a narrow region that is brighter than the regions on either side of it, and so shows up as a line marking the boundary between the solarized and unsolarized area—a Mackie line.

Further evidence for this explanation for Mackie lines is illustrated by the simple experiment shown in figure 4.7. For the top image I used GIMP to make a simple grayscale image, and then I drew over it with black lines made with a soft-edged brush tool. And so the dark marks have a range of gray levels at the edges. I then transformed the brightness of the pixel values to make the highlights negative while leaving the shadows unchanged, in a manner much like the Sabatier effect. Mackie lines appear just as one should expect.

4.2.4 Negative or positive?

The Sabatier effect can be employed at the development stage of either the negative or positive, and both choices have advantages and disadvantages. Solarizing a film negative in the darkroom is much more difficult than solarizing a print, and so it is done less often. But some of the best solarized images—most famously many of those by Lee Miller and Man Ray—were made by solarizing the negative.

The Sabatier effect naturally leads to an image with a very low range of tones. It is easy to make the darkest areas of the print or negative very dark. But the lightest areas are never white; the best one can hope for is a light-tone gray. And so using the Sabatier effect at the stage of the print has a disadvantage: *the final print will inevitably lack tonal range in the highlights.* When the Sabatier effect is used to solarize a negative, this is also true—for the negative. *But the added step of making a positive print from the solarized negative allows the possibility of producing a full tonal range in the final print.*

This distinction is evident in the two images in figure 4.4. I was able to accomplish the near-full tonal range in the left-hand image only because of adjustments I made in the process of making a positive from the solarized negative. The original negative had a very poor tonal range indeed. Teresa Patrick's direct print solarization on the right, however, has a much smaller tonal range; she was able to extend it somewhat with the added step of chemical toning.

Apart from the question of tonal range, there is another factor in choosing whether to apply the Sabatier effect directly to a final positive or instead to a negative that is then inverted to a positive. Which do you want reversed—the shadows or the highlights? Do you want the Mackie lines to be white or black?

> Applying the Sabatier effect *with no additional step* of inverting the image from negative to positive will produce a final image with normal shadows and reversed highlights, and with white Mackie lines.
>
> Applying the Sabatier effect and *then inverting the image* from negative to positive will produce a final image with reversed shadows and normal highlights, and with black Mackie lines.

It is important to note that the issue here is not, strictly speaking, whether one is dealing with a positive or negative. It is whether or not, after the Sabatier effect is applied, an additional reversal step (such as making a positive print from a negative) is performed.

And so, for example, consider the image in figure 4.8. It has reversed highlights and white Mackie lines, suggesting that I made this by applying the Sabatier effect to

Figure 4.8. *Examining the Moon*, John Beaver 2007. This image was made from a solarized instant film negative. But the Sabatier effect reversed most of the important parts of the picture (the figures) from a negative to a positive. And so I simply scanned the negative and adjusted the levels to make a final print, without inverting the negative to a positive. This gives the reversed-highlights effect of a solarized *print*, but the contrast can be stretched by digital image processing before making a final print.

a piece of enlarging paper while making a final positive print in the darkroom. But instead, I produced this image by applying the Sabatier effect to an instant film *negative*. In the scene I photographed, the figures and the trees were dark, while the sky and the ground were bright. Those dark areas would develop out light on a normally-processed negative, while the sky and ground would be dark. But when I solarized the negative, the light areas of the negative (corresponding to the darker figures and trees), reversed from a negative to a positive. The brighter ground and sky remained a negative.

I could have then applied the usual additional step of reversing the film negative to a positive print. But the Sabatier effect had already turned the important parts of the picture (the figures) into a positive image. So instead I simply scanned the negative and adjusted the contrast. And so I have a final image that looks like I have solarized a print in the darkroom even though it was made by solarizing a film negative. But the additional step of scanning the negative to make a digital image allowed me to increase the tonal range of the image through digital image processing before making a (digital) print.

There is another important question though, and it goes to the heart of much of the discussion in Part II of this book: do we want an *image*, or do we want an *object*? Part of the appeal of using the Sabatier effect in the darkroom to solarize a print is that the process is only partially under the artist's control, and so much is left to chance. *And every print is different*. But if instead one makes 100 identical prints from the same solarized negative, then *the element of chance comes in only at the level of the image*, not the final print. This sometimes leads to a tough choice. Do I want a more compelling *image*, even if it is only virtual, or do I want a more compelling *object* (a unique solarized print, for example), even if as an image it is inferior?

References

London B, Upton J, Stone J, Kobré K and Brill B 2005 *Photography* 8th edn (Englewood Cliffs, NJ: Prentice Hall)

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

Ware M 2016 *Cyanomicon II–History, Science and Art of Cyanotype: Photographic Printing in Prussian Blue* [http://www.mikeware.co.uk/mikeware/downloads.html]

Chapter 5

A brief diversion into the weird world of the photon

Let us pause for a moment and think a little about what happens when a 'photon' is absorbed by an 'atom' or 'molecule.' I put these words in quotes because there is an important sense in which photons, atoms, and molecules do not exist.

It is likely that our human brains come pre-equipped to think about the world in particular ways, including the idea that 'things' have solidity. That is, we have a 'common sense,' and we see the world as being made of objects that are solid and immutable particles. And so when we hear words such as atom or molecule or photon, we want to think of a solid something that either is or is not at a particular location in space at any given point in time.

But just because we naturally think in these terms, it this does not mean that the world is really like that. Consider: we have no direct experience, either as individuals or by means of our millions of years of evolution, of a subatomic world at the tiniest scales of existence. And so it would seem to be an odd coincidence if our minds happened to think in terms of concepts that were directly relevant to that minuscule world.

Quantum physics, developed by many individuals mostly over the years 1899–1929, provides an accurate description of this tiny world. When I say 'accurate' I mean that quantum physics works; one can use it to correctly predict, with great precision, the outcome of a wide range of experiments. But despite its predictive power, it makes little *intuitive* sense. That is to say, many of the concepts employed by quantum physics disagree with our (perhaps built-in) assumptions about how the world works.

The idea of a photon as a particle is one such concept; we use the word 'particle,' but by that we actually mean something very different from what our brains try to tell us a particle should be. For a photon does not really exist, in the ordinary meaning of the word, at any particular time or place. And it is certainly not 'solid.'

doi:10.1088/2053-2571/aaf0aech5

In order to learn quantum physics one must give up the idea that it will make intuitive sense. Instead, one must use the intellect to learn a formal theory that can be applied to many different circumstances. After much effort using the theory for many situations, a physicist gradually develops a second sense that allows her to get a feel for what the theory will say in general. And so we have a built-in intuition that (in the case of quantum physics) mostly fails us, but we can (with much effort) develop a different intuition to guide us in thinking about the subatomic world[1].

5.1 Young's double-slit experiment and the wave model of light

Light is a wave in the sense that, given the chance, light does the sorts of things that waves do. The phenomenon of interference is a good example. The first utterly convincing demonstration of the interference of light waves was performed in the early 1800s by Thomas Young.

Young allowed light to emerge from two tiny pinholes spaced a small distance apart. The light that emerged from these narrow restrictions was then allowed to fall upon a screen, and an alternating pattern of light and dark was observed. Young successfully explained this effect by appealing to a wave model of light. Since from any particular location on the screen the two holes are at slightly different distances, the two waves may interfere constructively or destructively, depending on the details of the geometry (see figure 5.1).

It may seem that this would be a trivial experiment to perform, but there were in fact several technical obstacles that Young had to overcome because a *coherent* source of light, such as a laser, was unavailable in his time. Young's experiment was key to the acceptance of a theory of light as a wave phenomenon, in opposition to the earlier view, expressed most famously by Isaac Newton, that light is a stream of

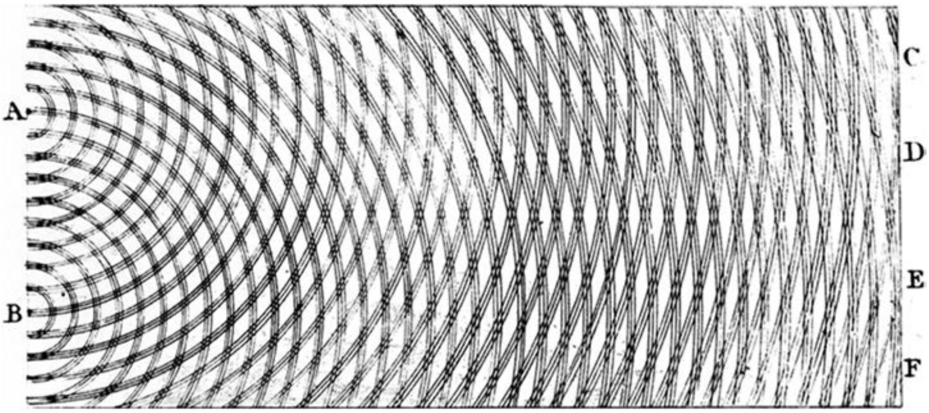


Figure 5.1. Young's diagram explaining the interference of light waves coming from two closely-spaced holes. Graphic: Public Domain.

---

[1] See Chomsky (2000, pp 108–12) for an interesting historical discussion of a different example (gravitation) of this process.

particles. Further developments, especially by Augustin-Jean Fresnel, provided detailed evidence for the wave-like nature of light.

The final nail in the coffin of Newton's particle theory of light was hammered in 1817. Poisson, a skeptic of the wave theory of light, showed mathematically that Fresnel's wave theory, if correct, makes a ridiculous prediction: if coherent light is blocked by a round obstruction, there would be a bright spot in the very center of the shadow. When the suggested experiment was immediately performed, however, it proved to everyone's surprise that 'Poisson's spot' is real.

For Poisson in 1817, the joke was on him. But perhaps in the end, he had the last laugh. The success of the wave theory of light demonstrates absolutely that light does the sorts of things that waves do. And so in that sense, it *is* a wave. But it does not mean light is *only* that, and many experiments, beginning around 1900, showed that there is more to it. For in certain other circumstances, light does indeed behave like a stream of tiny particles that transfer their action individually at specific times and places. Reconciling these two seemingly incompatible descriptions of light occupied many physicists for the first few decades of the 20th century, and the result is our modern theory of quantum mechanics.

5.2 The photoelectric effect and the particle model of light

One important historical example of light acting as a stream of particles is the *photoelectric effect.* If light is projected onto certain metallic surfaces, electrons are ejected from the metal. If this is accomplished inside a glass tube evacuated of air, these now freed electrons can be accelerated by an electric field to make a measurable electric current.

The surprising thing is that careful experiments showed the photoelectric effect to be, in many ways, an either/or thing, not a more-or-less thing. For example, if the wavelength of the light is longer than a particular threshold value (different for different metallic surfaces), then no electrons are freed by the light, *no matter how bright it is*. From the point of view of light as a wave, this makes no sense. It is true that these longer-wavelength waves, all else being equal, are of lower energy than shorter-wavelength waves. But we should be able to compensate for this simply by making the light brighter (so the waves have a larger amplitude). And so why is there a well-defined particular wavelength, longer than which no effect occurs at all?

Another odd feature of the photoelectric effect is that the energies of the individual electrons that are freed from the metal are not related at all to the brightness of the light. Instead, they relate only to the *wavelength* of the light. The brightness instead decides *how many* electrons are freed. From the point of view of waves, this makes no sense since the energy of a wave is related to *both* the wavelength and amplitude (brightness).

Finally, waves carry *energy per time*—power. This means that the energy transmitted by a wave depends on the length of time over which the wave is absorbed. A low-power wave absorbed over a long period of time transfers just as much energy as a high-power wave absorbed for a proportionally shorter amount of time. Thus, one would expect there to be a time delay for the photoelectric effect. That is, the atoms must absorb the light for a long enough time so that enough energy has been absorbed to free the electrons from the atoms. This also means we would expect a longer time delay for dim light than for bright light. But experiments clearly show that it doesn't work that way. The photoelectric effect either happens or it does not, and when it does happen, it is virtually instantaneous.

This mystery was solved by Albert Einstein in 1905. He showed that the experimental facts of the photoelectric effect *do* make sense if we think of the light shining on the metal not as a wave, but as a stream of particles. These particles (called photons) interact *one-on-one with individual atoms*. From this point of view, either an individual photon has enough energy to knock an electron off the atom, or it does not. And so if the light shining on the metal is made of photons that, *individually*, have too little energy, then nothing will happen no matter how many photons there are. There is no time delay because it is like a collision between two particles; all of the energy of the photon is delivered at once.

Einstein went further than this in an important way; he connected this 'particle model' to the existing 'wave model.' That is, light can act either as a wave or a stream of particles depending on the circumstance, but the two are not independent of each other; there is a *correspondence* between the two models. It works like this:

1. A bright light, which corresponds to waves of large *amplitude*, corresponds to a stream of *many* photons, while a relatively dimmer light (lower-amplitude wave) corresponds to a stream of relatively fewer photons.
2. Light of relatively *long* wavelength (low frequency) corresponds to a stream of photons, each of which has relatively *low* energy. Light of relatively *short* wavelength (high frequency) corresponds to a stream of photons, each of which has relatively *high* energy. Einstein proposed that photons have an energy given by the following simple equation:

   $$E = hf. \tag{5.1}$$

   Here E is the energy of each individual photon, f is the frequency of the light if it were instead acting as a wave, and h is a very tiny number called *Planck's constant*.

The constant, h, is called Planck's constant because it was first proposed by Max Planck in 1900, in his theoretical accounting for blackbody radiation. Einstein made the additional leap, five years later, that this tiny constant represents the energy per frequency of photons—individual particles of light[2].

The measured value of Planck's constant is 6.626×10^{-34} J s. It has this *particular* tiny value simply because … it does. It is one of the constants of nature, along with the speed of light and the constant of gravitation. It is unknown why these and other fundamental physical constants have the particular values that they do.

---

[2] The word 'photon' was first coined years later, in 1926, by Frithiof Wolfers and Gilbert N Lewis.

5.3 Young's experiment reconsidered

And so let us look at Young's experiment in terms of the particle model of light. How would we do that in practice? That is to say, how could we set up the experiment such that the light is acting like particles instead of waves? All we need to do is to allow the light shining on the two slits to be very dim, and to use extremely sensitive light detectors to record the light as it arrives at the screen.

For example, we could place a row of sensitive *photomultipliers* at the screen. These are glass vacuum tubes that use the photoelectric effect. But a single electron freed from the metal is accelerated (by an electric field) and collides with a second target, and more electrons are liberated by the collision. This process can be repeated several times within the tube, and so a single photon can result in a shower of many electrons, causing an easily measurable pulse of electric current. Thus, a photomultiplier is capable of detecting single photons as they arrive. If we then place a row of photomultipliers at the screen location in Young's experiment, we can monitor the arrival of each individual photon at the screen.

Next, imagine that we illuminate Young's double slit with light of such low intensity that *on average only one photon at a time arrives at the slit and at the row of detectors*. What happens? A moment's thought reveals that this situation makes little sense from the strict point of view of *either* waves *or* particles. For a single particle can only pass through one slit or the other, so how does it even 'know' that the slit is double? And a wave must be spread out in space and time; it is a pattern, not a thing. So how can a wave arrive at the detector at a single moment in time, at a single place?

In practice, each single photon that passes through the double slit arrives at one of the rows of detectors and triggers a signal. For any single one of these photons, *one cannot predict, even in principle, which of the detectors will be triggered.* For a single photon the interaction with the row of detectors is a *random event*.

But here is the curious thing. Although the arrival of each individual photon is random, if one keeps track of where each photon arrives, *after many thousands of photons have been detected, the original interference pattern emerges*. That is, more photons arrive at the places where Young's experiment showed the waves to interfere constructively, and very few arrive at the places where the wave experiment produced destructive interference. And it is the interference pattern for the *double* slit that arises, not the very different diffraction pattern one sees when light passes through a single slit.

This is very strange, indeed. For example, how does the first photon 'know' where it is supposed to go, such that in the end, after many thousands of photons, it is part of the overall 'picture' of bright and dark interference bands? If the photons are arriving only one at a time, the same question could be asked of each photon. How do they, individually, know what to do? Since a wave is spread out in space and time, there is no paradox there. The wave passes through *both* slits at once, and it stretches all the way from the slit to the screen.

Quantum physics can solve this paradox in different ways. One important example, the *wave mechanics* developed by Erwin Schrödinger in the 1920s, imagines a *wave function* that, although it cannot be observed directly, tells us how to calculate the *probability* that a particle will be detected. And so from this point of view, there *is* a wave that can interfere even when only one photon is detected at a time; it is the wave function. But constructive interference now means, ‘there is a high probability of detecting a photon here.’ *The photons themselves act randomly*, according to the laws of probability. So naturally, more of them arrive at the locations where detection is more likely to occur.

One interpretation is that the wave function is, in a sense, ‘what is really real,’ even though it cannot be directly observed, even in principle (see, for example, Penrose 2004, p 508). The photon, on the other hand, is just a sudden transfer of energy to the detector. In a sense, it is not a ‘thing.’ And so we have an interpretation that what is real cannot be observed—but it does tell us how to predict the *likelihood* of a particular observation.

Wave mechanics is not the only approach to quantum physics, and when other experiments are considered, it all gets weirder still, and the precise connection between our ordinary experience and the strange rules of quantum physics are still a subject of much debate (see, for example, Penrose 2004, chapter 29). I leave it to the interested reader to explore this in more detail in one or more of the many excellent books on quantum physics that are written for the lay person (for example, Feynman 1985).

References

Chomsky N 2000 *New Horizons in the Study of Language and Mind* (Cambridge: Cambridge University Press)

Feynman R P 1985 *QED: The Strange Theory of Light and Matter* (Princeton, NJ: Princeton University Press)

Penrose R 2004 *The Road to Reality: A Complete Guide to the Laws of the Universe* (New York: Vintage Books)

IOP Concise Physics

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Chapter 6

Digital photoelectric detectors

A digital detector produces some kind of measurable electrical signal upon exposure to light, and this signal is converted to a number. In and of itself, this is not so difficult, since many materials develop a slight electric charge upon exposure to light. The trick is that there must be some way to tell which signal corresponds to which particular *position* on the detector. Otherwise there would be no detection of an *image*, but instead only a single measurement of the total light striking the detector.

A sheet made of the element *selenium* is an example of a material that produces an electrical response to light. Focus an image onto a sheet of selenium, and a pattern of electrical charge will develop on its surface. Parts receiving a greater exposure will build up a larger electric charge. Many old camera light meters used selenium in this way, since the accumulated charge can be drained off to produce a tiny but directly measurable electric current. No battery or other power source is needed, since it is the energy of the detected light itself that is used to power the light meter. Many of these selenium light meters—most are now over 50 years old—still work today.

And so it might seem possible to simply place a sheet of selenium at the focus of a camera, and then measure the electric charge that builds up at each location on the sheet. The problem is that there is no easy way to do this. It is a simple task to measure the *total amount of charge accumulated on the whole piece of selenium*, but there is no easy way to tell how much of the charge came from which location. Thus, we have an excellent light meter, but not a detector suitable for recording *images*.

It turns out that there *is* a rather roundabout way to record an image with a selenium detector. It works like this:

1. Focus an image of a negative onto a sheet of selenium, so different parts will build different amounts of negative electric charge.
2. Dust the sheet of selenium with a very fine black powder (called *toner*) that will stick to the metal everywhere there is a build up of electric charge. If done just right, wherever the charge is greater, more toner will stick.

doi:10.1088/2053-2571/aaf0aech6

3. Use a high voltage (about 1000 V) to give a sheet of paper a *positive* electric charge.
4. Roll the paper onto the sheet of selenium, so the negatively charged toner sticks to the positively charged paper (opposites attract).
5. Heat the paper up so the toner bakes onto it.

This clearly sounds rather inconvenient for a detector in a camera. And furthermore, although the detector uses a photoelectric effect, there is nothing digital about this process. You may have guessed already that what I have just described is not a detector in a camera (let alone a digital detector), but rather a photocopier. But this illustrates the basic problem. If different parts of the detector build up an electrical signal, how do we *separately* measure those signals. The next section describes a practical way to solve this problem.

6.1 CCD and CMOS array detectors

The most common type of detector in a digital camera is either a *CCD* or *CMOS array*. CCD and CMOS are acronyms for, respectively, *charge-coupled device* and *complementary metal–oxide–semiconductor*. Fortunately, we can arrive at a basic understanding of what these detectors do without diving headfirst into the pit of technical details regarding the precise meaning of these terms.

The word ‘array’ means that the detector consists of many tiny individual detectors that are arranged in an evenly spaced rectangular pattern. And so, unlike the selenium sheet used in a photocopier, or the random pattern of silver halide grains in a silver gelatin emulsion, a CCD or CMOS array is actually many thousands (or millions) of individual detectors *arranged in a regular grid of rows and columns.*

The two types of detector arrays, CCD and CMOS, work in slightly different ways, but the overall effect is much the same. Photons are absorbed at the location of a particular light-sensitive site (called a pixel), and this produces an excess negative electric charge there. This negative charge is then measured, and the result converted to a digital number that represents the brightness of the light at that particular pixel.

Before considering how a CCD or CMOS array detects light, let us first imagine a simple detector array that uses selenium. We could simply take our sheet of selenium, and cut it into 100 × 100 tiny squares. Then we could put our selenium sheet back together by laying all the little pieces side by side in a grid pattern. If we were to attach a wire to each, and put insulating strips between the tiny squares so they couldn’t touch each other, we could then simultaneously (but separately) measure the signal coming from each part of our array of little tiny bits of selenium.

This could work, in theory, but consider that $100 \times 100 = 10\,000$; that is a lot of wires. And this would be a rather unimpressive digital detector, to put it mildly; *a modern CCD array contains at least several hundred times that many individual detectors*. And so we need a better way.

Instead of measuring the electric charge directly and separately from each pixel, a CCD *transfers* the charge from pixel to pixel, in a sort of ‘bucket brigade’ fashion.

Table 6.1. A simplified example of charge transfer in a CCD array, showing the nine steps in reading out the values in this imaginary nine pixel CCD. The nine pixel values of the exposed image are represented in the square at the upper left. The square to its right represents the first step in the charge transfer process; successive steps are represented by the other squares in order from left to right, and top to bottom. Each measured pixel comes out, one at a time, at the lower right of the CCD. The bottom row shifts right, and then the second from the bottom row shifts right and then down, followed by the top row. The pixels values thus come out one at a time in the order 5, 6, 2, 9, 5, 1, 4, 1, 3. A computer is needed to reconstruct the two-dimensional image from this one-dimensional string of pixel values.

3	1	4	3	1	4	3	1	4
1	5	9	1	5	9	1	5	9
2	6	5		2	6			2
3	1	4	3	1	4		3	1
	1	5			1			4
		9			5			1
		3						
		1			3			
		4			1			3

The electric charge is passed from one pixel to the next along rows, and then down the last column, one row at a time. Thus, the array is *read out*, one pixel at a time. As the tiny bits of electric charge, from each pixel, leave a CCD array, they are amplified to make a measurable voltage. It is a simple task for a computer to determine which measurement corresponds to what particular location on the array, and thus a picture can be reconstructed.

And so after the exposure is finished, an enormous number of pixels can be measured separately, one at a time, without the need for an equal number of individual electrical contacts. To illustrate this idea, consider a simplified CCD array of only 3 × 3 pixels. Table 6.1 illustrates an example of the different steps in the charge transfer process.

Real CCD arrays may take different approaches to this overall theme. Some transfer the values only along rows to a blank region that is kept in the dark. And thus the light-sensitive part of the CCD array can be used again while the previous image is being read out pixel-by-pixel from the covered area. Other CCD arrays are read out in sections, rather than the entire chip at once.

CMOS arrays differ in that they are read out one row at a time; one row is read out while the other rows are being exposed. And so different parts of the array are, in effect, exposed at different times. This can lead to odd effects for rapidly moving subjects, but it makes CMOS arrays more practical than CCD arrays for recording video. For this reason, CMOS arrays are increasingly common in digital cameras that double as a video camera.

6.2 The physics of CCD arrays

The light-sensitive sites on a CCD array are capacitors—individual microscopic spots on a silicon chip that can store electric charge in the form of electrons. When a bias voltage is applied, they become sensitive to light. The energy from absorbed photons raises electrons to higher energy levels, which allows them to move freely, and so gather at the capacitor site. The more photons that arrive at that part of the CCD array, the more electrons accumulate. By strategically altering the voltages applied to the different capacitor pixel sites, they can be induced to pass their charge to a neighboring site in order to read out the image.

The CCD is made from layers of silicon-based semiconductor materials that have been strategically tainted (*doped* is the technical term) with impurities. In addition, there are tiny pieces of conducting metal interlaced throughout the structure of the CCD chip to form microscopic wires and electrodes.

A semiconductor, as its name implies, has properties midway between those of conductors and insulators. While the positively charged nuclei of atoms remain intact, the negatively charged electrons can sometimes transfer from one atom to another. In a good conductor—a copper wire is an example—some of the electrons do not belong to any particular atom (one each per atom, in the case of copper). They can move around freely on the surface of the wire, randomly colliding with each other much like the atoms or molecules in a gas. And so these *conduction electrons* respond easily to external electric fields, and thus conductors are very good at easily transferring electric charge from one place to another. An insulator, on the other hand, is the opposite of a conductor; it is difficult to transport electrons, which remain stuck to their individual atoms.

If a steady current of electric charge is desired, then a good conductor is an excellent tool for the job. But a photoelectric light detector depends on the build up of *static* charge. There needs to be some way for that charge to move, and so gather at the light-sensitive site. But on the other hand, we want it to stay in place until the exposure is finished. For a good conductor, any static charge rearranges itself almost instantly until all of the electrical forces are balanced. And so if electrons are photoelectrically removed from their atoms on a conductor, to gather at one part of the image, they will not stay there. An insulator has the opposite problem; there is no easy way to make the charge move in order to measure it. A semiconductor provides the right balance; the charge can be better controlled, allowed to build up in one place during the exposure, but then encouraged to move in order to measure it after the exposure is completed.

But there is more to it than that, as semiconductors have many unique and complex properties. In particular, when two different semiconductor materials are placed in contact with each other, the door is opened to many unusual phenomena. Not only can some of the electrons move from atom to atom, but the empty spaces they leave behind begin to have many of the properties of particles. And so we have negatively charged electrons, and positively charged *holes* that interact with each other in complex ways that can be fully described only by the strange rules of quantum physics.

A photoelectric detector such as a CCD or CMOS array is, in a sense, a complex machine. This structure of different types of semiconductor and metal conductor is built up to form a patterned flat surface of several layers—like overlaying floor plans of a multi-story building. This overall flat, but still three-dimensional, design is engineered to produce an array of light-sensitive capacitor sites (the pixels) that produce a signal from the incident photons. But it also includes the semiconductor structures needed to receive input signals and to transfer a charge and amplify the signal as the array is read out.

The manufacture of CCD and CMOS arrays has in many ways overtaken the traditional photochemical processes described in chapters 2 and 3. And so it is perhaps ironic that photochemical photographic techniques are a crucial part of the manufacture of such digital detectors. The elaborate microscopic structures in the silicon chip of a CCD or CMOS array are vacuum deposited, one layer at a time. But the detailed structure of each layer is realized with *photolithography*. A light-sensitive coating is applied to the silicon wafer, a patterned mask is overlaid and it is exposed to light and then processed. The light alters the coating so that it can then resist chemical etching, or the vacuum deposition of additional layers. And so etching can dig out tiny patterns in the surface, or additional patterned layers can be added. One can compare this most high-tech and modern of industrial processes to the antiquarian gum bichromate process described in chapter 3.

6.3 Color digital detectors

The commonly used CCD and CMOS array digital detectors cannot, by themselves, distinguish between different colors of light. Thus they are, at their most basic, black-and-white detectors. But although the individual detectors cannot tell the difference between different wavelengths of light, they are sensitive to a broad range of wavelengths.

For most photographic purposes, CCD and CMOS arrays are sensitive to *too* broad of a wavelength range, since they are sensitive well into the near-infrared part of the electromagnetic spectrum. Typical camera lenses do not focus well at these wavelengths, and even if they did, their focus point is different for infrared than it is for visible light. Thus, if the full range of wavelengths were used, the picture would be blurry.

This problem is usually solved by placing a special filter directly over the detector, a filter that blocks infrared light while passing visible light (see figure 6.1). Equipped with this filter, a typical CCD array passes light of wavelengths stretching across the visible part of the spectrum.

Although the light-sensitive pixels in a CCD array are sensitive to a wide range of wavelengths, a given pixel has no way of sorting out those wavelengths; it has no way of ‘knowing’ after the fact what combinations of wavelength were detected. But a color image can still be produced from such an inherently black-and-white detector—simply use colored filters to record *three separate images*, through red, green, and blue filters. Digital processing can then be used to combine these three images into a single color image. See figure 6.2 for an example of the basic idea using black-and-white film. I took separate pictures, one after the other, each with a different colored filter placed over the lens of the camera.

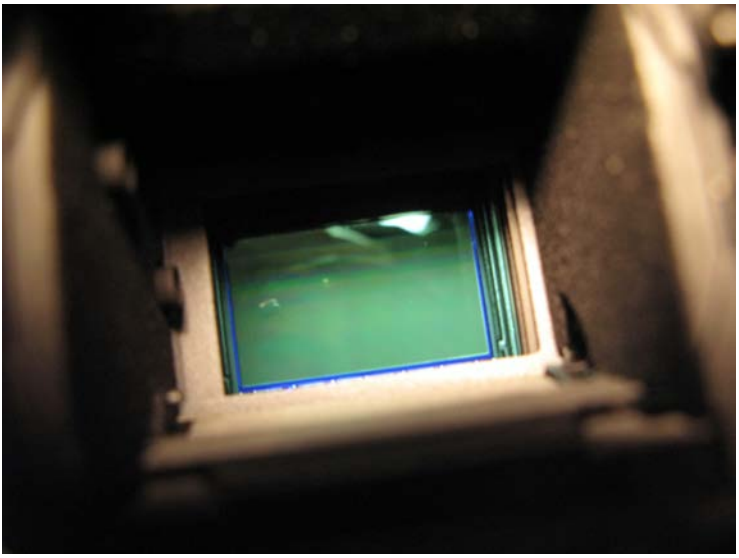


Figure 6.1. A CCD array detector from a digital SLR camera. This particular CCD array has 3000 × 2000 pixels on a silicon chip with dimensions of 23.7 mm × 15.6 mm. The individual pixels are far too small to see at this scale, and the green color is from a glass filter covering the sensor. The filter not only protects the delicate sensor, it also blocks near-infrared light, to which the sensor is particularly sensitive.

Figure 6.2. *Geranium*—John Beaver, 2014. Three photographs were taken with black-and-white film, each through a different colored filter (red, green, and blue). The three black-and-white negatives were scanned, and the resulting three digital images combined to form a single color image.

Clearly then, one could use any black-and-white detector, including a CCD or CMOS array, to perform this same trick. But of course this is not usually very practical; one must use a tripod to keep the camera in exactly the same position between exposures. And what if the subject moves?

For a CCD or CMOS array we can accomplish the same end result *by placing the colored filters on the detector itself, over the individual pixels*. The trick is that some pixels are covered by microscopic red filters, while others by blue, and others by green. If we arrange these *microfilters* in a regular pattern, groups of pixels can be combined to form a color image.

In practice, this is accomplished with what is called a *Bayer filter* or *Bayer mask*. The typical arrangement can be seen in figure 6.3. What is called a single 'pixel' in a color image from a CCD includes information from the measurements of a square of *four* neighboring pixels on a black-and-white CCD covered with a Bayer mask. Two of those pixels were located behind green parts of the Bayer mask, while the other two pixels were behind, respectively, red and blue sections of the mask.

One could simply combine these four measurements to produce one color pixel in the final image. But this would mean that our 3000 × 2000 pixel CCD array in figure 6.1 would produce a picture with only 1500 × 1000 pixels. And so instead, sophisticated computer *demosaicing* routines are employed to better take advantage of the information available in the image. This is usually performed by the micro-computer circuitry built in to the camera, immediately after the picture is taken. But many cameras also have the ability to store image data in an unprocessed, so-called *RAW* format, allowing the user to employ their own chosen demosaicing routine once the image has been downloaded to a computer.

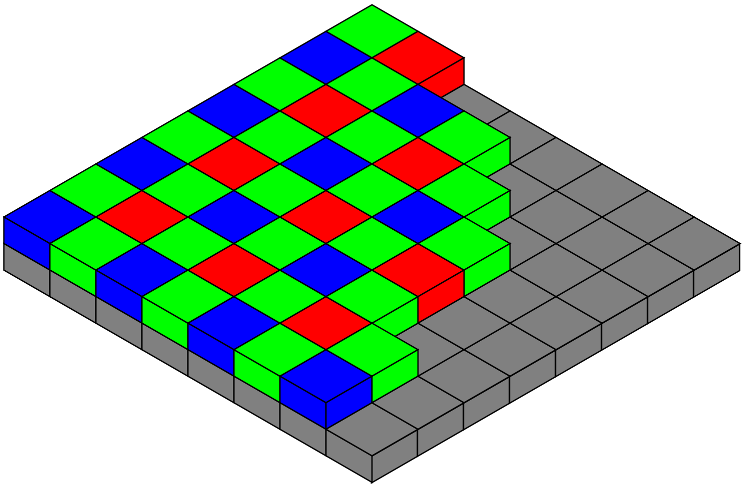


Figure 6.3. An illustration of the layout of a Bayer mask array of colored microfilters covering the light-sensitive pixels (represented in gray) of a CCD array detector. Graphic credit: Cburnett—Own work CC BY-SA 3.0.

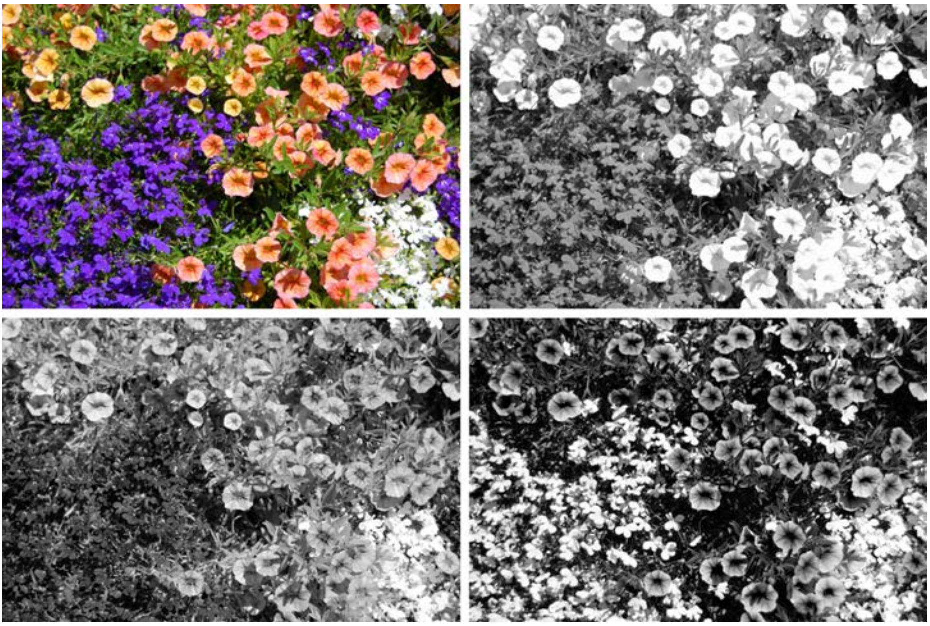


Figure 6.4. A color photograph from a digital camera is shown on the upper left, along with the three separate black-and-white images the camera took in order to make the color picture. The red-filtered image is on the upper right, the green-filtered image on the lower left, and the blue-filtered image on the lower right. One can easily see that, for example, the blue flowers appear brightest in the blue-filtered image, while the red flowers appear brightest in the red-filtered image.

Figure 6.4 shows a color photograph from a digital camera, along with three separate black-and-white images that were used to make it. It is evident that, for example, the blue flowers are much brighter on the image formed from the blue-filtered pixels.

IOP Concise Physics

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Chapter 7

Unusual detectors and 3D photography

7.1 Stereo photography

All of the detectors we have considered measure, in some way, the *brightness* of the light falling on each position of the detector. By using multiple detectors combined with colored filters, color detectors also record information about the *wavelength* of the incident light. But there is more to light than just wavelength and brightness.

Our brains use many cues in order to perceive depth. Some examples are atmospheric perspective, converging lines and the overlap of objects. But for most of us, there is another way our brain perceives depth that has no counterpart in a single, flat image. We have *binocular vision*—two eyes spaced roughly 2 inches apart that see the world from two slightly different vantage points, and our brains can combine this information to produce a sensation of a three-dimensional world (Kalloniatis and Luu 2018).

This ability is called *steropsis*, and it has been recognized since the earliest days of photography that one can trick the brain into producing the same effect simply by taking two pictures simultaneously of the same scene, but from two slightly different places. If these two photographs are then placed in a special viewing device, such that each eye sees only one of the two images, then a perception of depth can be achieved that is similar to that of viewing the real scene in-person.

This is called *stereo photography*, and its popularity has waxed and waned in cycles over the past 150 years. Special cameras have been made that use dual lenses and shutters to record both images onto different parts of the same roll of film. Some of these, such as the *Stereo Realist* made in Milwaukee, Wisconsin, were moderately popular, and working models can still be found on the used market. This particular camera was meant to use 35 mm slide film. The positive transparencies were then mounted side-by-side in special cardboard mounts designed to fit in a special two-lens viewer (see figure 7.1).

Older versions usually used specially made prints that were then mounted so as to be viewed with a device called a *stereoscope*. Figure 7.2 shows as an example my ugly-but-working version cobbled together with hobby plywood and a pair of cheap

doi:10.1088/2053-2571/aaf0aech7

Figure 7.1. A small viewer for mounted pairs of half-frame 35 mm stereo transparencies.

Figure 7.2. **Left:** A Speed Graphic camera for 4 × 5 inch film, with a dual lens and shutter mounted for stereo photography. The lens/shutter assembly is modified from an old passport camera. **Center:** My ugly, but functional, stereoscope for viewing the stereo image. **Right:** An example of a stereo pair photographed with the camera.

reading glasses, along with an example of a stereo photograph. The two pictures were taken simultaneously onto one piece of film, using a double lens and shutter assembly on a large format camera. The lens and shutter assembly were modified from an old passport camera.

Some popular accounts of our binocular vision compare it to the astronomer's use of trigonometric parallax to measure the distances of stars, or the parallax rangefinder in an old camera (see figure 7.3). In both cases, there is a view of the same object from two different vantage points—just as with our binocular vision. When the nearer object is seen against a more distant background, the two views are in slightly different directions; measurement of this *parallax angle*, along with some simple trigonometry, can then be used to find the distance to the object.

Our eyes do indeed have to point inward—they must cross slightly—in order to converge on a nearby object, as compared to a distant one. This may be a factor in

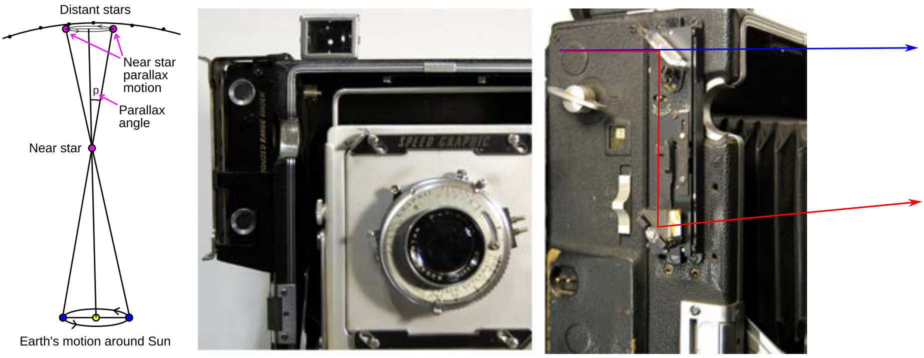


Figure 7.3. **Left:** Astronomers determine the distances to nearby stars by measuring the parallax that results from the Earth’s orbit around the Sun. Graphic: Public Domain, https://commons.wikimedia.org/w/index.php?curid=1903952 **Center:** The rangefinder on an old Speed Graphic camera uses a tilting prism to measure the parallax as seen through two different view windows simultaneously (seen on the left), to accurately set the focus of the camera. **Right:** The rangefinder assembly with the cover removed.

the brain’s processing, but the perception of depth that forms in the brain is clearly far more than simple mental trigonometry. Stereo photography is a demonstration of this mental complexity—the perception of depth in a stereo image can be quite striking, even though the eyes point in the same direction to view the two images.

7.2 Light-field photography

When light strikes a given place on an ordinary detector such as silver gelatin film or a digital CCD array, the detector only records the *brightness* of the light that hit that part of the detector. The use of a colored mask allows one to also record some wavelength information. But when a CCD or piece of film detects a photon, the detector cannot tell from which *direction* the photon came.

But digital detectors now exist that *can* tell not only the brightness of the light striking some part of the detector, but also how much of that light came from each part of the lens. This is done with a special mask, analogous to the Bayer mask that turns a CCD array into a color detector. The mask places tiny micro-lenses over assigned clusters of the detector’s light-sensitive sites.

The overall effect is that each pixel in the image is actually a tiny, very low-resolution photograph of the back side of the camera lens. And so each pixel contains information not only on how much total light arrived at that part of the detector, but it can sort out that light by which part of the lens it came from.

This information, combined with known properties of the lens in the *light-field camera*, can be used to reconstruct a three-dimensional model of the objects in the world from where the light rays arrived. It also means that a light-field camera can *focus after the picture was taken*. A disadvantage is that resolution is sacrificed in order to gain this additional depth information.

7.3 Autochrome Lumière process

It is interesting to note that the modern digital camera produces a color image in a way that is strikingly similar to one of the first successful methods of color photography. The *autochrome process*, invented over 100 years ago by Louis and Auguste Lumière, employed a single black-and-white emulsion on a glass plate, undercoated with a thin layer of tiny transparent globules of potato starch. The glass photographic plate was then exposed in a camera, through the back side of the glass. Thus, the image focused by the lens first passed through the layer of starch before reaching the light-sensitive emulsion.

The trick was that before being coated onto the glass plate, the starch was dyed in three separate batches of red, green, and blue and then mixed together. And so the starch globules acted like tiny colored microfilters during the exposure, very much like the Bayer mask of a digital camera. Only instead of a regular array, it was an essentially random arrangement of red, green, and blue starch filters.

After exposure, the photographic emulsion was then reversal processed, so as to form a positive transparency. When lit from behind, the black-and-white image is viewed through the same microfilter starch globules that were used to record the image, and so a color image appears. Surviving autochrome images are today highly prized because they are one-of-a kind direct positives, they have a remarkably subtle range of tones and hues, and they are some of the only directly recorded color images from the early 1900s.

7.4 Holography

At a superficial level, a hologram seems to be a type of photograph. A flat light-sensitive material is used to record light coming from objects. And when one looks at the processed hologram, an image of that object is visible. But there is an important sense in which a hologram is not a photograph at all. For if one looks closer and closer at a hologram, with greater and greater magnification, it becomes apparent that it contains only complex alternating patterns of light and dark. And this pattern at a given place on the surface of the hologram is unrelated to light and dark areas of the image.

A hologram contains no image on its surface at all; rather, the fine-scale pattern of light and dark is an *interference pattern*. When viewed in the proper way, the interference pattern reconstructs the three-dimensional structure of the light reflecting off the subject. A given part of the image—the eye of the subject, for example—is not located at any part of the hologram in particular. Instead, any given location on the hologram contains a complex interference pattern that carries information about the *entire* image, as seen from that particular location. The interference pattern at a different location on the hologram reconstructs the entire image as seen from that different vantage point. And so the hologram contains three-dimensional information about the subject.

This is possible because the process of making a hologram takes advantage of more than just the intensity of light coming from the subject. The hologram also uses the light wave's *phase*—the precise location in space, at a particular moment in time, of a given peak in the wave. In its simplest form, the coherent light of a laser—spread out by lenses

—is used to illuminate the subject. But part of the same beam is also split off with a partially-reflective beamsplitter, and sent directly to the film. So some of the laser light comes directly from the laser source, and some comes from reflections off the subject. These two waves interfere with each other—adding together if the wave peaks happen to match up, but canceling out if the peaks lie opposite each other. This produces a fine-scale interference pattern. When the hologram is illuminated by the same laser light, but without the object in place—the interference pattern produces a three-dimensional image of the subject.

Modern holograms are more sophisticated than the simple example I describe here, but they still involve some kind of geometrical arrangement between the light source, beam splitters and the photographic detector; a camera is not usually involved at all.

7.5 Lippmann process color photography

Gabriel Lippmann won the 1908 Nobel Prize in Physics for his invention of one of the earliest color photographic processes. In some ways, Lippmann photography is the only true color direct photography ever invented. Other methods—such as autochrome, the chromogenic process, and the Bayer mask of a contemporary digital camera—rely on a three-color process. Red, green, and blue colored filters sample three parts of the spectrum of the light. Colors are *synthesized*, taking advantage of the three-color nature of human color perception (see volume 2 of *The Physics and Art of Photography*). Lippmann photography was never made commercially viable, and it was overshadowed by the autochrome process patented about the same time. But it has been revived by at least a few contemporary photographers, notably Dr Hans I Bjelkhagen; his overview of the process (Bjelkhagen, 2018) provides much of the information below.

Lippmann photography uses no colored dyes or filters—the photograph is made with a black-and-white emulsion and consists of an extremely fine-scale *wave interference pattern* of only black and white. In this way it is like a hologram. But the hologram uses wave interference to record distance information while the Lippmann photograph uses interference to record wavelength information. When a Lippmann plate is viewed correctly, the same full spectrum of wavelengths is visible as in the original scene.

To accomplish this, Lippmann used a silver emulsion of extremely fine grain so the microscopic interference pattern could be recorded in detail. It was also necessary that the emulsion be *panchromatic*—sensitive to a broad range of wavelengths. One could not record the color red, for example, if the emulsion is not even sensitive to red wavelengths.

Wave interference is produced in a clever manner. The emulsion is coated onto a glass plate, to act as a stable, transparent base; a feature that was common for the black-and-white photography of the day. But the Lippmann plate is positioned in the camera such that the emulsion faces *away* from the camera lens. The emulsion is then backed by a mirror, so light passes through the microscopically-thin emulsion *twice*. The reflective backing of the emulsion must be in perfect contact, and Lippmann performed this feat by putting the glass plate in a special holder that allowed for a layer of shiny liquid mercury to contact the emulsion. The mirror-like backing of mercury allows the incoming and outgoing waves to interfere with each other, and produce either a dark or light band depending upon whether the two waves interfere constructively or destructively.

The key is that the condition for constructive or destructive interference depends not only on distance within the emulsion, but also on the wavelength of the light. Thus, color information is recorded as well. When the developed plate is viewed with the proper lighting, the same interference conditions are produced, and one sees the same wavelengths that caused the interference during the exposure.

Although both holography and Lippmann photography produce images in a black-and-white emulsion with wave interference, there is a fundamental difference. The interference pattern of alternating dark and light bands is on the two-dimensional surface of the hologram emulsion. In a Lippmann plate, however, the interference pattern is *within* the thin layer of the emulsion itself. At each point in the Lippmann plate image, dark and light bands of constructive and destructive interference alternate with each other in the tiny space from the front of the emulsion to its back. Where red light exposed the plate, these bands are farther apart, and they are closer together where blue light exposed the plate. When light reflects off the developed plate, the interference bands produce these same wavelengths in the reflected light.

And so the color is produced not by selective absorption and reflection by dyes or filters. Rather the color in the Lippmann plate arises from the wavelength-dependent constructive interference of a microscope regular pattern of physical obstructions. These are called *structural colors*, and it is how many blues and greens are formed in the animal kingdom—the intense, iridescent blue of the Morpho butterfly being one of the most famous examples (see figure 7.4). Structural colors are described in more detail in volume 2 of *The Physics and Art of Photography*.

Figure 7.4. The intense blue of the Morpho butterfly is a structural color. Photo credit: By Didier Descouens—Own work, CC BY-SA 4.0.

The interference that occurs in the making of a Lippmann plate is similar to what is known as *Bragg interference*. Crystalline materials consist of regularly spaced arrangements of atoms, and when illuminated by the light of a wavelength that is roughly similar to this spacing, an interference pattern appears. One can analyze this observed interference pattern in order to determine the atomic structure of the crystal. The relation between the spacing of the alternating layers that reflect light, d, and the wavelength of the light, λ, is given for Bragg interference by (ignoring angle factors):

$$d = \frac{\lambda}{2}. \tag{7.1}$$

For interference in a Lippmann plate, the same relationship holds, but for one difference; the wavelength of the light must be divided by the index of refraction, n, of the emulsion layer. As discussed in volume 1 of *The Physics and Art of Photography*, light slows down when it enters a transparent material, and n is the factor by which it does so. A consequence of this is that the wavelength of the light is decreased in the material by that same factor (as compared to air). And so for Lippmann photography, we have the following relation:

$$d = \frac{\lambda}{2n}. \tag{7.2}$$

Thus, we see that longer wavelengths produce a more widely-spaced interference pattern. If we choose the middle of the visible spectrum $\lambda = 550$ nm and an index of refraction of 1.4 for gelatin (choosing something between that of water and glass), equation (7.2) gives $d \approxeq 200$ nm. Thus, the photographic emulsion must have a grain fine enough to be able to record details down to significantly less than this size. That is very small indeed, and producing such an emulsion is one of the technical challenges of Lippmann photography.

The spacing between atoms in a crystal is much smaller than this, closer to only 0.1 nm. And so we can solve equation (7.1) to determine what wavelength would be needed to observe Bragg interference for a crystal. The answer is clearly about 0.2 nm, and this is in the *x-ray* part of the electromagnetic spectrum. And so *x-ray diffraction* is one of the principal tools by which scientists have probed the structure of crystals.

References

Bjelkhagen H I 2018 Lippmann colour photography [http://alternativephotography.com/]

Kalloniatis M and Luu C 2018 Depth perception *Web Vision* ed H Kolb, R Nelson, E Fernandez and B Jones ch VIII [https://webvision.med.utah.edu/]

Part II

Photography as an art and the meaning of digital

IOP Concise Physics

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Chapter 8

Comparison of digital and film techniques

It is usually considered an insult to say that a photograph has been 'Photoshopped,' implying that it isn't a *real* picture, but instead one that has been artificially produced. Although I often agree in specific cases, as a general statement this is certainly unfair.

All photographs are altered in some way after they are exposed; it is not clear what a pure and unaltered photograph would even be. There are always many necessary steps between exposure and a viewable image, even if they occur (as in a digital camera) automatically and in the blink of an eye.

At a more philosophical level, even the most purely representational photograph is always a picture, not a direct capture of some element of the real world. If one takes several pictures in quick succession of any moving subject, each may be quite different from the other (see figure 8.1), to the point of making very different statements about the subject. Yet the subject did not really change in any fundamental way in the one or two seconds between exposures. The photographer chooses to show us one of these and not the others. Why is that particular choice, out of many, the one that is 'real?' And so at one level digital (or darkroom) editing is not so different, in principle, from the editing one always does in deciding which picture to show and which to discard.

Figure 8.1. Photographs taken only seconds apart, if looked at in isolation, may make very different statements about the subject. So which one is 'real?'

doi:10.1088/2053-2571/aaf0aech8

But still, the use of digital technology often *is* different in many ways from pre-digital processes. In this chapter we try to untangle, at least a little, the old from the new to take stock of where photography is as an art in this modern digital age.

Many common digital manipulations have a direct correspondence to traditional darkroom techniques. Many tools in popular software packages for digital image processing are even named directly after darkroom processes. So here we step through some of the more common digital processing techniques found in, say Adobe Photoshop or GIMP, and compare them to traditional wet-chemistry darkroom techniques. Since GIMP is free and open-source, I will refer to it for my specific examples. If you know how to use either one of these software packages, it is fairly easy to learn the other.

I will first step through the process of making a black-and-white print in the darkroom, comparing the steps at each point to how one would accomplish the same tasks in GIMP. Then I will describe how the process is different for making a color print. Detailed descriptions of these basic operations are widely available, for example in London *et al* (2005, chapter 7, 11).

8.1 Borders and cropping

When making a print in the darkroom, the light-sensitive print paper is exposed to light from a negative, usually projected onto it by an *enlarger* (see figure 8.2). The enlarger consists of a projection head that contains a light source, a condenser lens, the negative, and a projection lens. This entire assembly can be moved up and down, closer or farther

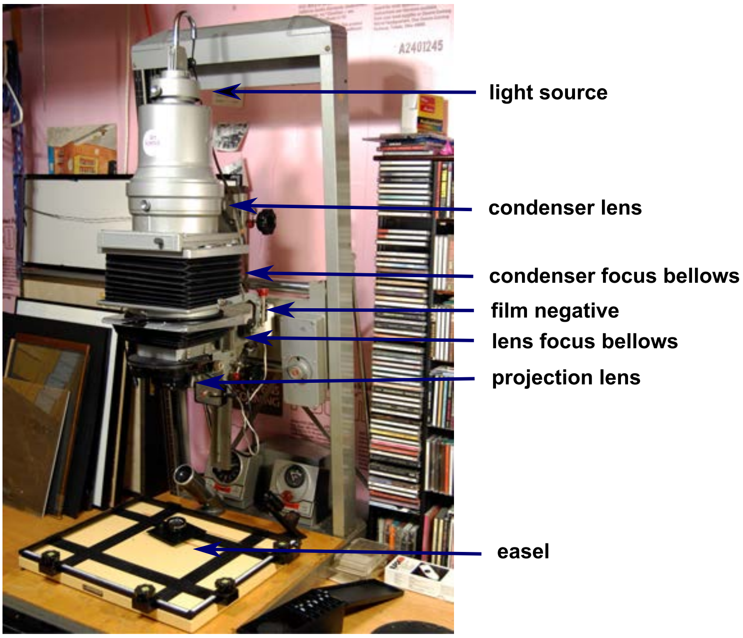


Figure 8.2. An enlarger for making black-and-white prints from negatives up to 4 × 5 inch in size.

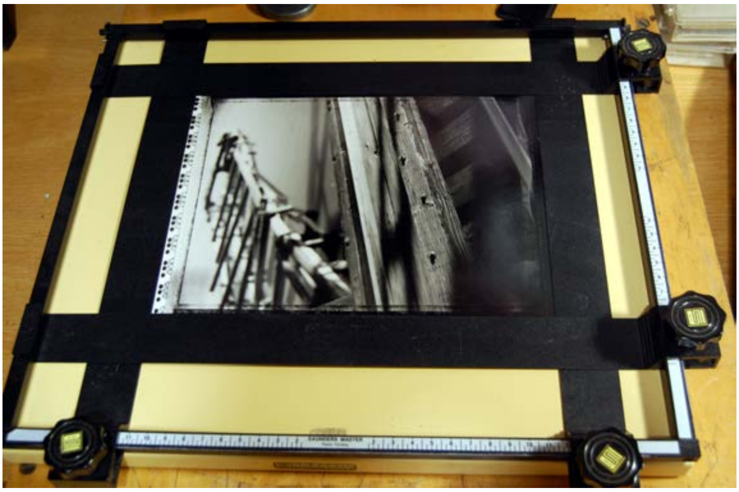


Figure 8.3. An easel is used to both hold the paper in place and to crop the borders as the paper is exposed to the image projected by the enlarger.

from the paper, to alter the size of the projection. The lens then moves independently in order to focus the image for the particular projection distance chosen.

The paper is placed in an *easel* to hold it in position. The easel typically has a white or yellow background, so one can see (and thus position and focus) the projection from the enlarger head *before* placing the light-sensitive paper. It also has a system of adjustable metal blades (see figure 8.3) to allow one to cover parts of the print paper so they will remain unexposed by the negative. Since enlarging paper is a negative process, it gets darker wherever it has been *most* exposed to light. Thus, areas of the paper covered by the easel blades will remain unexposed, leaving a white border around the print area. Black-and-white enlarging paper comes in standard sizes (in inches) of 5 × 7, 8 × 10, 11 × 14, and 16 × 20, and both the easel and enlarger can usually be adjusted to accommodate most of these sizes.

Sometimes an important part of composition is to *crop* the photograph by including only part of the original image in the final print. For example, a silhouetted tree on one side of an image can either be a vertical line cutting the image in two parts (if there is space to the right of the tree), or it can form one side of a framing element if the picture is cropped so the right edge of the picture passes through the middle of the tree.

To do this in the darkroom one first sets the easel to accommodate the size of the print paper, and to include whatever white border is desired. Then, with no enlarging paper in place, one adjusts the size of the image (by moving the enlarger head up or down) while repositioning the easel, until only the part of the picture one wants is included in the print area. The easel blades are black, so the image projected on them does not show up very well, and so one sees mostly only the part of the image that is to be left after cropping.

One of the appeals of digital image processing is that cropping is easy. GIMP, for example, includes simple tools designed expressly for that purpose. Furthermore, digital image processing saves previous steps and so allows one to instantly ‘undo’ an operation that did not go as planned. When cropping in the darkroom, on the other hand—there is no undo button; there is only 'start over and do it again.'

There is another fact that makes darkroom work, even for a relatively simple task such as cropping, more difficult than digital processing: the image projected by the enlarger is a *negative*, while with GIMP one works directly with the positive image. It is difficult to compose an image—or even to know if it is worth the bother of making a print—when one sees it only as a negative. For this and other reasons, photographers often first make a set of small positive *contact prints* that are the same size as the negatives.

8.2 Brightness and contrast adjustments

When one makes a print from a negative, each shade of gray on the negative is 'mapped' to a corresponding shade of gray on the print. Of course it is a negative process, and so a light shade on the negative becomes a dark shade on the print, and vice versa. But how does one tell precisely *which* shade from the negative ends up exactly *what* shade on the print?

For making a black-and-white print in the darkroom from a black-and-white negative, this question is usually answered in two parts. The overall lightness of the print is determined by controlling the exposure of the enlarging paper to the projected image. The projection lens in the darkroom enlarger has an *aperture adjustment* so that one may allow either more or less light to pass through the lens. But also, the enlarger lamp is always connected to some kind of timer (see figure 8.4)

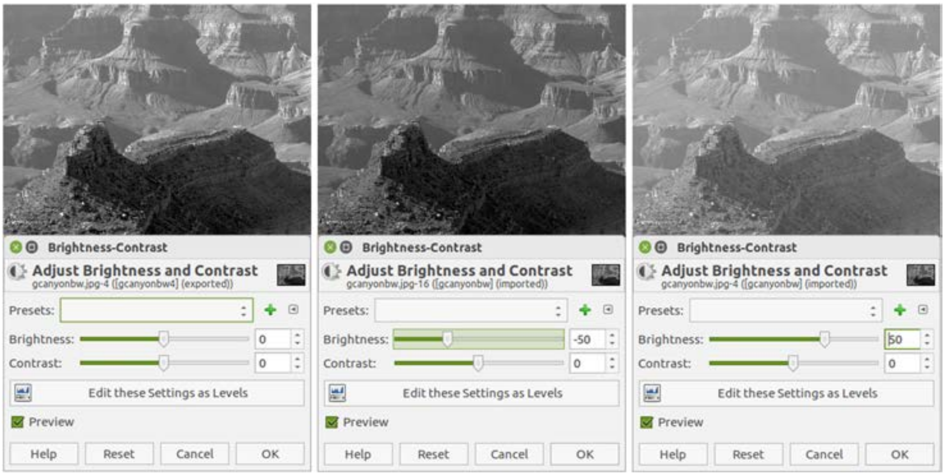
Figure 8.4. The overall lightness/darkness of a black-and-white print is controlled in the darkroom by adjusting the exposure from the image projected by the enlarger. A timer (rear left) allows one to vary the exposure time. Some enlargers also allow one to vary the brightness of the projection lamp (rear right). The two sighting devices in the foreground allow one to magnify the projected image in order to achieve a perfect focus.

that allows one to expose the print paper for a controlled length of time. Since the enlarging paper is a negative process, a longer exposure or a larger lens aperture results in an overall *darker* print.

So how does one determine exactly what exposure to use? The short answer is that one makes an initial guess, and then adjusts that guess by trial and error. There are ways to make this task easier, and experience helps a lot too. One common method is to cover all but a little strip of the print while the exposure is progressing, and every few seconds uncover a little more in steps. Thus, different parts of the print are exposed for different amounts of time, and one can then pick what looks best. A related method is to make a test print exposed through a *step wedge*—a sheet of thin plastic or film printed with different gradations of transparency. It can be pre-calibrated such that, for example, a 60 s test exposure shows, for each step, how many seconds of exposure would be needed to make the print have that same lightness.

But both of these methods apply different exposures to different parts of the print, and the final decision must be made in the context of the *whole* print. And so trial and error is still an important part. This takes time[1], since each print must be exposed in the enlarger and then run through about 5–10 min of chemical treatment in order to see the result. This is further complicated by the fact that as a wet print dries, it darkens slightly.

For a digital image on a computer this process is, of course, much easier since one can make adjustments interactively, seeing the result for the entire image as the adjustments are made. Photo editing software packages offer a choice of methods for adjusting the lightness/darkness of an image. In GIMP, For example, there is a simple *brightness–contrast* tool. Simply slide the *brightness* slider to the left with the mouse, and the picture becomes darker, corresponding to a longer exposure of the print paper in the enlarger. Slide to the right and the opposite happens; see figure 8.5.


Adjust Brightness and Contrast
gcanyonbw.jpg-4 ([gcanyonbw4] (exported))
Brightness: 0
Contrast: 0
☑ Preview

Adjust Brightness and Contrast
gcanyonbw.jpg-16 ([gcanyonbw] (imported))
Brightness: -50
Contrast: 0
☑ Preview

Adjust Brightness and Contrast
gcanyonbw.jpg-4 ([gcanyonbw] (imported))
Brightness: 50
Contrast: 0
☑ Preview

Figure 8.5. The effect of adjusting *brightness* in GIMP.

---

[1] And costs money—enlarging paper costs roughly $1/sheet.

There is a sense, however, in which the process in the darkroom is *more* straightforward than the process of adjusting a digital image. For the end result of all that work in the darkroom is a *print*, which is in fact the final goal of all that messing around. But when one interactively adjusts a digital image with GIMP or Photoshop, one is really adjusting *how the image appears on the particular computer screen one happens to be using at that moment*. It will likely appear somewhat different on another digital device. And making a digital print from that digital image is yet another step, the precise results of which depend on the make of printer, and the particular inks and paper that it uses. It is a complex procedure to guarantee that the print produced will look just like the image on the photo editor's computer screen. But that said, modern printers and computer screens are calibrated well enough that it is easy for a not-very-picky-person to produce a print that is at least a fair approximation of the image one adjusted with GIMP while looking at the computer screen. Technical details regarding computer monitor and printer calibration systems are available, for those who need them.

8.2.1 Digital contrast adjustments

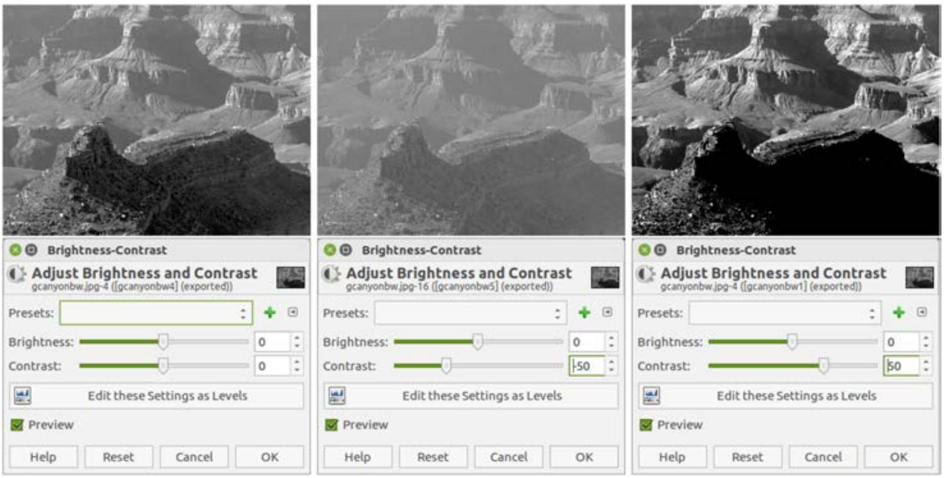
The fact that the 'brightness–contrast' adjustment dialog in GIMP contains an additional slider labeled *contrast* drives home the point that brightness, corresponding to overall exposure in the darkroom, is not the only issue. If some parts of the print are too dark, one can, if it is a digital image, slide the brightness slider to the right to make that part brighter. Or in the darkroom, one can choose a shorter exposure time, and that too will make that part of the print brighter. *But doing this will also make every other part of the print brighter as well.* What if some *other* parts of the print are now *too* bright?

It is at the very least a two-step process to make a connection between a particular shade of gray on the negative and a corresponding shade of gray on the print. And exactly the same is true when one wants to relate the numerical information contained in a digital image to shades of gray on the computer screen; it requires at the very least, two separate adjustments. And so the adjustment of *brightness* has a complement, called *contrast*.

A contrast adjustment alters the relationship between the total range of brightness levels on the original and the range of shades of gray that are displayed on the adjusted image. When one increases contrast in a digital image by, for example, sliding the 'contrast' slider to the right in GIMP, a given range of brightness levels in the unadjusted image is spread out to a *larger* range of brightnesses on the adjusted image. When lowering the contrast, the opposite happens; a given range of brightnesses on the unadjusted image is compressed to a *smaller* range of brightnesses on the adjusted image. See figure 8.6 for examples.

8.2.2 Contrast adjustments in the darkroom

But what about the darkroom? The print paper can be exposed either more or less. Greater exposure makes *all* of the printed values darker (except the ones that are already maximum black), while lesser exposure makes *all* of the printed values



Brightness: 0
Contrast: 0

Brightness: 0
Contrast: -50

Brightness: 0
Contrast: 50

Figure 8.6. The effect of adjusting *contrast* in GIMP.

lighter (except those that are already pure white). It might seem then that it is impossible to adjust contrast at all, But in practice there are several methods for adjusting the contrast of a black-and-white print:

Preflash exposure
The contrast of a print can sometimes be *lowered* by flashing the print paper with diffuse light before (or after) exposing it with the projected negative. Thus, one can give the paper two exposures, one with no negative in the enlarger, and one with the negative in place. The first exposure is of diffuse light that darkens all of the paper evenly. The second exposure produces the image on top of this even shade of gray. This method only works for *decreasing* contrast, and it has the (possibly undesirable) effect that no part of the print will appear completely white.

Chemical toning
Passing the processed print through a chemical *toning* bath, may alter the contrast of a print (London *et al* 2005, pp 156–7). One of the most common types is *selenium toning*, and it slightly increases contrast, while making subtle changes to the hue of the gray tones. Selenium toning, when properly done, also increases the life of a silver gelatin print, and so it is often considered a necessary step in producing *archival prints* guaranteed to last at least 100 years.

Adjusting the negative
One can instead adjust the contrast of the *negative*, by making careful choices during exposure and development. That is to say, by carefully choosing the exposure when the picture is taken, combined with how the film is developed, one can intentionally produce a negative of either high or low contrast. But of course one is then stuck with whatever choices were made while exposing and developing the negative.

Choosing a different negative

It may sound silly, but one can simply choose a different negative to print. To illustrate with a personal note, I have by now accumulated several times more film negatives than I will ever have time to print in three of my lifetimes—and surely I am not the only photographer my age for which this is true. Choosing which negative to print is always the first big step. Since, all else being equal, a higher-contrast negative will produce a higher-contrast print, I can simply keep that fact in mind while making the difficult choice of which negative to print and which to shun.

Graded enlarging papers

Graded photographic print papers are available, to produce final prints of more or less contrast from the same negative. These are assigned numerical values of 0–5 indicating their comparative level of contrast. A grade 0 paper produces a relatively low-contrast print, while a grade 5 paper produces a print of much higher contrast when exposed with the same negative. To use graded papers effectively one must, of course, have a selection of different grades on hand to choose from, and it can be difficult to find a source for grades 0, 1, and 5 since they are not used so much. But the very best quality print papers are usually available in graded form, and this is in fact the darkroom contrast-control method of choice for many discerning photographers.

Variable-contrast enlarging papers

Finally, one can use *variable-contrast print paper*. This is called, varying by different manufacturers, *VC* (Oriental), *polycontrast* (Kodak) or *MG* (Ilford). Variable-contrast black-and-white papers have two different photographic emulsions—one of high contrast and the other of low contrast—mixed together on the same paper. The trick is that although both emulsions are sensitive to blue light, the sensitivity of each emulsion to *green* light is different. And so colored filters can be used to change how much green light, compared to the amount of blue light, is projected onto the print paper. If the color is chosen carefully, one can control how much of each emulsion is exposed relative to the other, and this has a profound effect on the overall contrast of the print. Figure 8.7 shows a box of variable-contrast paper and a set of filters to go with it, along with a selection of graded papers.

And so with variable-contrast paper one can achieve prints of different contrast from the same paper. But to make use of it one must have a special set of filters designed to go with the paper. These usually come in a kit, and they fit in a holder either above or below the lens of the enlarger. The filters are labeled with numbers just like those for graded paper, but often with a greater range of contrast values than is now available for graded papers. Although many photographers prefer their favorite brand of graded paper, variable-contrast paper (with the one-time investment of the proper set of filters) is so convenient and easily available that it is the most common method for controlling the contrast of a black-and-white print in the darkroom.

One complication of variable-contrast paper is that increasing the contrast, by using a higher-contrast filter, also has the effect of decreasing the overall sensitivity of the paper, and so a longer exposure is needed to compensate. This means one

Figure 8.7. **Left:** An assortment of enlarging papers for making black-and-white prints from negatives. The top box is variable-contrast paper, while the others are contrast graded papers (larger numbers are higher contrast). **Right:** A set of filters to control the contrast of prints made with the VC paper.

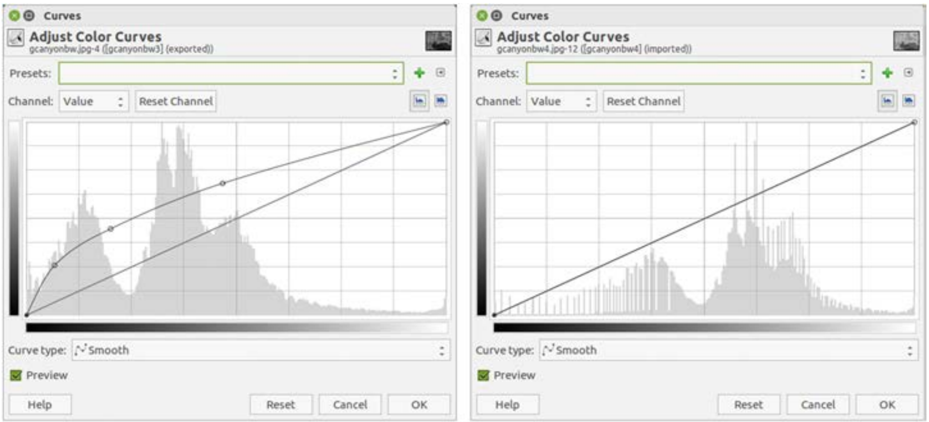
cannot simply adjust brightness and contrast of the print separately; adjusting one affects the other. This makes the trial-and-error process of getting the right exposure all that more complicated. There are, however, many tricks one can learn to make the process more systematic and less like a game of chance (see for example, *split-filter printing* in London *et al* 2005, p 151).

8.2.3 Levels and curves adjustments

The information contained within a digital image allows for far more sophisticated adjustments than simple brightness and contrast. A better approach is to directly control the *transformation* of the input gray levels in the unadjusted image to the output gray levels in the adjusted image. A graphical representation of this transformation is called an *adjustment curve*, and an example can be seen in the GIMP *curves* dialog shown in figure 8.8.

In the left-hand image, the different gray levels that correspond to the pixel values in the digital image—the *input*—are shown across the bottom of the graph. The ragged ‘mountain range’ in gray above is a *histogram* of those gray levels. Where the histogram is high, it means there are many pixels in the image that have that particular gray level. Where the histogram is low, there are relatively few pixels in the image that have that particular gray level.

There is another gray scale that progresses vertically on the left-hand side of the image; these are the *output* gray levels, after adjustment. The curved line cutting across the histogram is the *adjustment curve*. Pick a particular gray level on the input gray scale along the bottom, follow it vertically until reaching the adjustment curve, and then from that point move horizontally to the left until reaching the output gray



Channel: Value
Curve type: Smooth
☑ Preview

Channel: Value
Curve type: Smooth
☑ Preview

Figure 8.8. A GIMP *curves* dialog both before and after adjustment.

scale. The gray level at that intersection is the result—the output—if this particular adjustment curve is chosen (by clicking 'OK').

The corner-to-corner straight line is there for comparison, and represents no adjustment at all; any particular input gray level corresponds to exactly the same gray level for the output. And so we can see that this particular adjustment curve makes the darkest gray levels in the image brighter on the output than on the input.

The adjustment curve has much in common with the characteristic curve of a silver gelatin emulsion, discussed in chapter 1. It too is a graph that relates an input (photographic exposure) to an output (the resulting density of the film). But there is an important difference between the two; the density of photographic film is defined to be how *dark* it is, whereas the adjustment curve increases with *brightness*.

The right-hand image in figure 8.8 shows the result of applying the adjustment curve shown on the left. Since the adjustment curve made the darkest pixels brighter, there are now fewer dark pixels. The bulk of the pixels were darker than the middle tone in the input image, but the adjusted histogram now shows them to be more evenly distributed. Clearly the adjusted image would, overall, look brighter than the input image.

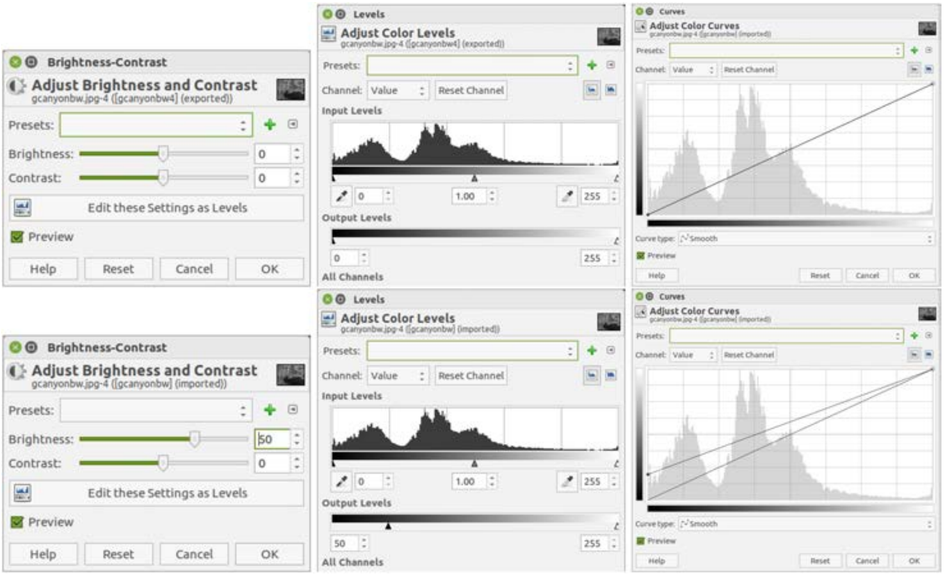
Consider another apparent feature of the output histogram shown on the right in figure 8.8. The shadow values now seem to come in discrete 'jumps.' That is, there are relatively many pixels in the image that have some of these gray values—but there are other gray values almost the same that have no pixels at all represented in the image. This seems to come in waves as we go along the gray scale from dark to light—the image has many pixels, no pixels, many pixels, no pixels, etc.

The reason for this is that relatively few input gray levels were transformed to a much larger range of output gray levels. A given digital image has only so many levels of gray—256 for this particular example. And there were even fewer of the dark gray levels that received the most adjustment. These few levels were 'stretched' to cover a much greater range on the output— and there were simply too few input levels to smoothly cover the much larger range of output levels. And so they appear in discrete steps instead. This may or may not be a problem. It can appear as sudden jumps in the brightness of the image, from one location to another, but if these jumps are in very dark regions of the image, they may not be noticeable. If they are, to make this adjustment successfully, one would have needed an original input image that had more than only 256 levels of gray. We further consider these issues of *bit depth* in chapter 9, section 9.4.

Let us now look at some examples of altering the adjustment curve of an image, and how this relates to the brightness–contrast adjustment in GIMP. In addition, we consider another dialog called *levels*. It represents an intermediate degree of control between brightness/contrast and curves. The levels dialog includes five adjustments and a histogram. Three of the adjustments are for the input gray levels, and two separate adjustments control the range of output gray levels. Input levels lower than the left-hand input slider are all transformed as the same (dark) gray level for the output. The right-hand input slider has a similar effect for the brightest levels. The range of output levels is set by two sliders on a separate gray scale for output. To illustrate the relation between these three ways of controlling the transformation between input and output gray levels, figure 8.9 shows both levels and curves for the adjustment of brightness portrayed in figure 8.5.

Figure 8.10 shows the examples of changing contrast from figure 8.6, in terms of both levels and curves. Notice that in all cases, if only the brightness and contrast is adjusted, the adjustment curve remains a straight line. Also note that the *contrast is related to the slope of the adjustment curve*. A slope steeper than the no-adjustment reference line (which has a slope of exactly one) increases the contrast, while a slope that is less than unity decreases the contrast.

The center slider of the levels dialog changes the *curvature* of the adjustment curve. Moving the slider to the right results in a concave-upward adjustment curve, while



Brightness-Contrast — Adjust Brightness and Contrast — gcanyonbw.jpg-4 ([gcanyonbw4] (exported))
Brightness: 0
Contrast: 0

Levels — Adjust Color Levels — gcanyonbw.jpg-4 ([gcanyonbw4] (exported))
Channel: Value
Input Levels: 0 1.00 255
Output Levels: 0 255

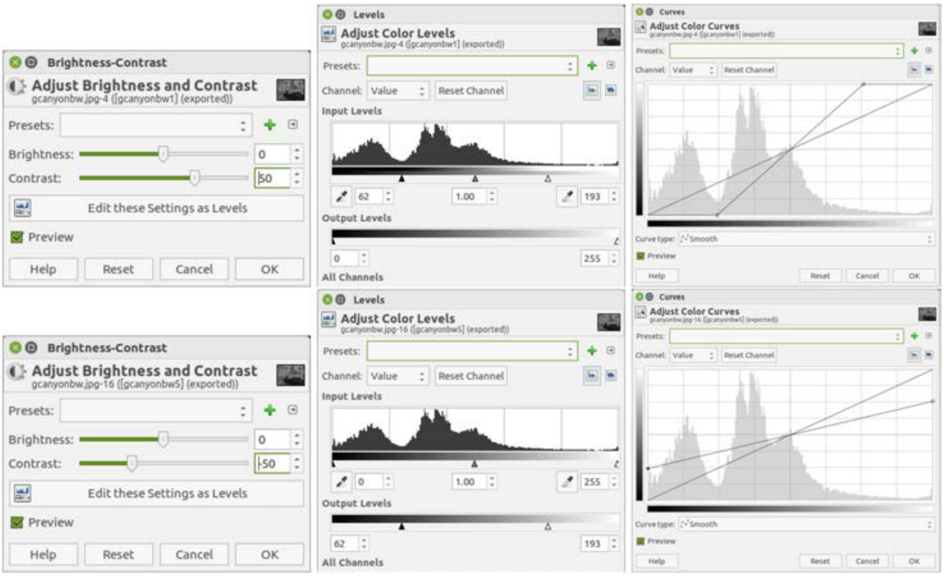
Curves — Adjust Color Curves
Channel: Value
Curve type: Smooth

Brightness-Contrast — Adjust Brightness and Contrast — gcanyonbw.jpg-4 ([gcanyonbw] (imported))
Brightness: 50
Contrast: 0

Levels — Adjust Color Levels — gcanyonbw.jpg-4 ([gcanyonbw] (imported))
Channel: Value
Input Levels: 0 1.00 255
Output Levels: 50 255

Curves — Adjust Color Curves — gcanyonbw.jpg-4 ([gcanyonbw] (imported))
Channel: Value
Curve type: Smooth

Figure 8.9. Adjusting brightness in GIMP with the brightness, levels and curves interfaces. The top dialogs are for the unadjusted image—the left-hand image in figure 8.5. The bottom dialogs show the adjustments that result in the center image.



Brightness-Contrast — Adjust Brightness and Contrast — gcanyonbw.jpg-4 ([gcanyonbw1] (exported))
Brightness: 0
Contrast: 50
Preview ☑

Levels — Adjust Color Levels — gcanyonbw.jpg-4 ([gcanyonbw1] (exported))
Channel: Value
Input Levels: 62, 1.00, 193
Output Levels: 0, 255

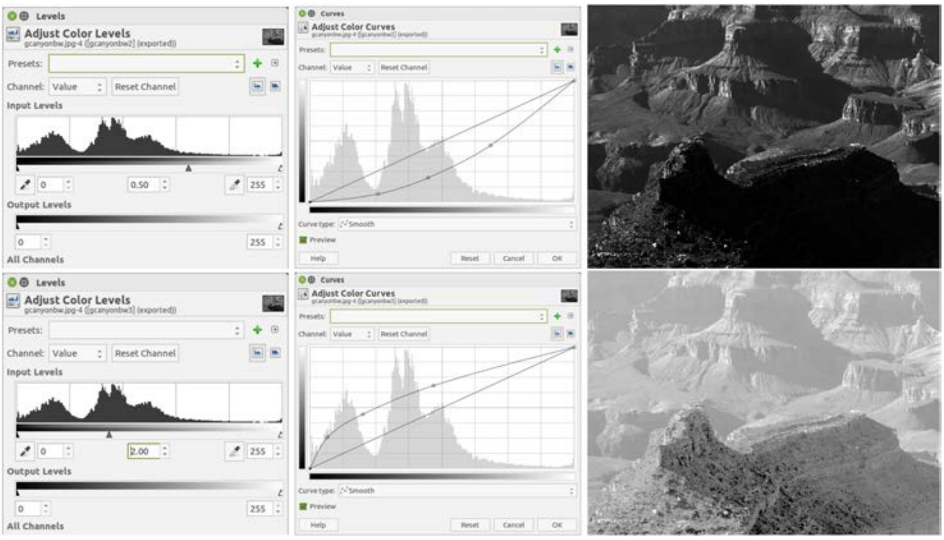
Curves — Adjust Color Curves
Channel: Value
Curve type: Smooth
Preview ☑

Brightness-Contrast — Adjust Brightness and Contrast — gcanyonbw.jpg-16 ([gcanyonbw5] (exported))
Brightness: 0
Contrast: -50
Preview ☑

Levels — Adjust Color Levels — gcanyonbw.jpg-16 ([gcanyonbw5] (exported))
Channel: Value
Input Levels: 0, 1.00, 255
Output Levels: 62, 193

Curves — Adjust Color Curves
Channel: Value
Curve type: Smooth
Preview ☑

Figure 8.10. Brightness/contrast and equivalent levels and curves adjustments of the left-hand image in figure 8.6 in order to produce the center and right-hand images.



Levels — Adjust Color Levels — gcanyonbw.jpg-4 ([gcanyonbw2] (exported))
Channel: Value
Input Levels: 0, 0.50, 255
Output Levels: 0, 255

Curves — Adjust Color Curves
Channel: Value
Curve type: Smooth
Preview ☑

Levels — Adjust Color Levels — gcanyonbw.jpg-4 ([gcanyonbw3] (exported))
Channel: Value
Input Levels: 0, 2.00, 255
Output Levels: 0, 255

Curves — Adjust Color Curves
Channel: Value
Curve type: Smooth
Preview ☑

Figure 8.11. Adjusting the middle slider in the levels dialog of GIMP causes the curve to deviate from a straight line. There is no equivalent to this more complex adjustment using only the brightness/contrast sliders.

sliding it to the left makes the adjustment curve concave-downward. The effect of either is to overall darken or brighten the image. But both the darkest and brightest gray levels are unaffected in either case. See figure 8.11 for examples of these adjustments and the results, both having the left-hand image in figure 8.6 as input. Such an adjustment is

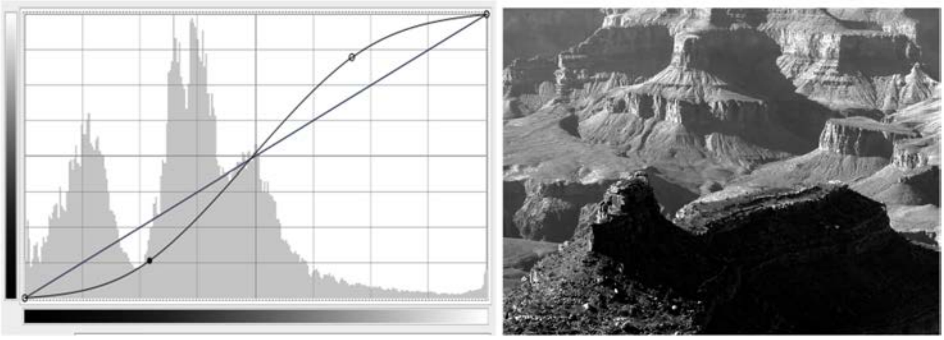


Figure 8.12. An s-shaped adjustment curve can be used to increase the contrast of the middle tones, while decreasing the contrast (and preserving) both the highlights and shadows.

sometimes called a *gamma* (γ) *adjustment*, because the shape of the adjustment curve on a linear scale is similar to that which alters the slope of the characteristic curve on a log–log plot. Notice the similarity between the adjustment curves in figure 8.11 and the power-law curves on the right side of figure 1.8 in chapter 1, section 1.6.

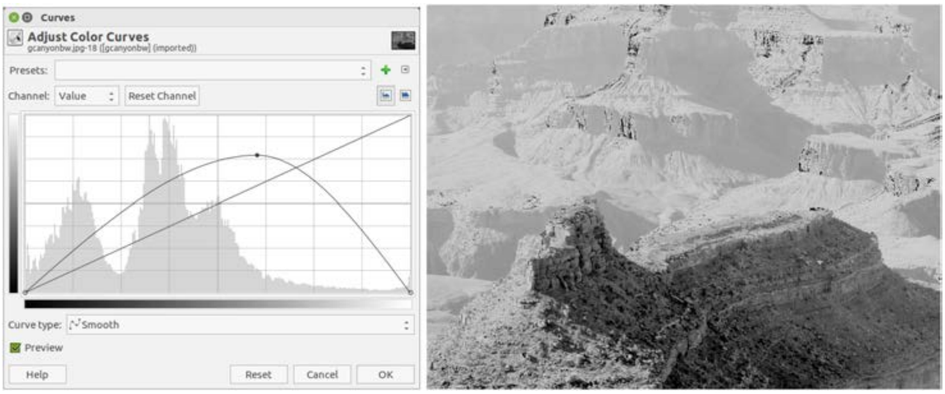
A very useful tool is to increase contrast with an s-shaped adjustment curve. At a given point in the histogram, the *slope* of the adjustment curve indicates the change in contrast. Where the adjustment curve is sloped greater than unity (the straight line, one-to-one pre-adjustment line), the contrast is increased, while the contrast is decreased wherever the adjustment curve has a slope less than one. The simple contrast adjustment tool increases or decreases the slope of the entire adjustment curve, which remains a straight line. When lowering the contrast, the result is that no part of the picture will appear either black or white. Increasing the contrast, on the other hand, has the effect that many highlight shades of gray are compressed to pure white, while many shadow shades of gray are compressed to pure black.

An s-shaped adjustment curve, on the other hand, increases (or decreases) the slope (and thus the contrast) for the mid-tones, but preserves the tonality in the highlights and shadows. See figure 8.12 for an example of using an s-curve to increase the contrast.

For all of the adjustment curves considered so far, each output level corresponds to *only one* input level. But that need not be the case. Figure 8.13 shows an adjustment curve for which there are *two different input levels* that produce a given output level. The upward-sloping part of this adjustment curve is similar to those already considered. But for the downward-sloping part, brighter input values result in *darker* output values. Thus, the brighter input values of the image have been inverted to a negative, while the darker values remained a positive. This is a digital version of the Sabatier effect, discussed in chapter 4.

8.2.4 Levels and curves in the darkroom: the zone system

The ability to interactively change the adjustment curve between the 'raw' digital image file and the displayed (or printed) image is one of the most important advantages of digital photography. Is there an equivalent to this with film photography? Well, of



Channel: Value

Curve type: Smooth

☑ Preview

Figure 8.13. This curve adjustment has two different input values for the same output value. This causes the highlights, where the curve slopes downward, to switch to a negative, producing an image much like the Sabatier effect (solarization).

course, one can always digitally scan a negative, and then use exactly the same type of level adjustments with the scanned digital image. Since digital scanners have been around longer than digital cameras have been in widespread use, photographers have been using these techniques from even before the time that digital cameras were practical.

But attempts to precisely control in a systematic way the shades of gray on a final photographic print long pre-date digital techniques. Probably the most important procedure is the *zone system*, formalized by Ansel Adams. The zone system relies on precise technical control at each of these steps:

1. The metering of the subject, ideally using a spot meter to measure the brightness at many different parts of the subject.
2. The exposure of the negative.
3. The development of the negative.
4. Measuring the transparency of the negative at different locations with a *densitometer*.
5. The exposure and contrast control of the print.

Each of these steps is linked to the others in order to determine a final print that has exact reproductions of light and dark areas according to the photographer’s plan. And so a photographer could, in theory, look at a scene and decide, before taking the picture, which areas they wanted to appear completely black, which areas completely white and which parts of the scene would appear the other shades of gray in between. See Stroebel *et al* (2000, pp 249–54) for an excellent description of precise tone reproduction with the zone system.

In practice, the zone system in its full incarnation requires an extraordinary amount of technical expertize. But apart from the mastery needed to accomplish an exact reproduction of tones according to a precise plan, one must have the artistic sense to know what is good and what is not. For both of these reasons, it is very special to see a print in the flesh by masters of printing such as Ansel Adams or Sally Mann.

8.3 Dodging and burning

An adjustment curve in GIMP ordinarily treats all of the pixels in the image equally. That is, every pixel with a particular brightness value gets changed by the adjustment curve in exactly the same way, no matter *where* it appears in the picture. But what if, for example, there is a part of the image that is overall too dark, while other parts are overall too light? What if one wants a *different* adjustment curve for different parts of the image?

Dodging and burning are two standard techniques from the traditional black-and-white darkroom that are also used in digital image manipulation. In the darkroom, we can consider this problem in light of exposure time. What if the proper overall exposure for the print is 60 s, but that 60 s exposure makes *one particular part* of the print too dark? One way to correct for this is to block the light to that area only, for part of the exposure. A small circle of cardboard held on a wire can be waved around over the area to be lightened in order to reduce the exposure (remember that this is a negative process, so reducing the exposure lightens the print).

The opposite of dodging is burning. To darken one particular area of the print, a second exposure is made—but using a cardboard mask with a hole that is waved over the area that requires additional darkening. Thus, the areas that project through the hole receive a greater exposure.

For both dodging and burning the mask is moved, typically in a circular motion of varying radius, in order to avoid sharp edges to the dodged or burned area of the print. And the mask is held at some distance above the print to make the shadow edges somewhat blurry. Doing this well requires practice, for one cannot see the result until the print is developed. And so every print that has been dodged or burned is slightly different, and this adds a hand-made element to darkroom prints.

8.3.1 Dodging and burning with GIMP

One can accomplish an effect very similar to dodging and burning in GIMP with digital tools of the same name. But the digital version has many advantages. In particular, one has instant feedback, and so the act of dodging and burning is much like drawing with the mouse. And since prior results can be stored in computer memory, it is possible to undo what one has done, and even to ‘back up’ part way and proceed from that point. This makes the digital versions of dodging and burning far easier than the corresponding darkroom methods, but there is a cost; the unique hand-made feature of a darkroom silver gelatin print is lost.

When dodging or burning in the darkroom one achieves different effects by choosing a mask of different size or shape, and by how close to the print one holds it. These choices are mimicked in GIMP by choosing a different *brush* and altering its size.

8.3.2 Spot healing and retouching

It is almost inevitable that a picture will have bad spots that need to be repaired. The most common example in the black-and-white darkroom printing is the white spots on a print produced by dust on the negative while the print was exposed. The process of removing mistakes such as this is called *retouching*, and it is a difficult process for a black-and-white print. Traditionally, one applies special dyes to the spot with a tiny paint brush, hoping to match the surrounding print area. To lighten an area is even more difficult; the spot on the print must first be bleached and then redyed to proper tone.

This can be a frustratingly difficult process, and it must be done by hand to each print. A mistake in the retouching process can easily ruin a print altogether. Apart from fixing mistakes such as dust specks, one might also want to alter the actual photographic subject, such as when a portrait photographer uses retouching to hide an unwanted skin blemish or to remove a feature that distracts from the composition. Highly-skilled techniques, using both water and oil-based paints, are needed to successfully retouch a traditional silver gelatin or color print in this way.

8.3.3 Digital retouching

Digital image editors come equipped with many tools that make retouching far easier than traditional brush work. For one thing, the ability to work on a copy of the image file, and to save work completed at any stage and to 'undo' previous steps means that one is much less likely to ruin many hours of work with a slip of the mouse. Below I list just a few of the more popular digital retouching tools (the examples are from GIMP).

- *Spot heal*—This tool is designed to easily cover over a small imperfection. It is especially useful for a tiny spot that is nearly round, such as a dust speck. When the mouse is clicked (while holding it steady), the region is replaced by sampling a ring around the mouse position. In some versions a different, featureless region can be used as the model for the underlying texture, and the hue, saturation and value are modeled from the region surrounding the mouse. It also can be used with a sweeping motion of the mouse, and in this case a region on either side of the mouse sweep is sampled.
- *Cloning tool*—Two regions are selected with the mouse. When the mouse is clicked and moved, the sample region is copied to the mouse location. This tool is more difficult to use, but when used with care it can accomplish complex retouching tasks not possible with the spot heal tool.
- *Drawing tools*—There are many tools for selecting colors and values, and then 'drawing' directly with the mouse, filling in selected regions with a uniform color, or deleting details directly with the mouse (as if using an eraser).
- *Masks*—A region of complex shape can be selected with several different tools, and adjustments applied to that selection alone, or to its inverse—everything but that selection. The selection can have *fuzzy edges* so that that there is a gradation between where the changes are applied and where they are not.

When editing a digital image, it is easy to magnify its every detail to a level that is much greater than what is seen by someone viewing a print made from it. And so if care is taken and time is spent, the photographer can retouch the image down to the finest detail.

8.4 Color darkroom vs digital

I have so far compared digital and darkroom techniques only for black-and-white photography; color adds many complications. This chapter is not meant to be a how-to guide for digital image processing, but I will here add just a few items specific to color photography.

To make a *chromogenic* color print (see chapter 2, section 2.2) in the darkroom, light is projected through the negative with an enlarger, onto light-sensitive color print paper. In these ways, the color darkroom is just like the black-and-white version. But there are many important differences:

1. Since the color print paper must be sensitive to all different colors of light, one cannot use a special safe-light; all handling of the paper must be carried out in total darkness. But there are special processing tanks that allow chemicals to be poured in and out, while keeping the light-sensitive paper in darkness.
2. A special light source, called a *color head*, is used to expose the paper. The color head combines three different light sources, one each for cyan, magenta, and yellow. The brightness of these can be varied separately and so the color of the combined light can be precisely controlled.
3. The chemical development of the light-sensitive paper requires additional steps.
4. For black-and-white printing, the primary considerations are exposure and contrast control. For color printing it is exposure and color balance that are most important.

8.4.1 Contrast control

We saw how the contrast of a black-and-white print can be controlled in the darkroom by using graded papers or variable-contrast (VC) paper. A moment's thought should make it clear that one could not use something like VC paper for color printing. The contrast of VC paper is altered by changing (with special filters) the color of the light falling on the print paper. Obviously, for color photography, such filters would also change the *color* of the print.

One could still, in principle, design color print papers that are analogous to graded black-and-white papers in order to produce prints of higher or lower contrast. This is, however, not all that practical since the steps in the color process are so much more complex. For these reasons it is difficult to exert as much control over the contrast of a color print in the darkroom as one can for black-and-white printing.

In fact, by now most film processing centers have abandoned altogether the traditional darkroom method of projecting light through the film negative onto the light-sensitive print paper. Instead, the processing machine scans the film negative, and the resulting digital image is then scanned onto the light-sensitive print paper with computer-controlled colored lasers. The print paper is then chemically developed in the usual way to produce what is, physically, a traditional chromogenic color print. Thus, contrast (and color balance) is controlled digitally *before the print paper is even exposed by the lasers*, and this allows for far greater control over contrast and color balance than is possible with traditional wet-chemistry darkroom techniques. This has the added advantage that the same processing machine can produce chromogenic prints from both color film and digital image files.

Changing the contrast of a color digital image is, in principle, the same as for a black-and-white image. The only difference is that the color image contains, in effect, three separate digital images, one each for red, blue, and green. Each of these three parts of the image is called a *channel*. We can control the levels and curves on each channel separately, or we can vary them all together.

8.4.2 Color balance

The default in GIMP, whenever one selects *levels* or *curves* is that all three red, green, and blue channels are varied together. But one can instead select only one channel to vary at a time. This will change not only the contrast of the picture overall, but also the color balance. Employing adjustment curves directly on individual color channels allows for sophisticated control of the overall color of the image. But there are simpler alternatives as well. Editing software such as GIMP also feature simple tools that allow one to interactively shift the overall color balance, with sliders that adjust the mix between the additive and subtractive primaries. And so one can separately adjust the balance between red–cyan, green–magenta, and blue–yellow. As an alternative, one can adjust the hue, saturation, and lightness separately with a different dialog. But this only scratches the surface of the many powerful digital tools available for adjusting color. More information on color adjustment of digital images can be found in volume 2 of *The Physics and Art of Photography*, and in many standard references both online and in print (for example London *et al* 2005, chapter 11).

The overall color balance of a chromogenic print in the darkroom is adjusted at the stage of exposing the light-sensitive paper to light. The photographic enlarger must have a special *color head* that allows for easy adjustment of the color of the projected light. This is usually accomplished with separate adjustments for the subtractive primaries—cyan, magenta, and yellow. Often, a numbering system is used for these adjustments so the photographer can easily reproduce the proper color balance, once they have found it. The cyan, magenta, and yellow values are simply 'dialed in' by the three knobs on the enlarger head. To make this mostly trial-and-error process easier, many color printing systems employ a set of correction filters. One looks at a finished print through these filters, selecting the one that makes the color balance look the best. The particular filter selected tells the photographer how many steps they need to add or subtract from the cyan, magenta, and yellow adjustments on the enlarger head, such that the next print will come out with the correct color balance.

References

London B, Upton J, Stone J, Kobré K and Brill B 2005 *Photography* 8th edn (Englewood Cliffs, NJ: Prentice Hall)

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

Chapter 9

The digital and the analog

It could be said that analog techniques are, *as a matter of principle*, more 'precise' than digital. Analog is smooth, continuous and infinitesimally divisible, whereas digital is discrete, and thus inherently limited in its precision. One could offer as an example the set of smoothly curving grooves on a vinyl LP recording, as opposed to the digitally-encoded sound information in a CD or digital sound file recording. The argument is that there is a clear limit to the amount of information in the digital file, while the grooves on the vinyl LP, since they are smooth and continuous, contain an essentially infinite amount of information.

But this view of digital-versus-analog is incorrect, at least as a general principle. It may turn out to be true in a particular instance, but the opposite could be true as well; the devil is in the details. Although any particular instance of digital encoding has limited precision, the digital process *in general* does not. If, in a particular case, it is the method of digital encoding that is limiting precision, one need only change to a different digital encoding to increase the precision. Furthermore, any particular instance of analog encoding, despite its appearance of an infinitesimally dividable range of possibilities, is also limited in precision. It is the *physical nature of the act of measurement* that ultimately decides how much information is potentially available, not whether it is recorded as digital or analog. And finally, many measurement processes that seem at first glance to be analog, with a smooth and continuous range of possibilities, are actually more like digital if one looks closely enough.

9.1 Pixels and granularity

One aspect of the digital is its *discreteness*—that it can be subdivided only so much, until a smallest unit of the image is reached. This fact is obvious for a digital detector such as a CCD array; a given example has only so many pixels. Indeed, the basic concept of discreteness is built in to the very word *pixel*—which is simply a shorthand for 'picture element.' By definition, there is no such thing as half a pixel.

But let us consider ordinary black-and-white film. We have already seen that a black-and-white silver gelatin emulsion consists of light-sensitive crystals of silver

doi:10.1088/2053-2571/aaf0aech9

halide suspended in a thin layer of gelatin. Each exposed and developed silver halide crystal acts as a sort-of 'atom' for the picture. That is, a given crystal is only one shade of gray; it can't be, for example, half black and half white. This means that the size of the crystals (and the spacing between them) determines the smallest amount of detail that can be imaged on the film. If the crystals are bigger (and farther apart), then less detail can be imaged in a given square centimeter of the film.

We call the microscopic pattern of developed silver halide crystals *film grain*, or simply *grain*. The quantitative measure of film grain is called *granularity*, while a subjective determination of its overall effect on the picture is called *graininess* (Stroebel *et al* 2000, p 258). See figure 9.1 for an example of a silver gelatin film emulsion as imaged through a microscope. The left-hand image is at lower magnification, and one can see both a part of the image detail, but also a hint of the film grain. A detail of the same image is shown on the right, at much greater magnification. No image detail is visible at this scale, only the pattern of the silver specks.

The left-hand image in figure 9.2 shows a similar magnified view of a *color* film transparency. The center image shows part of the same transparency at much higher magnification, and only the film grain is apparent. But now the grain is more complex, consisting of tiny clouds of cyan, magenta, and yellow dye, along with some black spots of silver.

The right-hand image in figure 9.2 is not of film grain at all; it is a view through a microscope (at much lower magnification than the film examples) of a digital inkjet print. It too consists of spots of cyan, magenta, yellow, and black. But it is a computer-controlled regular pattern, rather than a random one. The dots appear in regular lines because of the physical mechanism of the printer itself, but the position of the dots within a line appears more random, with some overlapping each other. This is by design, however, not by accident; the positions of the dots are carefully chosen according to a computer algorithm.

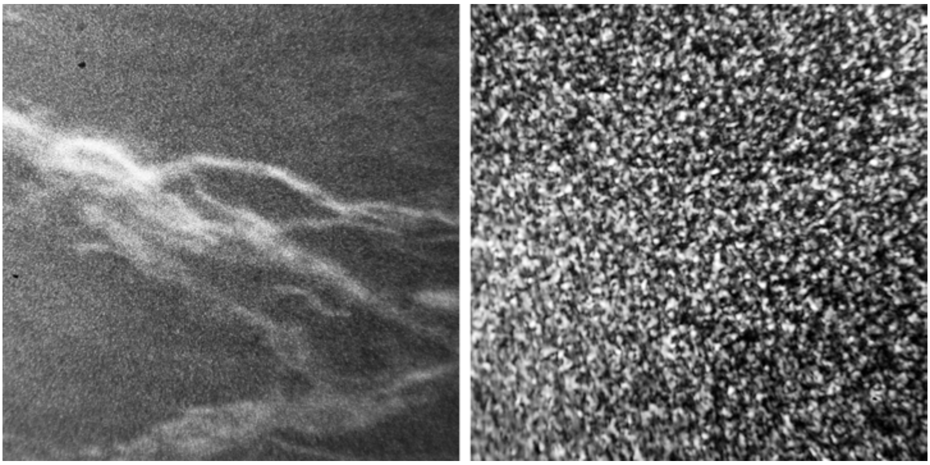


Figure 9.1. **Left:** Detail of a black-and-white film negative, as seen through a microscope. The film grain is just barely visible. **Right:** At much higher magnification only the film grain is visible, with no hint image detail.

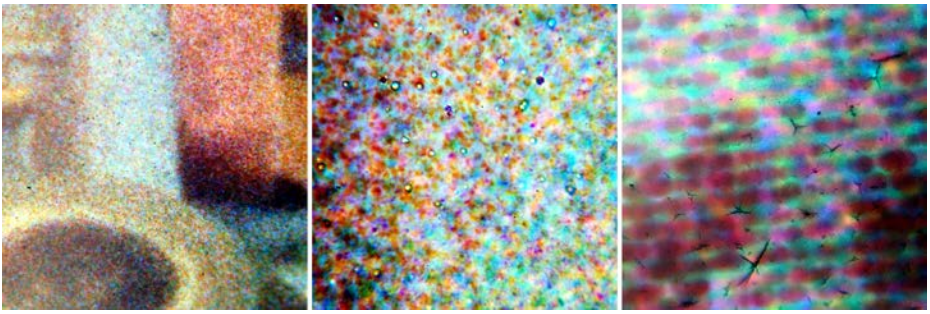


Figure 9.2. **Left:** Detail of a color film transparency, as seen through a microscope. The film grain is just barely visible. **Center:** At much higher magnification only the film grain is visible—small clouds of cyan, magenta and yellow dyes. **Right:** A magnified view of an inkjet print, made of a regular pattern of cyan, magenta, yellow, and black dots.

Figure 9.3. **Left:** *A Sunday on La Grande Jatte*, Georges Seurat, 1884/86. **Right:** A tiny detail from the painting. Art Institute of Chicago, CCO Public Domain Designation.

Figure 9.3 shows the famous painting *A Sunday on La Grand Jatte*, by Georges Seurat. The detail on the right is from a relatively tiny part of what is a huge painting. Notice the striking resemblance of the detail in the pointillist brush strokes of Seurat's painting to the arrangement of colored dots in the inkjet print in figure 9.2.

Figure 9.4 shows an image from a digital camera, along with a tiny detail, greatly magnified. The individual pixels are visible, and they are portrayed as simply square blocks. The digital pixels have no particular shape—they are simply image brightness values at different locations on a two-dimensional grid. Most would agree that this sort of magnified detail is both less interesting and more unappealing than the analogous detail in the silver gelatin emulsions or the Seurat painting.

9.2 Resolution

An important aspect of any detector is how much spatial detail is present in the image captured by it. If one were to look closer and closer at the image on a piece of film with magnifiers of higher and higher magnification, at what point would there be no more detail to be seen? Or, regarding a digital image, how much information is really there? At what point will zooming in on the image yield no additional detail?

Figure 9.4. The grid-like pattern of pixels at the right is from a tiny detail of the digital image at left.

This concept of *resolution* is complex and prone to confusion, and we must make a distinction between the *total* information contained in a picture, and the *information per surface area*. It is the *type* of detector material that usually limits the latter. The former is related both to the type and also the *size* of the detector. Confusingly, the word 'resolution' is often used inconsistently, sometimes meaning the total amount of information, and sometimes meaning the information per surface area. The process of printing adds even more complications.

Is there some kind of mathematical equivalence between the size of the pixels in the image from a CCD array on the one hand and the size of the grain in a silver gelatin emulsion on the other? There is, but the relation is far from straightforward. And so film with finer grain (smaller crystals) is capable of recording more information *per square centimeter* of film. We can talk meaningfully about the *average* number of silver halide crystals per square inch, and this can be used as a numerical measure of the resolution of the film. But there are other measures that relate more directly to visible detail in a real picture, and they are more useful. In particular, to record different values of light and dark, multiple grains—several at least—are needed, along with sufficient white space between them.

And so one useful measure of resolution is to imagine the size of the smallest 'bit' of the picture that can be discerned among the microscopic random pattern of film grain. We can think of this as an imaginary microscopic square, a tiny fraction of an inch across, and made of enough film grains (and the clear space between them) to record any value of light or dark. We can call this a *resolution element*, or *picture element*. The latter term is more commonly used, especially in its abbreviated form, the *pixel*. And so a pixel can be thought of as the smallest bit of a picture. For film it is an abstract idea; in a particular case, the actual size of the pixel depends somewhat on how one defines it mathematically. I will not consider those complications here. But whatever the precise definition, think of a pixel as the smallest indivisible element of the actual *image* on the film.

There is another fairly obvious complication. In practice, the resolution of an image depends not only on the physical properties of the detector, but also on the properties of the actual image that is focused onto it. What if one focuses a blurry image onto fine-grain film? And so in practice, it may be the *image that was focused onto the detector*, not the detector itself, that decides the finest amount of detail visible in the developed image. We will take up this question in more detail later. For now, let us consider the finest detail that *could* be imaged onto a particular detector, if a perfectly sharp image were focused onto it. And so what we are really talking about at the moment is the *maximum* resolution an image *could* have on a given detector.

For a given piece of film then, even though a pixel is a somewhat abstract concept, when considering resolution it is still meaningful to think of the film as having 'pixels' of a particular practical size, roughly equal to several times the size of the individual film grains. This is always a tiny fraction of an inch, and so it is more convenient to refer to the *number of pixels per inch*, rather than the size of an individual pixel in inches. This is often referred to (in the US) as *dots per inch*, or *dpi*. And so a given piece of film could be said to have a maximum resolution of, for example, 5000 dpi. This means we could imagine that a particular piece of film is made of an array of many individual pixels, each of which is 1/5000 of an inch across.

There is a trade-off between sensitivity to light and the size of the silver halide crystals in the emulsion. Details vary, but as a general rule, all else being equal *higher-speed film has lower resolution than lower-speed film*. So what is the answer? What *is* the maximum resolution, in dpi, for typical black-and-white film? Since the whole concept of 'pixel' or 'dot,' for a film emulsion, depends on one's precise definition, the *precise* answer to this question does also. And so, in short, the answer to this question is—not everyone agrees. But, roughly speaking, ISO 100 black-and-white film has a meaningful maximum resolution of somewhere between 2000–5000 dpi, depending upon whom one asks.

I have so far discussed the maximum information *per surface area*, and this is the concept for which I am attaching the word *resolution*. But just as important is the *total* information contained in the image on a particular detector or print. Since resolution can be measured in dpi, which means 'number of pixels per inch,' we can note the following:

$$\text{number of pixels} = \frac{\text{number of pixels}}{\text{inch}} \times \text{inches}. \tag{9.1}$$

And so to determine the total number of pixels along one dimension of a piece of film, we simply multiply dpi by the size in inches. For example, 35 mm film is approximately 1.5 inches across. If we assign a resolution of 3000 dpi to the film, then a single 35 mm frame would have a width of $1.5 \times 3000 = 4500$ pixels.

But just as important is the height of the film. In this case, 35 mm film has a height of approximately 1 inch, which would thus correspond to 3000 pixels high. And so we can think of the piece of 35 mm film with a resolution of 3000 dpi as being made of a total of 3000 pixels high by 4500 pixels wide, for a total of $3000 \times 4500 = 1.35 \times 10^7 = 13.5 \times 10^6$ pixels. I have put the final answer in terms of 10^6 pixels because that defines how many millions of pixels, and this is the most common measure of the total number of pixels in an image. One million pixels is called a *megapixel*, or *MP*. And so we can say the piece of film in our example has a total of 13.5 MP.

And so we have two different measures of the amount of detail in a picture. The total count of MP is a measure of the *total* amount of image detail that can be captured by the detector, while the dpi represents the amount of detail *per surface area* (what I am calling resolution). The total MP count is sometimes called 'resolution' too, and this can lead to confusion. This is almost inevitable because there *is* an important sense in which the total pixel count does relate to resolution, in a roundabout way.

Imagine I take the same image with two different cameras, both using film with the same resolution (dpi). But one camera uses 35 mm film (1.5 inch across), while the other uses medium format film that is 3.25 inches across. If, for the medium format camera, one uses a lens with a focal length that is proportionally longer than the one used on the 35 mm camera, then the image on the larger film will also be proportionally larger. For example, say I put a 50 mm focal-length lens on the 35 mm camera, but I put a 108.3 mm focal-length lens on the medium format camera. 108.3 is 2.167 times greater than 50 mm. But that is also how many times bigger the film is; 3.25 inch is 2.167 times greater than 1.5 inch. *Thus, images of the same subject made with both cameras will fill their respective pieces of film in exactly the same way.*

Since both cameras are using the same type of film, the resolution is the same; each has the same number of pixels per inch. But the same piece of the image (a hand, for example) covers proportionally more area on the medium format film than on the 35 mm film. Thus, although both images have the same dpi, the larger format film contains more pixels *per comparable part of the image*. To put it another way, even though the medium format film has the same number of pixels per inch as the 35 mm film, it has more 'pixels per hand.'

To further illustrate the distinction between MP and dpi, let us consider a slightly different case. Let us again set up the two cameras with the same film and again point them both at the same subject from the same vantage point. But what if this time we use lenses of the *same* focal length on both cameras? In this case *the hand would be the same physical size on both pieces of film*, and so would be imaged in exactly the same way. The larger format film includes a greater angle of view, and so the subject includes elements that don't appear at all in the 35 mm picture. *But for any given element of the picture, the resolution would be the same for both images, regardless of the total count of MP.*

Much of the point here is that what is important depends critically upon the question one is asking. For digital imaging, the concept of the pixel and resolution is less abstract—the pixel is a real physical site on the detector itself. Digital printing adds an extra complication, as most digital image formats allow one to specify default information regarding both the physical size and resolution of the *print*. The actual printing or display process, however, need not use this information. Since a digital image file only relates to a physical size when it is printed or displayed, it is the total number of MP that decides how much image information is contained in the image file. It is not uncommon for the 'resolution' or 'size' of a digital image file to be described in terms of *either* dpi or inches alone—but this practice should be avoided. Neither dpi nor inches alone has real meaning for a digital image file; it is only the product of the two—the number of pixels—that describes the real information in the digital image.

9.3 Signal and noise

The fundamental quantum nature of light becomes particularly apparent when a detector is subject to very low levels of illumination. Individual photons of light are either detected, or they are not; there is no in-between. A detector can detect, at a given location, one photon or two photons or three photons. But it cannot detect 1.3578 photons. This basic fact means that the mathematics of light detection is a problem of *statistics*.

Along with any measured *signal* from a light detector comes *noise*—statistical variations in the measurement that cannot be predicted ahead of time and that—in and of itself—carries no direct information about the image. But even though noise is, by its very nature, unpredictable, there is still much we can say about it. In section 9.3.1 I first work through a non-photographic example.

9.3.1 Pennies and Poisson

Over the course of the year 1997 I collected, from random sources, 1080 pennies and put them in a box. Apparently I had little else to do back then, and so I sorted the pennies by year, counting up the number of 1984 pennies on the one hand and 1985 pennies on the other. A graph of this data—pennies collected versus year—could then be called a *penny spectrum*. Figure 9.5 shows such a penny spectrum for my box of 1080 pennies.

What conclusions can one make from this data? Notice that, for example, there is a definite 'hump' of pennies around the year 1982. And there is a big drop in the number of pennies between 1970 and 1971; I collected 13 in 1971, while only 6 in 1970. Can I then, for example, conclude that there really were $\frac{13}{6} = 2.17$ times more 1970 pennies in circulation than 1971 pennies? To answer that question, we must

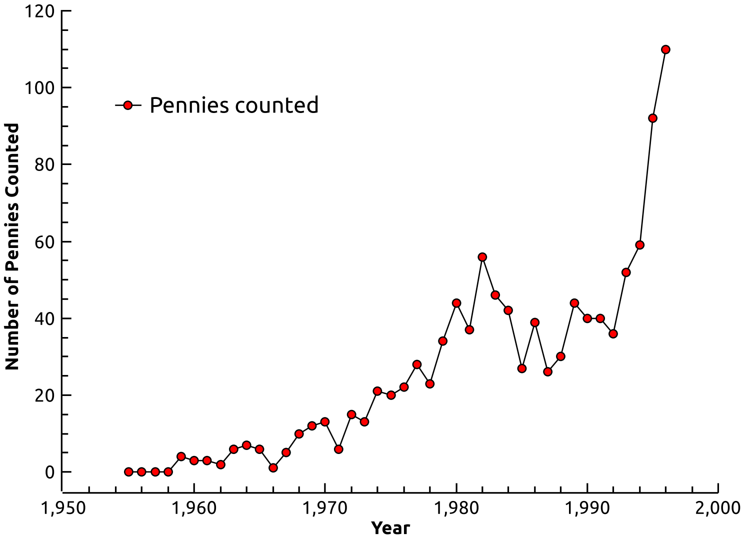


Figure 9.5. A graph of 1080 pennies, collected in 1997, with the number collected plotted versus year.

first ask ourselves a different one: if I had separately collected another random set of 1080 pennies, would it have made exactly the same penny spectrum?

The short answer to this question is—likely not. If the events are random, then *only the probabilities remain the same*, but the specific result will likely be only *approximately* the same. The detailed answer to this question is one of the most fundamental problems of statistics. It is beyond the scope of this book to work through that answer in detail (see, for example Bevington 1969), but it is not difficult to understand the basic result.

Let us imagine I had collected not one or even two boxes of 1080 pennies, but that I had collected 10 000 such boxes. I could then, for example, count how many 1970 pennies—out of the total of 1080—were in each box. With these numbers, I could make a different kind of graph, called a *relative frequency distribution*. Each point on the graph tells the percentage of the 10 000 boxes that contained that particular number of 1970 pennies.

Figure 9.6 gives an example, corresponding to our hypothetical tallying of 1970 pennies collected out of a huge number of 1080 penny samples. The distribution looks like a slightly asymmetrical hump, and the *peak* of the distribution (in my contrived example) lies between 12 and 13 pennies. This peak represents the experimental result that occurred most frequently. But even though 12 or 13 pennies were counted most frequently, this accounts for only 22% of the pennies counted. Clearly, for most of the 10 000 collections, either more or fewer than 12 or 13 pennies from 1970 were counted.

One can see from figure 9.6 that it was very common to count, for example, 10, 11, 12, 13, 14, or 15 pennies. But it was very rare to record as many as, for example, 30 pennies or as few as 3 pennies. A frequency distribution such as this is important,

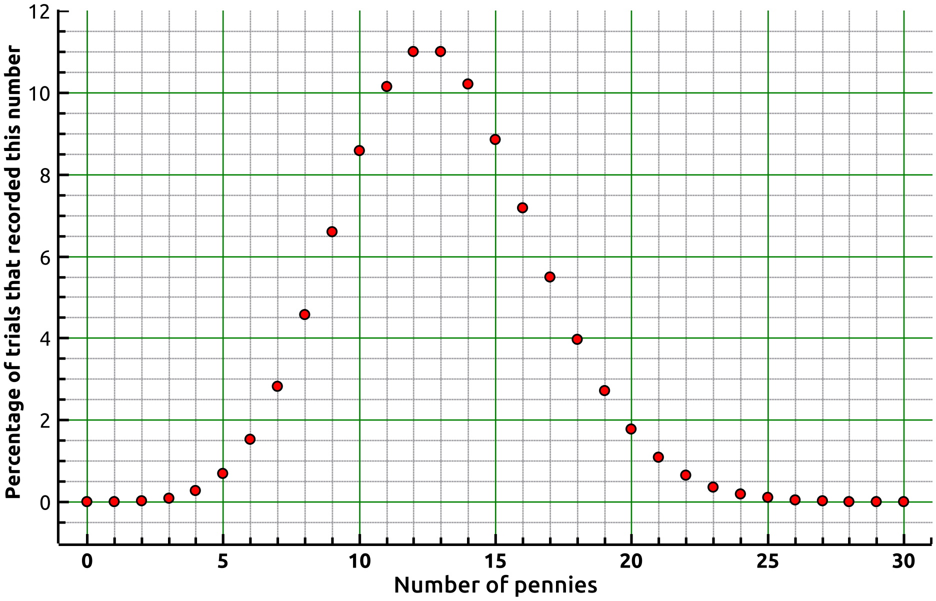


Figure 9.6. A Poisson distribution that peaks at between 12 and 13 events.

because repeating the same measurement many times tells us after-the-fact how likely it was to have obtained any one particular result. In this case there was, for example, a high probability of recording 12 pennies from 1970 but a very low probability of recording 28 pennies.

And so what would have happened if I had made one additional collection of 1080 pennies, for a total of 10,001 penny collections? How many 1970 pennies would it have contained? There is no way to know! *The distribution only tells us the likelihood*, not what happens. But a careful analysis of this common situation of counting randomly occurring events shows that, in the limit of more and more repetitions of the measurement, a frequency distribution similar to that shown in figure 9.6 would result. This *Poisson distribution* is an idealized mathematical distribution that describes the results of independent counts of a set of random events, in the limit as the number of trials approaches infinity.

An important feature of the Poisson distribution is its width, which clearly tells us something about the *uncertainty* in the expected results if we were to repeat the measurement again. If the distribution is very wide, it means that measurements either much greater or much less than our most-likely peak value are still very likely. But if the distribution is very narrow, then the opposite is true; it is very *un*likely to measure a value much different than the peak value. In figure 9.6, for example, we may note that the peak of the distribution, at between 12 and 13 pennies, is about 11%. At half of this peak, 5.5%, the distribution stretches from about 8.5 to 17 pennies. And so we could say the distribution has a *full width, half maximum (FWHM)* of 8.5, or ±4.25.

From my single box of 1080 pennies, all I can say is that 13 out of 1080 were from 1970. And so the best I can do is to generalize that result—to claim that when I collected these pennies, my best estimate is that 13 out of every 1080 pennies (1.2%) in circulation were from 1970. But the Poisson distribution for our thought experiment of performing the same measurement many, many times tells us something else—*it tells us the uncertainty in out best estimate*. And so because we counted only 13 pennies for 1970, we should expect that count to be representative of the total number of such pennies to within about ±4.25. This uncertainty is about 33% of our count of 13. And so if we use the 13 counts to form conclusions about the number of 1970 pennies in circulation, we should only expect our conclusions to be accurate to within ±33%.

A full statistical analysis of the Poisson distribution indicates that a better measure of its width, and thus the statistical uncertainty in our number of counts, is not the FWHM, but rather what is known as the *standard deviation*, σ (the Greek letter sigma). It has a particular mathematical definition, the details of which are not important for this discussion, but it turns out to be quantitatively similar to the FWHM. For the Poisson distribution shown in figure 9.6, the standard deviation is approximately $\sigma = 3.5$.

Happily, there is a very simple relation between a particular number of counts on the one hand, and the expected Poisson-statistics uncertainty on the other. *The best estimate for the uncertainty in the number of counts is simply the square root of the number of counts.* And so for my single box of 1080 pennies, I counted 13 from 1970.

Poisson statistics tells us that we should expect that number to be representative of the 1970 pennies in circulation only to within $\pm\sqrt{13} = \pm 3.6$.

And so let us return to my original question. My sample of pennies shows over twice as many 1970 pennies as 1971 pennies. Does this mean I should reasonably conclude that there really were twice as many 1970 pennies in circulation as 1971 pennies? Or can this difference instead be accounted for simply by the statistical nature of counting a limited sample? We can make a better judgment if we put *error bars* on the figure 9.5 data—vertical bars that encompass the likely uncertainty in the number of counts. Figure 9.7 shows the same data with these error bars in place, calculated simply by taking the square root of the number of counts. Careful inspection of the graph shows that the lower error bar for 1970 almost meets up with the upper error bar for 1971. This means that it is possible, but not extremely likely that the differences in those counts can be accounted for by Poisson statistics alone.

Analyzing the expected statistical uncertainty in measurements is of the utmost importance whenever we use measured data to form reasonable conclusions about models or theories. And so for example, can we form a reasonable theory to explain the ups and downs in the penny spectrum? The red line in figure 9.7 shows an attempt at explaining the data with the most obvious theory. I looked up the statistics from the US Mint, and simply plotted the number of pennies put into circulation each year—scaled to match up with the total number in my tiny sample. I scaled the U.S. Mint data to match my penny counts at 1960—but then the red line clearly falls well below the data in more recent years. Of even greater importance, this ‘theory’ lies well outside the error bars of the data.

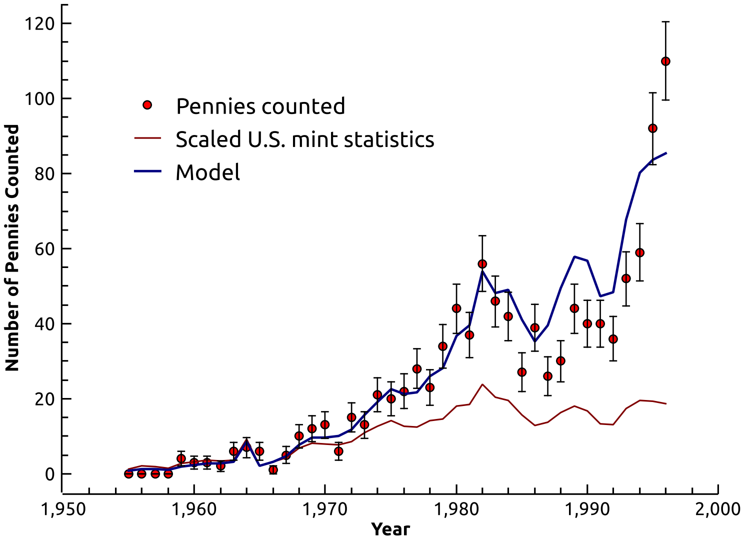


Figure 9.7. The same data as figure 9.5, but with Poisson statistics error bars added. The red line shows the number of pennies put into circulation each year by the US mint, scaled to fit the left end of the data. The blue line shows the same mint statistics as modified by a simple model that also includes pennies being taken *out* of circulation.

The blue line represents a slightly more sophisticated attempt at a theory for the data. It is the same as the red line, but also includes a model of how pennies might be taken *out* of circulation. My model assumes that pennies are taken out of circulation always at the same percentage rate. And so much of the reason there are so few 1960 pennies compared to 1990 pennies is that 1960 pennies have been going out of circulation for many more years. This model clearly 'fits' the data very well up to about 1985, but after that it lies at the edges or just outside the error bars of the data points.

Because the model does not completely fall within the error bars of the data, can we then conclude that the theory is incorrect? It would be nice if life were so simple. But it is also possible that the measured *data* is the problem. In particular, I could be mistaken in my assumption that I collected the pennies randomly; the data could be *biased* in various ways, meaning it does not directly represent the real distribution of pennies in circulation. In section 9.4 we consider the correction of biased data in the context of using digital detectors to make precise quantitative measurements of astronomical interest.

The Poisson distribution is named after the French mathematician and scientist Baron Siméon Denis Poisson (1781–840). Only one of his many accomplishments in mathematics and fields as diverse as celestial mechanics and electricity and magnetism, Poisson published his analysis of the statistics of counting randomly occurring events in 1837. There is some irony to this, as Poisson is also well-known for his public opposition to the new advances in the wave theory of light presented by his contemporary, Augustin-Jean Fresnel. Poisson was a firm believer in the *corpuscular* theory of light—that it was essentially a stream of particles, and Fresnel's work did much to lay the corpuscular theory to rest. And yet, a century later it was well established by quantum physics that in addition to its wave-like nature, light *does* in fact act as a stream of particles (photons). When photons arrive at a detector, they are either detected or they are not—just like counting pennies. Thus, Poisson's statistics of counting plays a key role in the measurement of light.

9.3.2 Photons, signal and noise

Because a particular photon either arrives at the detector and is detected, or it is not, Poisson statistics apply to the detection of light. This is not to say there are no *other* factors that add to this uncertainty—but those factors can often be mitigated by altering the design of the detector. One cannot, however, eliminate the uncertainty described in section 9.3.1; it is a fundamental part of the measurement process, due to the inherently statistical nature of the arrival of photons.

In practice then, there is an inherent random element to the physical detection of light. And so when exactly the same light falls on many nearby pixels, one might expect the detector to record exactly the same measurement for all of those pixels. But instead they will record, randomly, a range of values from pixel to pixel. Only the average of these randomly-fluctuating values may correspond closely to the actual brightness of the light. We call these random fluctuations *noise*, and in a photograph it shows up as a random speckled pattern where one would expect a uniform shade of gray (or a uniform color). The 'true' value that the noise deviates plus or minus from is called the *signal*.

This also means that *just because one pixel records more light than another, it is not necessarily true that the light was brighter there*. Consider this example: pixel A detects 9 photons while pixel B detects 7. When we combine these with their respective uncertainties ($\sqrt{9} = 3$ and $\sqrt{7} = 2.6$), it is clear that it is possible that the light was actually brighter at pixel B than at pixel A.

Thus, we have the *the actual brightness of the light* on the one hand, and on the other hand the often small and always finite sample of that light as actually measured by physical interactions with the detector. The measurement contains real information about the actual brightness of the light, but it always contains random fluctuations (noise) due to the statistical nature of the detection of light. In practice we can (and do) use our actual measurement as our best estimate of the signal, but we must always be aware that any signal comes with noise as part of the package.

9.3.3 Signal-to-noise ratio

Which has more noise, a signal of 10 000 photons detected or a signal of only 100 photons detected? Regarding the noise due to Poisson statistics, the answer is clear: the larger signal also has greater noise, for $\sqrt{10\,000} = 100$ while $\sqrt{100} = 10$. So we can be reasonably confident of our smaller signal to within about ±10, while we only know the larger measurement to within ±100.

But in most circumstances, we care about something very different. *It is usually not the noise itself that is important, but rather how that noise compares to the signal.* And so let us look at our example again, in this new light. We have, on the one hand, a measurement of $10\,000 \pm 100$ and on the other hand 100 ± 10. In the first example, the noise of ±100 is only 1% of our signal (10 000). In the second example, however, the smaller noise of ±10 is 10% of our also-smaller signal of 100. And so although the smaller signal has less noise that comes along with its measurement, that noise is a *larger* percentage of the signal itself.

And so we have this important but slightly counterintuitive result: more signal means more noise, but it means *less noise as compared to the signal itself.* A useful way to describe this numerically is with the *signal-to-noise ratio* (s/n), or simply the signal divided by the noise. Thus, our larger signal would have a s/n of $10\,000/100 = 100$, while the smaller signal has a s/n of $100/10 = 10$. The s/n represents then, how many times larger the 'true measurement' is than the random statistical fluctuations in that measurement. The 'noisy' or 'grainy' look of a photograph is closely related to s/n; a *higher* s/n means a more uniform, *less noisy* look to the picture.

For Poisson statistics, since the noise (n) is simply the square root of the signal (s), we have the following simple result:

$$s/n = \frac{s}{n} = \frac{s}{\sqrt{s}} = \frac{s}{\sqrt{s}} \cdot \frac{\sqrt{s}}{\sqrt{s}} = \frac{\cancel{s} \cdot \sqrt{s}}{\cancel{s}} = \sqrt{s}. \tag{9.2}$$

And so for Poisson statistics, not only is the noise the square root of the signal, but so too is the signal-to-noise ratio. Thus, for a greater s/n one simply needs more signal. This means that for a photograph the shadow areas (where less light was detected) are the parts that will always look the most noisy or grainy. The saving grace is that if we leave those shadow areas dark on our final print or display of the photograph, then although those areas have more noise (compared to the signal), it will be equally difficult to see that noise.

The problem arises when one alters the photograph. What if one wants to use digital photo editing software to 'brighten up' those shadows, to see detail that is otherwise hidden? A consequence is that when the signal for those shadow areas is re-scaled to make it look brighter on our photograph, it is inevitable that so too is the noise that comes along with that signal. And since an underexposed shadow area is likely to have a very low s/n, that brightened region will be made to appear very noisy. See figure 9.8.

We can make this effect a little more concrete with actual numbers. Say, for example, that in my two exposures of the tree in figure 9.8, a particular dark part of the image was recorded with a signal of 1000 photons detected in the left-hand picture, but that same part detected only 10 photons for the underexposed version on the right. I then re-scale the right-hand image, so it looks as bright as the left one.

Figure 9.8. The result of different s/n. The pictures on the left were taken with a higher s/n than the underexposed pictures on the right. The pictures on the right were re-scaled so they appear to have the same overall brightness as those on the left, but doing so increased the noise proportionally, leaving the low s/n unchanged. Compare the two magnified views at the bottom; the low s/n version on the right appears grainy and with much less contrast. The contrast could be digitally enhanced, but doing so would make the picture appear even more grainy still.

To do this I would have to multiply the values by 100, so the value of 10 photons would also be 1000. But now let's look at what happens to the noise. The left-hand picture has a s/n of $\sqrt{1000} = 32$, while the underexposed version has a s/n of only $\sqrt{10} = 3.2$. The problem is, when I re-scale the low s/n image so it looks as bright as the high s/n version, *the s/n remains the same*. This is because I will multiply *both* the signal and the noise by the same factor (100), so the ratio of the two remains the same. So now my altered underexposed image seems to have a signal of 1000 just like the unaltered image on the left. But it still has a s/n of only 3.2, and so it still looks 'noisy.'

For any real detector, there are many steps between the physical interaction of the photons and the ultimate realization of an image. A CCD array detector, for example, must transfer the signal from one pixel to the next, and pass it through an amplifier. All of the these physical interactions contribute noise in some way; the detailed analysis of all of the factors that contribute to noise is complex indeed, and well beyond the scope of this book. With these other sources of noise taken into account, the signal-to-noise ratio no longer has such a simple relationship to the signal as equation (9.2). But in most cases, the same lesson applies—a larger signal comes with more noise, but there is less noise compared to the signal. And so in most circumstances, more signal means greater s/n.

9.4 Digital photography and the data revolution in astronomy

Digital detector arrays such as CCDs are especially suited for precise, quantitative measurement of light. As such, astronomers were some of the first to widely use digital imagery, and it has transformed observational astronomy utterly. The graphs and images in this section are taken from data used for the preparation of Beaver *et al* (2013). In what follows I describe why these digital detectors are so suited to the quantitative measurement of light.

9.4.1 Digital detectors are reusable

Load a roll of 35 mm film into a camera, put the camera on a tripod, focus on a still subject under controlled lighting, and take 36 exposures exactly the same way. The idea that commercial photography has striven for is that all 36 pictures would be identical. And that is true, to a point. But in fact every single exposure was made with a *different* light detector. The pictures only seem to be identical because of the great pains film manufacturers have taken to insure that the film is uniform, from one frame to the next and from roll to roll, in its chemical and physical properties.

Of course, if one looks more closely—under a microscope for example—then the differences become evident. At the finest level of microscopic detail, the arrangement of particular silver halide grains on the film is random. And if one measures with enough precision, it becomes clear that the sensitivity of the film varies ever so slightly from one place to the next.

Much of this is true for a digital detector as well. Although there is nothing quite the same as film grain, the pixels on any real digital detector vary in sensitivity in a random way as compared to their neighbors. Furthermore, most digital detectors have manufacturing flaws that make certain regions of many pixels slightly more or less sensitive than other areas.

But here is the difference. Many of the flaws and imperfections of a digital detector are repeatable—the same flaws persist, and in the same way, from one picture to the next. But for film, the specific imperfections and random details in one piece of the film are unrelated to those in another. Light interacts with a new detector every time an exposure is made with film. But in a digital camera light interacts with the same physical silicon chip for every picture. *And this means that it is possible to measure that particular digital detector's imperfections, and to correct for them.*

Some of these corrections are performed automatically by many digital cameras, and the stored data file is most often pre-processed to take into account the features of the particular detector. More sophisticated corrections can be performed after the picture has been taken, by gathering additional data and using image processing software. Specific examples are discussed in section 9.4.5.

9.4.2 Linear response

Many photoelectric digital detectors—a CCD array is a good example—have over a broad range of exposures a true linear response. That is, if one plots the measured effect (electrons accumulated at a pixel for example) versus exposure, it is a straight line on an ordinary graph (not a log–log plot). One consequence of a linear response is that it is possible to *calibrate* the image—correct for systematic effects so that the measured signal for each pixel in the image is directly proportional to the intensity of light that was incident during the exposure (see section 9.4.5). This allows one to combine separate exposures of the same subject in order to increase the signal-to-noise ratio. Figure 9.9 shows graphs measuring pixel values along a particular row in two CCD images. The left-hand graph is from a single CCD exposure, while the right-hand image is from many separate exposures added together, pixel by pixel—and the random noise, compared to the signal, is greatly reduced.

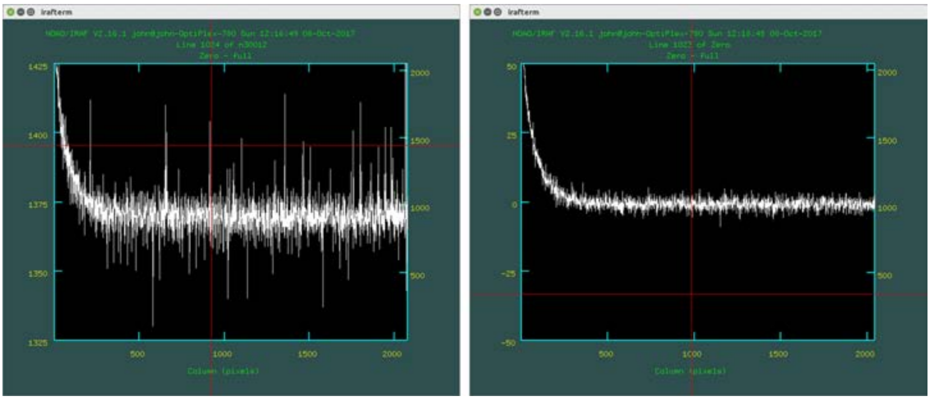


Figure 9.9. **Left:** A graph of pixel values across a single row of a low-signal CCD image. **Right:** The same, but here many low-signal images have been added together to form a new image with a much higher signal-to-noise ratio.

For nearly every photochemical detector, there are many complex steps between the original formation of the image and the final print. In the traditional silver gelatin process, for example, the latent image formed by light in the negative is first chemically amplified with a developer, then the remaining silver halide is removed chemically with a fixer. Then light is passed through the negative to expose the light-sensitive paper to make a photographic print, and the steps of development and fixing follow once again. All of these steps add complexity to the overall relation between exposure to light and the measurable effect of density. The end result is a *nonlinear response* to light, as discussed in chapter 1, section 1.6. Recall that even over the 'straight-line part' of the log–log characteristic curve, the relation between density and exposure is better described by a power law than a straight line.

A linear response is greatly desired if one wants to use a light detector for precise, quantitative measurements. But the advantage for ordinary photography—where the goal is to instead accurately reproduce human visual *perception*—is less clear. The s-shaped log–log characteristic curve of traditional film works well for its ordinary purpose—reproducing images from life, to be viewed by humans. The shallow slope of the bottom and top parts of the characteristic curve compress a lot of brightness levels in the world to relatively few in the picture. Regarding the shadows, that fact often makes little difference—the shadows of the picture are too dark to see much anyway. But regarding the highlights, this is a lucky thing—it seems to be similar to how we humans see high levels of brightness (Adler *et al* 2014).

9.4.3 Dynamic range

The *dynamic range* of a detector is the total range of brightness levels that it can record in the same image, while still retaining the ability to distinguish one brightness level from another in a useful way. There is no obvious precise definition for dynamic range, since what is a 'useful' range of brightnesses depends greatly upon the nature of the use. For astronomical applications however, the dynamic range refers to the brightness levels for which the detector response is linear.

Figure 9.10 shows two graphs taken from different rows of the right-hand image in figure 9.12, of the star cluster M 11. The left-hand image is across a particular row

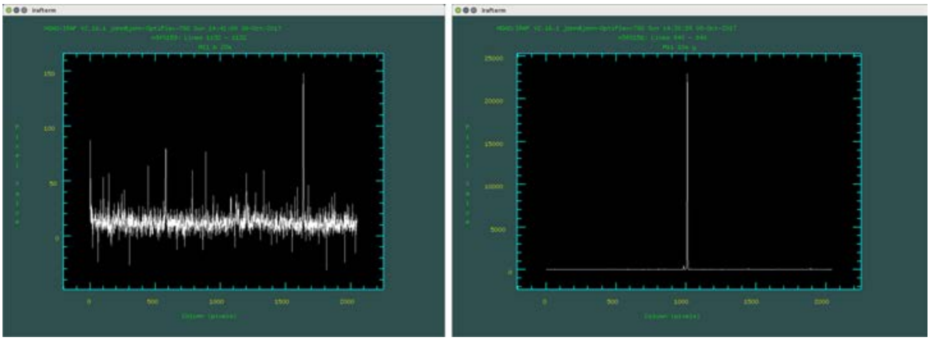


Figure 9.10. CCD detectors can have a large linear dynamic range. These two slices across different parts of the same calibrated image show usable data over a large range of values.

that contained only very faint stars. Much of the low-level up and down in this graph is a uniform signal from the dim sky background and the statistical noise that comes with that signal. Most of the spikes that shoot up to much higher levels on the graph are where the row cuts across individual stars. On this scale the sky background has a value of about 10 (the units are unimportant for this discussion), and that is a meaningful measurement for this fully-calibrated image.

The graph on the right shows the pixel values for a row that happens to cut across a bright star near the center of the image, and it produces a huge spike in the graph. Notice that at the scale of this graph, the sky background is negligibly small in comparison. The star's brightness on the same scale is nearly 25 000. *Both* of these measurements in this calibrated image are meaningful, and they can be compared to each other quantitatively. And so on this single image, we have a sky background measurement of 10, and a bright-star measurement of 25 000–2500 times greater—and both measurements are meaningful. Thus, we can say that this image has a dynamic range of at least 2500.

Without this large dynamic range, one would need to take many exposures of different lengths in order to adequately measure the brightnesses of both the faint and bright stars. And so the impressive *linear* dynamic range of CCD detectors is one of their principal attractions for astronomers.

A factor of 2500 corresponds to over 11 exposure steps ($2^{11} = 2048$). So, what is the dynamic range of film? What range of exposure steps can be meaningfully recorded with a single exposure of black-and-white or color film? The answer is not so straightforward for two reasons. First, the characteristic curve for film is very different from the linear response of a CCD. Secondly, the intentions are different—we expose black-and-white film not for its quantitative data, but rather to make a picture that is to be viewed by humans. And since we perceive brightness in a way that mirrors somewhat the s-shaped characteristic curve of film, we can use a large part of that highly nonlinear response to make a good picture. And so the useful dynamic range of film is a more-or-less thing, and answers vary. For a typical example, Stroebel *et al* (2000, p 101) present a film characteristic curve with a dynamic range of about 500, or about 9 exposure steps. Outdoor scenes, on the other hand, typically have a range of only about 100 in brightness, and so the dynamic range of film is more than adequate.

Dynamic range and bit depth

For a digital photoelectric detector, the signal at a given pixel is encoded by a computer or microprocessor (built into the camera) as a number. This computer encoding of the number must be accomplished according to some prearranged scheme—and that scheme sets a limit to the numerical information that can be recorded.

Digital information is stored by a computer in binary form—a string of ones and zeros. The *bit depth* of an image can be described in terms of the number of binary ones and zeros that are used to encode the level of signal at a given pixel. A single bit has only two possibilities—0 and 1. And so if only one bit per pixel is used to encode the image, then each part of the image has only two possibilities—black or white. But add a second bit, and the possibilities double; each bit added *multiplies* the possibilities by two. Thus, four different signal values could be recorded with 2 bits, but 8 values could be recorded with 3 bits. The *bit depth*, then, is the total number of bits used to encode all of the possible signal levels that can be recorded per pixel. If N is the bit depth, then 2^N different levels can be recorded.

And so, for example, an 8 bit black-and-white image can record $2^8 = 256$ different gray levels. Is that enough? What about 16 bits: $2^{16} = 65\,536$ possible shades of gray? Is that enough, or is it more than enough? The answer is that it depends on several factors. But first, we must consider another complication. The typical *color* image is essentially three black-and-white images that work together to synthesize color. Thus, if 8 bits are enough for a black-and-white image, it follows that we would need three times this many—24 bits, to record an equivalent color image. The separate red (R), green (G), and blue (B) portions of an RGB color digital image are often called *channels*, and so the terms '24 bit RGB image' and '8 bits per channel RGB image' mean the same thing[1].

And so, how many bits are enough? The answer depends on to what use the digital image is put. At one extreme, an image of simple text is often digitally encoded with only a single bit per pixel, since only black and white is needed. For an image that will serve no other purpose but to be directly displayed on a screen or printed, then 8 bits per channel is plenty. That corresponds to 256 possible shades per channel, and the human eye cannot even discern that many.

It is another matter entirely, however, if this same digital image is to be *adjusted*, as in the examples illustrated in chapter 8, section 8.2. In the course of applying an adjustment curve, some of the limited number of shades of gray are compressed, while others are stretched further apart. And so shades of gray that, in the unadjusted image, are too close to each other for the eye to discern as different, may be spread further apart by the adjustment. And so what appeared as a smooth gradation of tones in the original then shows sudden visible jumps in brightness from one area to another, leading to an unattractive *contoured* look. And so a digital image that is meant to allow for adjustments must have a greater bit depth than what is needed for the final image.

There is another important side to the entire consideration of bit depth—it also matters just what information is encoded by those bits. The left-hand image in figure 9.11 is encoded as 24 bit RGB. The right-hand image was made by transforming the 24 bit image to only 8 bits. This lower-bit color image has obvious contours; clearly 8 bits is not enough to encode all of the subtle levels of both hue and value present in the original image. This 8 bit image is not really an RGB image like the 24 bit image on the left; it does not have separate channels for red, green, and blue. Notice that 8 does not even divide evenly by 3. It is instead an *indexed* image; each of the 256 possible numbers in the 8 bit encoding is assigned its own color out of all of the full-color possibilities. The computer software decided how best to split up the vastly-more colors in the original to this much more limited color *palette*. But here is the point—one cannot then go backwards. One cannot get rid of the contours in the 8 bit image by simply transforming it (with computer

---

[1] RGB is not the only method for encoding a color image, but it is the most common.

Figure 9.11. On the left is a digital image stored as a 24 bit RGB image file. On the right is the same image, but after conversion to a 256 color (8 bit) indexed image. The subtle range of hues and values cannot be captured with a palette of only 256 colors, and so a contoured, paint-by-number effect results.

software) to a 24 bit RGB image. The image only has 256 colors, and so only 8 bits are needed to encode it, and the vast majority of possibilities in the 24 bit encoding would be unused.

Thus, it also matters how much information is in the image. Choosing a greater bit depth does not, in and of itself, put more information into the image; it only allows for the *possibility*. The amount of grayscale and color information in an image ultimately depends upon the physical means by which the image was originally formed. For a digital camera with a CCD detector, for example, the signal from each pixel is determined by measuring the number of electrons that accumulated at the light-sensitive site during the exposure. And so the bit depth that is need to count those electrons depends on the number of electrons that can accumulate before the site is 'full.' The general rule of thumb is that for a black-and-white image on a CCD detector, 8 bits (256 levels) are too few, but 16 bits (65 536 levels) are enough. The so-called RAW image formats on many CCD cameras are stored as 16 bits per channel—48 bits for an RGB color image.

But for most causal use, a 48 bit color image is overkill; it allows for far more adjustment than one is likely to perform. A good compromise is 24 bits—8 bits per channel, and this is the most commonly-used encoding for color images not meant for scientific measurement or sophisticated image processing. The standard jpeg (.jpg or .jpeg) color image format only supports up to 24 bits, and so a 48 bit image must be stored in other formats.

The CCD detectors used for astronomy usually do not have a Bayer mask to produce a color image. Instead, the astronomer takes separate images, one at a time, through different colored filters in order to record information related to color (see volume 2 of *The Physics and Art of Photography*). But the astronomer needs to use all of the information that the CCD is capable of recording, and so a 16 bit encoding is most-often used.

9.4.4 Quantum efficiency

Many different physical processes can be employed to make a detector of light. But whatever the method, a crucial question is *what percentage of the light falling on the detector is actually detected?* For it is inevitable that some of the light falling on the detector will result in no usable effect at all. And so, if 1000 photons strike the detector, how many of those photons result in a measurable effect?

The percentage of photons striking the detector that are actually detected is called the *quantum efficiency* of the detector. It is not difficult to see that a detector of higher quantum efficiency has some distinct advantages. Clearly a detector of 50% quantum efficiency should be able to accomplish the same exposure with, all else being equal, half the exposure time of a 25% quantum efficiency detector. This is in fact one of the big advantages of many photoelectric detectors as compared to photochemical detectors. A CCD array detector can achieve a quantum efficiency as high as 90%. The quantum efficiency of ordinary silver gelatin film on the other hand is only a few percent.

Since the Universe is a big place, objects of interest to astronomers are usually very far away, and more often than not, they are exceedingly faint. To compensate for this, astronomers must make long exposures with large telescopes (so as to gather a greater area of light). The high quantum efficiency of CCD arrays as compared to film is one of the primary reasons that amateur astronomers now take pictures with backyard telescopes that rival pre-digital photographs made decades ago with the largest telescopes in the world.

9.4.5 Image calibration

Because photoelectric array detectors such as CCDs have the ability to produce a truly linear response over a wide dynamic range, and because—unlike photochemical detectors—it is the same detector used again and again, it is possible to *calibrate* them. The left-hand image in figure 9.12 is a raw, uncalibrated exposure of

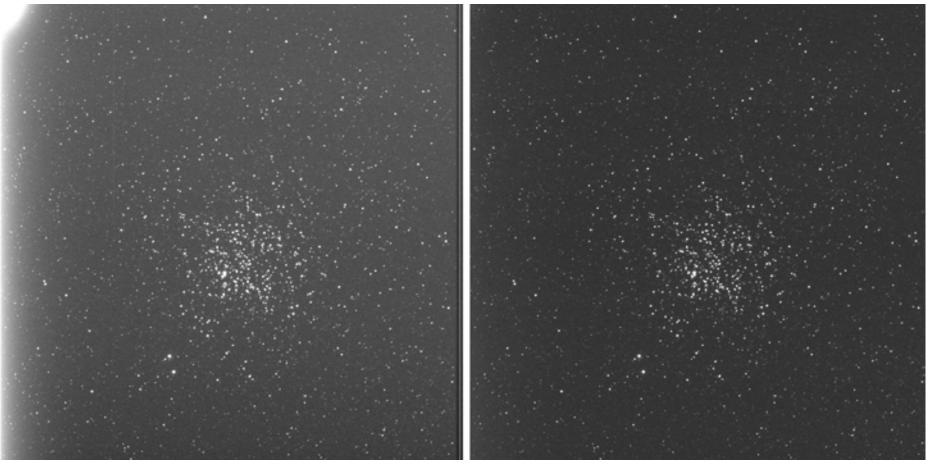


Figure 9.12. **Left**: Raw CCD image of the star cluster M 11. **Right:** The same image after calibration.

the star cluster M 11, in the constellation of Scutum, taken by astronomer Michael Briley with a CCD camera on a telescope at Kitt Peak National Observatory, as part of the data set for Beaver *et al* (2013). The right-hand image is a calibrated version of the same. To the eye, only slight differences are apparent between the raw and calibrated images. But the raw image by itself is unusable if the goal is to make accurate quantitative measurements of the brightnesses of the stars.

The linear response means that it is practical to *calibrate* the image from a CCD detector—correct for systematic effects so that an accurate quantitative measure of the exposure to light can be determined from the measured signal. A nonlinear response makes this process far more difficult; if the deviation from linearity is too drastic, accurate calibration may be impossible in practice. A linear response between the signal, s, and the exposure, E, implies that the following is true:

$$s = aE + b \tag{9.3}$$

where a and b are constants. Notice that the constant a is multiplied by the signal, while b is added. If this relationship holds, it means that some systematic effects are additive while others are multiplicative.

A good example of an additive effect is the CCD *bias signal*; even with no exposure to light, a signal is produced by the very act of applying a voltage to the pixels, reading out the CCD, and passing the result through amplifiers. This bias signal varies from pixel to pixel on the CCD; it must be determined separately for each pixel. The first image in figure 9.13 is an example of a *bias frame*—a zero-second exposure used to directly measure the bias of each pixel. If the response is linear, this image can be simply subtracted—pixel by pixel—from the image to be calibrated, in a process called *bias correction*.

The thermal motions of the atoms in the CCD also produce a signal that is independent of the exposure to light. This *dark signal* is (hopefully) proportional to the exposure *time* if the CCD is kept at a constant temperature. The linear response means that the dark signal for a given exposure time is simply additive, and so can be measured and simply subtracted. A so-called *dark frame* is simply an exposure with the shutter closed. Typically a dark frame would be taken immediately after the image to be calibrated, and would have the same exposure time. One then simply subtracts the dark frame from the 'object frame.' Some consumer digital cameras automatically do this for long time exposures; one can tell because a 30-second exposure keeps the camera busy for a full minute. Astronomers instead cool their CCD cameras with liquid nitrogen, so as to render the dark signal negligibly small.

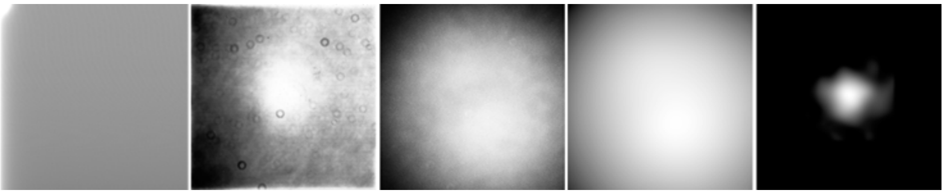


Figure 9.13. Some of the images used to calibrate the image in figure 9.12.

Once additive effects have been removed, the linear relationship becomes simply:

$$s = aE. \tag{9.4}$$

That is, signal is simply proportional to exposure. The constant of proportionality, a, then is related to the sensitivity of the detector; if a is larger, then a greater signal results from the same exposure. But the pixels of a CCD detector are essentially *separate* light detectors; they only have the same sensitivity insofar as they are manufactured exactly alike. Thus, the sensitivity of a real CCD varies significantly from pixel to pixel.

One can directly measure the differences in pixel sensitivity by taking a *flat field image*—an unfocused image of a uniformly illuminated, featureless white screen. If a particular pixel produces, for example, 1.37 times the signal of its neighbor, then it means that pixel is 1.37 times more sensitive. The second image in figure 9.13 is an example of a flat field image that was used for the calibration of the image in figure 9.12. Since sensitivity is a multiplicative effect, the image to be calibrated can be simply *divided* by the flat field image, in order to correct for these pixel-to-pixel variations in sensitivity.

It should be clear that if the response of the CCD were nonlinear, this simple flat field correction would not work. The sensitivity factor, a, in equation (9.4) is the *slope* of the response. But if the response is nonlinear, then there is no single slope, even for a single pixel. A nonlinear relationship between signal and exposure is a curve, not a straight line—and different parts of a curved line have *different* slopes. To put it differently, if a particular pixel produces 1.37 times more signal than its neighbor on a flat field image with a nonlinear response, it only means that pixel is 1.37 times more sensitive *for the flat field exposure*—but not for the image to be calibrated, which received a different exposure for those pixels.

The third, fourth, and fifth images in figure 9.13 are also multiplicative calibration images applied to the image in figure 9.12. They correct for such effects as uneven illumination of the CCD by the telescope optics and scattered light inside the CCD camera. The ability to accurately correct for these systematic effects depends crucially on the linear response of the CCD detector.

It *is* possible to calibrate a detector that has a significant nonlinear response, but in practice it can be too difficult to bother. Crucially, to correct for nonlinear effects requires far more experimental data. To correct for a nonlinear flat field response, for example, one would need to take many different flat field exposures, and fit a model to them. For an astronomer, this is a serious problem since the Earth rotates on its axis. Night does not last forever, and so only a limited number of calibration images can be recorded between the ending and beginning of astronomical twilight. A single night's data used in Beaver *et al* (2013), for example, required a total of sixteen hours to record 187 CCD images. Only 28 of those images (for a total of 30 *minutes* of exposure time) were of the actual astronomical objects being studied; the rest were for calibration. And that was for a detector with a mostly-linear response.

Once an image has been calibrated, it can be used for measurements. And so, for example, the brightnesses of the individual stars can now be measured from the right-hand image in figure 9.12. Scientific image calibration and measurement is usually performed with specialized software that maintains the mathematical integrity of the data, and keeps a running record of what calibrations have been performed. As discussed in chapter 9, section 9.3, the software must also keep track of the statistical noise in the data if measurements are to be meaningfully compared to each other.

References

Adler M, Mayo A and Alon U 2014 Logarithmic and power law input-output relations in sensory systems with fold-change detection *PLOS Comput. Biol.* **10** e1003781

Beaver J, Kaltcheva N, Briley M and Piehl D 2013 Strömgren H-β photometry of the rich open cluster NGC 6705 (M 11) *Publ. Astron. Soc. Pac.* **125** 1412

Bevington P R 1969 *Data Reduction and Error Analysis for the Physical Sciences* (New York: McGraw Hill)

Stroebel L, Compton J, Current I and Zakia R 2000 *Basic Photographic Materials and Processes* 2nd edn (Waltham, MA: Focal Press)

Chapter 10

Is digital manipulation cheating?

I have shown that digital manipulation is, in many ways, not as new as it is often made out to be, much of it being simply a different technical means to alter photographs in pretty much the same way photographers have been doing so for a century. And we have seen that there is much in common between digital and photochemical detectors such as film.

And yet, there is still something about it that rankles many. When certain pictures are disclosed to have been digitally manipulated, one sometimes feels cheated or let down. And so here I try to lay out a few of the issues as best as I can see them, to better identify where, when, and how that reaction arises, the better to judge under what circumstances it is or is not justified.

10.1 Paying one's dues

As we have seen in chapter 8, many of the techniques for digital image processing are far easier to accomplish than their traditional film counterparts. When a lot of effort is required to accomplish a task, that fact provides something of a brake. One thinks long and hard before beginning a task that is sure to be difficult. And so when doing something is *too* easy, one is not forced to think so much about whether it *should* be done, even if done well. And thus a common criticism of some examples of digital manipulation is a lack of good taste. The digital adjustment may have been done with expertize, but it makes the photograph worse and should not have been done at all.

This basic point can hold in a larger sense, however, regarding a photographer's body of work rather than just a specific picture. A lot of time is required to master anything beyond basic film and darkroom technique. The digital analogs of many of those same techniques, on the other hand, can be mastered quite easily. And so with much less knowledge one can produce photographs that are acceptable from a technical standpoint. And this may mean that one has spent less time contemplating

doi:10.1088/2053-2571/aaf0aech10

one's own work before showing it to others. And so one might say that digital photography in general is too easy.

But of course it need not turn out that way, and for many photographers it does not. Anything done well, including digital image processing, requires care and hard-earned skill. One can argue that the ease of many digital processes frees one to spend time on what really matters for a photograph. And so many photographers who use digital exclusively, *have* paid their dues, just in a different way than the traditional photographer.

10.2 Honesty

When one sees a picture and knows that it is a photograph, many expectations come with this. We often expect a photograph to be a literal record of an event—the taking of the photograph itself. We want to believe that a photographer was holding a camera and clicked a shutter while stuff was happening. And so how is one supposed to look at a photograph that could in theory have been taken just as shown, but was not?

As an example, there is a much-reproduced photograph from 1999, by the nature photographer Ralph A Clevenger, of an iceberg showing both the parts above and below the waterline (see figure 10.1). A few years after the original image was released, in a classic example of an urban legend in the form of an internet hoax, a version of the image appeared that was accompanied by a completely fanciful story of how the photograph had come to be. This hoax version of the picture was widely distributed, without the photographer's permission, on the internet along with the false description that the photograph had been snapped as-is by a rig manager from an oil drilling company. Every few years the picture reappears on social media with a similar false description (Snopes.com, 2018).

In fact, Clevenger made a composite of multiple photographs to portray a dramatic scene that could not have occurred in real life. The digital image processing used to make the composite photograph was an impressive technical feat at the time. And although the final image is digital, it was all done with scans from *film* photography. Clevenger was always completely honest about this; he described repeatedly and in detail that the photograph had been extensively 'Photoshopped,' and just how it had been done.

And so in this particular case, we really have two photographs—an honest one distributed by the photographer, and a dishonest one. The pixels are identical but we see them very differently.

To my eye, the iceberg picture, as presented honestly by the photographer, is an excellent photograph. And I am not at all put off by the fact that the photographer 'faked' real life. If this is somehow illegitimate *in principle*, then much of painting is also illegitimate. It is important to keep in mind that whatever else a photograph might be, it is still a bunch of marks on a flat surface that makes stuff happen in our brains when we look at it. No matter how representative the subject, we always 'make stuff up,' at least in part, when we make any picture. And although

Figure 10.1. *Iceberg*. Ralph A Clevenger, 1996. A composite of different photographs. As Clevenger explained upon the publication of the photograph, it would be impossible to view this scene in real life. *Photograph used with permission of Ralph A Clevenger*.

Clevenger's picture shows something that could not have been seen from any vantage point, it does nonetheless depict something real. It is an excellent picture partly because it manages to capture the essence of the subject (an iceberg), even as it shows a particular view that is not possible in a literal sense.

In a way, what Clevenger did with the iceberg has much in common with the long history (predating photography) of artistic illustration of scientific works. Look at the paintings of birds in the field guides of Roger Torey Peterson or David Sibley (see, for example Sibley 2002, Peterson 2002). They look unlike any particular individual view of a bird; there is no sense of a light source, for example. Yet any one of those paintings captures the essence of what that particular species of bird looks like in a way that no single straight photograph could, which is why many birdwatchers prefer these illustrated guides to ones that use actual photographs from life.

For me, not only is the internet-hoax version of the picture dishonest about its origin, the all-too-perfect story encourages the viewer to *stop* thinking. And it does worse than tell a lie about nature; it implicitly states that neither nature nor the truth matter a whit. The natural scientist in me finds this not only dishonest but also unspeakably arrogant. Clevenger's original version, on the other hand, is *evocative*. Even with no accompanying description of how the picture was taken, one cannot see it without immediately asking questions. And it is this asking of questions that leads us to see the world in a deeper and richer way.

10.3 Retouching

It can happen that a tiny detail of a photograph ruins the overall composition of a picture—'It would be perfect if not for that edge of a sign in the upper left corner ···' And so, it can be tempting to fix a picture by removing or altering minor details. This has always been a part of photography, and individuals differ as to whether (or to what degree) *retouching* destroys some important notion of purity for a photograph. But even some very famous photographers retouched shamelessly, not only removing detail they didn't like, but sometimes adding in new detail as well.

As an example, I have seen an original print of one of Bill Brandt's photographs which features a gull caught in mid-flight. The gull is slightly blurred by motion, and also just barely out of perfect focus, but Brandt used retouching to enhance the sharpness of the leading edge of its wing, so it forms a high-contrast sharp line. In my opinion, it was necessary for that particular detail to be sharp, and so Brandt's aggressive retouching improved the photograph greatly. Brandt in fact goes far beyond this retouching with paint; the photograph itself is a montage. The seagull was photographed separately, and added later to a landscape photograph that was taken under lighting conditions that would have rendered the sharp and distinct capture of the bird impossible.

Was it cheating? I think not. For one thing, he was not trying to trick us into believing he achieved this with his camera alone. Anyone curious as to how he managed to get that particular detail to be so sharp and high-contrast would do just what I did—move up close to the print and look at it carefully. And from a close inspection it is obvious that it was retouched with paint. Furthermore, he did not hide the fact that the photograph was a montage, for he unabashedly superimposed exactly the same gull photograph onto other landscapes as well.

But his techniques were considered controversial by some at the time, as they did not square with the purist conceptions of photography advanced by other influential photographers. Brandt's response to these criticisms was that, 'I am not interested in rules and conventions … photography is not a sport' (Brandt, 1948). It should be clear from my own photography that I essentially agree with his response. But it should also be clear that I would not take it to be *literally* true that 'anything goes,' and I doubt that Bill Brandt meant it that way either. Rather, he meant (I suppose) that whatever rules there may be, they are always open for debate, and it is the art that matters in the end.

All of these digital retouching techniques have their antecedents in the traditional black-and-white darkroom. But many find digital retouching to be even more controversial, and mostly for the reason I already discussed: it is too easy. Retouching a silver gelatin print with paint brushes and dyes requires a lot of skill and patience, and there is the very real risk that one will instead destroy the print. Digital editing is, in a sense, cost free; if you don't like the result, simply click the 'undo' button. And so one must always keep in mind that just because you *can* do something, it doesn't mean you should.

10.4 Digital filters and cliché

Digital filters have the ability to greatly alter the look of a photograph, sometimes in ways that are quite interesting. But since a digital filter is a mathematical algorithm that always does the same thing when applied in the same way, too heavy use of a digital filter can give a look of sameness to one's photographs. This is not to say that it is invalid for a photographer to specialize in a particular kind of look. Far from it! But there must be room for originality, and the particular look must have depth to it; ideally it should give the appearance that there is *more* there than what one literally sees in the picture.

And so, for example, when one makes an instant film transfer (see chapter 2, section 2.6), it usually looks like an instant film transfer. But every transfer looks different in unpredictable ways. It is possible to design a digital filter to make any digital image look like an instant film transfer. But when that same filter is applied to different images, they will likely have a sameness to them that will not stand up well, when compared to a set of real instant film transfers. Furthermore, the instant film transfer is not only an image; it is an *object*, made of dyes and silver on good watercolor paper.

When someone looks at a photograph and instantly recognizes the technique that was used to make it look as it does, different things can happen. If the technique is difficult, and the result of a not-completely-controllable *physical* process, recognition of that fact by the viewer can sometimes add to the power of the photograph. But if on the other hand, the viewer suddenly realizes that much of the photograph's appeal was achieved by simply clicking a button in GIMP, then the opposite effect is likely. It goes from 'That's so cool!' to 'Oh … that's the Van Gogh filter ….' See figure 10.2 for an example. *Cool* can easily turn into *cliché* if there is too much of it and it always looks the same.

Figure 10.2. The same image as in figure 6.4, run through the GIMP digital filter called 'Van Gogh.' Digital filters can alter the look of a photograph in interesting ways. But sometimes the result is more cliché than captivating.

References

Brandt B 1948 A statement on photography *Camera in London* (London: Focal Press)

Peterson R T 2002 *Birds of Eastern and Central North America* 5th edn (Boston, MA: Houghton Mifflin Company)

Sibley D A 2002 *The Sibley Guide to Birds* (New York: Alfred A. Knopf)

Snopes.com 2018 Giant newfoundland iceberg [https://www.snopes.com/fact-check/giant-newfoundland-iceberg/]

Chapter 11

The image, the object, and the process

11.1 Some preliminary ideas

Part of the goal of the three volumes of *The Physics and Art of Photography* is to explore the possible connections between physics and the *physicality* of a piece of visual art. Artists use the term physicality in a variety of ways, but it often relates to evidence of the physical in the medium itself. And so throughout, I have more often than not depended on examples of photographic processes that result in a rather blatant physicality.

But the *image* is important too, especially in photography, in a sense that is wholly separate from the physical embodiment of the medium. If the physical object *Mona Lisa* were to be tragically destroyed in a fire and so cease to exist, would the Mona Lisa cease to exist? Only a small percentage of the human population has ever seen the Mona Lisa. But it is also true that a rather large percentage of the human population has seen the Mona Lisa. The image is real too, apart from the physical object.

Finally, the *process* is intimately connected to both the image and the object, especially whenever either of the two has an obvious physicality. This is nowhere more evident than in the community of photographers who identify with so-called *alternative process* photography. And it is part of the motivation for my own term 'ephemeral process' as a (perhaps pretentious) name for the techniques described in chapter 2, section 2.5, later in this chapter, and the first two volumes of *The Physics and Art of Photography*. The process is the means, and both the image and the object are the ends. And an explicit focus on process recognizes the inevitable connection between means and ends—what a physicist might call cause and effect.

The following sections in this chapter can be seen as a set of *Gendankenexperiments* on the connections between process, the image, and the object in photography. The German term translates to 'thought experiment' in English, and it is often associated with the birth of modern physics in the late 18th and early 19th century. Sometimes one can learn much from carefully posing a question as a detailed description for *doing something*—even if that thing is impossible to do in practice—because the question

doi:10.1088/2053-2571/aaf0aech11

itself is clarified. This technique was used famously in the pedagogy of Einstein, Schrödinger and others in explaining the development of relativity and quantum physics.

In what follows I provide no answers, but I hope the examples call attention to, and better clarify, the questions. In this section, I begin with descriptions of some concepts that I believe to be useful for the discussion. I apply these concepts to some examples of work contributed by five artists in section 11.2, and in section 11.3. I consider these ideas in light of examples from ephemeral process photography and its close relative, the lumen process. In sections 11.4 and 11.5, I consider two examples that blur the border between photography and drawing.

11.1.1 Photographic and representational content

Some of the elements of a photograph relate directly to the straight-line rays of light passing from things in the world, redirected in an organized way by the lens and arriving on the surface of the film or other light detector. I call this *photographic content*.

But photographs can and often do contain elements that are not directly related to the photographic image. There may be *flares* of light that actually come from reflections and scatterings within the lens itself. The detector may not respond to light in a uniform way. Other processes may have added elements after the detector was exposed, as when we solarize a negative by flashing it with light.

- *Photographic content:* elements of the photograph that correspond to rays of light from the world, redirected in an organized way to an image on the detector.
- *Non-photographic content:* elements of the photograph that *do not* correspond directly to rays of light from the world, redirected in an organized way to an image on the detector.

This is not the same thing as ‘abstract’ versus ‘representational.’ Actually, I prefer the terms representational and non-representational, because the term *abstract* has a more general meaning. For example, one can identify abstract elements in any two-dimensional image, as discussed in Volume 1 of *The Physics and Art of Photography*. By *representational content* and *non-representational content* I mean the following:

- *Representational content:* elements of the photograph that correspond in a recognizable way to forms in the real world.
- *Non-representational content:* elements of the photograph that *do not* correspond in a recognizable way to forms in the real world.

Photographic content may or may not be representational. For example, an out-of-focus background is photographic, but it may represent only abstract shapes that are unrecognizable as forms in the world. Non-photographic content on the other hand is usually non-representational, since it does not arise from rays of light connected to objects in the world. But accidents happen, and it is possible even for non-photographic elements to make, or at least alter, representational elements. See figure 11.1 for an example. The 'Sun' in the picture is actually from a flaw in the hand-brushed emulsion.

Non-photographic content can be introduced either at the level of the detector, at the level of the printmaking, or both. Part of the appeal of alternative process techniques is that they naturally introduce non-photographic elements in ways that are only partially under the photographer's control, and this introduces an element of chance and the possibility of happy accidents. The photographer must still, of

Figure 11.1. The 'Sun' in this photograph is representational, since it seems to represent a form in the world. But it is non-photographic; it did not arise from rays of light redirected by a lens, but rather from an imperfection in the hand-coated emulsion.

course, have a good enough eye to recognize the many bad accidents when they see them, and throw them away.

11.1.2 The picture plane

The picture plane is the flat two-dimensional surface of the image or print itself and the marks thereon, independent of what forms in the world those marks might represent. It has long been recognized that there is a power in calling attention to the picture plane in a photograph, as a painter does with obvious brush strokes. Inexpensive plastic cameras have been popular among art photographers in part for this reason. They often leak light, for example, and this imparts a look that is as if some source of light is shining from a low angle onto a flat surface—which is exactly what happened to the negative inside the leaky camera. Another old trick is to file the machined edges of the film frame in a 35 mm camera, so the image has an irregular edge and maybe extends onto the sprocket holes of the film. An even simpler trick is to simply include the sprocket holes and edge markings of the film in the print itself, instead of cropping them off. See figure 11.2 for a rather clichéd example.

Other photographers would intentionally damage the negative after-the-fact, scratching it with sandpaper or placing pieces of cellophane tape onto it. All of these techniques put *into the image itself* a reminder that a flat surface was used to record

Figure 11.2. *Prairie Path* John Beaver, 2011.

the image. Not only is there the image, there is also a superimposed image of the flat negative itself.

An emphasis of the picture plane calls attention to the fact that there was a physical interaction between light and matter. This may or may not always be a good thing. But for the times when it is a wise choice, it is one of the goals of this book to offer some new tools for doing so, even if these new tools sometimes introduce the virtual into our aesthetic of the physical.

11.1.3 Control and happy accidents

It is possible to seek out processes that take away some of the photographer's control, that lead to 'happy accidents.' The idea of an accident implies that something *random* has been introduced (usually as non-photographic detail) to the picture. Sometimes this is a good thing, sometimes not, but there can be value in allowing for the possibility of the unplanned.

Figure 11.3 shows two examples. The image on the left was made from a transfer of Polaroid instant film onto watercolor paper; the bands at the top and bottom of the picture come from the fact that the Polaroid emulsion is larger than the print area. In a regular Polaroid print, the edges of the emulsion are masked off to form an even edge; for this picture I removed that mask before I made the transfer. It wasn't until later that I realized how well that feature works for this particular image, as it is echoed in the pattern of the coats of the two animals.

The ephemeral process photograph on the right was a long time exposure. I was unlucky enough to start the exposure right before the church service ended, and so people began to stream out the door and down the sidewalk, right past the camera. Their motion made them invisible, but one figure apparently stood almost perfectly still, right in the doorway for nearly the entire 10 min duration of the exposure.

Not all accidents are happy; most are, at best, uninteresting. But with an emulsion or digital detector that produces a 'perfect' image every time, there are no accidents at all.

Figure 11.3. **Left:** *Pig and Goat*. John Beaver, 2006. **Right:** *The Doorway*, John Beaver 2018.

11.1.4 Negative versus positive

It is often said that the invention of the first photographic negative process was important because it allowed one to make many identical prints from the same in-camera original. It is true that a negative of a negative is a positive—and so light can be projected through a transparent negative, in a second step, to make a positive print. But the key here is that the in-camera original is *transparent*, not that it is a negative. For it is also true that a positive of a positive is a positive. And so a *positive* transparency can also be used, in a second step, to make a positive print. It turns out, however, that the early direct-positive processes—daguerreotype, ambrotype, and tintype—were all opaque. And so two-step photographic processes have mostly been negative on negative.

Apart from the ability to make multiple prints from the same original, there are two additional important features of two-step negative processes. First, the step of transforming from in-camera original to the final print allows one to make adjustments; the print can be better than the original. Second, and less well understood, the process of transforming from a negative to a positive adds some psychological distance between the in-camera original and the print.

It is this last feature that I am most concerned with for this particular discussion. And so consider figure 11.4. Both images use the technique of transferring peel-apart instant film directly to watercolor paper. The transfer process introduces much non-photographic content, with all of the possibilities for happy accidents, and an emphasis of the picture plane.

The image on the left is from a digital scan of a 3.25 × 4.25 inch instant film transfer. It was a transfer directly from the camera to watercolor paper. Seen in the

Figure 11.4. **Left:** *The Red Table*, John Beaver 2013. Instant film transfer on watercolor paper. **Right:** *Valeria*, John Beaver 2007. Digital image made by scanning an instant film transfer on watercolor paper of a *negative*. The digital image was then inverted to a positive.

flesh, it is an intriguing (but small) image/object. The example on the right, however, was made by exposing instant film to a projection of a color film negative. The exposure was then transferred to watercolor paper in the same manner as the image on the left. But this image transfer was of a *negative*—and so I scanned the negative image transfer, and used GIMP to invert the resulting digital image to a positive.

One could make a large digital print from either of these images. But the result for the left-hand image would likely look just like what it is—a *photographic reproduction* of a small image transfer on nice paper. But for the right-hand image, the act of moving from a negative to a positive establishes psychological distance from the original (negative) image transfer on paper. And so it has many of the features of the left-hand image—non-photographic content, accidental features, emphasis of the picture plane. But it is now its own thing; it no longer looks like a magnified reproduction of a piece of paper with an image on it.

The image on the right is just an image, and the original instant film transfer negative used to make it is not particularly interesting in and of itself. But on the other hand, there exists an object in a frame, on the wall of a friend's house, that looks much like the image on the left. And that object is a nice piece of Arches Aquarelle, underneath an image made of the very molecules of silver and dyes that were in the piece of Fuji FP100 film that was in my Polaroid 195 camera, when I pointed that camera at a red table and tripped the shutter. Which is best?

11.1.5 Order, complexity, and randomness

It would seem that there is a range of possibilities between the completely uniform and predictable on the one hand, and the completely random and unpredictable on the other. Consider the silver gelatin emulsion that is used for traditional film photography. It is engineered and manufactured to be completely uniform in sensitivity and thickness at the scale of the photographic image. But when seen under a microscope, it is comprised of individual grains of silver halide that are arranged *randomly*.

But it is not so simple as that. Figure 11.5 shows four patterns. The first example is simple in the sense that it would require few instructions to describe how to reproduce it. The second example was generated by a particular mathematical algorithm—a set of step-by-step instructions—called a Julia-set fractal. It has the appearance of complexity, but the algorithm used to generate it is surprisingly simple, when expressed in the appropriate mathematical language. And so a kind of complexity arises from a set of simple rules.

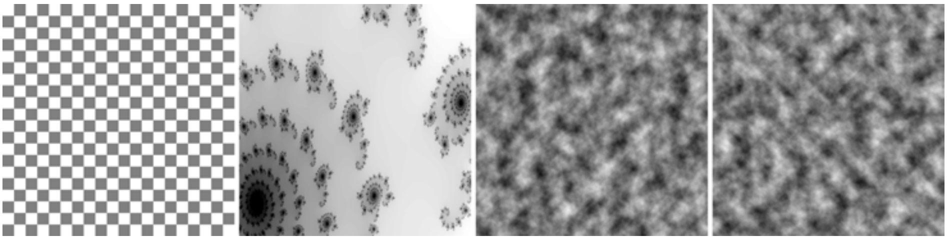


Figure 11.5. Examples of complexity and randomness generated by different mathematical algorithms in GIMP.

The third example in figure 11.5 is fundamentally different from the first two; it has an element of randomness. A simple algorithm takes as an input a list of random numbers in order to generate the specific details. If one were to use the same algorithm, but with a different input list, a different pattern would result. And so there is a sense in which this pattern is more complex than the others; to reproduce it one would need not only the algorithm, but also that particular long list of random numbers.

The fourth example was generated with the same algorithm as the third; only the list of random numbers was different. The third and fourth patterns are very different from each other; to reproduce each in detail would require a long set of instructions, and most of those instructions—the long list of random numbers used—would be different for the two patterns. But we see those differences as unimportant, and so the patterns appear to us as *essentially* the same, even though the overwhelming majority of the details are different.

Randomness is associated with unpredictability—but the relation between the two ideas is subtle. In the last two examples in figure 11.5 the lists of random numbers were generated by a computer. Given one of the lists, one cannot predict the other, and in that sense they are random. But given exactly the same instructions, the computer will generate exactly the same list.

Nature, however, can produce true randomness via the rules of quantum physics, as described in chapter 5. And so it is possible (but quite difficult) to make a device that takes advantage of the laws of quantum mechanics to generate a string of truly random numbers (see, for example Chen 2018). But it is easy to produce a near-randomness—simply find a physical system for which the end result depends on many overwhelmingly-complex critical details along the way. This is called *sensitive dependence on initial conditions*, and flipping a coin is a good example. The result of a coin flip is not *truly* random. If all of the forces acting on the coin were reproduced in precise detail every time, the same side of the coin would come up with every flip. But those details are so complex and, in practice, *uncontrollable* that the result is, at casual inspection, as good as random. Thus, a complex physical system can be used to produce a seeming randomness even when the strange, mostly subatomic, world of quantum mechanics is not a significant factor.

The left-hand image in figure 11.6 shows a tiny detail of the image on the right-hand side of the same figure (changed to black and white for comparison with figure 11.5).

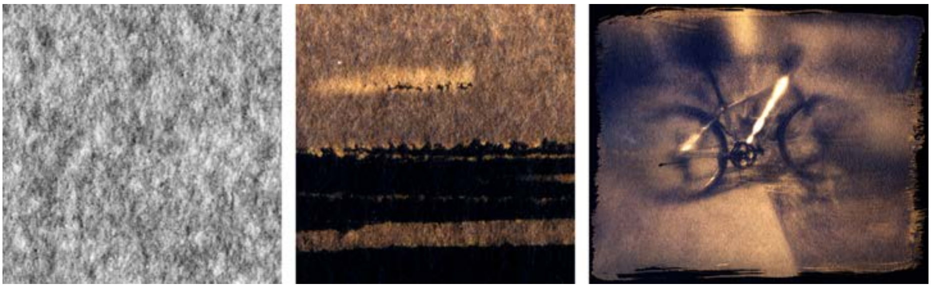


Figure 11.6. Magnified details from the cyanonegative image at far right, showing non-photographic detail that is unpredictable, but not quite random.

There is certainly a kind of randomness to the complex pattern of light and dark. But that randomness comes not from computer-generated random numbers, but rather the almost unimaginable—and thus unpredictable—*complexity* of the arrangement of the fibers in this piece of high quality watercolor paper.

The center image in figure 11.6 is also cropped from a tiny portion of the right-hand image. It has added elements that arose from the hand-brushing of the cyanotype sensitizer used to make the picture. Although these elements of the brushwork are unpredictable in detail, they clearly depart significantly from 'random.' In their complex connection to a physical act (human-controlled in this case), they could not have been predicted exactly as shown. But they can be *understood* after the fact.

11.1.6 The new antiquarian movement

Much of the history of photography has consisted of attempts to exert greater and greater conscious control and to *eliminate* randomness and non-photographic content as much as possible. Digital photography has taken this process to even higher levels, using the same detector over and over to give precisely repeatable results. But for photography as an art, this is a mixed blessing.

Lyle Rexer in his important book, *Photography's Antiquarian Avant-Garde: The New Wave in Old Processes* (Rexer 2002), describes how the standard gelatin silver negative photographic process became, over the decades, more and more flawless and exactly reproducible. His book showcases contemporary art photographers who consciously seek out techniques that subvert that perfection and exact reproducibility—a modern renaissance of sorts in pre-gelatin-silver techniques from the earliest days of photography. Thus, we have a *new antiquarian* movement, seeking out quirky and partially uncontrollable techniques that can provide not only a new look, but also the happy accidents that subvert the inherent sameness of the photographic process.

Compared to digital photography, even the traditional silver gelatin process has a physical elegance to it; there is a continuous flow of interactions between light and matter. The paths of light rays from the world are altered by geometry and the glass of the camera lens, to fall in an ordered manner onto a flat surface. And that flat surface undergoes an invisible physical change as a result in the formation of a latent image. The direct action of chemistry brings these hidden changes forth as a visible image, and a transparent negative is made. In the darkroom, this negative alters light passing through it from the photographic enlarger—a sort of camera in reverse. Modulated by the negative, this light is focused by a lens to form an image onto another piece of light-sensitive material. And it too is physically changed by that process, and a print—marks of dark silver on white paper—is made. Part of the appeal of the New Antiquarian is that this continuous interplay between artist, medium, and process is even more direct still.

11.1.7 The archival ethos

Any particular photographic print or direct positive is an *object*. But what kind of object will it be in 200 years? The idea that a photographic print at its best should be *archival*—relatively permanent—has long been a value among photographers. There is no official definition of archival; there is no standard answer to the question how permanent is permanent enough. But great pains are often taken to ensure that a photographic print can last as long as possible. The silver gelatin printing process, for example, requires special materials and processing techniques to produce a print that is considered archival (see for example London *et al* 2005, p 155).

Some of the appeal of alternative processes is undoubtedly the relatively permanent physical objects they make. I was fortunate enough to see firsthand (thanks to the wonderful staff of the Royal Botanic Garden in Edinburgh, Scotland) one of the 17 known copies of *Photographs of British Algae: Cyanotype Impressions*, published between 1843 and 1853 by Anna Atkins. Widely held to be the first photographically-illustrated book, these direct cyanotype photograms from the mid 19th century are still achingly beautiful, comparable in tone and contrast to many cyanotypes made today.

Some alternative processes do not follow this ethos of the archival. Or perhaps instead, they subvert that ethos, and therein lies the appeal. The anthotype, for example, is faded by the very light one must use to see it, and this gives it an endearing, ephemeral quality.

Archival considerations are also considered to be important for digital printing. After all, no one wants to buy an expensive art print only to have it fade or fall apart in a few years. But one can also consider the archival nature of the digital image itself. The central idea of digital storage is that the physical storage mechanism is only used to save a code—for a computer it is usually a string of ones and zeros. This means that the *physical* storage medium can be allowed to degrade up to a point—so long as the ones are still recognizable as ones and the zeros as zeros. Since the image is in the code, not in the physical medium itself, a degradation of the medium may produce no degradation at all in the image.

Of course if the medium degrades *too* much, then it may be impossible to tell ones from zeros, and then the image is lost. But unlike an image that is also its own physical medium—a cyanotype for example—it is possible for a digital image to be immortal. Simply copy the digital code to a fresh medium every now and then, before the medium is allowed to degrade too much. If one were to do this often enough, the image could be preserved forever.

This is nothing new. Text that is hand-printed on paper, for example, is essentially *digital* storage; the content depends not on the precise shape of any given character, but rather on which letter of the alphabet it represents. Paper and ink degrade, and so any individual piece of written text will not last forever. But one can simply recopy the text by hand onto fresh paper before the old piece becomes unreadable. This process of regular hand-copying is in fact why many ancient texts still exist today.

11.2 Four photographers and a musician

I here consider several examples from five artists whose work I admire, in light of the issues raised in section 11.1. All of these artists do work that is outside the

Figure 11.7. Chlorophyll print from *Growing Concerns*, Almudena Romero 2018 (https://www.almudenaromero.co.uk/growing-concerns). Photograph used by permission of Almudena Romero.

commercial mainstream of photography, and they employ techniques that highlight the interplay between the image, the object, and the process.

11.2.1 Almudena Romero

Almudena Romero is a London based visual artist who works with a wide range of photographic processes, including many antiquarian printing and image capture techniques such as cyanotype, salt printing and wet collodion, but also new technologies such as 3D scanning and printing. Romero's work has been exhibited widely over the past decade in the UK, Spain, France, Italy, and China.

Figure 11.7 shows a chlorophyll print (see chapter 3, section 3.6) by Romero. She has used this process to print directly onto the leaves of plants that are native to the countries that provide the subject matter for the photographs. She thus forms intimate links between the photographic process, its context in the world, and the object of art that results.

Although enlarged reproductions of Romero's chlorophyll prints are compelling as images, they have even more significance when seen as the objects they are. Figure 11.8 shows one of her chlorophyll prints in the context of her 2018 London installation Growing Concerns. Setting the prints among tropical plants calls attention to the process used to make them.

Chlorophyll prints are a variety of anthotype, and so remain somewhat sensitive to light. They are ephemeral, and to display them is to damage them. In contrast to this, Romero's *Growing Concerns* exhibition also included ambrotypes—the direct-positive variation of the wet collodion process (see chapter 3, section 3.2). These original image captures were made with an 8 × 10 view camera on sensitized glass plates. They go from the camera directly to the gallery wall, and are about as heavy, solid, and permanent as a photographic process can be.

11.2.2 Caitlin Noll

Caitlin Noll is a photographer based in Jacksonville, Florida. Her experience as a commercial photographer has implications for her art photography; she is equally

Figure 11.8. Chlorophyll print from an installation of *Growing Concerns*, Almudena Romero 2018 (https://www.almudenaromero.co.uk/growing-concerns). Photograph used by permission of Almudena Romero.

Figure 11.9. *Untitled*, from the series Hypocritical Herbology, Caitlin Noll, 2018. Image used by permission of Caitlin Noll.

comfortable with analog and digital technologies, and uses both according to the needs of the project.

Noll uses the camera-less lumen process (see chapter 2, section 2.4) to make source material for up-front digital compositions. In figure 11.9, the strict, digitally-imposed symmetrical forms play off the slight asymmetry of the natural leaf object, and the much more random complexity arising from the overt physicality of the lumen process. This blend of (and tension between) science and art emphasizes the relationship between humans and nature, a core focus of her work.

The combination of digital and analog techniques emphasizes process. The end result is a digital image and digital print—but the physicality of her variations on the lumen process source is part of that image too. She describes her process as follows (Noll 2018):

> My process for lumen prints is tedious, but effective. I … start with the basic lumen process by taking a desired plant and placing it on the light-sensitive paper. I will add texture and create different hues by taking rubbing alcohol, peroxide and/or any man-made household product that's easily available and splash it over the plant and on the paper; each chemical reacting differently to the plant and paper. Once I've exposed the paper, I scan it twice—both before and after fixing. Taking the two scans, I'll combine them on Photoshop and they create a symmetrical and saturated image. I do tweak a few more things digitally, but I don't over-manipulate it. Rarely, do I ever stick with the simple structures of the older processes, I like to expand on what I'm doing and make my images more contemporary by including digital manipulation.

11.2.3 Diane Fenster

Based in California, Diane Fenster is a wide-ranging art photographer and photo-illustrator who for nearly 30 years has exhibited extensively, both nationally and internationally. She has won many awards for her work, and is widely sought as a lecturer and exhibition curator.

Fenster has worked with a wide variety of computer and photo-based processes, but figure 11.10 shows one of her more recent lumen-based works, *The Holly King*

Figure 11.10. *The Holly King and the Oak King*, Diane Fenster. Image used by permission of Diane Fenster.

and the Oak King. There are many process-oriented features evident here. Lumen process in its 'pure' form is object-oriented; the result is a one-of-a-kind, archival silver gelatin emulsion with an image on it. The process includes both intentional hand work (the placing of objects) and an uncontrolled (by the photographer) physicality arising from the process itself. But there is much more going on in Fenster's work.

There is a complex interplay between the image and the object in this piece. Digital negatives were used for the original lumen exposures; the images themselves could have any photographic process as their origin. In addition to the contact lumen print exposures made with the negatives, Fenster did hand work, and apparently included hand-placed botanical objects as well. But the lumen prints were not fixed; they were scanned unfixed to make digital prints. Part of the motivation for this might be because an unfixed lumen has a very different look. But for Fenster, the concern is also about the process and its relation to the artist as a human in the world; to chemically fix a silver gelatin emulsion requires a significant amount of water in drought-ridden California. But the object is important too, and so the digital print is physically altered in this work by encaustic techniques (Fenster 2018).

11.2.4 Chrystal Lea Nause

Chrystal Lea Nause is a New Orleans, Louisiana based award-winning visual artist and anthropologist who does both commercial and art photography. Her photography has been juried into many exhibitions, both online and in galleries in Illinois, Vermont, Oregon, Kentucky, Tennesee, Missouri, and China.

Her fine art photography usually starts with a traditional 4 × 5 large format view camera. She uses ordinary large format film, but also (no longer made) Type 55 Polaroid instant film. She includes the edges of the instant film negative in her prints, and the image detail there is overlaid with many imperfections. This not only provides an interesting frame but also draws attention to the picture plane, a fact noticed by many other photographers who used (and now very much miss) this film (see O'Brien and Waits 2011, for another good example).

But Nause sometimes makes atypical use of this film. In the left-hand image of figure 11.11, the slight tilting of the photographic image and the irregular corner resulted from a flaw in the instant film processing as it was pulled through the metal rollers of the special film holder—a not-uncommon occurrence with this type of film. But also, the film partially and unintentionally solarized. In this case, both of these accidents are happy ones, as she explains (Nause 2016):

> I think that the image of McAndrew Stadium is a great study of transitory places, since it was demolished less than a week after the image was created. The negative solarized spontaneously, so I was never able to do a direct print of it. I'm guessing that the solarization was a by-product of the film's age.

Nause uses traditional wet-chemistry darkroom techniques to make, directly from her negatives, archival, toned silver gelatin prints of the finest quality. But this

Figure 11.11. **Left:** Chrystal Lea Nause, *McAndrew Stadium, Southern Illinois University. Carbondale, Illionis,* 2011 (courtesy of the artist). **Right:** Chrystal Lea Nause, *Vecchia Stazione Ferroviaria. Manfredonia, Italia.* | *Old Train Station. Manfredonia, Italy,* 2007 (courtesy of the artist).

particular negative was essentially unprintable in this way. And so a print of this negative is available only by way of the intermediary of digital scanning and printing. She has this to say about the right-hand image in figure 11.11 (Nause 2016):

> I made the photo about 10 minutes before my train left the station. From there, I was rushing back to my apartment to gather all of my possessions and begin the less-than-glamorous journey back to Rome for my departing flight. The negative sloshed in the Sodium Sulfite fix and unfiltered water for about 3 days. All of the chemical deterioration and scratches are a result of these 'rough' travels. The perfectionist in me cringed at my mistakes—but I've really come to love that this image alone can represent my exploration of southern Italy; rough travel conditions, missed connections, sleeping on my backpack in empty stations, exploring the rural mundane of the south, and really getting to know myself as a photographer.

11.2.5 Hal Rammel

Hal Rammel lives in Wisconsin and is an accomplished photographer specializing in a variety of inventive, process-oriented analog techniques: pinhole, stereo, and camera-less photograms being a few examples. His wide-ranging visual art includes unique, functional objects as well, including handmade stereoscopes for his stereo images, and his work has been exhibited in galleries both nationally and internationally.

But Rammel is equally well-known as an innovative improvisational musician and inventor of musical instruments. Two examples of one of his musical inventions can be seen in figure 11.12. His *amplified palette* is both visual art and musical instrument. He describes it as follows (Rammel 2008):

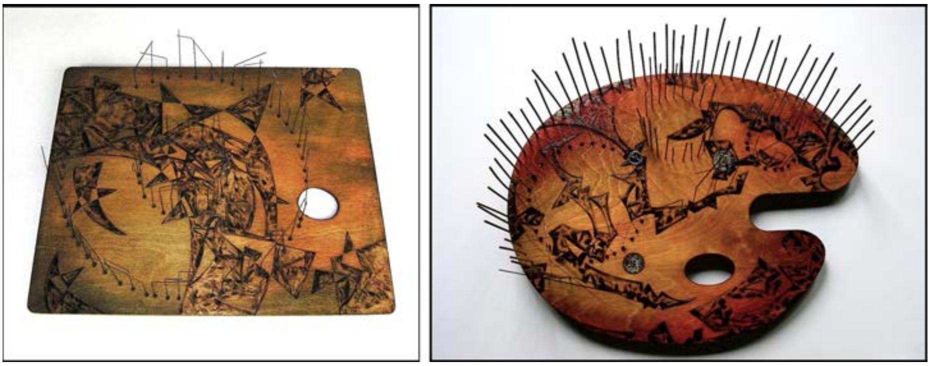


Figure 11.12. Amplified palettes by musician and visual artist Hal Rammel. Left: Hal Rammel, 2010. Amplified palette, wood and metal, 11 × 9 inch. Right: Hal Rammel, 2014. Amplified palette, wood and metal, 13 × 10 inch. Images used by permission of Hal Rammel.

> This new invention took the form, obviously, of a painter's palette, ergonomically designed to be hand-held, light, and overall perfectly fitted for—in this case—a miniature version of table-top electro-acoustics. A contact microphone is fixed to the back in order to amplify sounds that otherwise would be too quiet and too insubstantial for music-making. Amplified, these sounds—produced by thin wooden and metal rods—are rich and full, capable of making expressive music. I'm not interested in exploring many of the conventions of music form, like chords, harmony, melody, and so on. Therefore, I cut the rods in random lengths, juxtapose different lengths and sizes in intense proximity, add elements to the design that make it difficult to play individual pitches or melodic lines; and then attempt to wrestle musicality out of the result.

The construction of Rammel's instruments and his methods for playing them have expressive power, but he has less *conscious* control than most performers. Instead, many of the details of the music arise naturally out of the physical nature of the instrument and the natural vibrations of the often-flexible mallets he uses to play them. His performances (even on recordings) are improvised, but the music is nonetheless highly complex and structured, and closely integrated to the *process* of designing, building, and altering the instruments themselves (Rammel 2007):

> I've let the form of the palette be my launching point for opening possibilities and followed the road of possibilities. That is the advantage of mixing media, that music might be shaped by a sculptural form rather than the instrument shaped to play any preconceived notion of what is musical.
>
> I can't isolate improvisation simply to performances. The entire process of taking an idea, building, refining, playing, performing, redesigning, rebuilding to create obstructions to newly defined playing methods are all part of a fluid process of invention and discovery.

Photography is an art of the interplay between three-dimensional and two-dimensional space, while music is the art of time. Throughout *The Physics and Art of Photography* I have paid special attention to photographic processes that intentionally seek out possibilities for the physical to have its own say in the art, apart from the conscious control of the artist. Hal Rammel does this with both his visual art and his music—and he is not alone. Apart from the contemporary improvisational musicians he often performs with (such as Linda Binder and Matt Turner), Sun Ra and Harry Partch would understand. The popular, but difficult to classify, musician Tom Waits was quoted as saying (I can no longer find the source), 'I like to hear the hair on the music.' Some photographers, perhaps, like to see the hair on the negative.

11.3 Examples from lumen and ephemeral process photography

The ephemeral process (EP) technique described in chapter 2, section 2.5 and volumes 1 and 2 of *The Physics and Art of Photography* allows for many ways to investigate the interaction between the image, the object, and the process. It also has a deep connection to lumen prints, and the similarities and differences are worth exploring, in light of the issues raised in section 11.1.

11.3.1 EP pictures from pictures

Like many photographers my age, I have boxes and boxes of small chromogenic prints, printed commercially from the hundreds of rolls of color film I have taken over the years. Ephemeral process techniques allow me to mine these prints as source material for new images, by making EP contact *negatives* from the positive prints. The result is almost certain to be different from the original. The print is placed face-down, in contact with the accelerated enlarging paper, and the light source passes through the back of the color print for the exposure. A brief illustrated guide to making EP images in this way can be found in appendix A.

As described in more detail in volume 2 of *The Physics and Art of Photography*, EP photography can result in a duotone color image, even though black-and-white enlarging paper is used. With the right source of light and choice of paper and accelerator, there can be an odd synergy between the printed-out scattering colors that may appear in the enlarging paper and the colored dyes used in the chromogenic printing process. The resulting color palette, although clearly related to the colors of the source print, has a quite different feel to it.

Since the exposure is only intense wherever the accelerator is applied, the original picture can be re-composed to make something new. Accidents are easy to come by, and some of them are happy (many are not). The accelerator must be hand applied to the paper in dim working light, working quickly. And so it is easy to end up with something that is not quite as planned.

Figure 11.13 shows two examples. For the image on the left I intended the accelerated area to be entirely underneath the color print during the exposure. But some strayed beyond that boundary, making a stark white mark at the bottom of the picture from the directly exposed accelerated region (remember, this is a negative

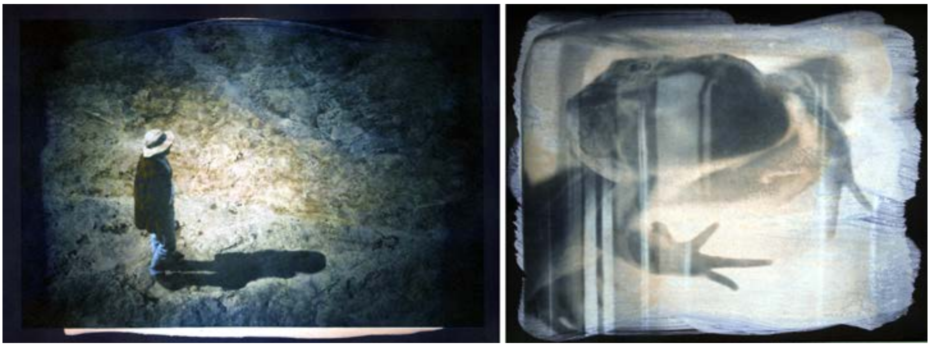


Figure 11.13. **Left:** *Doug on Mars*, John Beaver 2018. **Right:** Preserved Frog, John Beaver 2017.

Figure 11.14. **Left:** *St. Mary*'s Church #2, John Beaver 2018. **Right:** *Blue Dandelion*, John Beaver 2016.

process). That feature may or may not be a good thing; if I really wanted to find out, I could do it again many times, and each time would be different. But its unexpected addition to the image forced me to see the picture in a completely different way than I otherwise would have.

The original print for the image on the right was very different. It had a portrait orientation, not landscape, and it was of a preserved frog upside down in the bottom of a jar. By selective placement of the accelerator, I used only the bottom portion of the original picture for this image, and then turned it upside down.

The photogram (shadow print) technique can be combined with the exposure of the print. And so before beginning the exposure for the left-hand image in figure 11.14, I sandwiched blades of grass between the color print and the light-sensitive paper. The colors come out as a duotone with the hues correct on the negative, which makes them backwards upon inversion to a positive. They can be switched back, as in both images in figure 11.13, or they can be left reversed as in those of figure 11.14.

A good picture can sometimes be made from a bad print. The left-hand example in figure 11.15, was made from a disappointing picture that I had stashed away and nearly forgotten. But none of its flaws mattered for the result shown here; in particular, non-photographic content is added to what was a featureless,

Figure 11.15. **Left:** *The Wildcat*, John Beaver 2016. **Right:** *Vali and the Leaves*, John Beaver 2015.

overexposed sky in the original. On the other hand, a color print that is excellent on its own terms may not produce anything interesting when used as a source material for an EP image.

The original print used as source for the right-hand image in figure 11.15 also had its problems. There was much I liked about it, but the composition did not quite work. This EP image derived from it has more unity, with the simplified color palette better drawing the connection between the leaves and the clothes of the figure, and the overall shape of the illuminated region (from the hand-brushing of the accelerator) echoing the pose of the subject. Neither of those features were part of my conscious intent when I made the picture.

The color prints from commercial photo finishers typically have back printing—the date and watermark of the brand of paper used, perhaps, or maybe a code that makes it easier for them to find your negative if you ask for additional prints. Since the exposure is made by shining light through the print (from the back), then these features will end up in the EP image. (Close inspection reveals a faint example in the right-hand image of figure 11.15.) The result may be simply distracting. But in certain cases it may add to the image, grounding it by calling attention to the picture plane and the world of thoughts and words that exists outside the image.

Figure 11.16 was made from a contact print with a drawing, rather than a print. A photographic contact print made with a transparent (or translucent) handmade drawing is often called a *cliché verre* (James 2016, p 506) and the technique can be used with any contact-printing process. But there is a twist in this case; the drawing was made directly from a black-and-white photographic film negative, using the drawing-from-negatives technique described in section 11.4. And so it is a print in a book (or an image on a screen if you are reading the e-version) of a digital positive

Figure 11.16. An EP image made in the same manner as those of figures 11.13–11.15, but with a pencil drawing on paper, instead of a color print, used as the source material.

made from a scanned paper negative made from a drawing made from a film negative exposed in a camera. So, is it a photograph?

11.3.2 Limited edition prints from EP negatives

The right side of figure 11.17 shows an EP photograph, made from a scan of an in-camera 3 × 4 inch negative, exposed in a homemade camera, in the manner described in volume 2 of *The Physics and Art of Photography*. Or rather, it is a reproduction of that digital image on either a printed page or a lighted screen, depending upon whether you are reading a printed copy or an ebook.

Digital image capture is the only way I have found to effectively capture the subtle range of values and hues and make an image from these unfixed paper negatives. Because the unexposed silver halide crystals have not been removed from the emulsion by fixing, the light areas of the negative *reflect* a lot of light, but they don't *transmit* very much.

Thus, one cannot simply place the EP negative in an ordinary photographic enlarger and project it onto light-sensitive paper to make a traditional black-and-white silver gelatin print. But there is a way, and I have tried it; one of my most successful results can be seen on the left in figure 11.17, made directly in the darkroom from the same EP negative after it was scanned to produce the image on the right.

I was able to make a silver gelatin print directly from the EP negative by retrofitting my enlarger for opaque projection. I removed the light source and condenser from the enlarger (see figure 8.2), and placed the paper negative on top of the condenser focus bellows, instead of beneath it as ordinary film is intended to be

Figure 11.17. The image on the left is a reproduction in this book made from a scan of a silver gelatin print made in the darkroom from an EP negative. The image on the right is a digital image from a direct scan of that same EP negative.

placed. I then mounted small LED lights inside the condenser bellows, pointing up toward the negative and away from the lens. And so the light source reflected off the EP negative, instead of passing through it.

Once the exposed EP negative is washed and dried, it is far less sensitive than the developed-out enlarging paper used to make the print. And so although the negative is exposed to much brighter light than is the paper for the print, it is still possible to make a print without damaging the still light-sensitive negative too much. That is, if one makes only a few prints in total, *including the test exposures* that are usually necessary to determine the proper exposure and contrast. This is made far more difficult by the fact that it is a moving target. The second print is made with a negative that has been altered by the light used to make the first print. And even the first print is made from a negative that has been altered by the light from the test exposures.

My initial experiments proved that it is possible to make at least a few silver gelatin prints from an EP negative. For the example in figure 11.17 I succeeded in making two prints that are perhaps at least non-awful, before the negative became too fogged by the light of the exposures. I know that with more work, I could do better. But to make a series of *very good* prints in this way would require much planning and experimentation; it is probably well beyond my modest skills as a darkroom technician. One intriguing feature, however, is that it would be—by necessity—a limited edition; the act of making the edition of prints destroys the negative used to make them. And even with the best of technique, each print would almost certainly be different.

The version on the right side of figure 11.17 is only an image, while that on the left is of a physical object, made of silver and gelatin on pressed wood pulp (paper). Which is better, and why? Can one confidently answer that question without both holding the physical version in hand, and seeing the digital version rendered in various contexts other than this book? Does it matter that only two physical prints exist, and that I have no way to make more (because the act of making them destroyed the negative)? Does the *very existence* of the two (and only two) physical silver gelatin prints render a print made from the digital version to be less valuable, because it would be only a reproduction of the two 'originals?'

11.3.3 Ephemeral prints

Ephemeral process techniques can be used to make positive prints, by exposing them in contact with negatives. These EP prints could be scanned to produce a positive digital image that could be printed much larger than the original negative. One can, however, instead think of the resulting EP positive print as the final image—an image that is also an ephemeral object.

Figure 11.18 shows an EP contact print made directly from a 35 mm color negative. I deliberately chose an enlarging paper that produces strong duotone colors in order to make a color print from the color negative. It is kept in a little match box because it is, of course, still sensitive to light; to look at it is to damage it. Even the match box has too many light leaks to keep the image safe for long term storage. So it and several others are kept in a larger light-tight box; I take them out every now and then to peak at them in dim light.

Ephemeral process prints can be made from any black-and-white film negative, but I find the results most intriguing if color can be reproduced. As with other examples of color images in EP photography, even though it is a negative process the

Figure 11.18. *Fútbol, Ipanema*, John Beaver, 2016.

hues match those of the exposing light directly, rather than the complements of those colors. A color negative uses complementary colors to the final result, because these will revert to the color of the original light when the negative is used to make a positive in the ordinary chromogenic process. And so the colors in a color film negative are the complement of what they are in the world. And this means that an EP positive print has reversed colors if it is made from direct contact with a film negative. This hue reversal is evident from the blue sand and pink sky in figure 11.18.

There is a rather serious impediment to making EP prints from color film negatives. An ordinary color negative has an orange mask to the plastic film, and this transmits little of the short wavelength light to which the enlarging paper is sensitive. And so the result tends to be of very low contrast and completely monochrome. The success of figure 11.18 is because of a trick. The negative was made not from ordinary color negative film, but rather from positive transparency film (slide film) that has been *cross processed*.

As described in chapter 2, section 2.3, the silver gelatin process is inherently negative, but it can be made positive by reversal processing. And so the same emulsion can be processed normally to produce a negative, or it can be reversal processed to form a positive. But there are many other technical details regarding the specific engineering of photographic emulsions, and so there are design differences between chromogenic films meant for reversal processing and those intended to be processed as negatives. In particular, color films meant for reversal processing do not have the orange mask of negative films. The negative used for the print shown in figure 11.18 was made from film that was intended by the manufacturer to be reversal processed to a positive; but instead I had it processed as if it were negative film. This produced a color negative without the orange mask.

Even with cross processing, the precise colors that result depend critically on a complex combination of the specific brand and type of film used, details of the negative processing, the accelerator formula, the light source for the print exposure, and the type of enlarging paper used. After many experiments, I have had far more failure than success, regarding the production of an ephemeral print with intriguing colors made directly from a film negative.

A much easier way to make an ephemeral color print is to make one's own negative by printing on transparency film with an inkjet printer. High quality transparency films are available, and inkjet prints on these films are often used to make contact prints with alternative processes such as cyanotype or platinum printing. Furthermore, one can print the negative large to make a large print, and it can be made from any digital image regardless of its origin. With photo editing software, the image can be inverted to a negative, and then the hues can be reversed before printing; the hues then come out correctly on the print. See figure 11.19 for an example comparing the original digital image, and an EP print made from it.

Unlike for a film negative, the color and contrast can be controlled digitally at the negative stage in order to produce good results with a particular combination of light source and enlarging paper. The contrast—in both hue and value—can be enhanced by carefully stacking two identical negatives for the contact print. For the examples I have tried, the most intense colors on the print come from the blues and

Figure 11.19. A digital image (left) can be inverted to a negative and the hues reversed before printing on transparency film with an inkjet printer. An EP contact print (right) can be made from the negative. With the right choice of enlarging paper and light source, the print may have subtle duotone colors that relate directly to the original image color.

Figure 11.20. Left: A color transparency printed with an inkjet printer from a digital negative image. The greens were removed from the image before printing. Right: Direct EP contact print made from two layers of the transparency at left.

yellows in the negative; greens produce the least intense colors. One can take advantage of this by reducing the green in the color negative before printing; more intense duotone colors result in the print. See the right side of figure 11.20, and at left, the negative transparency used to make the print.

I sometimes imagine, half tongue-in-cheek, an exhibition of these ephemeral image/objects, of various sizes and subjects. The visitor to the gallery is greeted by

Figure 11.21. Left: Unfixed, ephemeral process print made from a color digital negative. Right: A similar print, but after fixing. The exposure was greater than for the left-hand print, because the fixing process reduces the density considerably.

nothing but blank walls and a scattering of simple boxes on floor pedestals. Next to each *sealed* box is a label—with the usual listing of title, medium, artist, and date—and a framed critique of what is inside. But to see the picture, one must buy it.

It is possible to fix an EP print to make it permanent, and an example can be seen on the right side of figure 11.21, with an unfixed example from the same negative on the left. We could call this an 'accelerated, unfixed lumen print,' instead of a 'fixed ephemeral process print.' In all of my tests, the image hues disappear as a result of the fixing process, although some papers retain colors that are unrelated to the colors in the negative. And so fixed EP prints are better suited to black-and-white negatives. These have their own appeal, and it is an easy technique if one wants to produce a positive print with the appearance of a hand-brushed emulsion.

11.3.4 To …, or not to …

There are many possible variations on the theme of the ephemeral process, lumen process and combinations thereof. We can choose to:

- Use an accelerator, or not.
- Expose with a positive to produce a negative, or expose with a negative to produce a positive.
- Produce normal color hues, or inverted color hues.
- Scan the result to make a digital image, or let the exposed enlarging paper be its own image/object.
- Fix the printed-out enlarging paper to make it permanent, or leave it unfixed.

Which combination is best? It should be clear that I believe there is no correct answer, in general. Rather, we should look at these possibilities as artistic tools—some more appropriate, some less so—for whatever task may be at hand.

Consider figure 11.22. The same EP photogram has been simply scanned (left), inverted to a positive with the contrast stretched (center), and fixed (right). The image on the left is of a physical object that looks like this. But the original object would need to be kept in darkness most of the time. It would be changed by the act

Figure 11.22. Three versions of the same original ephemeral process photogram exposure. Left: A faithful scan of the original, unfixed photogram. Center: The previous digital scan inverted to a positive and with the contrast appropriately stretched. Right: A faithful scan of the original photogram, but now it has been chemically fixed so it is no longer sensitive to light.

of viewing it, because it is still sensitive to light. For the center image, subtle variations in hue and value appear that were not obvious in the original photogram. But this is only a digital image. The right-hand image is now a permanent image/ object, but the act of fixing it reduced the subtle range of hues and tones to a much flatter, near monochrome. And because of the fixing process, the image/object reproduced at the far left no longer exists; there is only the digital image of it in this book. *Which of the three versions is best, and why?*

In figure 11.23 I show several iterations of the same photogram concept, both with and without a hand-brushed accelerator. For each I compare the result of fixing, not fixing, and making contrast-enhanced digital images that are either inverted to a positive or left as a negative. Which is best, and why?

11.3.5 EP accelerator transfers

Another ephemeral process approach is to use objects in the world to transfer accelerator to only selective parts of the enlarging paper, which can be then simply exposed to light with no object in place. Wherever the accelerator was applied, it turns dark. And so, for example, the accelerator can be brushed onto a leaf, and the leaf pressed onto the paper. If done carefully, the accelerator can be transferred to the paper according to the shape and texture of the leaf. When the paper is held up to the light, the pattern of transfer appears.

An intriguing variation is to mist an abandoned spider web with accelerator, and then let the web stick to the enlarging paper. Since the web is sticky, it efficiently transfers the accelerator to make a direct and unique print of the web (which is destroyed in the process). Since I usually do this to the webs of orb weaver spiders (which typically build new webs each day), I call them *araneitypes*, after the Latin name for this family of spiders (araneidae). Three examples can be seen in figure 11.24.

Figure 11.23. The same photogram concept executed eight different ways, all using printed-out enlarging paper. For the top row I used no accelerator, but for the bottom row I hand-brushed vitamin C accelerator onto the paper before exposing. From left to right: (1) after exposure, (2) after fixing, (3) enhancement of pre-fixing scan, (4) enhanced negative version of scan.

The araneitypes can be made permanent with fixer; even though there is an inevitable loss in density, the high contrast of the technique yields satisfactory results. Depending on the paper used, the colors tend to shift toward the brown, but that too can be appealing. But I usually leave mine unfixed to better draw a connection to the ephemeral nature of the spider web.

11.4 Drawing from negatives

Figure 11.25 demonstrates a technique I call *drawing from negatives*. A single frame of a 35 mm black-and-white *negative* is mounted in a glass slide mount, and then projected with a slide projector directly onto a piece of drawing paper. Graphite or charcoal is then used to make the light areas of the projected negative dark, until all of the picture appears as the same dark value. Thus, one draws in until the picture disappears. When the slide projector is turned off and the room lights turned on, a positive appears.

I have done many of these (and held workshops to teach others to do it), and the experience is always odd. At one level it is very tedious. While the projector is turned on, it is almost impossible to tell how the picture is progressing, since the areas already darkened with the pencil simply merge together with neighboring areas that are dark because the negative is dark there. One starts with an image of a negative, and one tediously, bit-by-bit, obliterates it with a pencil. It is always tempting to turn

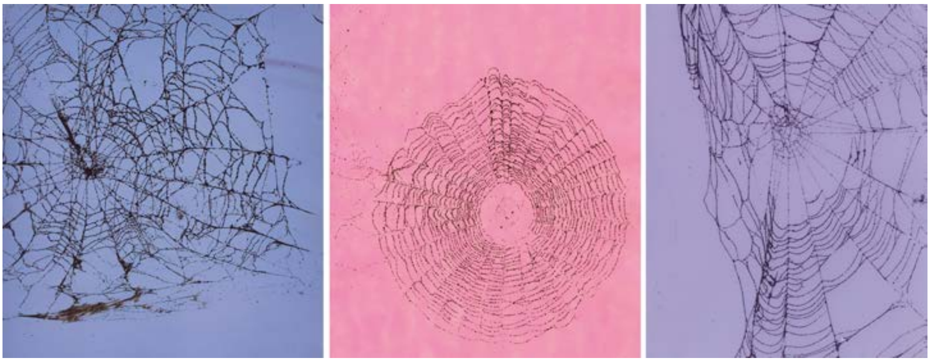


Figure 11.24. Araneitypes: accelerator was sprayed onto spider webs, and the wet webs stuck to the paper immediately before exposure to sunlight. The background colors are a result of the different varieties of enlarging paper used.

Figure 11.25. Artist Judith Baker Waller drawing from a negative. The projected black-and-white film negative is darkened in by hand, wherever it is light, resulting in a positive image.

off the projector to check one's progress, for only then can the actual drawing be seen.

And yet, one *can* make choices. For the example in figure 11.26, I spent a lot of time on the human figure, carefully coloring in the tiniest details in the light parts of the negative. For the trees, on the other hand, I made the marks very quickly and

Figure 11.26. A negative was projected onto a piece of drawing paper, and the light areas darkened in by hand. Is it a drawing or a photographic print? Does it matter?

with a lot of gesture, trying to darken the light areas as best I could with fast movements of the pencil. But I didn't go back and touch up every little detail. For the field in the foreground I made marks that were only in the same spirit as the original negative (which consisted of a sharply focused field of dandelions). I made quick gestures with the pencil that attempted to mimic the ball-shaped flowers and vertical stems and also to agree with the broader areas of light and dark. But I did not try at all to line up my dark marks with the specific light areas of the negative. If I had tried to do that, I would probably *still* be there to this day, tediously drawing.

Clearly, I could not make a second *identical* 'print' from this same negative, even if I tried. And in making a second print, I could make different choices regarding the making of marks, with *the intention of producing very different results*.

Figure 11.27 shows two drawings, both made by the artist Judith Baker Waller from one of my negatives. This particular negative is such that it is very difficult to tell, simply by looking at it, what the picture even represents. Waller made the second drawing *without having seen the result of her first drawing*. That is, I made her turn away while I turned off the projector and removed the drawing, and then attached a second piece of paper.

And so Waller made two drawings, each with different choices regarding how to make her 'mark' on the paper. But she did not know what the drawings even represented as she made them. So are these drawings or prints? Was she simply an underpaid and overworked, sometimes complaining but often cheerful, self-propelled and expensive-to-feed inkjet printer? Or was there more to it than that?

Part of the central idea of 'the print' is that, once the negative (or image file or printer's block) is chosen, one can no longer decide which parts of the paper will be

Figure 11.27. Two pictures drawn from one of my negatives by the artist Judith Baker Waller. Images used by permission of Judith Baker Waller.

light and which parts will be dark. And the process of making the print is mechanical, and so it can be done again and again, producing multiple versions of the same picture. A painter, on the other hand, gets to look at the whole picture at once, and alter their decisions based on the work taken as a whole.

But these thought experiments show, I think, that it is not quite so simple. The matter of human agency is important, even if it is at the level only of the mark and not regarding the entire picture. It matters that a human mind is controlling a hand that is making marks, and that mind is responding to what it has done, thus altering future marks. This happens only a little when one makes a straightforward silver gelatin print in the darkroom, and it happens not at all when one presses the 'print' button on a computer.

11.5 The *camera stupida*

The *camera lucida*, as originally conceived by William Hyde Wollston (see chapter 1, section 1.5), is in many ways misnamed. The name means, literally, 'light box.' Clearly, the name was a play on *camera obscura*, the 'dark box' to form images widely known since the 16th century, that evolved into the modern camera. But in Wollaston's design, there is no *camera*, no box.

To the contemporary ear, the name would seem to imply that when one uses a camera lucida, one is basically 'taking a picture'—only the image must be drawn to be captured, since there is no light-sensitive material to record it directly. This is how I imagined the device, when I first heard it described in *Herschel at the Cape*—the annotated letters and diaries of photography pioneer John Herschel, describing his intellectual exploits between 1834 and 1838 in South Africa (Evans *et al* 1969). I pictured it as a box with some arrangement of lenses and beam splitters, which one could carry around and point like a camera. When looking into the box one would see, simultaneously, the paper and the image, and thus one could simply mechanically trace over the image.

For anyone who has used (or tried to use) a *real* camera lucida[1], however, it is clear that this is wishful thinking. The traditional camera lucida is a drawing aid more than it is a device for 'capturing an image' by drawing. For one thing, it must be set up and clamped to a drawing table, and the paper and device carefully positioned before one can even begin to draw. Secondly, it is difficult to see both the image and the drawing paper simultaneously because in most circumstances the lighting is vastly different for both. One usually compensates for this by holding some kind of screen in the non-drawing hand, to cast just the right shadow onto the drawing paper so it doesn't drown out the image superimposed on it, or by flipping a filter in or out of the light path when needed.

Another difficulty is that the drawing paper is only a foot or so away, while the subject may be at effective infinity. Thus, one cannot necessarily focus the eye on both simultaneously. This is particularly troublesome if one is, like me, not young and so needs reading glasses to focus on a nearby object.

And so the camera lucida is not a camera-by-drawing. It is really a device that helps an artist, someone who already knows how to draw, get the proportions, locations and angles exactly correct in their drawing. The typical use is to employ the camera lucida to mark the locations of key points in the drawing—the ends of lines, points to mark the rough borders of objects. Then the drawing is drafted in the conventional way, using these markings as guides.

And so I was disappointed when I first sought out the details of the design, as there was a practical reason for my wishful thinking. I have had no training in drawing, and neither have most of my photography students. But I wanted a device that would give them (and me) a visceral connection between an *image* and its permanent rendering. Towards this end, I developed a variation on the camera lucida that really is somewhat similar to what I ignorantly first imagined. The result borrows from both the camera lucida and the camera obscura, and adds a few new ideas of my own. It is simple and inexpensive to build, and because the end product is in some ways a bit silly, I dub it the *camera stupida*:

1. I have put the *camera* back in the camera stupida. The device is built around a box that one can put on a tripod or rest on a table-top, and point at the subject like an ordinary camera.
2. It is designed for one size of drawing paper, and other design features necessitate that it is a *small* piece of drawing paper. The drawing paper is analogous to the film format in any particular camera. My prototype model uses sheets of paper that are 4 × 4 inch square.
3. The small format means that one can quickly sketch the drawing, because there is not that much to draw. *And so the intent is completely different*. One is not making an art drawing from life. One is using the device and a pencil to quickly capture an image in a format not unlike the picture spit out by a Polaroid SX-70 of decades ago. The result is more likely to appear pinned

---

[1] There are at least a couple of companies that manufacture and sell modern versions.

Figure 11.28. The working prototype for my 'camera stupida' design. One looks through the lens at the top, and sees an image from the lens in front projected onto the drawing paper inside. Both images are simultaneously in focus, and one can adjust their relative brightnesses by rotating the polarizer over the viewing lens.

under a magnet on the refrigerator than framed, wired and hung on the wall of an art gallery.

4. Establishing a fixed format and putting it all in a box allows other technical problems to be solved. In particular, the relatively dimly-lit box (open only on the side facing the drawing hand) means that a simple piece of thin glass or clear plastic can be used in place of an expensive partially-silvered beam splitter, and *polarization* can be used to make the brightness of the subject image adjustable by simply rotating a circular filter. The fixed format also allows one to use lenses to ensure that the images of both the nearby drawing paper and distant subject are in focus simultaneously. With this arrangement of lenses, one sees a relatively wide-angle view of the subject superimposed on a magnified view of the drawing paper.
5. The entire device is easily portable. Ideally, it would close up into its own box with all of the optical parts stowed safely inside. There would be room left over for spare pre-cut sheets of drawing paper, as well as pencils, erasers, etc. The box would close with bi-parting doors, which provide adjustable sun-shade 'wings' when the device is in use.

And so one simply puts my camera stupida onto a tripod and points it at the subject. Peering through a lens at the top of the box, one sees both the drawing paper and the subject, and both are in focus[2]. As one traces over the image, a polarizing filter over the lens can be rotated to brighten or dim the projected image. And so the image can be dimmed for relatively bright parts of the subject, and then brightened for relatively dim parts of the subject, with a simple twist of a dial. If the light is too dim inside the box to see the paper well, the built-in LED lamp can be turned on. See figure 11.28 for

---

[2] This is true, but only within the limits in my design, which is optimized for relatively distant objects. An acceptable focus is achieved for objects more distant than about 1 meter.

Figure 11.29. Two drawings made with the camera stupida. Each 4 × 4 inch sketch was made in about 5 min, and completed entirely while looking through the camera stupida. The traditional camera lucida was primarily used only to mark key points in the drawing in order to establish scale, proportions, and angles; the drawing was then drafted by eye in the usual way, using these marks as guides.

images of my working prototype, constructed easily with simple and inexpensive materials. More details on the optical design can be found in appendix B.

I find the camera stupida very satisfying to use, if my goal is to quickly capture the essence of an image. I usually get the best results when I try to make a particular drawing as fast as I can; usually, only about five minutes is required to make the 4 × 4 inch sketch. I find this exercise enjoyable, but also it forces me to think hard about just what it is in a given picture that makes it a picture. Anyone who ever makes any drawing from life must do this too. But it is my contention that, for the photographer, it is very instructive to draw while seeing like a camera sees. See figure 11.29 for two examples of drawings I made with the prototype device.

The camera stupida brings up the issue again regarding what is the distinction between a drawing (or painting) and a photograph? What if your best friend were to lead you blindfolded to a camera stupida they had already set up and pointed at the subject. You then look only through the device and not at the world, and you carefully and exactly trace what you see. Did you make a drawing? Or did your best friend take a photograph—using you as the light detector?

References

Chen S 2018 Quantum mechanics could solve cryptography's random number problem *Wired* [https://www.wired.com/story/quantum-mechanics-could-solve-cryptographys-random-number-problem/]

Evans D S, Deeming T J, Evans B H and Goldfarb S (ed) 1969 *Herschel at the Cape: Diaries and Correspondence of Sir John Herschel* (Austin, TX: University of Texas Press)

Fenster D 2018 personal communication

James C 2016 *The Book of Alternative Photographic Processes* 3rd edn (Boston, MA: Cengage Learning)

London B, Upton J, Stone J, Kobré K and Brill B 2005 *Photography* 8th edn (Englewood Cliffs, NJ: Prentice Hall)

Nause C L 2016 private communication

Noll C 2018 personal communication

O'Brien M and Waits T 2011 *Hard Ground* (Austin: University of Texas Press)

Rammel H 2007 *Like Water Tightly Wound (Crouton, Milwaukee, WI)* (Liner notes to audio recording)

Rammel H 2008 The amplified palette Sweet Pea Review [http://www.sweetpeareview.com/sweets10.html]

Rexer L 2002 *Photography's Antiquarian Avant-Garde: The New Wave in Old Processes* (New York: Harry N. Abrams)

IOP Concise Physics

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Chapter 12

Towards an art and science of nature

12.1 A personal note

Rays of light from the world are altered—blocked or redirected—by matter, and then allowed to fall onto some flat surface that is *light-sensitive*. That is, the light striking that surface disappears, and the surface is forever changed as a result. These two processes taken together—the altering of the paths of light rays and the interaction of light with matter—might reasonably be taken to be the defining principles of *photography*.

But photography is also a form of two-dimensional art. And as such, a photograph is a set of marks on a flat surface—and stuff happens in our minds when we look at those marks on that surface. Thus, it is of crucial importance to recognize that there is also an *image*—something that exists apart from its physical representation, presumably identifiable only because of the common nature of our human minds. If Martians, with their green blood and differently-structured minds, were brought into the mix—who knows what the 'image' would mean to them?

Throughout *The Physics and Art of Photography*, I have been trying to find ways to make the connection explicit between the physical part of photography—the part that is separate from the image—and the image itself. Part of the trick has been to look for processes and techniques that include evidence of this physicality within the image.

What if one could make a photograph that is an exact replica of any image that appears in one's own mind? It seems to me that this is where some digital photography has mostly positioned itself in recent years. Modern digital detectors and lenses enable the easy and precise capture of representational photographic content as never before. And digital image processing allows not only for the easy manipulation of content, it allows for the creation of new content. And so it is unsurprising that many photographers want to make hay with this new toy. This precisely-controlled capture of representational content in any circumstance,

doi:10.1088/2053-2571/aaf0aech12

combined with digital processing to alter it in almost any imaginable way, brings one closer to the ideal of making a direct record of the photographer's mind.

Some of the results are truly impressive and instantly compelling. But I sometimes wonder, will they remain compelling for long? For I believe there is a sameness that pervades some current photography, even as the accurate capture of photographic content has been (laudably) democratized and made far more accessible. But for a beginner, proud of their perfect photograph of a monarch butterfly on a purple aster, it can be a sobering exercise to perform an internet image search on 'monarch and aster.'

So which is better? To put oneself in a situation such that one is able to take advantage of the lucky shot when it happens, however it happens? To use processes and techniques that encourage the physical nature of the photographic process to have its own say? Or instead to control every element such that the picture is exactly according to plan?

I do not pretend to have a general response to these questions—but I can answer it for myself. I do not want perfect control over my art. I want instead to make use of the tension that naturally arises when the artist's conscious control is tempered by a technique that is at least in part uncontrollable. Or rather, I want my art to be partly controlled not consciously by myself, but by the world itself. Perhaps part of this comes from the natural scientist in me. The natural sciences are human endeavors, and there is often a particular 'style' that comes with that. But on the other hand, the scientist doesn't get to just make stuff up; there is *The World*, and it would still be there even if we were not.

And so perhaps this is why I am drawn to an art that allows me to make my mark in my own way, but only within limits. I can arrange things as in a still life, but I am still photographing something that exists in the world, not only in my mind. And the non-photographic content I seek out is a way to let nature (through complex physical processes only partly under my control) touch the surfaces of my pictures directly; they are not made solely by my own hand. And so for me it comes down to two beliefs:

1. My mind is much like the minds of those who look at my photographs.
2. Nature is much bigger than any of our minds.

The first point brings with it the possibility of art. For it is the commonality of our minds that allows us to see something of the same thing, when two of us look at the same picture. And thus we can use our art to communicate with each other. But it also brings a dilemma. For if our minds are alike, then when I use my art to say to you *only* what is in my mind—then I am, in a sense, telling you only what you already know. And so I find myself bored with photography that allows me to portray whatever is in my mind with perfect precision. Why would anyone want to see something so mundane as that?

I believe that the second point above is the way out of this dilemma. For it means we have *nature*, something vastly bigger and older than any of us, to both talk about and listen to. And this is why I want nature to somehow have its own say in my art.

By ‘nature,’ I do not mean that which is not human. A straight portrait of a human face against a black background may have more of what I am talking about than many so-called nature photographs of polar bears in the Arctic or lions in Africa. And one might reasonably object that we *can* have an art that is under the complete control of the artist if it is used to *say* things of importance *about* the world. That is, the artist can control the *technique* completely without risk of self-indulgent irrelevance, so long as the *content* of the art is about the world and not simply about the mind of the artist.

But for whatever reason, that doesn’t seem to work for me; perhaps I am not so sure of myself. For what could I ever say about nature that is more interesting than what nature has to say about itself? Perhaps as photographers we are simply stuck here, trying to arrange lenses and boxes and light-sensitive materials such that nature makes interesting marks on flat surfaces, in the hope that someone will see those marks and in so doing, new thoughts and feelings—new things—are created. And so we change a tiny part of our only world.

Appendix A

Making ephemeral process (EP) negatives from chromogenic prints

Ephemeral process (EP) photography uses silver gelatin enlarging paper, intended for making black-and-white prints in the darkroom from black-and-white negatives. But instead of exposing the paper to faint light, and then treating it with a chemical developer, the paper is instead allowed to *print out*. And so it turns dark with the action of light itself, with no chemical amplification after the exposure.

This requires orders of magnitude more light than the normal developing-out process, but printing out can be made much more sensitive, temporarily, by brushing an *accelerator* onto the paper immediately before exposure. The paper is then washed and dried, and it returns to its very low sensitive state, and so it can be easily scanned to capture the negative image.

A.1 EP accelerator formula

The chemistry of the accelerating process is described in some detail in chapter 2, section 2.5, but the key ingredient is water. In fact, water alone has a significant accelerating effect on most enlarging papers. The accelerating effect is, for most papers, greatly enhanced over plain water by adding sodium sulfite or ascorbic acid (vitamin C).

I recommend ascorbic acid, as it is safe and can be easily purchased in powdered crystals, sometimes even at the grocery store. Sodium sulfite is also an excellent, inexpensive and easily-available accelerator, and it is considered to be mostly non-hazardous. It has some advantages over ascorbic acid, but it does cause an allergic response in some people, and direct contact to the skin (or inhalation of the dry powder) should be generally avoided.

Finally, in order to make the accelerator brush more easily onto the paper, I use xanthan gum as a binder. It works better (for this purpose) and is far less expensive than a traditional art-medium binder such as gum arabic. It can be easily found online or in the gluten-free baking section of many grocery stores; the smallest package will last a lifetime for this purpose. My preferred formula is this:

doi:10.1088/2053-2571/aaf0aech13

1. Mix together dry:
    (a) 1/8 tsp xanthan gum powder,
    (b) 1/8 tsp ascorbic acid (dry powdered crystals).
2. Mix the dry ingredients with 1/2 cup water and shake well. The xanthan gum will want to form lumps. This can be mitigated somewhat by carefully sprinkling tiny bits onto the surface of the water, shaking, and then repeating the process. But even if lumps form, they should dissipate within 24 h. The accelerator solution should work for at least a week.

Ascorbic acid will stain some papers brown, especially if either the exposure time or concentration is too high. The ascorbic acid solution also dries fast, and so it can be problematic for very long exposures. And it can turn into something like a glue as it dries, and so the acetate (see below) must be carefully peeled off under running water.

Sodium sulfite has none of these problems, but it has the disadvantage that its use is much less benign. It will fog some types of papers, and the citric acid in the recipe below is to counteract that tendency. But like ascorbic acid, citric acid may stain some papers brown. So far, I have not found a paper that both *needs* citric acid to prevent fogging, but also is stained by it. My working sodium sulfite formula is as follows:

1. Mix together dry:
    (a) 1 tsp sodium sulfite powder, Na_2SO_3. Note that this is *not* sodium *sulfate* (Na_2SO_4).
    (b) 1/8 tsp xanthan gum powder.
    (c) Optional: 1/4 tsp citric acid (dry crystals).
2. Mix the dry ingredients with 1/2 cup water and shake gently. This mixes more easily, with fewer lumps, than the ascorbic acid formula. The large amount of sodium sulfite keeps the xanthan gum particles separated from each other when the water is added.

One should consider these recipes as starting points for experimentation. Some papers require more (or less) of the ascorbic acid or sodium sulfite, and the amount of xanthan gum can be adjusted to make the solution either thicker or more watery. In my experience, the ascorbic acid recipe brushes onto the paper more smoothly than the sodium sulfite version.

A.2 Choosing the paper

There are many varieties and sizes of black-and-white enlarging paper that can be purchased online, or possibly at a nearby camera store. The cost is usually about $1 per 8 × 10 sheet—less if bought in larger quantities, more if bought in larger sizes. There are two basic categories:

1. Resin coated papers (RC): The light-sensitive silver gelatin emulsion is coated onto paper that is waterproof, as it is sealed with a plastic resin. This is usually the least expensive type of paper, and it is the easiest to use. But for this purpose it may be less satisfying, as the paper surface has a plastic-like perfection. Furthermore, of the RC papers I have tested, they are all *more*

sensitive while dry than the fiber-based papers I describe below. But they are, in general, no more sensitive when the accelerator is applied. And so they tend to show less contrast between the unaccelerated and accelerated parts of the paper. RC papers also tend to have a very high unaccelerated sensitivity when the humidity is high, and this means one must be very careful when scanning the still light-sensitive photogram whenever the humidity is high.

2. Fiber-based papers (FB): FB papers are the go-to choice for the art photographer. The silver gelatin emulsion is applied directly to good quality paper, with all of its subtle micro-texture. It is usually more expensive and difficult to handle than RC paper. For EP photography, I have found FB papers, in general, to be more interesting and useful than RC papers. But the results vary widely from one type of paper to another.

For EP photography, it is not necessary to use newly-purchased enlarging paper. My favorite papers, in fact, have been unavailable for decades outside of the used market. Papers that are long expired and nearly useless for their original purpose may give outstanding results for EP photography. Almost any black-and-white enlarging paper will produce results that are at least interesting in *some* way. Experiment!

A.3 Preparing the paper

Figure A.1 illustrates the steps in making an EP negative from a chromogenic print. The enlarging paper should be handled for as little time as possible, and in light that is just barely bright enough to work under. Incandescent lighting will produce less

Figure A.1. Clockwise from upper left: (1) Materials. (2) The accelerator is brushed onto the enlarging paper and overlaid with acetate. (3) The enlarging paper is positioned face down on the print. (4) The print with enlarging paper underneath is turned upside down and held flat under glass. (5) Exposed to light from an LED flood. (6) The resulting negative. The entire procedure (except for the exposure) was carried out under the light of a single 60 W light bulb.

exposure than daylight, fluorescent or white LED lighting of the same brightness. A *red* LED headlamp, on the other hand, can be used with no fear of exposure at all.

Brush the sensitizer onto the emulsion side of the paper; any type of paintbrush can be used. Gently lay on a thin sheet of acetate, and smooth out the bubbles. In order to determine how the final image will be affected by air bubbles and unevenness in the application of the sensitizer, you will have to experiment. The answer depends on too many details to describe here, but that is much of the fun.

It is important to keep in mind that once the accelerator is applied, the paper will be far more sensitive to light. And so one can be much more casual (sometimes *very* casual) about unwanted exposure *before* the accelerator has been applied. After that, however, one should work both quickly and under light that is as dim as possible.

Lay the print image-side-down on top of the accelerated paper and acetate, and place a sheet of glass on top to hold it in place. Alternatively, one can dispense with the piece of acetate, and simply stick the print directly to the accelerated paper. Depending on the combination of enlarging paper, accelerator formula, and whether the print has a glossy or matt finish, they may be difficult to separate without soaking in warm water. But a typical chromogenic print is not permanently harmed by the process.

Expose through the back of the chromogenic print. sunlight works well, and an exposure time of roughly one minute should be expected. But indoor flood lights can also be used, although the exposure time is likely to be longer. The colors produced depend critically on the choice of enlarging paper, print subject, accelerator, and light source. Experiment!

Figure A.2. The image resulting from the negative shown in figure A.1, after washing, drying, scanning, and simple digital processing with GIMP.

A.4 Washing, drying, and scanning

Once the exposure is complete, carefully separate, under running water, the acetate from the paper (this is most important with the combination of FB paper and ascorbic acid accelerator). A smooth, clean surface and a shower-stall squeegee can be very helpful for removing most of the water from the paper. RC paper can be simply hung to dry from a corner, with a spring clothespin. FB paper should be left to dry upside down on a clean porous surface (a plastic window screen works well). Washing should be carried out in dim light, and the paper should be left to dry in total darkness. See figure A.2 for the positive image that resulted from the procedure shown in figure A.1.

Scan the paper only after it is dry. The scanning process will damage the negative, but if done carefully, the image can be captured well. A higher resolution scan exposes the paper more, as does repeated scanning. Most papers (especially RC papers) will suffer more damage from scanning if the humidity is high.

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Appendix B

The optics of the camera stupida

The camera stupida uses a mirror–beamsplitter arrangement similar to that used in some versions of the camera lucida. One looks directly at a sheet of drawing paper, but with a superimposed view of the subject to be drawn. The view of the subject undergoes two reflections, so the image is not reversed. One of these reflections is off a normal mirror, while the other is off a beamsplitter—which allows some light to pass through and some to reflect. A beamsplitter that divides this light into equal parts uses a semi-transparent layer of vacuum-deposited metal. In the camera stupida, I use a simple piece of thin glass, taking advantage of the partial reflection off the surface of an air–glass interface.

If the mirror and glass beamsplitter are angled at 45° to each other, then the straight-through and reflected images are 90° apart. Thus, one can look straight down at the drawing paper, and simultaneously straight forward at the subject. The difference is that I angle the mirror–beamsplitter combination at an angle of 11° to the vertical. This allows the final reflection off the beamsplitter to be at an angle of 34° to the surface of the glass, or 56° to its normal.

When light reflects off the interface at a sudden change in index of refraction between two transparent materials, it is strongly *polarized* if the reflection is at an angle to the normal called the *Brewster angle*, given by:

$$\tan\theta = \frac{n_2}{n_1} \tag{B.1}$$

where n_1 is the index of refraction of the material for the incident ray, and n_2 is the index of refraction of the material the ray reflects off. For light in air ($n_1 = 1$) reflecting off a glass surface with a typical index of refraction of about 1.5, this corresponds to an angle of about 56° with the normal, or 34° with the surface of the glass. See figure B.1 for the geometry of the camera stupida, incorporating this angle into the design.

doi:10.1088/2053-2571/aaf0aech14

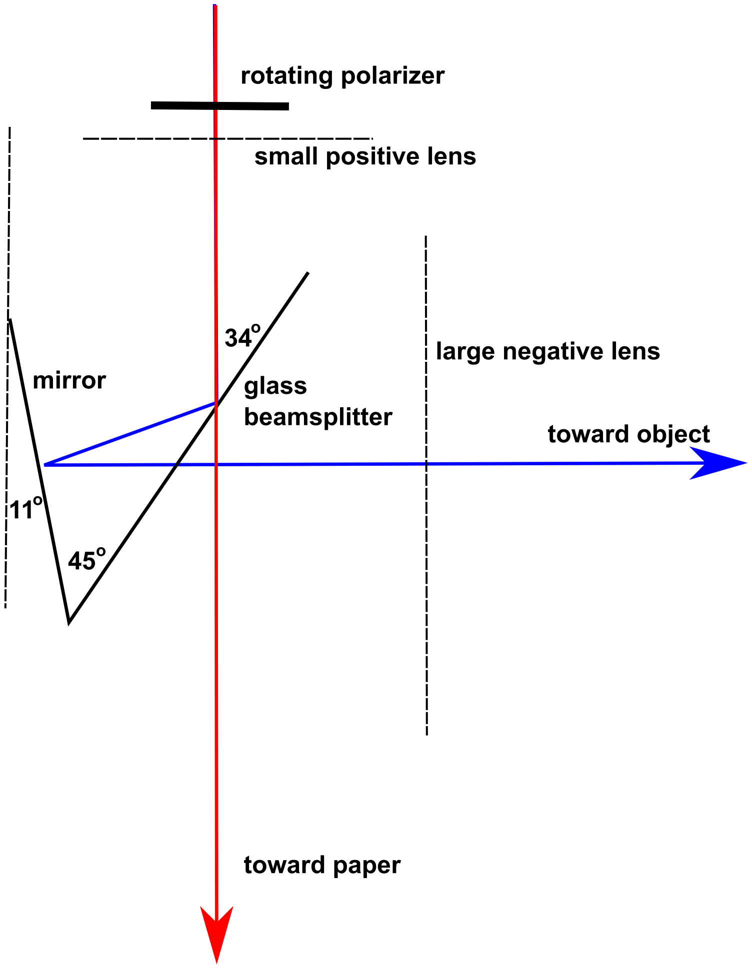


Figure B.1. The arrangement of the optics for the camera stupida. One looks downward through the polarizer and small lens at the top to see a simultaneous view of the drawing paper directly below and the subject in the distance horizontally.

And so one looks from above, directly through the glass beamsplitter at the drawing paper, but one also sees through two reflections to the subject. The reflection off the glass is at the Brewster angle, and so the light from the subject is strongly polarized while that from the drawing paper is not.

All that is needed then, is a polarizing filter at the top viewing lens to adjust the brightness of the subject image, while leaving the view of the drawing paper unaffected. This works because a polarizing filter passes only light of a particular polarization. If the filter is oriented so that it passes only light polarized perpendicularly to that of the polarized light of the subject, then no light is passed.

The fixed format allows one to view the paper through a small positive focal length (convex) magnifier lens. For the view of the subject, a large negative focal length (concave) lens is placed in front of the beamsplitter–mirror assembly. When looking at the subject, one sees through *both* lenses, while the view of the drawing paper is only through the convex lens. The particular arrangement of lenses for the subject view is called a *reverse Galilean viewfinder*, and it shows a somewhat wide-angle view of the subject. The focal lengths of the two lenses, combined with the fixed distance to the drawing paper, can be chosen such that both images are in focus.

The Physics and Art of Photography, Volume 3

Detectors and the meaning of digital

John Beaver

Appendix C

Units, dimensions, and scientific notation

C.1 Units and dimensions

When we refer to a physical quantity, it must always have a set of *dimensions* associated with it, and also in many circumstances, a set of *units*. In this context the word ‘dimension’ refers not to spatial dimensions, but rather to the *type* of physical quantity. For example, length is a fundamentally different type of quantity than time. One cannot add a length to a time, nor can one subtract one from the other, because that would equal nonsense. Note that this is not the same thing as apples and oranges. Unlike length and time, one *can* add apples and oranges (it equals fruit salad).

But on the other hand, it is just fine to multiply or divide a length by a time. This produces something with different dimensions, that are a combination of the two. For example, if one divides a length by a time, the result is something that has dimensions of length/time (‘length per time’). Often these combined dimensions have special names. This example of length/time has the special name of velocity or speed. And so whenever one divides a length by a time, a new thing with dimensions of length/time results.

But what about the actual numbers one plugs into the calculator in a specific case? What if one has a specific length, and a specific time, and wants to calculate a specific speed? Whenever actual numbers are involved, there must also be *units*.

A length of 12.0345 is ambiguous. Is it 12.0345 meters or 12.0345 furlongs? The meter and the furlong are examples of *units*, which are agreed-upon standards for attaching a numerical value to a particular physical quantity. And so the meter is a unit of the dimension of length, and so is a furlong. One can convert between units of the same dimension, by establishing an equivalence between them. And so 1 m = 3.280 feet = 39.37 inches = 0.004 97 furlongs, etc.

In the physical sciences we mostly use a particular international system of units, called SI, which stands for ‘International System’ (in French). The SI unit of length

doi:10.1088/2053-2571/aaf0aech15

Table C.1. Common SI units.

Dimension	Unit	Abbreviation
Length	meter	m
Time	second	s
Mass	kilogram	kg
Temperature	kelvin	K
Force	newton	N
Energy	joule	J
Power	watt	W

is the meter, while the SI unit of time is the second. Every SI unit has an official abbreviation. The abbreviation for the meter is *m*, and for the second it is *s* (it matters that they are lower-case). Table C.1 lists some common SI units, with their dimensions and official abbreviations.

Just as we can derive new dimensions by multiplying or dividing dimensions by each other (length/time, for example), we can do the same for units. And so we can divide meters by seconds to get a newly derived unit, which we write m/s (called 'meters per second'). What if we want to divide m/s by seconds? We can do that just fine, and we get $\mathrm{m/s/s} = \mathrm{m/s^2}$ (called 'meters per second squared'). Many of the units in table C.1 are actually derived combinations of other units. For example, the newton is actually a combination of kilograms, meters, and seconds:

$$1\ \mathrm{N} = 1\ \mathrm{kg}\,\frac{\mathrm{m}}{\mathrm{s}^2}. \tag{C.1}$$

These base units can be modified by any one of a number of official prefixes, which then multiplies the unit by some power of 10. These prefixes and their abbreviations are listed in table C.2, although some are more commonly used than others. For example, 'milli' means '×1/1000'. And so a millimeter (abbreviated mm) is one thousandth of a meter.

C.2 Scientific notation

We have used scientific notation for the values in table C.2. Physical quantities in nature can vary by many powers of 10. And so, for example, the light given off by the Sun, its power, P, is many times greater than the light given off by a 60 W light bulb:

$$P_{\mathrm{sun}} = 667\,000\,000\,000\,000\,000\,000\,000\,000 P_{\mathrm{lightbulb}}. \tag{C.2}$$

After the 667, there are 24 zeros there. What if I had mistyped (or you miscounted) and you found 23 zeros instead? Well, that number would be *ten times* too small. And so clearly, when dealing with numbers like this, we need a better way.

Table C.2. Prefixes for SI units.

Prefix	Abbreviation	Meaning
Femto	f	$\times 10^{-15}$
Pico	p	$\times 10^{-12}$
Nano	n	$\times 10^{-9}$
Micro	μ	$\times 10^{6}$
Milli	m	$\times 10^{-3}$
Centi	c	$\times 10^{-2}$
Deci	d	$\times 10^{-1}$
Hecto	h	$\times 10^{2}$
Kilo	k	$\times 10^{3}$
Mega	M	$\times 10^{6}$
Giga	G	$\times 10^{9}$
Tera	T	$\times 10^{12}$

And so we use what is called scientific notation. Written this way, the above equation becomes:

$$P_{\mathrm{sun}} = 6.67 \times 10^{26} P_{\mathrm{lightbulb}}. \tag{C.3}$$

The $\times 10^{26}$ part means, $\times 100\,000\,000\,000\,000\,000\,000\,000\,000$. But in practical terms this also means, ‘take the decimal point in 6.67, and move it 26 places to the right, filling in with zeros as needed.’

Raising something to a negative power means the same thing as dividing 1 by that same thing, but raised to the same *positive* power. For example:

$$27^{-3} = \frac{1}{27^3}. \tag{C.4}$$

And so we can also use negative numbers in scientific notation; it means simply *divide* by the power of 10 instead of multiplying by it. And as with positive powers, we can also express this as a decimal equivalent:

$$3.27 \times 10^{-5} = 3.27 \times \frac{1}{10^5} = \frac{3.27}{10^5} = 0.000\,032\,7. \tag{C.5}$$

Here we can see that 3.27×10^{-5} means, ‘take the decimal place in 3.27 and move it 5 places to the *left*, filling in with zeros as needed.’

This has a couple of advantages. For one thing, we can see at a glance the most important part numerically: how many powers of ten. Secondly, when we write it this way, we don’t need the zeros for place holders. And so if I put them there, it means I believe that they are significant.

And so, 6.67×10^{26} and 6.670×10^{26} are not really the same number, although they will both appear the same on a calculator. 6.67×10^{26} could possibly be 6.673×10^{26} or even 6.668×10^{26}. If I do not include any more decimal places, then

I am making a statement that, based on my uncertainty in the measurement of that quantity, I have no idea what the value of the next decimal place would be. If, on the other hand, I write 6.670×10^{26} then I am saying that I believe (even if with some uncertainty) that it really is 6.670×10^{26} and not, say, 6.673×10^{26}.

Note that one *could* use scientific notation to write the same number in several different ways. You should verify for yourself that the following is true:

$$9.75 \times 10^{7} = 975 \times 10^{5} = 0.009\,75 \times 10^{10} = 97\,500\,000\,000 \times 10^{-3}. \tag{C.6}$$

Clearly, the last two possibilities look a bit silly, but we try to avoid even the second version. When using scientific notation, it is customary to pick whatever power of 10 is needed in order to have one and only one digit to the left of the decimal place.