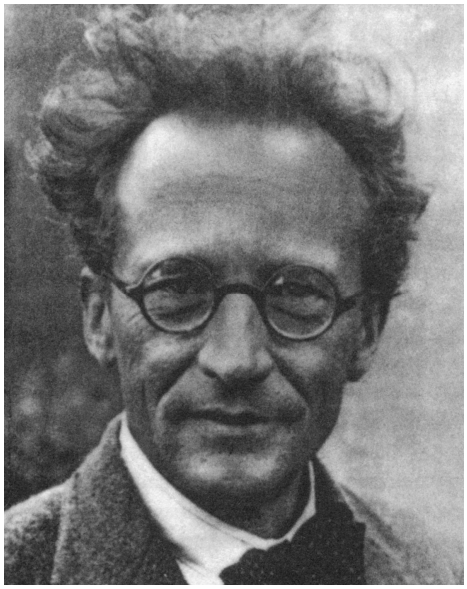
Erwin Schrödinger

Theorie der Pigmente von grösster Leuchtkraft (Theory of pigments of greatest lightness)

Annalen der Physik 4, 62 (1920) 603-622

An English translation with a short biographical introduction by Rolf G. Kuehni and a technical introduction by Michael H. Brill

Erwin Schrödinger 1887-1961



Erwin Schrödinger was born in Erdberg, Austria, to a father of Austrian and a mother of mixed Austrian-English decent. He studied physics in Vienna under Franz Exner whose assistant he became in 1911. He was influenced early by the writings of the German philosopher Arthur Schopenhauer resulting, among other things, in his interest in color theory. All three of his papers on color were written in Vienna after he concluded his military service in 1918 and before assuming a position at the University of Zürich in 1921. There he did his most important work, on quantum wave mechanics, for which he shared the 1933 Nobel Prize in physics with Paul Dirac. In 1940 Schrödinger moved to Ireland where he remained until his retirement in 1955 after which he returned to Vienna. Among his achievements is the thought experiment known as Schrödinger's Cat.

Retrospective Introduction to Erwin Schrödinger's "Theory of pigments of greatest lightness"

Since the investigations by Wilhelm Ostwald 100 years ago [1], the tristimulus envelope of all object colors under a given light (the object-color solid) has been known to comprise reflectance spectra that at all wavelengths are either 0 or 1 with at most two transitions. To Erwin Schrödinger is attributed the first mathematical proof. In his brief proof, Schrödinger implicitly used the convexity of the spectrum locus (which is by no means guaranteed from the axioms of colorimetry). David MacAdam [2] generated a geometric proof and brought the Ostwald/ Schrödinger theorem to the English-speaking world. MacAdam's proof used the *de facto* convexity of the spectrum locus without noting it as a limiting property. West and Brill [3] restated MacAdam's proof so as to show the importance of convexity, and explored generalizations of the theorem to nonconvex spectrum loci. Later work [4, 5] used spectrum-locus convexity for other applications. The first foray, though, was Schrödinger's---representing one of his first scientific efforts, made at the fairly young age of 33.

The goal function used by Schrödinger was greatest lightness for a given chromaticity, and "lightness" could be any tristimulus value deriving from an all-nonnegative color-matching function. Equivalently, one could imagine a goal function that, given dominant wavelength and lightness, would find the greatest colorimetric purity. Such equivalence is guaranteed because of the convexity of the object-color solid. In turn, that solid's convexity is enforced by the fact that it is a three-dimensional projection of the (convex) reflectance hypercube (0 to 1 for each wavelength and N wavelengths). Convexity of the object-color solid is, however, not the same as convexity of the spectrum locus, and the latter only happens to be (approximately) true for human vision.

Schrödinger defined *bivalent* reflectances as 0 or 1 at each wavelength, but may have any number of transitions between 0 and 1 over the visible wavelength range. Schrödinger showed that optimal reflectances are bivalent. This property is independent of the convexity of the spectrum locus (although Schrödinger does not remark on this fact). However, the maximum of two transitions between 0 and 1 depends on that convexity. Schrödinger implicitly used this condition in his very brief proof: "*If a pigment has three transition points that in the chromaticity diagram are not located on a straight line, by moving the transition points its reflectance can be changed in a manner that results in a lighter pigment of the same chromaticity coordinates. Therefore, it cannot be optimal.*"

In other words, three transition wavelengths would comprise three places where the r = 1 portions could be shaved or augmented, allowing an increment of light that has the same chromaticity as the original reflectance, and hence preventing optimality. These positive or negative slivers of spectrum can act as primaries to match any chromaticity, but only (as Schrödinger said) when the chromaticities of these slivers (hence of the wavelengths of transition) do not lie in a straight line. If three transition wavelengths lie on a straight line, then the spectrum locus is not convex. Thus, in the few words above, Schrödinger conditioned his proof on spectrum-locus convexity.

It was later shown geometrically [3] that the number of transitions of an optimal reflectance is the number of spectrum-locus crossings of a straight line in chromaticity space. Each straight line through the spectrum locus generates two optimal reflectances, a reflectance that is 0 for wavelengths on one side of the line and 1 on the other side, and the reverse. In [3], G. West and I illustrated the result using a fictitious spectrum locus comprising two concentric circular arcs and radii between them; a line crossing that locus four times indicated four transition wavelengths for two of the optimal reflectances.

Although Schrödinger did not treat nonconvex spectrum loci (such as occurs for the bee [6]), he spent considerable effort on what happens when the spectrum locus has a straight-line part, as is true for human vision. (He thought the spectrum locus had two flat regions, at low and high wavelengths, whereas we now recognize only the latter.) For such a space, bivalent, two-transition reflectances certainly are on the tristimulus envelope of reflected colors, but they have metamers (or what Schrödinger calls physiological duplicates) that are bivalent with more than two transitions, and even metamers that are not bivalent. The nomenclature might be awkward, but the concept is not: A yellow optimal color is easily understood to be metameric to an additive combination of red and green.

As you read Schrödinger's article, you should be aware that he combined the illuminant and the color-matching functions to create his sensitivity functions $x_j(\lambda)$. Fortunately the convexity property of the spectrum locus doesn't depend on multiplication of all the color-matching functions by the same function of wavelength, so Schrödinger's elision in this case was a compact way of proving theorems that apply across illuminants.

You will also note that Schrödinger avoided doing calculations of the shape of the object-color solid. He saw them as superfluous to the mathematical theorems he was proving, but he also realized that others would soon render such calculations obsolete by arriving at more exact specifications for human color-matching functions. Others, starting with MacAdam, indeed performed the calculations based on the 1931 CIE Standard Observer, and they appear in various color spaces, e.g., in the book by Wyszecki and Stiles [7].

Even now, generalizations of the optimal reflectance formalism are being published. For example, D. Couzin [8] embarked on generalizing the formalism to fluorescent colors. That generalization is not complete, and will need more theoretical analysis perhaps instructed by a linear-programming solution.

As a parting note, we can speculate whether Schrödinger's habitual use of the word "quantum" in his article prefigured his later and profound impact in quantum mechanics.

Michael H. Brill
Datacolor

[1] Ostwald W. Neue Forschungen zur Farbenlehre, Phys. Z. 17**,** 322-332 (1916).

[2] MacAdam DL, The theory of the maximum visual efficiency of colored materials, J. Opt. Soc. Am. 25, 249-252 (1935).

[3] West G and Brill MH. Conditions under which Schrödinger object colors are optimal, J Opt Soc Am 73 (1983)

[4] Brill MH. Color-science applications of the Binet-Cauchy theorem, Color Res. Appl. 27, 310-315 (2002).

[5] Brill MH and Larimer J. Avoiding on-screen metamerism in N-primary displays. J. Soc. Info. Displ. 13, 509-516 (2005).

[6] Menzel R, "Spectral sensitivity and color vision in invertebrates," in *Handbook of Sensory Physiology* Vol. VII/6A, *Comparative Physiology and Evolution of Vision in Invertebrates A: Invertebrate Photoreceptors,* H. Autrum, ed. (Springer-Verlag, New York, 1979), pp. 564-565.

[7] Wyszecki G and Stiles WS, Color Science, 2nd Ed., Wiley Interscience, 1982.

[8] Couzin D, Optimal fluorescent colors, Color Res. Appl. 32, 85-91 (2007).

Theory of pigments of greatest lightness

by Erwin Schrödinger

Annalen der Physik 4, 62 (1920) 603-622

#1 Statement of problem

As is well known, the color of the light reflected from a painted pigment layer never reaches the same degree of saturation that pure spectral lights have, but always appears more or less whitish compared to the pure light of the same hue. It can be generated from the pure light with the addition of a certain amount of white light. The impossibility of realizing colors of spectral saturation with pigments is not a technical problem but to a degree fundamental. Its cause is that the mixture of two spectral lights that in the spectrum are located not too far from each other have, in mixture, a specific hue that lies between those of the two components but, in general, is more whitish than that of the spectral light. To obtain the full saturation of a spectral light the pigment would have to reflect only an infinitesimally narrow wavelength range and absorb all others. But, as already Helmholtz mentioned, its appearance would be very dark, in the limiting case black.

In general, pigments of spectral saturation can only be produced with minimal lightness (we will get back to the necessary limitation shortly). −

The reason for the whitishness of all pigment colors becomes more evident when considering the Newton-König color diagram. It lies in the convexity of the spectral curve *RGV* (see Fig. 1). The color of the pigment is represented by the center of gravity of a certain linear mass distribution along the spectral curve, a mass distribution that is defined by the reflectance function (reflectance coefficient as a function of wavelength) and the illuminant. The center of gravity *P* generally falls into the interior of the real color segment (limited by the spectral curve and the straight "purple line" *RV*), thereby consisting of a certain spectral light *S* and white light *W*, or possibly of *W* and a certain purple mixture.

The only exception to this situation occurs when the reflection is limited completely to the short-wave or completely to the long-wave end of the spectrum, such that the mass distribution is limited either to the range from *V* to *J* (λ approx. 475 nm) or to the range *R* to *O* (λ approx. 630 nm). These end regions of the spectrum, the indicated limits of which are by nature approximate, are according to König completely straight. The center of gravity points of such pigments would therefore fall onto the spectral curve; they would not have less saturation than the corresponding spectral lights.

Red-orange and indigo-violet pigments can be produced with limited (although not very high) lightness but perfect spectral saturation.

It must be kept in mind that points *R* and *V* each represent a finite wavelength range, specifically the two monochromatic end sections of the spectrum (until λ = 655 and λ = 430, respectively), each absorbing a *finite* mass of points (not just a *Liniendichte* [density of line]) if there is reflectance in the corresponding end section.

It is possible to raise for any point on the orange-red or the indigo-violet sections, and equally for any point on the purple line, the question in *which* maximal lightness it can be realized with pigments. It is also possible − while sacrificing in the interest of light

intensity maximal saturation that can be reached with pure lights − to ask this question for points located *near* the boundaries of these segments. This, in the end, leads to the following question applicable to any point of the real color diagram:

In which maximal lightness can this point be realized with pigments and what must the properties of the pigments, i.e. their reflectance functions, be to make this possible. The purpose of the present article is to answer this question.

Considering three arbitrary points not falling on a straight line in the chromatic plane, such as König's fundamentals or points representing three real primary lights, or any three points within or outside the real color diagram, and considering that the unit quanta for these points are defined in some manner, any color is commonly defined with three numbers, that is, by the quanta of the selected three primary colors from which it can be mixed in the real or non-real[1] sense. Viewed as ratios, the three tristimulus values represent the projective, barycentric triangular coordinates[2] of the color of the corresponding point relative to the basis triangle as a coordinate triangle. The sum of the tristimulus values is the mass of the point to be represented and is designated the quantum or quantity of the color. It is, for colors of different location of the representing point (colors of different chromaticity coordinates [*Reizart*] as von Kries aptly says), not a measure of their brightness ratio, except − possibly − in case of very specific selection of the unit quanta of the selected primaries; it is an as yet undecided question that, however, does not need to be raised here. [FN1]

For colors of the *same* chromaticity coordinates (falling on the same point and with the potential to be made identical by simple change of the objective intensity) the quantity of the objective intensity is proportional, as a result certainly − ceteris paribus − a monotonic measure of brightness. Extending our spatially two-dimensional representation of the manifold of color to a three-dimensional one by plotting for every real color point its quantity vertically as an ordinate over the chromaticity diagram, pigments of greatest quantity or highest light intensity or of − ceteris paribus − highest brightness, those that we have been looking for, represent a *surface* of a particular shape over the segment of real colors. Along the *curved* part of the border its ordinate value declines to zero and along the three straight-lined border sections to small ordinate values of continuous curvature. This surface, together with three vertical walls generated by the border ordinates and the standard planar chromaticity diagram shown in Fig. 1 as a basis, *limits the region of colors that can be represented with pigments* in our three-dimensional model. By the way, this form has been selected only for the present for representation purposes and will not be considered further (it is not practical because the absence of objective light is not represented by a *single* point but by the complete basis area.

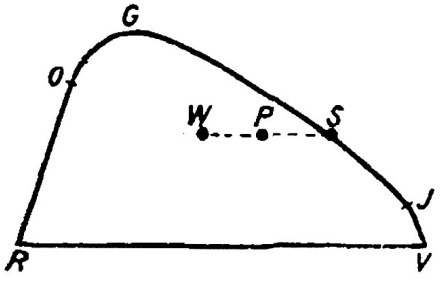


Figure 1

FN1 Schrödinger's exact meaning of *Reizart* is not completely clear. In 1940 M. Richter defined it as 'chromaticity coordinates,' applicable primarily to unrelated colors (M. Richter, Grundriss der Farbenlehre der Gegenwart, Dresden: Steinkopf). This is the English term used in this translation.

All that has been said so far applies for any kind of *illumination* of the colored material, but the light must be selected in advance according to its physical composition and intensity, to be kept in mind in all further considerations. The position of the surface that in our representation limits the upper boundary of the region of pigments varies with the illumination. Not only do all its ordinates grow according to the *intensity* of the illumination, but the *form* of the surface varies with the spectral *composition* of the light. For example, blue object colors are relatively easier produced in considerable saturation and brightness with a bluish light than with a white or reddish one. But the latter simplifies the production of saturated, bright red object colors, etc. The result that we will obtain regarding the pigments of − at given chromaticity coordinates − highest intensity, the optimal object colors as I am calling them, will nevertheless in a certain sense be completely independent of the illumination. It will become apparent that they are those pigments that for any kind of illumination always have optimal character; or, at least, that a two-dimensional diversity of pigments, i. e., of reflectance functions, can be specified independent of the illuminating light and even independent of the specific form of the spectral curve of Fig. 1. For any light they represent the diversity of the optimal object colors for that illumination.

#2 Identification of the diversity of pigment colors implicit forming the bordering surface

The following special coordinate selection for pigments is used for standardization purposes. The points of reference are those of König's fundamentals. The illuminant is one of choice, but for the present it should contain all visible wavelengths; an example might be sunlight. The center of gravity of the triangle represents the color of an ideally white pigment and of all pigments neutrally gray in that light, i.e., of all pigments with constant reflectance functions. The ideal white pigment with reflectance 1 is to have the coordinates 1, 1, 1. If $x_1(\lambda)$, $x_2(\lambda)$, $x_3(\lambda)$ represent König's fundamental sensitivity curves for the interference spectrum of the illuminating light scaled so that

(1) $$\int x_1(\lambda)d\lambda = \int x_2(\lambda)d\lambda = \int x_3(\lambda)d\lambda = 1,$$

the coordinates of a pigment with the reflectance function $r(\lambda)$ are

(2) $$p_1 = \int x_1(\lambda)r(\lambda)d\lambda, \quad p_2 = \int x_2(\lambda)r(\lambda)d\lambda, \quad p_3 = \int x_3(\lambda)r(\lambda)d\lambda,$$

and its quantity is

(3) $$q = p_1 + p_2 + p_3 = \int (x_1 + x_2 + x_3)rd\lambda.$$

The variable r can only have values between 0 and 1, including the limiting values. The p values have the same range, while those for q range from 0 to 3. The coordinate representation is independent of the *intensity* of the illuminating light because due to the definition of the coordinates for the white pigment the "unit quanta of the fundamentals" are automatically adjusted.

The possible forms of the reflectance functions are for the present limited by the following statement that is fundamental for this small investigation:

If a pigment in the vicinity of three locations in the spectrum that in the color triangle do not fall on a straight line has a reflectance different from 0 or 1, that is, lying

between 0 and 1, the reflectance at these three locations can be changed in such a manner that a lighter pigment of the same chromaticity coordinates results.

Because from these three pure lights a *positive* quantum of color inherent in that pigment can be obtained by mixture, at least in the non-real sense, that is, potentially with 1 or 2 negative mixture coefficients. If the reflectance in the three areas falls *between* 0 and 1 one can change them to small degrees (increasing or reducing) in such a manner that the added color quanta are in the right proportion resulting in a lighter pigment of the same chromaticity coordinates. — Symbolically, $\lambda = a$, $\lambda = b$, $\lambda = c$, represent the three places in the spectrum. Then, the assumption is that

$$\begin{vmatrix} x_1(a) & x_2(a) & x_3(a) \\ x_1(b) & x_2(b) & x_3(b) \\ x_1(c) & x_2(c) & x_3(c) \end{vmatrix} \neq 0 \tag{4}$$

The equations

$$\begin{cases} x_1(a)\delta_a + x_1(b)\delta_b + x_1(c)\delta_c = p_1\delta \\ x_2(a)\delta_a + x_2(b)\delta_b + x_2(c)\delta_c = p_2\delta \\ x_3(a)\delta_a + x_3(b)\delta_b + x_3(c)\delta_c = p_3\delta \end{cases} \tag{5}$$

then have, for a preselected small $\delta > 0$, solutions in δ_a, δ_b, δ_c. Changing in the small regions ($\varepsilon > 0$)

$$a \leq \lambda \leq a + \varepsilon, \quad b \leq \lambda \leq b + \varepsilon, \quad c \leq \lambda \leq c + \varepsilon \tag{6}$$

the reflectance $r(\lambda)$ to the following related values

$$r(a) + \delta_a, \qquad r(b) + \delta_b, \qquad r(c) + \delta_c \tag{7}$$

the pigment coordinates change according to equation (2) by the amounts obtained by multiplying ε with the left side of equation (5). The chromaticity coordinates do *not change* because of equation (5), but the quantum changes according to equation (3) by the positive amount

$$\varepsilon\,\delta\,(p_1 + p_2 + p_3) = \varepsilon\delta q,$$

quod erat demonstrandum.

It follows that the reflectance of an optimal pigment can in none of the *finite* regions of the curved section of the spectral curve, and identically on finite regions of *both* straight-lined end portions, be other than zero or one, but certainly in the curved and one of the straight sections can only have one of these two values.

For the purpose of abbreviation I will name these pigments *bivalent* in a spectral region where their reflectance is always zero or one; I will quite simply name those bivalent whose reflectance in the complete spectrum has *only one of these two* values.

The above proof does not apply to one of the two straight ("dichromatic") spectral regions because the determinant (4) disappears. In addition, it must be considered that each of the end points *R* and *V* of the spectral curve corresponds to a finite

("monochromatic") wavelength region. But the present proof also considers the case where in both monochromatic regions simultaneously − but nowhere else − deviations from bivalence occur. Such deviations remain possible either

a) in *one* dichromatic region, including the adjacent monochromatic region, or
b) in both monochromatic regions simultaneously.

Regardless, we can *limit* the considerations *to bivalent pigments* if we only want to experience at least *one* representative each for every optimal pigment and we are not interested in excluding *physiological duplicates*. Because it is easy to see that in cases a) and b) bivalence can be achieved with successive change of the reflectance function, *without change in the appearance of the object color*.

If the three points *a, b, c* are located in the chromaticity diagram on a straight line determinant (4) disappears and the equations (5) have non-disappearing solutions in δ_a, δ_b, δ_c for $\delta = 0$. The corresponding change (7) does *not* change the appearance of the pigment and can be continued in the same sense (or immediately selected finite and of *such* a magnitude) until one of the three numbers $r(a)$, $r(b)$, $r(c)$ reaches the value zero or one. The process can be continued, as long as $r(\lambda)$ continues to contravene bivalence at least in three places, that is, the pigment can be replaced by a bivalent one without change in its appearance. *Quod erat demonstrandum*.

We now limit our examination to bivalent pigments and have to find among them those that are optimal. The reflectance of a bivalent pigment is an unstable function of λ; it jumps at one or more places from zero to one or from one to zero. I call such a place a transition point [*Sprungstelle*] ($1 \rightarrow 0$) or ($0 \rightarrow 1$), the arrows to be understood as indicating increasing wavelength. It is quite obvious that in general *the number of transition points* for optimal pigments *cannot be larger than two* − but this again excludes duplicates. For now, the following statement applies, being analogous to the first one in many respects:

If a pigment has three transition points that in the chromaticity diagram are not located on a straight line, by moving the transition points its reflectance can be changed in a manner that results in a lighter pigment of the same chromaticity coordinates. Therefore, it cannot be optimal.

The proof is completely analogous to the former one and therefore does not need to be extensively detailed. It is based on the fact that the color of the pigment can be mixed, at least in the non-real sense, using the three transition colors so that an appropriately, according to direction and magnitude, selected shift of the transition wavelengths produces the desired increase in lightness without change in chromaticity coordinates.

Three or more collinear transition locations appear again to be possible. But as long as there are at least three present they can be adjusted without change in the appearance of the pigment until they join up so that eventually there are at most two. Even easier than via the earlier analog calculation, this can be directly comprehended in the following manner, for example for the long-wave region *RO*.

Let's assume a pigment that in the remaining spectrum is bivalent but in the spectral region *RO* (including the borders) has an arbitrary form. (Compare the peculiar but easy to comprehend Fig. 2 in which the reflectance coefficients have been plotted on the outside of the spectral curve.) The problem to be solved is to replace the existing

pigment with one that is also in the region *RO* bivalent, has a minimal number of transition points, and is physiologically identical.

So that the latter requirement is met, the "partial color" generated by the reflectance in the region *RO* must — according to Grassmann's rule that when mixing equal-appearing lights the result is equal-appearing lights — be the same as the original one. In any case, the point corresponding to this partial color lies somewhere on *RO*. So as to have in the end as few transition points as possible we "produce" the partial color in different ways, depending if the reflectance function of the pigment in question perfectly reflects or absorbs at point *O* in the violet direction.

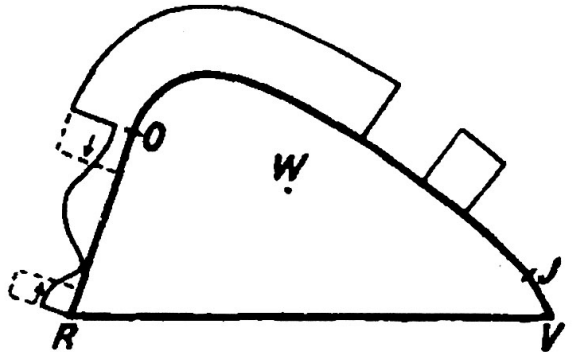


Figure 2

In either case we first assume complete absorption in the region *RO*. Then we move in the *first* case (corresponding to the illustration) a transition point ($0 \rightarrow 1$) from the red end of the spectrum in the direction of violet, and at the same time the transition ($1 \rightarrow 0$) from *O* in the direction of red, at such a relative speed that the in this fashion generated reflectance is always in agreement with the chromaticity coordinates of the partial color. This is apparently possible and can be continued until also the quantum of the partial color is reached. Continuation of the process is only made impossible when one of the transitions passes over the point of the partial color. But this is not possible before the quantum of the partial color is reached because in such a case between the transition points only spectral lights would remain "unused" that are either all redder or all yellower than the partial color and from which no quantum of the partial color can be mixed anymore; the partial color could not be mixed from the remaining available light in sufficient quantity, contradicting the initial assumption. —

In the second case where there is a region of absorption from point O on in the direction of violet we generate the partial color in increasing strength by with a single reflectance range that contains the point of the partial color in its interior. We move, again starting with an identically minute *r* in *RO*, a transition location ($1 \rightarrow 0$) from the partial color in the direction of red and at the same time one ($0 \rightarrow 1$) from the partial color in the direction of violet, at such relative speed that the chromaticity coordinates of the partial color is matched. For reasons comparable to those presented earlier also the quantum of the color can be matched in this manner either before, or in the worst case at the same moment where the continuation of the process is thwarted because either the first transition reaches the red end of the spectrum or the second the point *O*. —

Thereby we have solved our task.

If in this manner only one transition results, a situation that rarely occurs, a second can be introduced somehow in the spectrum; if two result, they must be the only ones if the pigment is to be an optimal one.

Only one case can occur: the two necessary transitions both are located in the monochromatic end region. They thereby fall on the same point R of the chromaticity diagram and there is the possibility of a third transition somewhere in the spectrum. But it is immediately apparent that the isolated monochromatic reflectance or absorption region involved in this situation can be pushed completely to the end of the spectrum so that there is only *one* monochromatic transition point present. —

Reviewing what has been said, *all that remains*, after excluding many duplicates, *is the two-dimensional diversity of bivalent pigments with one or two transitions*. Among them every optimal pigment color must have at least one representative. We now only need to exclude a few smaller groups that certainly are not optimal. It will be possible to demonstrate for the remainder that among them are no two pigments of identical chromaticity coordinates. The result will be a proof that all these pigments are truly optimal and that they exactly represent the originally mentioned diversity of borders, always in one example, without duplicates.

In the next step we place the one- and two-transition bivalent pigments into the following groups, made quite certainly easily comprehensible in Fig. 3. The abscissa lines represent the visible spectrum, the ordinate reflectance.

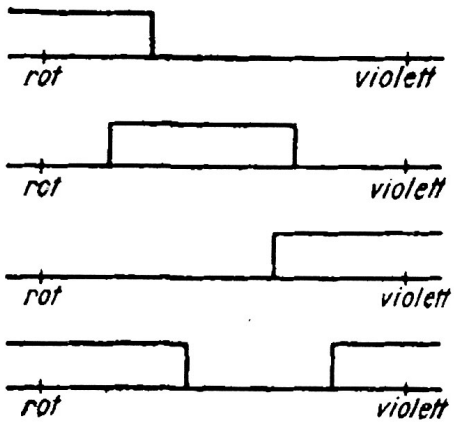


Figure 3, Top: Long-end pigments; second: Middle pigments; third: Short-end pigments; bottom: Middle-fail pigments

The following can be excluded as certainly non-optimal:

a) of the long-end pigments those whose reflectance does not reach up to the short-wave border of the monochromatic red;
b) of the middle pigments those whose transition points are both located either *between* the red spectral end and *O* or *both between J* and the short-wave end;
c) of the short-end pigments those whose reflectance does not reach to the long-wave border of monochromatic violet;
d) of the middle-fail pigments those who have one transition *within the region* of monochromatic red, the other *within the region* of monochromatic violet, because the related purple can be strengthened by expansion of the two reflectance regions until one of the two transition points reaches the border of

the monochromatic region. (Those with two identically colored monochromatic transitions have been already earlier excluded as duplicates and been replaced with end pigments.)

It is further necessary to prove that all the pigments left over after excluding those from a) to d) have different chromaticity coordinates. We will first consider those three sub-groups the chromaticity coordinates of which fall on one of the three straight-lined border segments. They only are in competition among each of them separately. The truth of the claim is clear without further consideration and has just been established using exclusions a) to d).

For the rest the proof − not very elegant − must be established by group, i. e., first each group must be investigated for its members and then compared against each other group and all possible combinations of position of transitions must be taken in consideration. At any rate, the end pigments can be neglected because they are after all considered as degenerated versions of the middle and the middle-fail pigments. We therefore, first compare

A. Middle pigments among themselves.

In case of totally separated reflectance regions coincidence of centers of gravity is clearly impossible. The same applies in case of connected reflectance regions. Because in that case (according to the above mentioned treatment of spectrally saturated pigments) the reflecting regions must also contain curved portions of the spectrum locus, the center of gravity of the non-common external parts of the more extensive reflectance region falls outside of the segment belonging to the smaller reflectance region and must displace the center of gravity of the latter. The same applies if the regions of reflectance cross over. The common middle part combined in each case with a non-common outer part cannot result in the same point in both cases.

B. Middle-fail pigments among themselves

Here the basis in all cases is a pigment that contains both absorption areas and it is to be kept in mind that the center of gravity of such a "difference pigment" is changed in different ways by the corresponding enlargement.

C. Middle with Middle-fail pigments

α) The reflectance region of the first and the absorption region of the second are completely separated. The middle-fail pigment is generated from the middle pigment by addition of reflectance regions, a situation that makes it impossible that its center of gravity remains unchanged.

β) The reflectance region is located within the absorption region. This case does not require consideration.

γ) The absorption region falls within the reflectance region (see Fig. 4). In this case the total color region consists of three strips the centers of gravity of which (that is, of the bordering parts of the spectral curves), if the pigments are to have the same chromaticity coordinates, must lie on a straight line in such a way that the center of gravity of the middle strip does *not* lie in the middle; this is evidently impossible.

δ) Absorption and reflectance regions overlap (see Fig. 5). It would have to be possible to locate within each of the shaded regions I, II, and III a point of such kind that

the three points fall on a straight line, with point III *not* between points I and II. To achieve this, the addition of curve II to curve III would have to bring the center of gravity of III to the same point as the addition of the curve pair I. This is evidently impossible. —

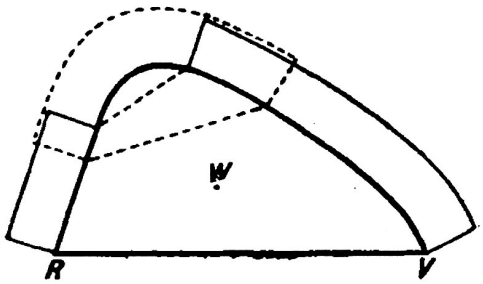


Figure 4

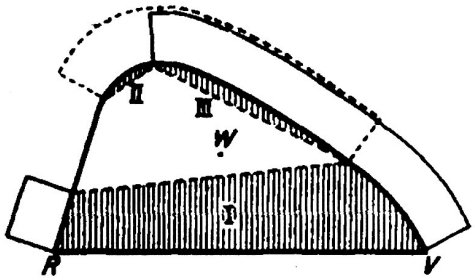


Figure 5

As a result, the two-dimensional diversity of pigments to which we have been led completely represents the optimal pigment colors, and always in *one* example. As has been mentioned earlier, this applies for *any kind of illuminant* as long as it contains all wavelengths. Because our definition of pigments has neither any relationship to the illuminant, nor have we in the process of the investigation made use of any other property of light except silently positing that no homogenous light of any wavelength should be completely lacking.

That our pigments do not loose their optimal character in cases where the illuminant has spectral gaps can be recognized immediately by a border crossing from an illuminant that is different only to a small degree, in which the gaps are filled with small ordinate values being reduced toward zero according to a particular rule. In such a situation a pigment will generally change its position in the chromaticity diagram but not its physical properties, therefore remaining optimal in the borderline case. But the unequivocal separation between pigment multitude and optimal colors does not stay in place because large groups of pigments assume identical colors, that is, all those where the transition point falls into a spectral gap of the illuminant; of course in such a case in general the course of reflectance within such a gap is completely without effect on the appearance of the pigment.

For a short set of conclusions free of the special terminology introduced above concerning the so far determined main results of our investigation I refer the reader to the last page of this article.

#3 Concerning the answer to the questions regarding the highest possible lightness, the highest possible saturation, and the necessary conditions for pigments of highest lightness.

It is evident how one needs to proceed in calculation to answer in any specific case the first part of the earlier posed question: *In which maximal lightness can a certain point of the chromaticity diagram be realized with a pigment.*

This question is only meaningful if the illuminant is specified. The preliminary first step is the recalculation for the interference spectrum of this light of the three fundamental sensitivity curves in form of a table of the following three integrals

$$\int_{\lambda 0}^{\lambda} x_1(\lambda)d\lambda \ , \qquad \int_{\lambda 0}^{\lambda} x_2(\lambda)d\lambda \, , \qquad \int_{\lambda 0}^{\lambda} x_3(\lambda)d\lambda$$

where λ_0 represents the short-wave end of the spectrum, with the upper integral limit varying from λ_0 all the way to the long-wave end. In such a table the coordinates of the optimal colors can be determined with little effort. If used regularly the coordinates can be combined in a table with two entries. For sensible experimentation − I do not see an easier way − a graphical method might be best, plotting *one* integral as a function of the other and looking on the resulting curves for cords of the appropriate slope − in this way one can determine *those* transition points (borders of integration) for which the three coordinates of the pigment have those values predetermined by the point in the chromaticity diagram. These three coordinates indicate which fractions of the red, green, and blue sensations caused by an ideal white pigment illuminated by the selected light can be maximally caused by a pigment that in this illumination realizes the desired dominant/complementary wavelength.

Of course, the same considerations and calculations can be made employing any other calibration functions in place of the fundamental functions, for example those representing three real primary lights. In such a case the optimal coordinates have a more concrete meaning, outwardly free of any hypotheses concerning the generation of color perceptions.

For this reason I would consider it more or less wasted effort to actually do the described calculations because it is very uncertain if the first person who would want to make practical use of such data would necessarily want to do it for daylight and for König's fundamental sensitivity functions in the form in which they exist today. —

I would only like to briefly touch on a question related to the one just discussed and perhaps of greater interest to people with practical concerns: in which maximal *saturation* can a color of a given *hue* be produced with a pigment if its *lightness* is not to fall below a given value.

Without doubt the desired pigment is to be looked for in the border manifold, to be specific on a radial line originating in the white point,[3] very far away from it, just *far enough* so that it is not located below the desired lightness. It would be necessary to raise over the range of colors, in the same manner as earlier the "quantity surface", a "lightness surface" the ordinates of which at any point show the lightness of the optimal color. So that its meaning is truly unique, it must be uniquely calculable from the found three coordinates that uniquely define the pigment color. But the ideas on *how* to do this today deviate still widely. Some consider lightness to be a linear function of the coordinates, with constant coefficients, the "specific lightnesses" of the primaries; others, among them Helmholtz in his treatises on the integration of Fechner's law into the color system, believe that there is not just a simple additive link.

The question of the correct definition of the term heterochromatic brightness is extremely important and of much greater consequence than the current investigation; I will concern myself with it soon in some detail in connection with another matter. Here it

obviously plays only a subordinate role. If one would produce the optimal light mixtures of the hue to be obtained with increasing saturation according to the mentioned procedures, one could quickly find among them the one that just meets the requirement, if such requirement has been posed reasonably and is not just a verbal one.

Concerning the *second part of the earlier posed question*: *what is the spectral structure of the optimal pigments, i. e., what are their reflectance functions*? − It is not at all sufficiently answered with our two-dimensional manifold of pigments. All excluded duplicates are also optimal. An optimal pigment does not at all just be bivalent; it can have more than two transitions, etc.

All occurring uncertainties have their basis in the existence of dichromatic and monochromatic regions in the spectrum; in these regions the reflectance function may in certain circumstances be arbitrary in the widest sense. But what is permitted and what not can in any case be comprehended on basis of our earlier considerations and I consider a complete listing of all possible cases to be without interest and unnecessary. The question is always the following: can a reflectance function be appropriately changed into one of the specified forms or not.

For example, on the one hand a pigment that completely reflects light from the short-wave end to point *O* (border of orange) will be optimal, regardless of its reflection in the long-wave region, because it can be changed into a middle-fail pigment. On the other hand pigments where the reflection is limited to the long-wave end up to point O, deviations from bivalence in the dichromatic portion are inadmissible, in the monochromatic portion only admissible if the dichromatic portion is reflecting completely. Otherwise the pigment, when converted, would result in an inadmissible middle pigment. (see pg. [7] above). A corresponding situation applies to the analogous cases at the short-wave end.

#4. Comparison with empirical findings

Wilhelm Ostwald, based on his extensive experimental investigations of pigments, has purely empirically drawn the conclusion that to obtain highest *purity of color* (Farbenreinheit), pigments should only have reflectance values of zero and one, with a sharp transition and that there should be only *one* uninterrupted region of reflection or *one* uninterrupted region of absorption.[4]

So far Ostwald's empirical findings are in the main in complete agreement with my theoretically developed conditions for optimal pigments.

To achieve highest purity Ostwald also requires that absorption and reflectance, respectively, exactly encompass a "semichrome" (*Farbenhalb*), that is, exactly from a spectral color all the way to its complementary color. So as to develop at least a qualitative understanding of this requirement we have to remember that Ostwald's understanding of *purity* involves that *fraction of pure color* that must be contained in the total perception caused by the mixture and that it should be possible to conceptually separate it from the mixture; to this fraction of pure color are added certain fractions of *white* and *black* that contaminate or dull the color.

According to Ostwald the blackness of a color can be increased for example by mixing it on a disk mixture apparatus with an ideally black, i. e., non-reflecting pigment,

or, even better, with a black aperture. Considering Talbot's law, it can be concluded from this that − regardless of how one thinks about the perceptual nature of black − the objective correlate of what Ostwald calls blackness content is in any case a *relatively low lightness*.

It is clear that those optimal pigments, the reflection of which is limited to a very *small* region of the spectrum, have very low lightness. That is, according to Ostwald, they have a high blackness content and for this reason they will have low purity. — Alternately, pigments where the reflecting region extends over a very *large* portion of the spectrum have high lightness, thereby containing little blackness but much whiteness — the latter according to the known general laws of light mixture. — That Ostwald's pigments of highest purity can only be found among my optimal pigments naturally follows from the consideration that a pigment of reduced lightness but identical chromaticity coordinates has the same whiteness content but higher blackness content and thereby reduced purity.

Limitation of the reflectance range to the region between two complementary colors is apparently an empirically proven compromise between the Scylla of whitish and the Charybdis of blackish dulling. Expressed in a manner that follows Helmholtz and that is pointed more in the direction of the objective composition of the radiation mixture: it is a tool for achieving as high a color saturation as possible without a lot of loss of lightness from absorption.

As a result of this kind of compromise one would expect that the quality of dullness attached to Ostwald's best pigment colors is that of a middle gray, mixed from similar amounts of white and black. According to Ostwald's purity determinations[4] this applies as an approximation to a number of them. But for many, primarily the blue and green ones, the blackness content is significantly higher than the whiteness content.

I have to mention here that I used all of Ostwald's terminology only to be able to compare the facts found by him with my theoretical results, not because I am already convinced that terms like 'purity", "blackness content" and "gray" have the same quantitative validity as those of, for example, Helmholtz-König's physiological color metric. Even though I have highest esteem for Ostwald's valuable successes, obtained with hard work, I do not consider his absolute determinations of "purity" and "gray" from the reflectance values at only *two* locations, even though they are unique (maximum and minimum) but as at best a good rule of thumb, in no way suitable for exact definitions of these concepts.

Conclusion

1. Pigments of a given chromaticity coordinates have the highest lightness if they have the following constitution:
   a) In no region of the spectrum do they have a reflectance value other than 0 or 1.
   b) Their reflectance undergoes change at most at two places (transitions from 0 to 1 or from 1 to 0) and cannot be zero everywhere.
   c) If reflectance is limited to one of the two dichromatic spectral regions, including the bordering monochromatic one, it reaches at least to the one end of this region.
   d) If reflectance is limited to the monochromatic regions it covers at least one of the two completely.

e) If *absorption* is limited to *one* monochromatic region it begins at the spectral end.

2. The described pigments have the indicated property for any kind of illumination, that is, under any kind of illumination none of the pigments is exceeded in lightness by any other pigment that remits in the *identical illuminant* light of the same chromaticity coordinates.

3. If the illuminant does not have spectral gaps all the pigments among themselves are physiologically different and uniquely cover the segment of real colors, including its surface.

4. There are pigments of a given chromaticity coordinates with highest lightness in addition to those mentioned in section 1. *All* pigments whose *absorption* is limited to one mono- and dichromatic end of the spectrum have this property, with the course of absorption arbitrary. In general the allowed deviations from the properties mentioned in section 1 are in all cases related to the mono- and dichromatic regions. Of course, each such pigment with a deviating structure will match *one* of the pigments mentioned in section 1, — but *which one* will generally depend on the illumination.

5. The above theoretical results are in agreement with a few of Wilhelm Ostwald's empirical results.

Vienna, December 1919, II. Physical Institute of the University.

(Entered December 22, 1919)

Notes

1. Mixable in a non-real sense is my designation for a color located outside the selected basic triangle for which, therefore, one or two tristimulus values are negative. The concrete meaning of this is well known.
2. I designate as "*barycentric*" a triangular coordinate system where the "unit point" falls on the center of gravity of the coordinate triangle. That fact is not to be confused with the arbitrary but convenient and therefore frequently practiced transition of the *white point* to the center of gravity.
3. It should be noted that in the chromaticity diagram in use for our pigments the color at the center of gravity is not that of sunlight-white but that of an ideally reflecting surface so that the color at the center of gravity is that of the illuminant. *Here* "white point" does not indicate the center of gravity but the locus of sunlight-white.
4. W. Ostwald, Königl. Sächs. Ges. D. Wiss., Abh. d. Math.-Phys. Kl. **34**. No. 3, p 471 ff. 1917; Phys. Zeitschr. **17**. p. 328 ff. 1916.