Colour

Section introduction

This section covers a topic of huge practical importance, colour. You'll find many optics textbooks that don't discuss colour, perhaps on the grounds that it introduces applications of ideas rather than new ideas. Whatever their reasons might be, they're wrong to miss it out. Colour transforms our lives and it is certainly worth knowing about some of the science relevant to colouring and colour reproduction. Some phenomena of colour vision involve physiology and psychology, and investigating these has led people as notable as Goethe and Schopenhauer to reject the physicists' account of the origin of colour phenomena. Time and measurement have proved that the physicists' descriptions are correct, though there are certainly perception phenomena that need explaining over and above the descriptions given in this section.

I haven't copied most of the lecture slide pictures into the text in this chapter, both on the grounds that the colours are better seen on the slides and because of the expense of printing off a version with all the pictures. Quite a number of the slides have diagrams that will be activated during the lectures, allowing you to see how the phenomenon illustrated varies with colour.

Colour mixing

Most colours can be made by adding together three primary colours in varying proportions. See the slide diagram on how to make yellow, cyan and magenta (and white) from red, green and blue. Java applets accessible on the course web page allow you to vary the colours being overlapped and see the results.

3-colour matching

An important consequence of the 3-colouring mixing experience is that any colour can be matched by the addition of appropriate amounts of three primaries. This is known as 3-colour matching. It is a remarkable phenomenon. One of the remarkable details is that the three primaries are not themselves unique. You can choose the primaries that you use. An implication of this is that although the colour sensation of light is determined uniquely by its spectral content, the reverse isn't true. The same colour can be produced by many different spectral compositions. This phenomenon is called **metamerism** and it underlies the workings of 3-colour matching. You don't have to reproduce the exact spectral content of a colour to reproduce the same colour sensation.

Mathematically, three colour matching can be written:

$$(C) \equiv x(R) + y(G) + z(B), \quad \text{where } x + y + z = 1 .$$

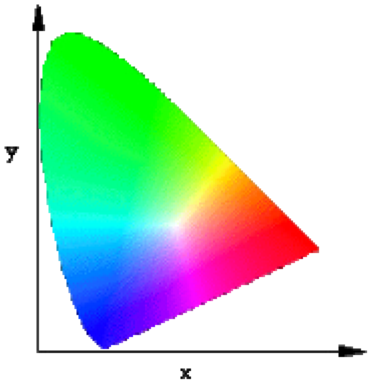
(C) stands for the colour to be matched; (R) the red primary; (G) the green primary; (B) the blue primary. The fraction of each of these primaries is given by the coefficients x, y, z. The fractions are always written out for unit intensity of the colour (C). Nothing new in terms of colour matching is gained simply by making the light stronger. In other words, if you double the light from each primary, then you double the total amount of light and the colour match remains the same. Therefore x, y and z are not three independent quantities, since $z = 1 - x - y$.

Maxwell's colour D

One of Maxwell's great achievements while he was at Aberdeen in the 1850s was to invent a device for measuring the x, y (z) coefficients in 3-colour matching when three narrow spectral colours were chosen as primaries. Maxwell not only found how to represent the sensation produced by any other spectral colour but also found the coefficients for a wide range of colours. He realised how the underlying mathematical relationship given above could be represented by putting all the colours in a triangle, with the primary colours represented by the corners.

The lecture will activate the *demo* “MaxwellTriangle.exe”. A mixture of two colours lies on the straight line joining them. Finding the result of a mixture of three colours is like finding the centre of gravity of three weights of different size. Size in this case is the intensity of each colour, more properly known as its *irradiance*. Maxwell’s work wasn't just abstract science for its own sake. He realised how to make and reproduce a colour picture photographically. His scheme was to make three black and white transparencies, one each through red, green and blue filters. He then simultaneously projected the image from each transparency back through the filter it was made with, superimposing the three images on the one screen. A full colour picture would result, he said. It did. Maxwell demonstrated this in the year after he left Aberdeen, in 1861, with the help of a London photographer. The class can see the result in a framed reconstruction of the original picture, made from the original negatives, which I can’t show on the web for copyright reasons. The very same principles are used to generate colour images all the time nowadays by remote sensing imagers and also in astronomical colour photography.

CIE Diagram



Maxwell's triangle has a practical problem: a few spectral colours lie outside the triangle, which ones depending on the exact choice of primaries. For these colours the situation is messy. You have to add one primary colour on top of the colour being matched and match the result with the remaining two primaries. It was 70 years and a lifetime later before an international commission standardised the measurement of colour in 1931 by choosing a new set of primaries, called (X), (Y), (Z), that everyone has used since. The very odd thing about this choice is that you cannot make colours that are exactly these primaries. They are a mathematical construction from known colours. The story is too involved for today. The result is that colours are plotted against the first two co-ordinates, x & y, just like the Maxwell colour triangle, and the resulting plot is a sail-like shape called the **CIE chromaticity diagram** within which all real colours are found. In this system the colours are represented by the relationship:

$$(C) \equiv x(X) + y(Y) + z(Z), \quad \text{where } x + y + z = 1 .$$

You can see that the primary colours in this system are going to represented by the points ($x = 1$, $y = 0$), ($x = 0$, $y = 1$) and ($x = 0$, $y = 0$), the last point being just $z = 1$. If you look at a properly drawn CIE diagram with the co-ordinates marked along the x and y axes, then you'd see that none of these points correspond to a real colour, emphasising the statement made before.

The 1931 diagram is still probably the most commonly met with chromaticity diagram but later versions redefined the primaries yet again, making later diagrams look a bit different in shape. This was done because the unequal emphasis given to different parts of the spectrum in the 1931 diagram is at odds with how the normal observer perceives colours.

Spectral wavelengths

Single wavelengths provide the envelope of pure colours, around the outside of the chromaticity diagram. The spectrum isn't uniformly spread along the chart, which just reflects an empirical feature of human colour vision. Our capacity to distinguish the colours of neighbouring spectral lines varies from a discrimination of about 10 nm at either end of the spectrum to around 2 nm at best. This leads to just over 100 distinguishable spectral colours for a normal observer.

An example

Any facts to do with colour perception can be, and probably are somewhere, plotted on the CIE chromaticity diagram. For example, you can plot the colour range of filters allowed in red, amber and green traffic lights; the colour changes of vegetables as they become less and less fresh; the colour changes of bodies as they are heated up; how well we distinguish neighbouring colours, the range of colours reproduced by a video projector, the consistency of colour coding on electronic resistors, and so on. Anyone serious about colour measurement will probably end up plotting their results or information on a CIE chromaticity chart. Here on the slide is a different example, showing colours produced by the components used in fireworks. Remember that if two colours are present together, the result is somewhere along the line joining them. Exactly where along the line is determined by the relative intensities of the two colours.

The rgb colour system

The most obvious practical system of representing colours that are going to be reproduced by projecting three primaries is the **rgb system**. Each coloured pixel has three component primaries: (R), (G), and (B). If 256 steps in illumination is allowed for each primary, then the amount of each primary can be represented by a 1-byte number, or equivalently an integer from 0 to 255. r is the integer for red, g, for green and b for blue, allowing a point to be specified by the integer triple r,g,b. 255, 255, 255 represents the brightest white. Decrease the integers in proportion and you get the same point on the CIE diagram but a different luminosity. Thus 100, 100, 100 is a grey and 30, 30, 30 is in practice hardly different from black. 255, 0, 0 is full red. 50, 0, 0 is the same hue but very dark.

Sometimes the 'colour space' described by the rgb system is called cubic, since the colour values are distributed at equal intervals within a cube whose opposite corners lie between the origin (0, 0, 0) and the white point (255, 255, 255).

The technology of colour monitors is closely tied to the *rgb* system. Their displays contain millions of *red*, *green*, and *blue* pixels, each so small that your eye at normal viewing distance can't see them individually. To represent any colour the monitor can display, the electronics must cause these primary colour pixels to send out their light in exactly the right proportions according to the colour matching equation. How this is done depends on the technology and is different for the traditional CRT tube (that operates via electron beams), the LCD display and the plasma display. Whatever the technology, the colour principles used are the same.

Colour TV

The phosphors used on colour TV screens to emit light aren't quite spectrally pure. Their light is represented by three points within the CIE chart. To produce a desired colour, the coded signal at one instant must specify the relative amount of each colour emitted and this is part of an international standard. Britain uses the EBU (European Broadcasting Union) standards. Manufacturers can't change the characteristics of the phosphors. If they did so, the reproduced colours would be wrong. Broadcasters can't change the colour representations either. The only colours that can be reproduced on our colour TVs are those within the triangle of points whose corners are the standard EBU phosphors.

There is a further complication in practice that the colour signals transmitted for colour television aren't the raw signals in the *rgb* system, or indeed in the *hsv* system that I'll mention soon. A different system is used for analogue TV at least, known as the YIQ system. Y is the luminance signal, essentially what is seen on a black-and-white TV. This is transmitted in the greatest detail. Two colour signals I and Q are both combinations of *r*,*g*,*b* that together with Y add the colour information. I is transmitted in more detail than Q. All this was introduced when it was important for transmissions to be compatible with monochrome TV and to save as much bandwidth as possible. The method of coding also has some advantages when a noisy signal is received. Everyone has seen how a weak TV signal can still show a recognisable black and white picture because, essentially, the monochrome luminance signal Y occupies the greatest share of the transmitted signal.

Also shown on the lecture diagram is the set of CIE co-ordinates for bodies heated to increasing temperatures. They follow a curved line passing close to the white point. The appearance of colours along this line was shown on a slide in the previous course section.

This is perhaps a suitable place to put into the notes a fact that was mentioned before in connection with images recorded by one or more sensors that operate outside the visible spectrum. I'm thinking of IR, UV or even X-ray sensors. Images taken at these wavelengths are shown in *false colour*. The earlier section of this course on *spectra* showed some false colour geological imaging maps. The false colour is very simply achieved. One channel (or sometimes more than one channel) in the *r*,*g*,*b* system is fed a signal from the sensor recording outside the visible spectrum. In the geological map examples the red channel was fed a combination of signals from more than one IR channel. Looking at the resultant picture on a monitor we see it as coloured but the colours aren't what anyone would see in real life because one or more of the channels is showing a signal from beyond the visible. The technique is very effective.

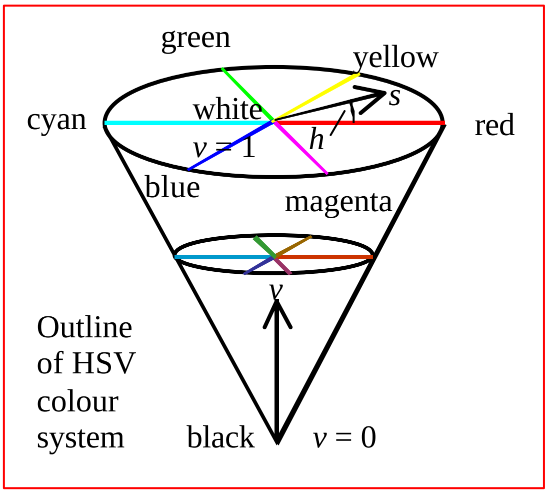
White point, dominant wavelength and purity

Notice the **white point** near the centre of the diagram. What the eye perceives as white depends on the illumination, because we adjust our sense of white to the source of light. For a reasonable choice of primary colours, the most obvious definition of white is an equal amount of each primary. This is known as the equal energy white, S_E. Clearly its x, y, z coordinates are 0.333, 0.333, 0.333. A lot of everyday illumination is produced by tungsten lamp sources and an internationally defined white is the illuminant S_A, which is a blackbody that has a temperature of 2856 K, representing a tungsten lamp. Finally, the illuminant S_C has a colour temperature representing light from an overcast sky. There are other standard illuminants, too.

Points on the CIE chart going out in a straight line from the white point tend to have a similar tint of colour, which becomes more and more saturated as you go towards the perimeter of the chart. For any point **P** (see the accompanying slide), this leads to the useful concepts of **dominant wavelength** (λ_D) and **spectral purity**, usually expressed as a %. Purity is defined as $(WP/W\lambda_D)\times 100$, again as on the slide. Spectral purity is another name for **saturation**. The colour of the dominant wavelength defines the **hue** of the point P. The colour in the opposite direction from the dominant wavelength is called the **complementary colour**.

For example, if you want to reproduce all possible hues of colours using three primaries, you must choose these primaries to define a triangle on the CIE chart that encloses the white point. That is all. If your triangle is small, then you will only be able to reproduce colours of moderate purity, i.e. colours that are not saturated and always have a significant white component. The bigger you can make your triangle, the nearer you get in places to the perimeter of the CIE diagram and the more saturated some of the colours you reproduce will be. You can now see why the CIE charts I have been projecting in recent slides aren't truly accurate. Our projector is a 3-colour reproduction device. It can only produce colours within the triangle of its own primaries and no triangle can cover the true CIE colour space.

An alternative colour system to the *rgb* system represents pixels by **hue, purity and luminosity** (or lightness), as have been defined above. In computing circles this is also know as the HSB (hue, saturation and brightness) system or HSV system (**hue**, **saturation** and **value**). Changing the luminosity changes how light the pixel is; changing the purity changes how far away the point is from white, changing the hue rotates the point on the CIE diagram around the white point. Changing hue is typically represented in graphics packages by a changing angle around the white point, angles being measured from the red line. A change in angle by 180° turns the colour into its complement.



Outline of HSV colour system

The following algorithmic steps sketch how r,g,b coordinates, expressed as 1-byte integers, are changed to HSV values h,s,v, where h is an angle in degrees, s and v are numbers in the range (0,1). The steps are shown in a notation that is hopefully self-explanatory, rather than that of any specific computer language.

$v = \max(r,g,b)/255$;
$\delta = \max(r,g,b) - \min(r,g,b)$; $s = \delta/\max(r,g,b)$;
$h = \{\text{if } (r = \max) \text{ then } (g-b)*60/\delta \text{ else if } (g = \max) \text{ then } ((b-r)*60/\delta + 120) \text{ else if } (b = \max) \text{ then } ((r-g)*60/\delta + 240)\}$

h is adjusted to be in the range (0,360) and if s is zero then h is undefined.

The HSV system is closest to how we naturally perceive colour information. It is represented above as a conical colour space rather than a cylinder because as the black point is approached, ever darker slices perpendicular to the value axis can be represented by fewer and fewer combinations of rgb. You might get a hint from the word ‘value’ that the system has a non-scientific origin. It is closely related to the Munsell colour system, introduced in 1905 by the American artist Albert Munsell, who used the word ‘value’ to describe the parameter measuring variation from white to dark. The steps in the Munsell system were intended to be related to the normal observer’s visual perception skills, unlike the algorithmic steps defined above.

Colouring by selective absorption

Most colouring takes place because of selective absorption across the spectrum by one or more of the component materials. In the picture on the slide, you can see the white wall reflecting the entire spectrum of the incident light. The green leaves preferentially return light in the green part of the spectrum. Red light and blue light is preferentially absorbed. You might be tempted to think that the reflectivity of the leaf material varied with wavelength but this isn’t the case. Reflection is a phenomenon that takes place at the surface of a material. The reflectivity is almost independent of colour. The light being coloured has to penetrate beneath the surface into the material itself. I’ll say more about this soon. In the picture, the flowers return purple and a little red light to produce their magenta tint.

Colour printing

Ink puts colour on paper by absorbing some of the incident light and hence returning a reduced spectral range. The process is called **subtractive colouring**. You should not be surprised that the **subtractive primary colours** are:

- **cyan** = (white – red), i.e. cyan ink absorbs red
- **magenta** = (white – green), i.e. magenta ink absorbs green
- **yellow** = (white – blue), i.e. yellow ink absorbs blue.

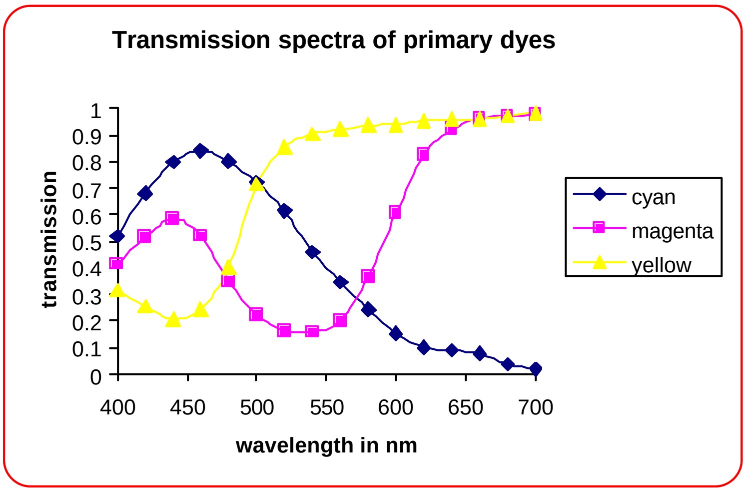
These are the colours of printers’ inks. Clearly if you put all three colours together then the total absorption will be red+green+blue, i.e. the whole spectrum, and the result will be black. Such a colour model is naturally called the CMY system. In fact, printers use these three colours plus black ink, the better to define shapes and grey tones, and also to save on the use of the coloured inks. The examples shown in the lecture illustrate the four stages in printing a coloured sheet.

Electronically, the printers' system is known as the CMYK colour model. c',m',y' are related to r,g,b in the obvious way ($c' = 255 - r$, etc.) and then the coordinates c,m,y,k are:

$k = \min(c',m',y')$
$c = c' - k$
$m = m' - k$
$y = y' - k$.

Spectra of subtractive primaries

The transmission spectra for typical inks of the primary colours for printing are shown in the slide. You can relate these to the descriptions of the colours given above.



Mixing of inks

Inks are usually put on white paper. If the ink is thin, its colour is pale and represented by a point on the CIE chart near the white point. What you see is mainly white light reflected from the paper beneath, with only mild preferential absorption. As the ink is made more concentrated, the colour becomes more saturated. Mixing pairs of inks produces a coverage of the CIE chart along the lines shown in the slide. For example, a line labelled '2:1' drawn between the lines of two subtractive primaries refers to a concentration of 2 parts of the nearest primary to 1 part of the next nearest primary.

The appearance of things

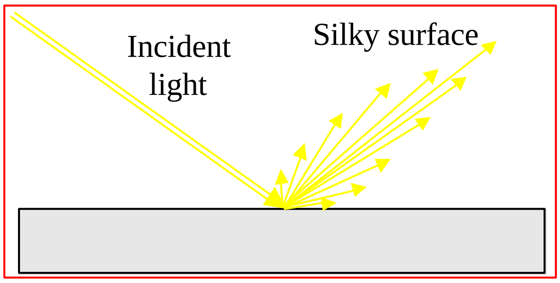
We are used to objects being matt or shiny, being white, coloured or various shades of grey, being opaque, translucent or transparent. These characteristics all relate to what happens to light when it hits the surface of objects. The rest of this section looks at the range of possibilities.

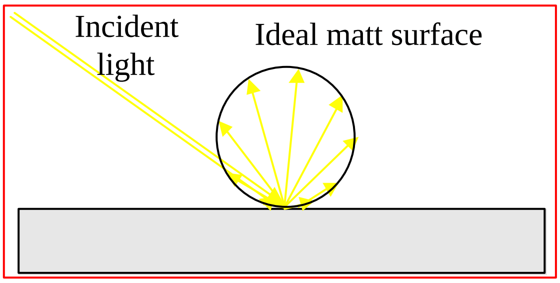
Reflection – 1

We've discussed the **specular reflection** of light in the section on mirror reflection, without looking at all at the mirror itself. Reflection is a surface phenomenon, or practically so, with the light coming back from the first few hundred nanometers of the reflecting surface. The surface must be optically smooth, otherwise the mirror laws break down. Undisturbed liquids have suitably smooth surfaces but the reflectivity of transparent materials only gets high at very large angles of incidence. Mercury is the obvious highly reflecting liquid but it has the disadvantage that it gets covered by a thin coating when left out in the atmosphere for a long time, not to dwell on the toxicity of mercury fumes. We don't see mercury mirrors much these days. Susceptibility to atmospheric corrosion is the bane of most mirror materials. Corrosion spoils the smoothness of the surface and alters its chemical composition, probably in an irregular way. Since reflection is a surface phenomenon, it quickly suffers. You can see why totally internally reflecting prisms are an attractive solution for optical instruments that need a reflecting surface.

For most materials, the reflectivity is pretty well independent of the frequency of the incident light. Glass, water, perspex, for example, reflect for a given angle of incidence much the same from the red end of the spectrum to the violet end. There is a slight variation, because the reflectivity depends on refractive index and we've already covered the dispersion of refractive index. This effect does not produce a noticeable colouring of the reflected light.

Reflection – 2

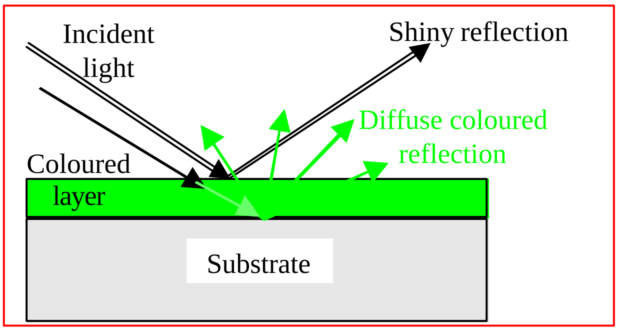




Diffuse reflection also takes place with little surface penetration, the diffuseness arising simply because the reflecting surface is optically rough. This means that on a scale using the wavelength of light as a ruler, the surface is not flat. By suitable preparation, you can alter the surface roughness and so change the apparent texture of the surface from a silky finish to a full matt finish. Paint manufacturers exploit this. Notice also that the gauge against which roughness is measured is the wavelength of light. A surface that may be optically rough can be smooth for longer wavelengths. For example, satellite dishes on the sides and chimneys of many houses are smooth at microwave frequencies but optically rough.

Watercolours, inks, etc.

We've already met coloured inks. They behave like coloured filters, colouring the light that passes through them by preferential absorption. Watercolour paints are the same. The thicker the layer, the darker the colour. The greater the concentration, the darker the colour. These effects can be summarised in the quantitative relationship given below.



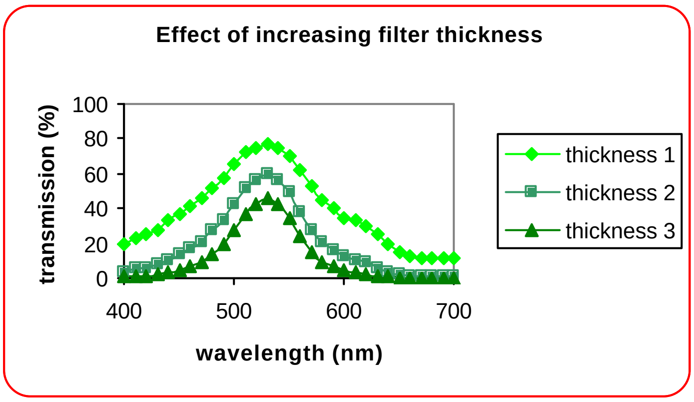
- Let T_0 be the fractional transmission through thickness d_0 at standard concentration. The transmission T through thickness d at concentration c% is:E.g. the transmission through

$$T = T_0^{(d/d_0)\times(c/100)}$$

0.1 mm at standard concentration is 80%. What is the transmission through 1 mm when the concentration is 30%?

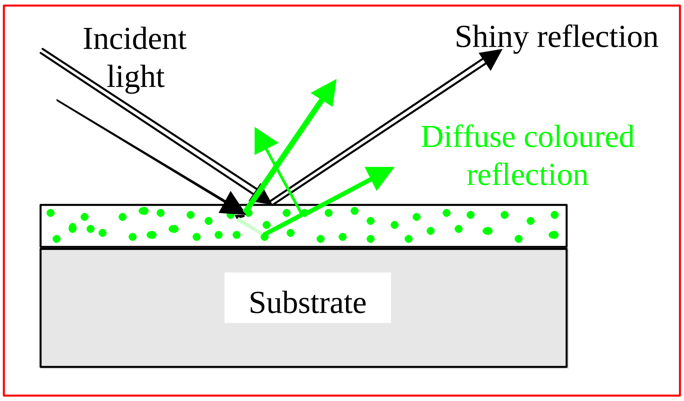
$$T = (0.8)^{(1/0.1)\times(30/100)} \times 100 = 51.2\ \%$$

Changing filter thickness



The slide shows the transmission curves of a green dye for a single thickness layer and for double and treble thickness layers. You can see how the single thickness contains a noticeable amount of white, whereas the treble thickness is a more concentrated green. This is simply the effect of cubing the absorption factor at each wavelength.

Oil paints



In a typical oil paint, the colourant particles are closely mixed in a fairly transparent medium that doesn't add much colouring of its own. The colouring process occurs by scattering from these particles under the surface. The surface itself is likely to have a shine, a shine that reflects the incident light because reflection is independent of wavelength. You can see that effect well in the picture shown on the slide, where in one place the reflection from the illuminating light is visible. The more matt the surface, the more this reflected light adds its own whiteness to the view, desaturating the colours of the painting itself.

Wood varnish

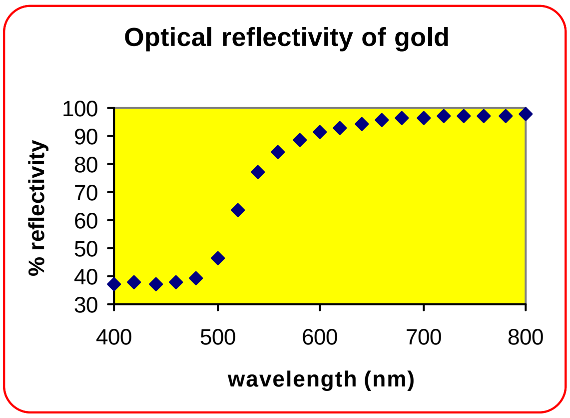
Everyone must have sanded a piece of wood and noticed how the fine roughness of the resulting surface and the residue of powdered wood leave a matt surface that shows little of the real texture of the wood. Just wet the wood and the grain shows up. Varnish does the same job, without evaporating. The varnish provides an optically homogenous surface that allows the light to penetrate through to the wood cells, letting them do their own colouring.

Whiteness

Whiteness is big business - big business in the soap powder industry, in paints, in papermaking, in fabrics. Whiteness is matt reflection independent of wavelength. How can you make a surface as white as possible? The answer is to get rid of as much preferential absorption as possible, exploit the fact that pure reflection is pretty colourless. Paper is made of cellulose fibres that are not intrinsically very white. To make them white, they have to be coated with a material that has a high refractive index, for maximum reflection, a material that is mechanically and chemically stable, that has the correct texture and is affordable. Titanium dioxide is a popular choice in the papermaking industry.

What helps to make surfaces white is to have a highly discontinuous surface, one that is broken up so that multiple reflection is encouraged and there are no thick layers of material for the light to penetrate and become preferentially absorbed. This effect of fractured surfaces is quite powerful in turning them white. Everyone knows how the head on a beer is white, yet the liquid it is made of can be strongly coloured. Foam produced by industrial discharge of waste can look almost as white as soapsuds, though the bulk liquid is a filthy dark colour. Fine powders of materials that are quite highly coloured in bulk can be white.

Metallic colouring



The colour of gold, copper, bronze and other distinctive metal colouring is the one effect that is achieved by true wavelength dependent reflection. Because the effect is inherent to the very material itself, it is the same all over the object, independent of surface finish. This distinguishes metallic colouring from other colouring processes. The data for gold is shown on the graph and you can see immediately, I hope, why gold is yellow. In the IR, gold has a pretty constant reflectivity exceeding 98% right through the near IR and deep into the far IR. That is why it is used to coat astronauts' visors and many other objects exposed to thermal radiation.

Concluding comment

There are other colouring mechanisms, found both in nature and in technology. We'll come to some of them later. The concepts discussed in this section may not be very deep concepts but they are all immensely important in the real world. What was good enough for Maxwell to study is certainly good enough for me. Getting colouring to work properly in detail involves non-trivial physics and chemistry. In my youth, newspapers, TV and textbooks, to name just three kinds of media every student is very familiar with, were in black and white. The cinema had colour but there were plenty of 'old' films around in black and white and newsreels were black and white. Colour in the diversity we have today is just taken for granted, but it is a modern phenomenon, relying on advanced technology to deliver the effects.

JSR