Color naming: color scientists use Munsell Sheets of Color

Giordano B. Beretta, Nathan M. Moroney

Abstract:

With the advent of high dynamic range imaging and wide gamut color spaces, gamut mapping algorithms have to nudge image colors much more drastically to constrain them within a rendering device's gamut. Classical colorimetry is concerned with color matching and the developed color difference metrics are for small distances. For larger distances, categorization becomes a more useful concept. In the gamut mapping case, lexical distance induced by color names is a more useful metric, which translates to the condition that a nudged color may not cross a name boundary. The new problem is to find these color name boundaries. We compare the experimental procedures used for color naming by linguists, ethnologists, and color scientists and propose a methodology that leads to robust repeatable experiments.

Color naming: color scientists use Munsell Sheets of Color

Giordano B. Beretta, Nathan M. Moroney

Hewlett-Packard Laboratories, Print Production Automation Lab
1501 Page Mill Road, Palo Alto, USA

ABSTRACT

With the advent of high dynamic range imaging and wide gamut color spaces, gamut mapping algorithms have to nudge image colors much more drastically to constrain them within a rendering device's gamut. Classical colorimetry is concerned with color matching and the developed color difference metrics are for small distances. For larger distances, categorization becomes a more useful concept. In the gamut mapping case, lexical distance induced by color names is a more useful metric, which translates to the condition that a nudged color may not cross a name boundary. The new problem is to find these color name boundaries. We compare the experimental procedures used for color naming by linguists, ethnologists, and color scientists and propose a methodology that leads to robust repeatable experiments.

Keywords: color naming, categorization, crowd-sourcing, print automation

1. INTRODUCTION

1.1 Communicating color

The characteristic that sets *homo sapiens* apart from other animals is the capability of highly structured rich communication, and humans have always been willing to make large investments in communications technology, guiding the direction of industrial research in color science and other technologies. Color naming, being at the intersection of color science and linguistics, has stimulated a particularly rich cornucopia of research.

When people are interested to invest in a technology, companies rush to create products in their R&D laboratories. These days, product ideas come mostly form strategists and marketeers, so in reality the labs quite often solve hard detail problems rather than creating technologies *ex*

novo. In the remainder of this section we will present some relevant problems we encountered. In the next section we will discuss prior work and in the remaining sections we will present our contribution, while relating it to the prior work.

1.2 Some real world problems

A few years ago we developed a number of Web services (today we would call them *EaaS* for **E**verithing **as a S**ervice or we would invoke cloud computing) for small and medium size businesses (SMB) that allowed them to create all sort of printed materials using a "graphical artist in the box" metaphor. A customer sent us the problem illustrated in Fig. 1: a number of store displays had been designed, but then the customer decided to change the background color in the template to make the text in the displays more conspicuous. Unfortunately, after the job was completed, some of the products in the display fused with the background.
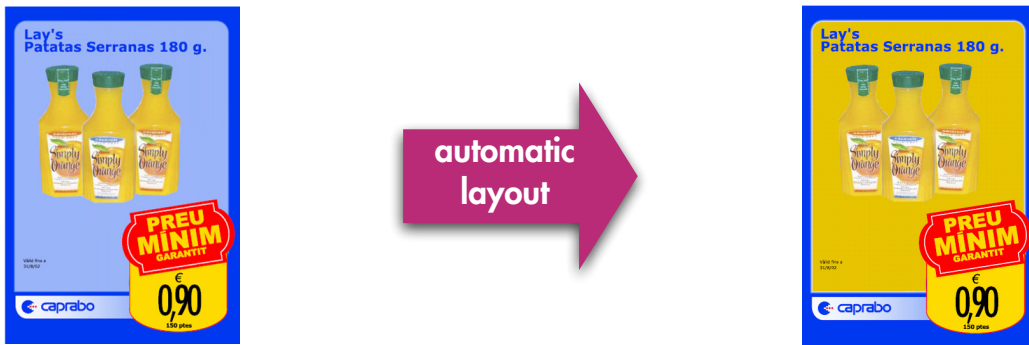

Figure 1. A store display, where the poster's color scheme was changed and suddenly the depicted product became invisible.

The Web service called for a kind of "spelling checker for colors." The algorithm is straightforward, all that is needed is comparing each page element's color with the color of the element behind it. However, the value for the threshold is not obvious — clearly a just noticeable difference (JND) would be far too small. We decided to opt for a heuristic based on how easy it is to name a color unequivocally. Sturges and Whitfield[30] had determined the foci and centroids of the basic color terms in the Munsell system; these terms are defimed as follows (see also [6, p. 143]):

DEFINITION 1.1 (FOCAL COLOR). *Focal colors are defined as the fastest named consensus sample (all subjects name that sample consistently with that color term) for each basic color sample.*

In other words, the focal color is in the region where the probability of that name being used is greatest.

DEFINITION 1.2 (CENTROID COLOR). *Centroid values are calculated by taking an average across trials, of hue, chroma, and lightness from all responses that elicited the particular color name in a color naming experiment, weighted according to the number of times it was used.*

In other words, the centroid color is the middle of the region of color space in which the term is used.

The heuristic was to consider the distance between the focal and the centroid color of a same color term to be an indication of how clearly a basic color is discriminated without hesitation, and we took the maximum as the threshold. We also introduced a threshold on the lightness difference to add robustness for people with color vision deficiencies (CVD).[2] This heuristic was then replaced with proper psychophysics experiments and statistical analysis by Zuffi and Brambilla.[34]

Another Web service we had developed was for the creation and management of marketing collaterals. A company like HP has a large number of products and therefore also of brochures. We had a Literature Distribution Center (LDC) in Campbell, and before a sales call the account managers would order the collaterals for the proposed products from the LDC. Brochures were designed incorporating HP's corporate color palette. We implemented a document management system in which a sales person could select a number of materials for a particular prospective customer through a graphical user interface (GUI) based on the shopping cart metaphor. Our system would customize the components for the customer and create a booklet with a uniform look.

Later the corporate palette was improved and to preserve the uniform look we had to constrain the functional colors to the new palette: while the left chiclet in Fig. 2 is readable, the right chiclet in the new palette has become hard if not impossible to read for persons with CVD. This requires first assessing the readability of colored text on a colored background and second determining how the colors should be nudged to restore readability.
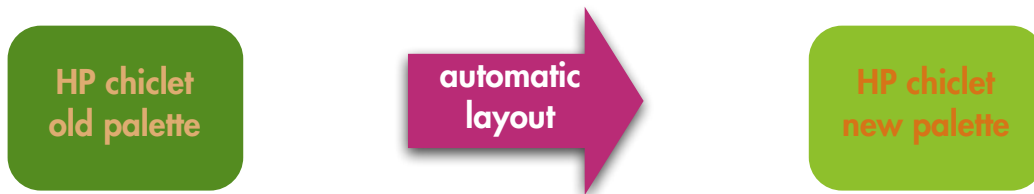

Figure 2. Our company changed it corporate color palette, and after the colors in a brochure where nudged to the new palette, some colored text an a colored background became unreadable.

The solution for the store display example was too aggressive and we needed a new solution that would minimally nudge the colors. To this purpose we implemented the Coloroid system,[23] which has a comprehensive color naming system based on hue leaves as shown in Fig. 3 for the hue $A = 20$. This allowed us to introduce a lexical metric requiring that between a foreground color's name and that of the background there be at least an intermediate name.[3]

1.3 New problems on the horizon

In the 1997 edition of this conference[1] and in a follow-up panel discussion at the Color Imaging Conference later that year in Scottsdale, we had argued that color consistency is more important than color fidelity, where the latter is defined as all pixels in an image differing less than $1\Delta E$ between original and reproduction, and the former referring to the color palette of an image remaining consistent, specifically, the color error vector field between original and reproduction
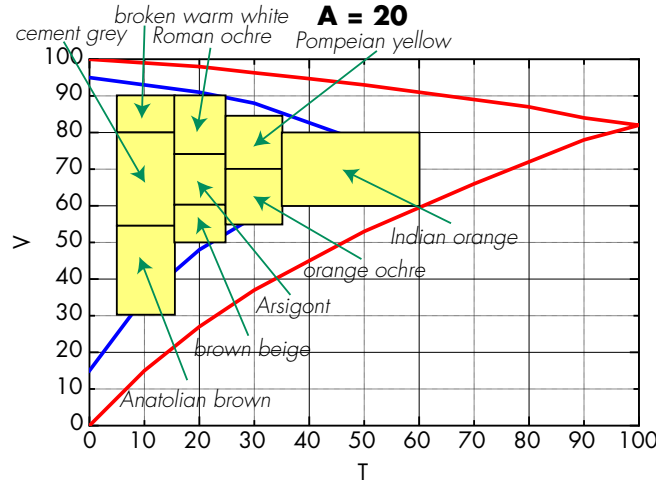
Figure 3. A hue leaf in the Coloroid system and its subdivision in regions with the same color name.

be free of divergence (no spurious light sources are introduced) and no color name boundaries be crossed.

However, since then the relative gamut difference between displays and printers has not changed much and gamut mapping has remained a relatively benign — albeit unsolved — problem not requiring large geometric changes. But in the last couple of years we have experienced the advent of imaging devices with very large color gamuts, like consumer cameras with 14 bits per pixel and displays with the xvYCC gamut. When images captured and edited on such devices are printed, a more radical gamut mapping is necessary than heretofore. This introduces new problems, like for example when a yellow pixel is made much darker, its color name changes to olive and object recognition mechanisms in the human visual system (HVS) require an additional cognitive effort, which is perceived as a reduction of the image's quality.

EaaS Web services have also evolved into a new ecosystem where users without training in graphical arts can use a Web tool to design collaterals for their products and print jobs are dispatched to nearby print service providers (PSP) for fulfillment. For the service's success it is important to assist the user with practical advice on how to improve their design to obtain a professional result. In the example shown in Fig. 4, Mr. Hans Müller, a consultant based in Munich, wants to explain Mrs. Hanako Yamada, a customer based in Kyoto, that some of the lines have insufficient contrast with the background. When Mr. Müller is a person, he can efficiently label the gray arms (at least in German). However, when Mr. Müller is a synthetic character in the computing cloud, an algorithm is required to label the line segments, and "gray" in this case is easier to elicit than "arm."

Ideally, the color naming algorithm should be multilingual, so Mr. Müller or his implementor can work in German while Mrs. Yamada can rely on the cloud to work as she would be communicating with somebody over in Gion. This is illustrated in Fig. 5. "Multicultural" here entails more than just translating color names. For example, In Japanese there are many loan words called "garaigo." In the case of color terms, the Japanese term and the loan word are not
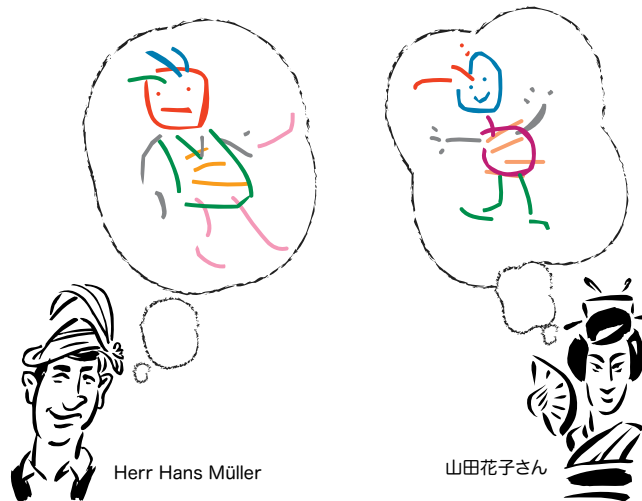
4

Figure 4. Communicating about color across cultures or from machine to human can be challenging.

synonyms, but the term of English derivation is generally brighter.[29]

Interestingly, already in the mid 1990s, when the first low-cost color printers became available but Windows 3.1 was using color just for the GUI and not for content, Canon was shipping its bubble jet printers with drivers that could accomplish to some extent this sophisticated kind of operations.[20]

Our final imminent problem is print quality control. As we explained in detail about preflight checking in variable data printing,[2] when every page is different quality control is a big challenge because pages cannot be sampled from the press stacker (Fig. 6). When hundreds of distinct pages are printed every minute, print quality can only be checked through computer vision
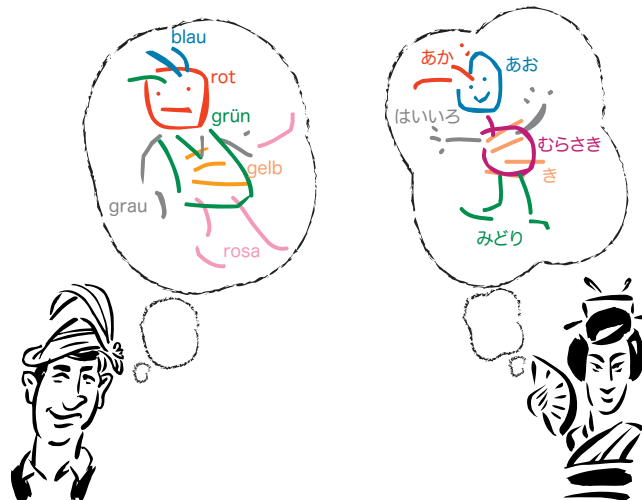

Figure 5. When services are in the cloud, they are multilingual and customers can remain oblivious of language and culture. They just work!
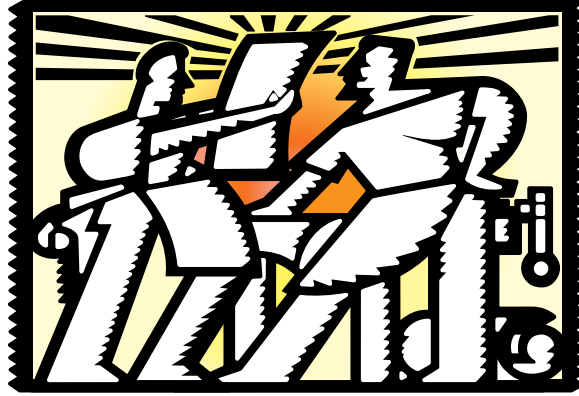
Figure 6. In conventional printing without variable data, every page is the same and print quality can be checked by sampling printed sheets from the stacker. With variable data printing, quality assurance requires machine vision solutions that can ensure every single sheet meets a quality threshold.

algorithms deployed between the marking engine and the stacker. The challenge is to determine the threshold beyond which the press output is no longer acceptable.

In the discussion at last year's session of *The Dark Side of Color* in this conference, we offered another argument against optimizing for color fidelity in print. We noted how in the 1960s and 1970s people were infatuated with HiFi stereo sets, single lens reflex cameras, and fine grain silver halide films. For example, phase distortions were considered very critical and exotic speakers made sure sound from all instruments in an orchestra was delivered at the correct time. Where are we now? Today people listen to their music wherever and with MP3 players that have no notion of phase; when people take pictures, they do it with their phone cameras based on microscopic sensors and cheap lenses.

In this light we should understand that — as we pointed out at the beginning — what people pay for is communication. If we can print more efficiently and at a lower cost, people are willing to live with a little less quality. The art of specmanship is to determine what is good enough. For what concerns quality control at the press, the same rules apply as for gamut mapping. We suspect that for most print jobs, good enough is that there are no color name changes, regardless of the error in terms of ΔE.

All the problems in this section require algorithms for color naming. For some problems we must know where the boundaries of color names are; for others we must be able to assess large color differences, for which a lexical metric is more adequate than one based on just noticeable differences (JND); and for the Web service problem we must be able to translate a colorimetric color specification into a label easily understood by a person without domain knowledge in the graphical arts. We should be able to do this for any culture.

2. COLOR NAMING APPROACHES

We omit an overview of the literature: the interested reader may want to start with Ronchi's overview of research in color and language,[26] which includes an extensive bibliography. The

proceedings from the 1992 Asilomar conference[16] revisiting Berlin and Kay's work are also an authoritative overview with in-depth discussions. Finally, for a brief and more general overview, Derefeldt et al.[11] have written an excellent survey paper on cognitive color.

2.1 The linguist's approach

The typical approach linguists use to study color names it to ask a number of subjects, for example the students in a lecture hall, to write on a sheet of paper all color names that come to their mind within a given amount of time.[28] The resulting lists are then analyzed statistically.[9]

The quandary we have with this approach is that the subjects are never shown an actual color sample. While some correspondences between languages can be made through the color ontogeny of the World Color Survey (WCS) as depicted in Fig. 7, not having an actual color sample as a reference impedes translations or the study of how focal colors change over time. This important limitation prompted the title of this paper. If the subjects are not shown at least something as basic as Munsell Sheets of Color or samples from the Natural Color System (NCS) atlas, the resulting inferences are not of much value for us color scientists in industrial laboratories.

Before we continue with the details of research in color naming, this is the appropriate place to define the terms of saliency, [16, p. 4] and consistency [6, p. 141].

DEFINITION 2.1 (SALIENCY). *A color term is salient if it is readily elicitable, occurs in the idiolects of most speakers, and is used consistently by individuals and with a high degree of consensus among individuals.*

DEFINITION 2.2 (CONSISTENCY). *Consistency relates to the probability that a color name, if used by a given subject on the first presentation, will be used again on the second one.*

2.2 The World Color Survey

The cornerstone of research in color naming is Brent Berlin and Paul Kay's proposal of eleven basic color terms and the study of their universality and evolution.[5] This work evolved into the World Color Survey. Fig. 7 summarizes their results in a single graph.

Essentially, it postulates that there are 11 basic color terms that are universal and complete, i.e., there are no further salient color terms. The WCS is based on physical color samples that are shown to the subjects. The specifications are available on the WCS Data Archives[8] Web site, where you can also find the results of many experiments. Fig. 8 is a rendition of the color samples on that site.

The use of Munsell Sheets of Color allows to subdivide the grid of the Munsell system color gamut shown in Fig. 8 to outline each region where the same color term is used. More precisely each patch in the grid is labeled with the modal color term, i.e., the color term that is assigned to that patch by the largest number of subjects. The region for a term may be required to be topologically connected.[25] When we use a graph coloring algorithm to paint each grid location with a same color when it has a same label, then we obtain a *mode map*.

DEFINITION 2.3 (LEXICAL COLOR CATEGORY). *A the region on the Munsell gamut surface formed by the patches that have a same label in a mode map, i.e., the patches that are assigned a same basic color term.*
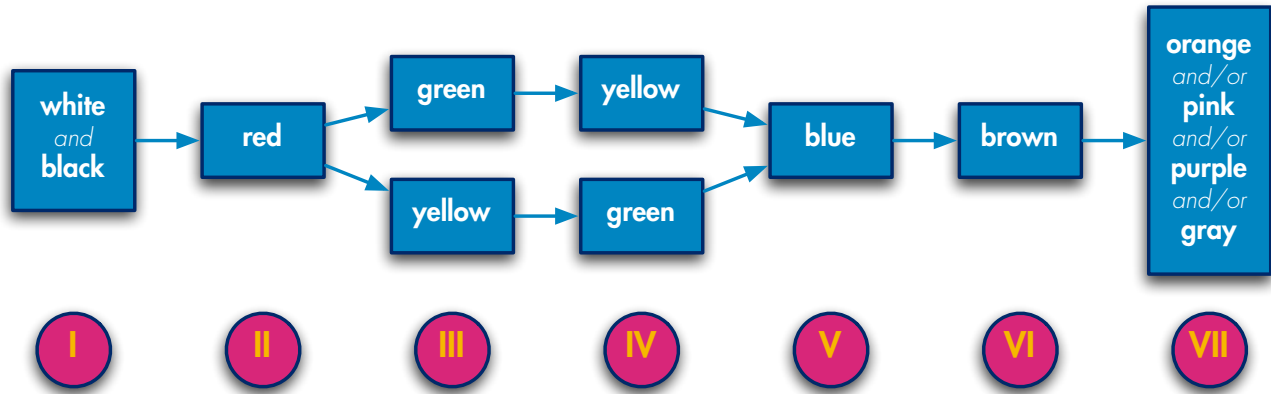
Figure 7. This graph summarizes Berlin and Kay's work, listing the basic color terms and their ontogeny.

Despite its authority, the WCS has a number of problems. For example, only samples on the surface of the Munsell system color gamut are used. Thus, less vivid terms like *peach* and *tan* can never be elicited [6, pp. 144–145]. Boynton and Olson[7] have solved the gamut surface limitation with a careful psychophysics experiment conducted under controlled conditions in a laboratory and using the OSA color atlas.

Although using the OSA atlas they were able to present a full gamut volume, this volume is smaller than the Munsell gamut volume. Sturges and Whitfield[30] repeated the experiment in the Munsell color space. In both cases, the constraints on consistency, consensus, and response time lead to large unnamed gaps, which require the introduction of artifacts like *links* and *bridge* colors [6, p. 144].

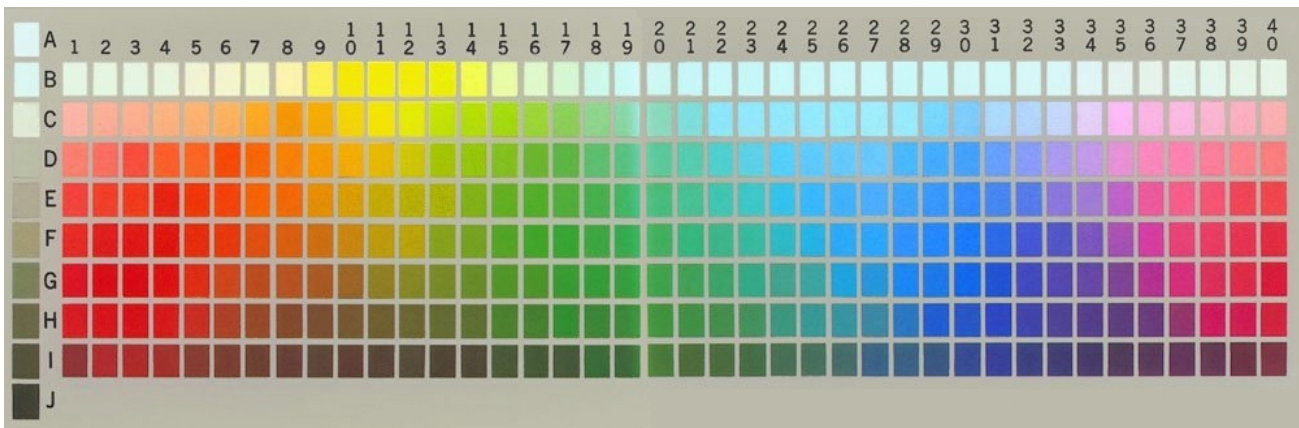These gaps have sparked much controversy. For example, as described earlier *peach* (some-

Figure 8. The colors on the surface of the Munsell system color gamut are used in the WCS. The shown colors are very approximative, for experiments you should use either Munsell Sheets of Color or create instances from the specifications in the WCS Data Archives, `http://www.icsi.berkeley.edu/wcs/data.html`.

times labeled *tan* or *flesh*) is never elicited [6, p. 145] and [16, p. 362], despite having become very common terms after carpets of this color became ubiquitous in American homes. Most discussion has been around the yellow-green (*chartreuse*) and blue-green (*turquoise*) regions [16, pp. 6, 12] and [6, p. 144].

For example, in 1984 Zollinger and Zimmer had a diatribe in the journal *Psychological Research*[31,32] over the need of a *turquoise* category. More recently, Ronchi summarizes thirteen papers suggesting that in the Far East *yellow-green* and *blue-green* are basic color terms. [26, section 2.4.2, pp. 18–20]. In Chinese, Japanese, and Korean both colors are monolexemic and very common, unlike in English where *chartreuse* is a rather refined term and *turquoise* is not commonly used as a color name (note that *cyan* is a technical term from the printing industry) and therefore both are non-basic terms albeit monolexemic.

2.3 Space subdivisions

Fig. 9 from Boynton and Olson's paper[7] illustrates the gaps in color space left by the WCS approach. The figure is a projection in the OSA space along the lightness axis L. The chromatic axes are very roughly red–green (g for *green*) and yellow–blue (j for *jaune* or yellow in French). Such a sparse color categorization is not very useful to solve our problems. Also, we need a much finer subdivision: even a non-specialist uses up to 50 color names to cover the whole color space without overlaps and confusions [16, p. 363].
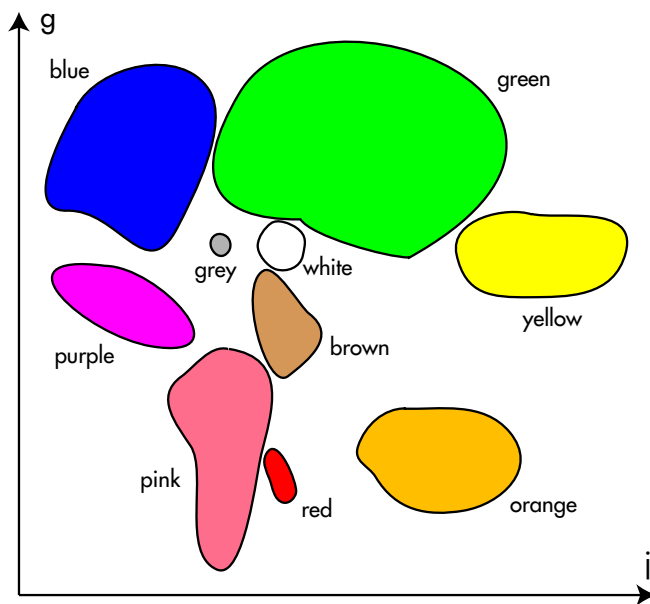

Figure 9. The constraints on consistency, consensus, and response time lead to large unnamed gaps that require artifacts like *links* and *bridge colors* (after Boynton[7]).

How can we achieve a complete tiling of color space in many named regions? We have to replace the strict saliency requirement (see Def. 2.1) with the weaker engineering concept of

operational consensus:

DEFINITION 2.4 (OPERATIONAL CONSENSUS). *The colors in a region have the same name if the participants in a conversation agree on the use of that name.*

This engineering concept is less arbitrary than it may seem from a cursory look. In fact, in the discussion at the Asilomar meeting reported in [16, pp. 353–354], color categories may change in time — but more about this later in Sec. 3. Let us first survey some examples of color space subdivisions by color name.

In 1955, the National Bureau of Standards (NBS, now National Technical Information Service NTIS) published circular 553 with a vocabulary of 7,500 color names with the purpose of assisting scientists, businessmen, and laymen to understand the different color vocabularies used at that time in the many fields of art, science, and industry in the USA.[19]

A subcommittee of the Inter-Society Color Council (ISCC) defined the boundaries of 267 color name categories in terms of the Munsell renotation. This categorization was based on some recommendation made in 1933 by I.H. Godlove, a scheme of hue modifiers shown in Fig. 10, and heuristics.
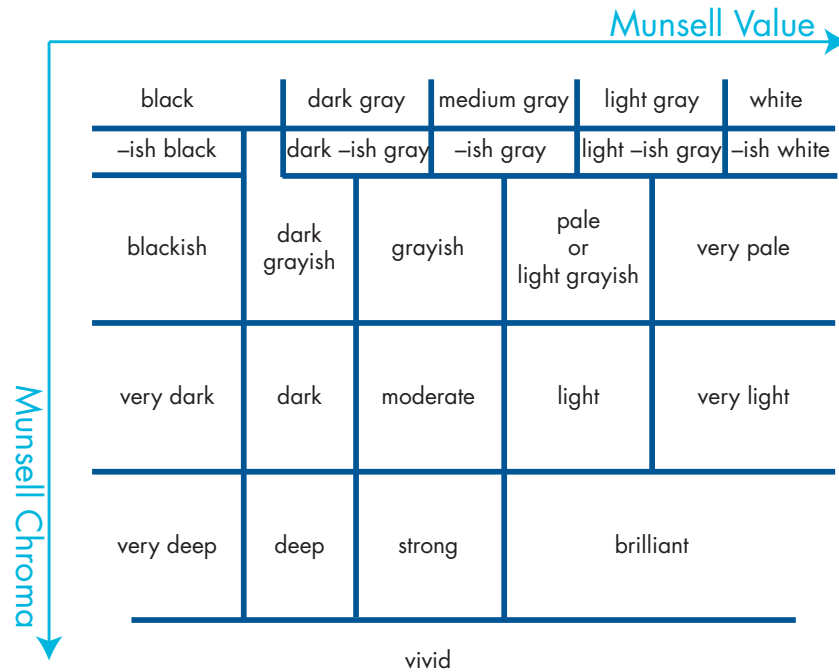

Figure 10. Scheme of the hue modifiers, the "-ish" grays and the neutrals with their modifiers.

The ISCC subcommittee checked the color boundaries by observations of all the color standards obtainable for which Munsell renotations were available at the time. The final charts were much more complicated than the original ISCC–NBS system shown in the figure and differ significantly from one level of Munsell value to another.

The resulting compilation is a true color name thesaurus. To date, the 1932 vision of ISCC's first chairman E.N. Gathercoal of the University of Illinois College of Pharmacy to develop

a means of designating colors has been unparalleled. This is the more astonishing as it was accomplished before the availability of automation tools and by a relatively small group of people.

For the problems described in the introduction, even with 7,500 color names, the dictionary is very limited. The ISCC–NBS system is:

- a snapshot in time (1955)

- mostly government and industry related

- only in English

Nemcsics compiled a color dictionary for his Coloroid system to help color designers communicate with their customers [23, p. 118]. He introduced a hierarchy of names: the Coloroid system has 48 *primary hues A*, numbered from 10 to 76 with gaps. Each set of tens (e.g., from 10 to 16) is called a *color domain* with names yellow, orange, red, violet, blue, green 1, and green 2. Within each domain the primary hues are then assigned names like yellow 1, yellow 2, yellow 3, warm yellow 1, warm yellow 2, orange yellow 1, orange yellow 2 (see Fig. 16).

For each Coloroid hue, Nemcsics then gives a table of names with the ranges specified in terms of rectangles of width in T (saturation) and height in V (lightness). No reference is made on how the names were compiled and how the category boundaries were determined. An example for hue 20, *yellowish orange 1*, is shown in Fig. 3. The red curve indicates the boundary of the Coloroid color space and the blue line indicates the boundary of the surface colors.

From a conversation with Nemcsics, we suspect that he just made a substantial intellectual effort to come up with a consistent system, rather than a more formal psychophysics experiment as was done in the ISCC–NBS system; it is more akin to alchemy than to chemistry. For more information see Ref. [26, Appendix III]

When one browses the Coloroid dictionary, many names that may be well known to designers, are rather arcane to the general public. While in a design process a designer can educate a customer in the naming of colors, this is not possible when one communicates anonymously with the public in general, as when a user interface is designed for Mr. Müller and Mrs. Yamada in our example.

David Post and his collaborators[24] performed a number of experiments from 1985 to 1989 in which they collected data for two symbol sizes presented on a CRT under a wide range of background colors and ambient illumination conditions. The two sizes were 2° representing area fills, and 20′ representing symbology.

As shown in Fig. 11, Post used a fixed 12-name vocabulary consisting of the 11 basic color terms plus *peach*. The main goal of these experiments was to identify robust colors for avionic applications.

Paul Green-Armytage's work[15] presented at the AIC meeting in Rochester has the advantage that it is based on the more modern NCS atlas. His experiment was conducted in a number of phases:
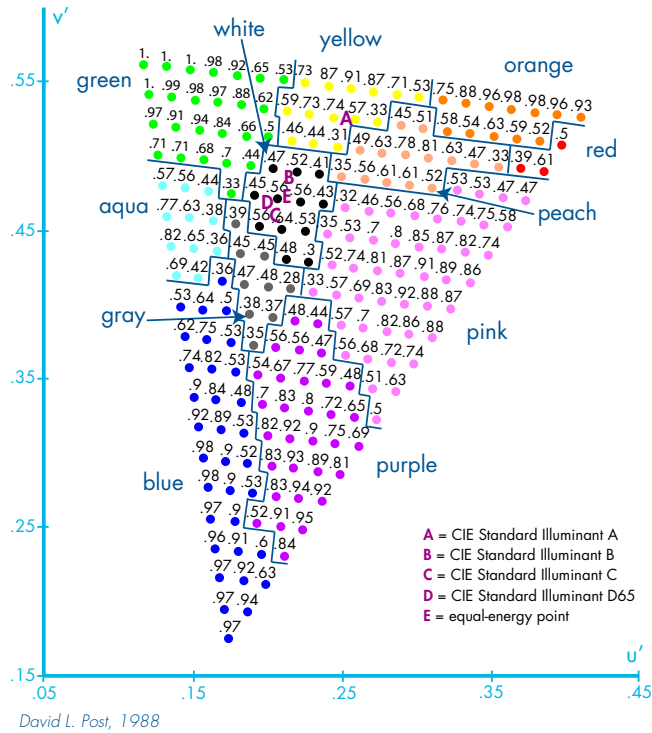
Figure 11. Color-naming data for stimuli presented on a CRT as 2° circles on a black background. The boundaries enclose areas within which the modal color-name response corresponds to the color name shown (after Post[24]).

1. Ask 247 subjects to write a list of color names in 5 minutes; select the 216 names written by more than one subject person

2. Select the 183 names defined in more than 1 English dictionary

3. Ask 148 subjects which names they prefer for 27 focal colors

4. Ask 54 subjects for each name to select the swatch in the NCS atlas best representing it

5. Ask 35 subjects if the names are credible

In the last step, where there is no consensus: the experimenter makes the final decision. Sometimes a name can be replaced at the experimenter's discretion, e.g. *aubergine* instead of *grape*. A similar procedure is performed for modifiers to cover the complete the NCS space.

This experiment has many problems. First, it starts with the linguist's approach of eliciting names without showing examples, in this case swatches from the NCS system should have been shown. After that there is a lot of discretion from the side of both the subjects and the experimenter.

In conclusion of this section, a lot of good research has been conducted by many excellent researchers, but none delivers a color dictionary based on color coordinates and concomitantly

good enough to solve the engineering problems elucidated in the introduction. Before we can present our proposed solution leading to a tiling of color space in a rich set of terms for an operational consensus, we have to discuss some findings in the literature about the ephemerality of color naming.

3. EPHEMERALITY OF COLOR NAMING

In a discussion at the Asilomar meeting reported in [16, pp. 353–354] it was noted that the color categories may change in time. As described by Davidoff,[10] color naming is an acquired skill. A convincing experimental proof came with Franklin's research in the Surrey Baby Lab.[13] The data shown in Fig. 12 suggests that whereas color categorical perception is stronger in the left hemisphere than in the right hemisphere — which suggests that color categorical perception in adults is caused by the influence of lexical color codes in the left hemisphere — prelinguistic color categorical perception in infants is lateralized to the right hemisphere. This suggests that language-driven categorical perception in adults may not build on prelinguistic categorical perception, but that language instead imposes its categories on a left hemisphere that is not categorically prepartitioned.
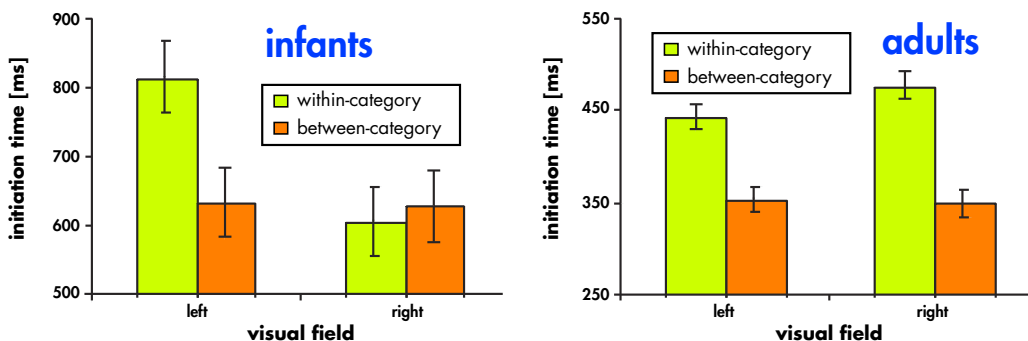

Figure 12. Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults (after Franklin et al.[13]).

This has two consequences. The first is that since everybody grows up in a different environment, everybody has a different color lexicon, and this undermines the Universalist's position. The second consequence is that the color lexicon will have a strong dependence on the socio-economic status (SES), because in most societies poor people tend to be less well educated or less exposed to color decisions.

The color lexicon also evolves in time. For example, in classical Greece the quality of light was more important than colors. Similarly up to about a thousand years ago, the Japanese used only *black* (*kuro*) and *white* (*shiro*) to designate color, with a term for *colorful* (*aka*), which now is *red*. About a thousand years ago indigo dye was introduced to Japan and a new color name came into use, namely *ao* for *dye*. *Ao* was used for *grue*, i.e. *green or blue*, and only in the 14th to 15th century *green* and *blue* were categorized into *midori* and *ao*.[32]

In fact, Stanlaw[29] describes how *midori* is used for a static *green* while *ao* is used for a dynamic *green*, like an unripe apple still green or complexion that became green from being

frightened. Because the go-light in a traffic signal is something dynamic, the color term for it is *ao*, though it has the same colorimetric coordinates as American go-lights. The interesting experiment is when Japanese expatriates, who see traffic light every day, are asked to select the color chip best representing a Japanese go-light. Stanlaw found that the longer a subject had lived in America, the bluer a chip he would select. In the Western culture we do not have an equivalent to the concept behind the verb *naru* (to become), so for us this subtlety is hard to fully appreciate.

Looking at modern cultures in general, as we are exposed to a more colorful environment at an earlier age, we learn earlier in life how to name colors. For example, if in 1900 we learned the basic four colors by age 8, in 1950 we learned them 3 years earlier at age 5.

An interesting experimental finding [33, p. 149] is that that art students are slower in naming colors than chemistry students, because the former are more inclined to describe color impressions rather than the perceptions of physical sensations. Specifically, they use more modifiers, a trend that actually decreases as art students proceed to their senior years.

This teaches us that the richness or size of a color lexicon is not a measure of color fluency. What is important is the operational consensus (see Def. 2.4). This consensus can be reached in a relatively short time; for example, Drivonikou et al.[12] have shown that subjects can learn to use the two perceptual categories of *chartreuse* and *turquoise* across four days.

The global village is a large consolidator of archetypes, and with it, color names are becoming more universal and less tied to the availability of local natural dyes and pigments. This can be seen in the general increase of the use of color names from foreign languages. For example *pink* is finding adoption in several languages, but the Japanese *pinku* is equivalent to the English *pink* and brighter than the Japanese[29] *momoiro*, and in German[14] *pink* is different from its dictionary translation *rosa*.

As a further comment on the cultural effect on color naming, we report on an experiment conducted by Zollinger. His protocol was as follows.[33] First, the observers were asked to write down a number of color terms, which they were to divide into two groups

1. a first group consisting of words considered absolutely necessary for a minimum color lexicon

2. a second group including words considered to be of secondary importance

The total number of words allowed was *arbitrarily* set to twelve. Next the subjects were shown and asked to name a set of 113 to 117 Munsell Sheet of Color samples. This system was made up of 20 Munsell hues at three to four levels of brightness and three to four levels of saturation. Each sample had to be named within 20 seconds. Each subject could describe a particular sample by using a word from his or her chosen lexicon, or by using any other words. If the sample could not be described within 20 seconds, the corresponding space in the questionnaire was left blank.

The frequency of occurrence is the percentage of subjects mentioning a specific term for a specific sample. The certainty of determination is the sum of all color terms given to a specific

sample, or to all twenty hues at specific levels of value and chroma. The graphs in Fig. 13 and 14 indicate that certainty of determination is medium for German and and low for Japanese. Zollinger writes [33, p. 147]:
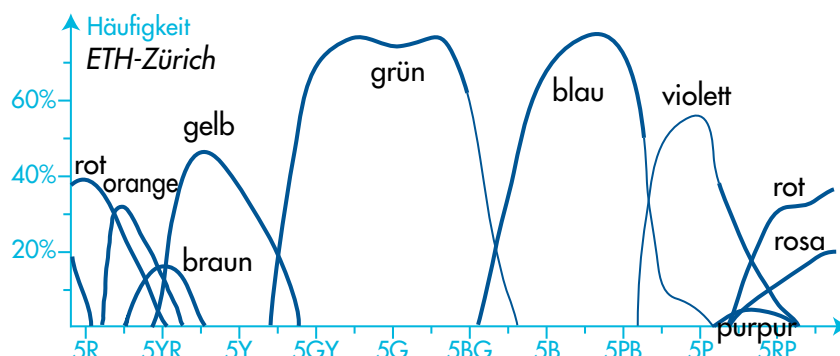

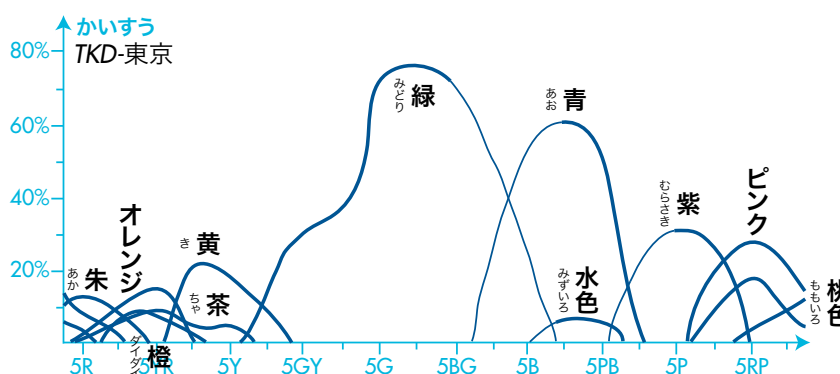Figure 13. Frequency of occurrence of color terms with German-speaking science students in Zürich (after Zollinger[33]).

Figure 14. Frequency of occurrence of color terms with Japanese-speaking science students in Tōkyō (after Zollinger[33]).

"Science students — test subjects with fairly comparable backgrounds of schooling, professional interests, and age — were studied for five different mother-tongues. The results for native speakers of German, French, English, and Hebrew cannot be differentiated further, although the probability is that French has a higher certainty of determination of 60–70% as revealed by statistical tests.

"The certainty of determination of Japanese students is, however, clearly lower (probability > 90%). Drawing on my own experience of Japanese culture, I assume that the tasks in these tests are more difficult for Japanese students than for their Western counterparts. Japanese etiquette requires very subtle and intricate forms of addressing the person to whom one is speaking and is much more important (and difficult) for a Japanese in all situations; this applies also to color naming."

15

4. STRUCTURE

4.1 Categorization

One of the key characteristics of human cognition is categorization. The introduction of structure allows the distillation of data into knowledge through the reasoning about relationships. For example, we can talk about broccoli, strawberries, Emmenthaler, carrots, and grapes, but when discussing food values, it is more efficient to talk about vegetables, milk products, and fruits, i.e., it is more efficient to categorize the items being discussed, as illustrated in Fig. 15.

Figure 15. One of the cognitive methods we use in daily life to communicate more efficiently is to group items into categories, so we can for example speak about vegetables, cheeses, and fruits, instead of having to enumerate the items individually each time.

Categorical perception refers to the perception of different sensory phenomena as being categorically different, while continuous perception refers to sensory phenomena located on a smooth continuum. Categorical perception occurs whenever perceived within-category differences are compressed and between-category differences are magnified, relative to some unit comparison threshold. This has caused some misunderstanding when we presented our paper at the AIC meeting in Sydney,[4] but it is important not to confuse "WCS color categories" with "categorical perception." In fact, each subdivision in Sec. 2.3 yields categories; they are all different, but they are all categories: *category* is a category (note the semantic of emphasis in this paper — this is not a Zen moment). Stevan Harnad[17] has proposed the following definition of categorical perception:

DEFINITION 4.1 (CATEGORICAL PERCEPTION). *A categorical perception effect occurs when*

1. *a set of stimuli ranging along a physical continuum is given one label on one side of a category boundary and another label on the other side and*

2. *the subject can discriminate smaller physical differences between pairs of stimuli that straddle boundary than between pairs that are entirely within one category or the other*

The HVS is faster at discriminating between two colors from different categories than two colors from the same category. In the discussion of the Asilomar meeting [16, p. 358], Maffi and Hardin report how perceptual color category boundaries can be determined by recording eye movements, and this is the technique used in the Surrey Baby Lab experiment mentioned in the first paragraph of Section 3.

It is fundamental to understand the difference between WCS categories and perceptual categories. The color categories in the WCS can be determined with a psychological experiment, while the color categories in categorical color perception can be determined with a physiological measurement. From here on, we are interested in the color categories from categorical color perception, because these are the ones that allow us to solve the problems described in the introduction. The perceptual category boundaries are what introduces a "bump" when we cross them in a reproduced image.

Miller[21] famously stated that short-term memory can hold 7 ± 2 sequential elements, therefore we chunk longer sequences, like we communicate telephone numbers as 650-857-6713 instead of 6508576713. When the number of chunks becomes too large, a hierarchy is introduced. Color is at least 3-dimensional, not sequential, so Miller's work and its refinements do not apply directly (10 ± 5 has been suggested[11] for color), but it is useful to remember

- short-term memory groups items in a small number of chunks

- when there are more chunks, a hierarchy is introduced

- what is a chunk depends on a person's knowledge

Especially for the applications in EaaS Web services, we do not just want to tile the color spaces in regions corresponding to color names, but we need a hierarchy of names, like this was done to some extent in the Coloroid sytem and is illustrated in Fig. 16.

5. IMPLEMENTATION

To solve the problems described in the introduction, we need to compile a large hierarchical dictionary of color names and determine their boundaries in a color space. The dictionary must be based on actual color samples, like Munsell Sheets of Color, NCS swatches, or colorimetrically specified colors presented on a computer display. For example, introducing a Crayola box like in Fig. 17 could help Mr. Müller communicating with Mrs. Yamada, but sRGB coordinates encoded for HTML are not very useful to humans.

Compiling a large hierarchical dictionary of color names is a daunting task to solve, but we are computer engineers and we can use the Web to assist us. Instead of forcing friends and relatives into a laboratory to perform long and complex tasks, we can recruit hundreds of thousands of subjects on the Web and ask them to perform simple short tasks for us. Such experiments are called *crowd-sourcing* and can be used successfully for psychophysics experiments.[34]
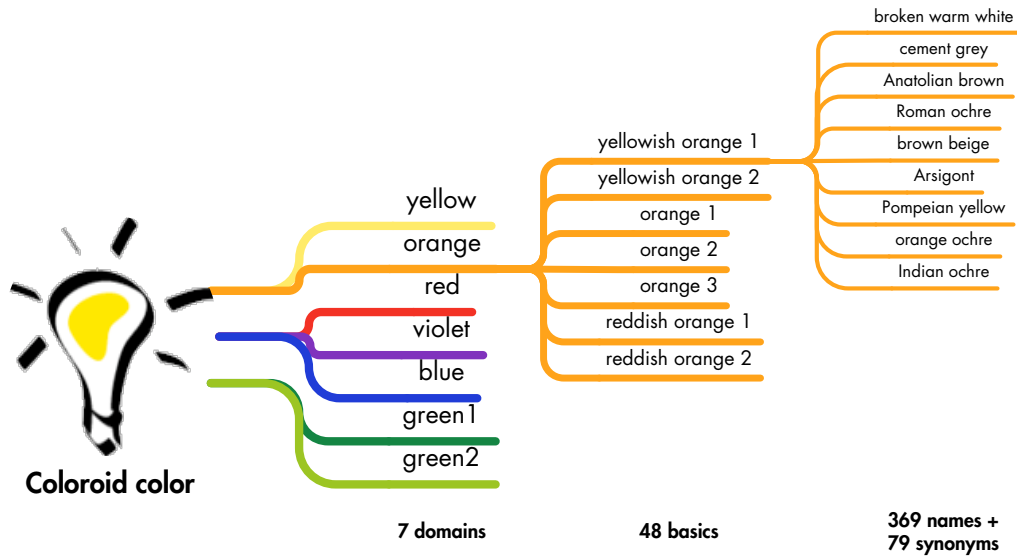
Figure 16. In the Coloroid system the color names are organized in a hierarchy. There are 7 domains, which is each subdivided into 7 ± 2 basic colors for a total of 48 basic colors. The basic colors are further subdivided for a total of 369 names and 79 synonyms.
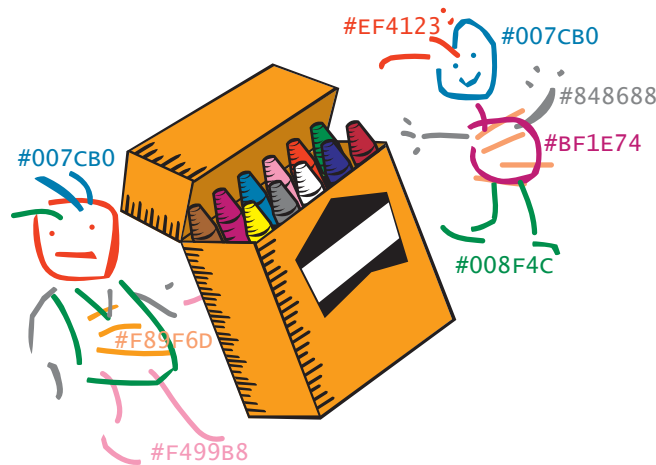

Figure 17. Color samples can bridge the color names between two languages, providing the basis for multilingual EaaS Web services.

An immediate complaint might be that experiments not conducted under strictly controlled laboratory conditions bear no scientific value. We beg to differ. For example, the WCS experiments are conducted under uncontrolled conditions, but Boynton did not have a problem with that, although he did his own experiment in the laboratory:

> The mechanisms of color constancy work so well that, within limits, the intensity and spectral distribution of the light used to illuminate the experimental materials make surprisingly little difference [6, p. 138].

Something is scientific if it can be measured and analyzed. Physicists do not travel to Mars to perform an experiment on Martian rocks — they use remote sensing. A large amount of data must be collected, but if it can be modeled it provides a valid experimental result. In the past 30+ years excellent robust statistical methods have been developed to deal with outliers. Crowd-sourcing is just a form of remote sensing.

Experiments based on crowd-sourcing can last in perpetuity, or at least be continuously ongoing and the data is harvested freshly when needed. This automatically solves the ephemerality problem discussed in Section 3, yielding a color naming corpus that is alive like an organism and always conveys the *Zeitgeist*.

Color names are just labels, and the actual name of a color is not important for user performance. For example, Smallman and Boynton[27] performed a visual search task experiment using named colors, then they changed the color names. After a short learning time, the subjects had the same performance as with the old names, and only the perceptual color distance was a factor. (We reported on the experiment by Drivonikou et al.[12] earlier.) Therefore, our experiment must contain an instructional element.

5.1 Multilingual color naming experiment

Our basic experiment was described in the 2003 edition of this conference.[22] From then, there has been a number of objections, which we will address here. The first is that data from experiments not conducted in a laboratory is not meaningful. We just addressed this objection with Boynton's help.

The second objection is that samples created by selecting random points on a grid in the sRGB color space yield to biased sampling. It is true that a grid in CIELAB or an appearance space would yield better results, but the color regions for a name are relatively large compared to the grid spacing. The current implementation is less elegant, but we just saw that the WCS is based only on the surface colors of the Munsell gamut, and Boynton's experiment is based on the OSA color atlas, which has a smaller gamut.

The third objection is that the samples are presented on a white background. This is a valid objection and we will implement a grey background as soon as we have the resources (we blame it on the economy). Fig. 18 shows the page of the Italian version of the experiment.

Fig. 19 is a flowchart of the processing that occurs after a subject submits a data set. The data scrubbing eliminates the most common errors, like typing errors, misspellings, and disruptive data. So far, tens of thousands of subjects have contributed their names to the experiments. The URL is `http://www.hpl.hp.com/personal/Nathan_Moroney/mlcn.html` and we welcome any and all participants.

5.2 A color thesaurus

We mentioned that an instructive element is required for the experiment to stay alive. This is accomplished by offering a color thesaurus, which is also the main vehicle to lure subjects to contribute data. The English thesaurus can be found at the bottom of Nathan's home page `http://www.hpl.hp.com/personal/Nathan_Moroney/`. At the time of this writing, the
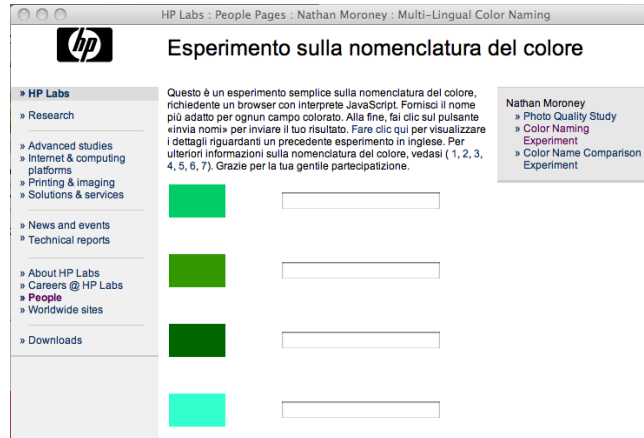
Figure 18. User interface of the experiment in its Italian version. Seven random color patches are presented, on the right of each patch is a field to type a name for that patch's color.
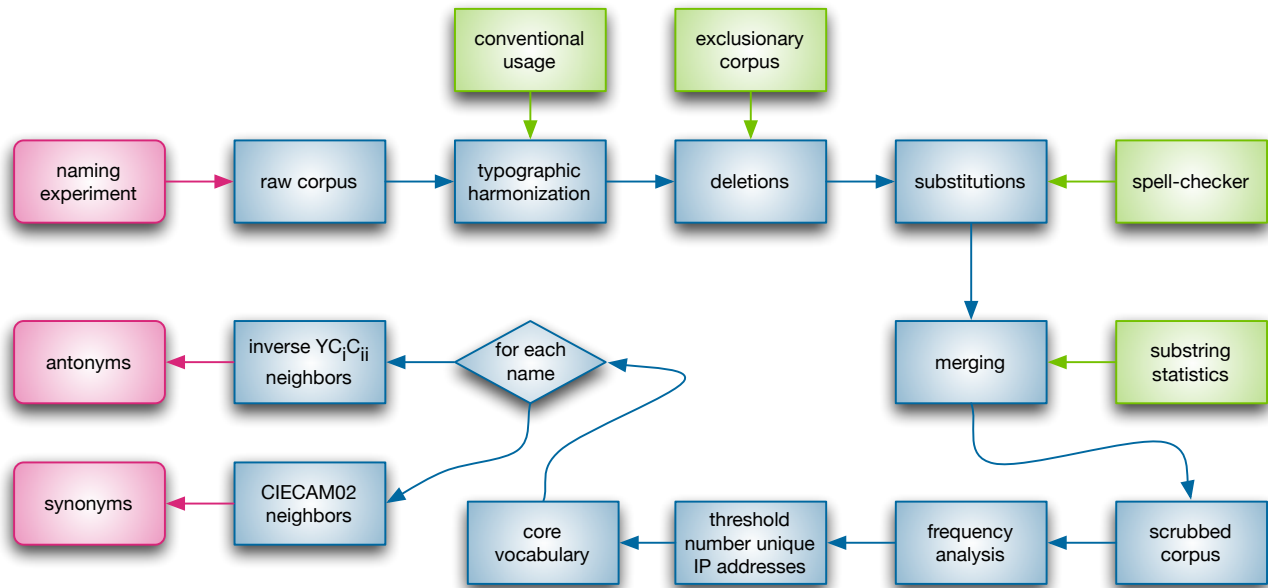

Figure 19. The collected data is vetted and cleaned before it is admitted to the core vocabulary at the bottom center of the flowchart.

thesaurus has been accessed successfully 205,362 times. Fig. 20 shows the result of a query with the term *celeste*.

The current implementation is shown at the left of the flowchart Fig. 19. Obviously the algorithm is not completely meaningful at this stage of the experiment. We have not yet implemented the statistical methods to build a hierarchy of categories and currently roughly approximate the synonyms by considering the names for colors in an interval around the queried color. The antonym formation is even more speculative, as it is not the name of the opposite color.[18]
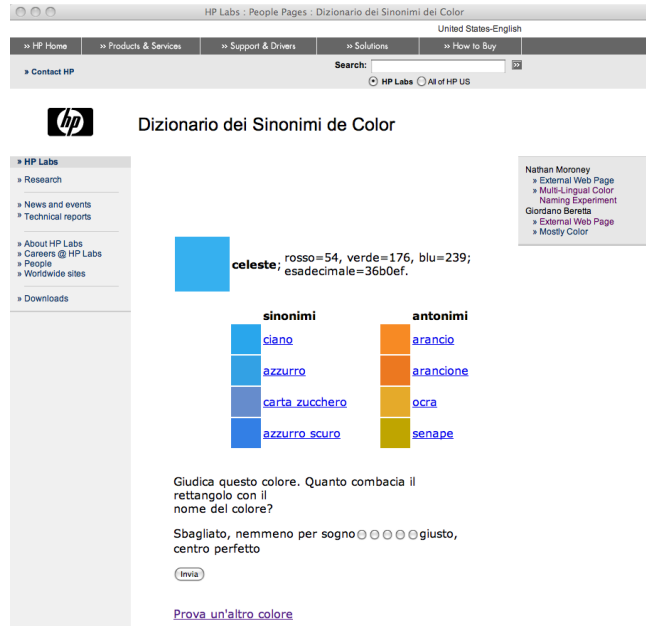
Figure 20. Result of a color name query in the Italian thesaurus. The user enters a color name and is presented with a representative color patch and the sRGB color coordinates.

5.3 Improving the corpus quality

One of the advantages of crowd-sourcing is that the crowd can also be recruited for a feedback mechanism to improve the data corpus quality. Fig. 21 shows that at the bottom of the response screen the user is solicited to rate the quality of the response. This feedback data is used as a weight in the data corpus, indicating the confidence for a contributed color name.
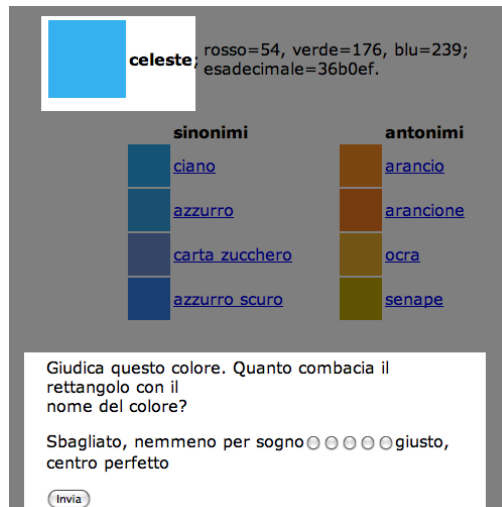

Figure 21. After each successful delivery of a sample for a color name, the subject is asked to rate the quality of the response.

Fig. 22 shows the distribution of the ratings. While the corpus does contain a small amount of bad and poor data, for an experiment of this size with subjects having only a superficial understanding of the experiment, the ratings are quite encouraging.
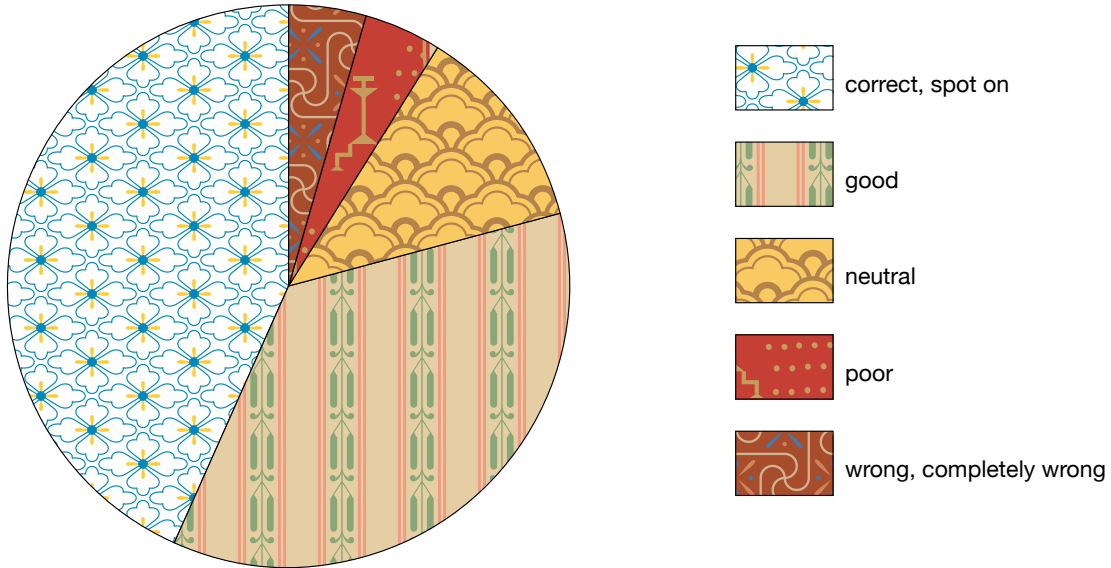

Figure 22. The users of the color thesaurus are quite satisfied with the quality of the data corpus. Only a small number of the color name and color sample pairs is rated poorly.

5.4 Expanding the corpus

Fig. 23 shows the distribution of the contributed names. It is clearly a heavy-tail distribution, but we want to make clear from the beginning that it is not a Zipf distribution. Indeed, Zipf's law states that more frequently occurring words tend to be shorter (less characters), and it is known that Zipf's law does not hold for color names [9, p. 210]. The long-tail distribution just means that non-basic terms like *eggplant* are used more rarely (but more frequently than *grape*, lending to some justification to Green-Armytage's manipulation[15]).

The challenge for building our color name corpus is how we can coax subjects and elicit those less frequent names, while not compromising the statistics. We may be skating on thin ice, but we assume that if a subject queries for a name that is not in the corpus, then the name should be in the corpus if we have had a sufficient number of subjects, because the color space from which we draw the color patches is sampled with a uniform random distribution, even when the color space is not uniform.

When the thesaurus service does not find a name in the data corpus, instead of an error page it presents the GUI shown in Fig. 24, which allows the subject to dial in a color for the name he or she was looking for. The hope is that the subject will use some other means (e.g. our blog at `http://www.mostlycolor.ch/2008/12/transitioning-colors-glaucus.html`) to find the color for that name and will then report it back to the data corpus for all to share.

Figure 23. Distribution of the contributed names. On the abscissa are the color names in the insert and on the ordinate is the number of times that name has been elicited.
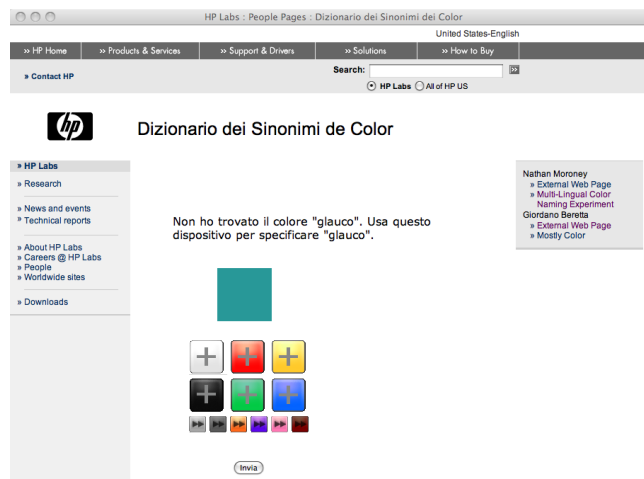

Figure 24. When a color name is not found in the corpus, the subject is asked to contribute the color for that name.

Finally, the simple flowchart in Fig. 25 shows how the newly harvested color name is added to the corpus. Of course, the elicited name is processed like those from the multilingual color naming experiment to maintain consistency in the experimental procedure.
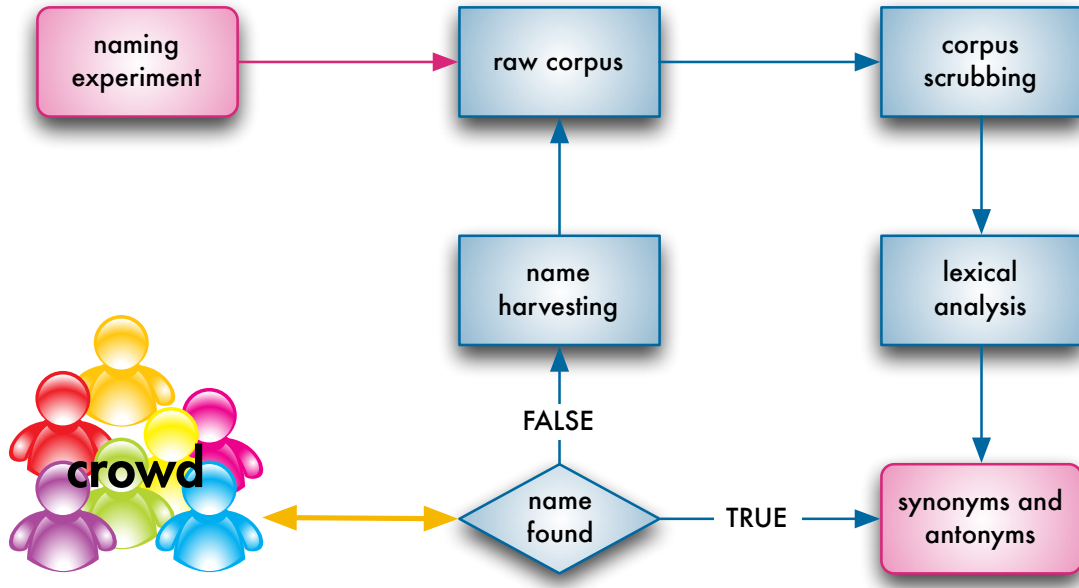
Figure 25. A new color name is elicited from the crowd and is added to the corpus.

6. CONCLUSIONS

We have presented some problems occurring in the domain of commercial print automation, notably in variable data printing, EaaS Web services, gamut mapping, and print quality control. These problems can be solved using lexical metrics, which in turn require a comprehensive color naming system.

We have given an overview of some of the research in color naming and shown why we cannot use its results to solve our problems. We have proposed a new methodology based on crowd-sourcing to perform very large scale color naming experiments. We have shown that this methodology is consistent with the other research in this field when one understands that *category* is a cognitive category.

Although at the time of this writing we had 205,362 subjects using the color thesaurus, we have only implemented the beginning of the experiment. There are some dubious elements, like not presenting the color samples on a neutral grey background. But more importantly, we have not yet implemented the most important portions of the experiment, namely those that will allow us to find the boundaries of the names related to categorical color perception. We have also not yet implemented the tools to infer a hierarchy of color categories.

One of the interesting open research questions is that of whether the hierarchy has to be elicited explicitly from subjects, or it can be inferred statistically from analysis of the data corpus, or a tool can be implemented to elicit the hierarchy implicitly as a side effect. An algorithm to compute antonyms is a very hard problem, requiring semantic information; this will call for a substantial refinement of the crowd-sourcing experiment.

24

REFERENCES

[1] Giordano B. Beretta, *Internet's impact on publishing*, Color Imaging: Device-Independent Color, Color Hard Copy, and Graphic Arts II (San Jose) (Giordano B. Beretta and Reiner Eschbach, eds.), vol. 3018, SPIE, April 1997, DOI: 10.1117/12.271592, pp. 200–211.

[2] ———, *Color aspects of variable data proofing*, Internet Imaging VI (Simone Santini, Raimondo Schettini, and Theo Gevers, eds.), vol. 5670, SPIE, January 2005, DOI: 10.1117/12.592489, pp. 175–182.

[3] ———, *Pre-flighting variable data print jobs using the coloroid system*, The 10th Congress of the International Colour Association (Granada (Andalusia, Spain)), vol. 2, May 2005, pp. 1231–1234.

[4] Giordano B. Beretta, Nathan M. Moroney, and John Ludd Recker, *Towards robust categorical color perception*, The 11th Congress of the International Colour Association (Sydney) (Dianne Smith, Paul Green-Armytage, Margaret A. Pope, and Nick Harkness, eds.), AIC, September 2009, p. 0909015Final00251.

[5] Brent Berlin and Paul Kay, *Basic color terms: Their universality and evolution*, University of California Press, Berkeley, 1969.

[6] Robert M. Boynton, *Insights gained from naming the osa colors*, Color categories in thought and language (Clyde L. Hardin and Luisa Maffi, eds.), Cambridge University Press, 1997, pp. 135–150.

[7] Robert M. Boynton and Conrad X. Olson, *Locating basic colors in the OSA space*, Color research & applications **12** (1987), no. 2, 94–105.

[8] Richard Cook, Paul Kay, and Terry Regier, *World color survey data archives*, `http://www.icsi.berkeley.edu/wcs/data.html`, November 2006.

[9] Greville G. Corbett and Ian R. L. Davies, *Establishing basic color terms: measures and techniques*, Color categories in thought and language (Clyde L. Hardin and Luisa Maffi, eds.), Cambridge University Press, 1997, pp. 197–223.

[10] Jules B. Davidoff, *Cognition through color*, The MIT Press, Cambridge, 1991.

[11] Gunilla Derefeldt, Tiina Swartling, Ulf Berggrund, and Peter Bodrogi, *Cognitive color*, Color research and applications **29** (2004), no. 1, 7–19.

[12] G. Vicky Drivonikou, Alexandra Clifford, Anna Franklin, Emre Özgen, and Ian R. L. Davies, *Category training affects colour discrimination*, Progress in Colour Studies 2008 (Carole Biggam, Carole Hough, Christian Kay, and David Simmons, eds.), University of Glasgow, July 2008, p. 11.

[13] Anna Franklin, G. Vicky Drivonikou, Ian R. L. Davies, Paul Kay, and Terry Regier, *Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults*, Proceedings of the National Academy of Sciences **105** (2008), no. 9, 1097–1102, `http://www.icsi.berkeley.edu/~kay/infant.pdf`.

[14] Claudia Frenzel, *Rosa Schätze — pink zum kaufen: Stylistic confusion, subjective perception and semantic uncertainty of a loaned colour term*, Progress in Colour Studies 2008 (Carole Biggam, Carole Hough, Christian Kay, and David Simmons, eds.), University of Glasgow, July 2008, pp. 13–14.

[15] Paul Green-Armytage, *Colour zones — explanatory diagrams, colour names, and modifying adjectives*, The 9th Congress of the International Colour Association (Rochester) (Paula Alessi, ed.), Proceedings of SPIE, vol. 4421, AIC, SPIE, June 24–29 2001, pp. 917–920.

[16] Clyde L. Hardin and Luisa Maffi (eds.), *Color categories in thought and language*, Cambridge University Press, Cambridge, 1997.

[17] Stevan R. Harnad, *Psychological and cognitive aspects of categorica perception: A critical overview*, Categorical Perception: The Groundwork of Cognition (Stevan R. Harnad, ed.), Cambridge University Press, 1987, pp. 535–565.

[18] Kimberly Jameson and Roy G. D'Andrade, *It's not really red, green, yellow, blue: an inquiry into perceptual color space*, Color categories in thought and language (Clyde L. Hardin and Luisa Maffi, eds.), Cambridge University Press, 1997, pp. 295–319.

[19] Kenneth L. Kelly and Deane B. Judd, *The ISCC–NBS method of designating colors and a dictionary of color names*, vol. 553, National Bureau of Standards, Washington, November 1955.

[20] Larry Lavendel and Tim Kohler, *The story of a color advisor*, Sixth Color Imaging Conference: Color Science, Systems and Applications (Scottsdale (Arizona)) (Sabine Süsstrunk and Andras Lakatos, eds.), vol. 6, IS&T and SID, The Society for Imaging Science and Technology, November 1998, pp. 228–229.

[21] George A. Miller, *The magical number seven, plus or minus two: Some limits on our capacity for processing information*, Psychological Review **63** (1956), no. 2, 81–97, `http://psychclassics.yorku.ca/Miller/`.

[22] Nathan Moroney, *Unconstrained web-based color naming experiment*, Color Imaging VIII: Processing, Hardcopy, and Applications (Reiner Eschbach and Gabriel G. Marcu, eds.), vol. 5008, SPIE, January 2003, pp. 36–46.

[23] Antal Nemcsics, *Colour dynamics: Environmental colour design*, Akadémiai Kiadó, Budapest, 1993.

[24] David L. Post, *Applied color-vision research*, Color in electronic displays (Heino Widdel and David L. Post, eds.), Defense Research Series, vol. 3, Plenum Press, 1992.

[25] Terry Regier, Paul Kay, and Naveen Khetarpal, *Color naming reflects optimal partitions of color space*, Proceedings of the National Academy of Sciences **104** (2007), no. 4, 1436–1441, `http://www.icsi.berkeley.edu/~kay/WFness.pdf`.

[26] Lucia Rositani Ronchi, *La ricerca di avanguardia vista dall'aic nel terzo millennio: Colore e linguaggio*, vol. 81, Fondazione Giorgio Ronchi, via San Felice a Ema 20, 50125 Firenze, 2004, `http://books.google.com/books?id=BYBOPjY-ExgC`.

[27] Harvey S. Smallman and Robert M. Boynton, *Segregation of basic colors in an information display*, J. Opt. Soc. Am. A **7** (1990), no. 10, 1985–1994, `http://josaa.osa.org/abstract.cfm?URI=josaa-7-10-1985`.

[28] J. Jerome Smith, Louanna Furbee, Kelly Maynard, Sarah Quick, and Larry Ross, *Salience counts: A domain analysis of english color terms*, Journal of Linguistic Antropology **5** (1995), no. 2, 203–216.

[29] James Stanlaw, *Two observations on culture contact and the Japanese color nomenclature system*, Color categories in thought and language (Clyde L. Hardin and Luisa Maffi, eds.), Cambridge University Press, 1997, pp. 240–260.

[30] Julia Sturges and T. W. Allan Whitfield, *Locating basic colours in the Munsell space*, Color Research & Application **20** (1995), no. 6, 364–376, `http://www3.interscience.wiley.com/journal/114186803/`.

[31] Alf C. Zimmer, *There is more than one level in color naming — a reply to Zollinger*, Psychological Research **46** (1984), 411–416.

[32] Heinrich Zollinger, *Why just turquoise? remarks on the evolution of color terms*, Psychological Research **46** (1984), 403–409.

[33] _____, *Color: A multidisciplinary approach*, Helvetica Chimica Acta, Zürich, 1999.

[34] Silvia Zuffi, Carla Brambilla, Giordano B. Beretta, and Paolo Scala, *Understanding the readability of colored text by crowd-sourcing on the web*, External HPL-2009-182, HP Laboratories, August 6 2009, `http://www.hpl.hp.com/techreports/2009/HPL-2009-182.html`.

http://www.hpl.hp.com/personal/Giordano_Beretta/
http://www.hpl.hp.com/personal/Nathan_Moroney/
blog: http://mostlycolor.ch