https://doi.org/10.2352/ISSN.2470-1173.2019.6.MAAP-202

Material Appearance: Ordering and Clustering

Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen, Jon Yngve Hardeberg;
Department of Computer Science, Norwegian University of Science and Technology; Gjøvik, Norway

Abstract

Appearance is a complex psychovisual phenomenon impacted by various objective and subjective factors that are not yet fully understood. In this work we use real objects and unconstrained conditions to study appearance perception in human subjects, allowing free interaction between objects and observers. Human observers were asked to describe resin objects from an artwork collection and to complete two visual tasks of appearance-based clustering and ordering. The process was filmed for subsequent analysis with the consent of the observers. While clustering task helps us identify attributes people use to assess appearance similarity and difference, the ordering task is used to identify potential cues to create an appearance ordering system.

Finally, we generate research hypotheses about how people perceive appearance and outline future studies to validate them. Preliminary observations revealed interesting cross-individual consistency in appearance assessment, while personal background of the observer might be affecting deviation from the general appearance assessment trends. On the other hand, no appearance ordering system stood out from the rest that might be explained with the sparse sampling of our dataset.

Introduction

"Appearance is the visual sensation through which an object is perceived to have attributes as size, shape, colour, texture, gloss, transparency, opacity etc." [1] For a better understanding, appearance was split into several attributes. CIE defines four key appearance attributes: color, gloss, translucency and texture [1, 2].

Color is the most understood and measurable of those attributes. According to the colorimetry, we can compute pseudo-perceptual distances between two colors in a specific context. [3, 4]. On the other hand, understanding appearance difference and similarity is still at the very basic level leaving many questions to be answered. We conducted an experiment where observers were asked to cluster objects by their appearance, in a way *"that looks natural"*. The objective of this task is to identify what attributes are prioritized in appearance similarity judgment and to get clues how people understand appearance similarity and difference. Also far from perfect perceptual uniformity, color is well-organized and described by various color spaces and color ordering systems [4, 5, 6]. However, to the best of our knowledge, there is no overall appearance ordering system currently in existence. In our experiment we ask observers to arrange objects in a space by their appearance, in any way *"that looks natural"* is the first step to understand how appearance ordering system might look, or whether it is possible to create one at all.

Appearance has been usually studied either using computer graphics, or with physical objects under controlled laboratory conditions. Despite some advantages of using synthetic images, or laboratory conditions, the appearance and perception still differ from that of real-life situations. Therefore, we used physical objects for our study and decided to make the setup natural and close to real-life situations allowing observers to freely interact with the real objects used for the study.

The tasks discussed below are part of a larger experiment. Quantitative results of the experiment are summarized in our previous work [7]. However, comprehensive behavioral and procedural analysis in order to understand how and why people get to those results has not been conducted. In this paper we focus on investigation of appearance ordering and clustering process though the analysis of the video recordings from the experiments. Furthermore, we analyze semantic information to reach this goal and generate research hypotheses for future studies.

The paper is organized as follows: in the next section we introduce the experimental setup, followed by the analysis description. Afterwards quantitative and qualitative results are summarized, discussed, and directions for further work are outlined.

Experimental Setup

In these experiments, we used resin objects of the *Plastique* artwork collection described by Thomas *et al.* [8]. In order to make procedures close to real-life situations, the experiments were conducted in uncontrolled conditions, under daylight illumi-

Figure 1: A screenshot from a sample video.

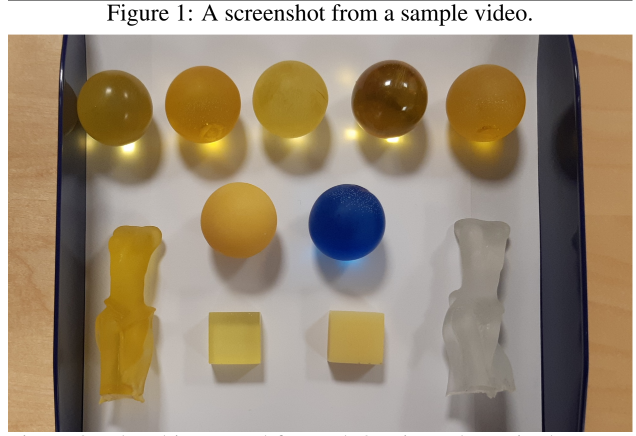

Figure 2: The objects used for Task 2. Five spheres in the upper row are used for the first sub-task.

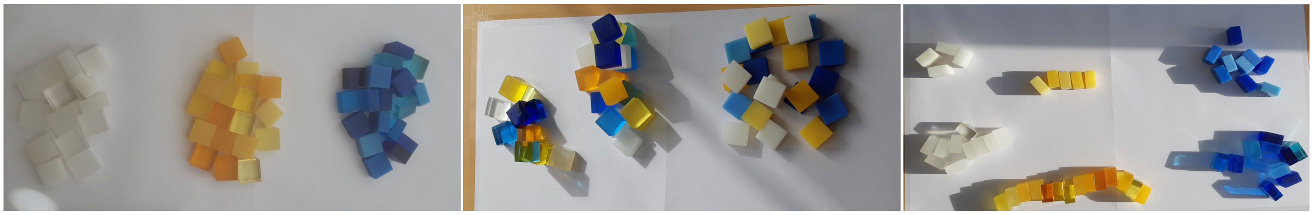

Figure 3: Examples of objects being grouped by color (left), translucency (middle), and color & translucency (right). The images also demonstrate the difference in lighting conditions among different interviews.

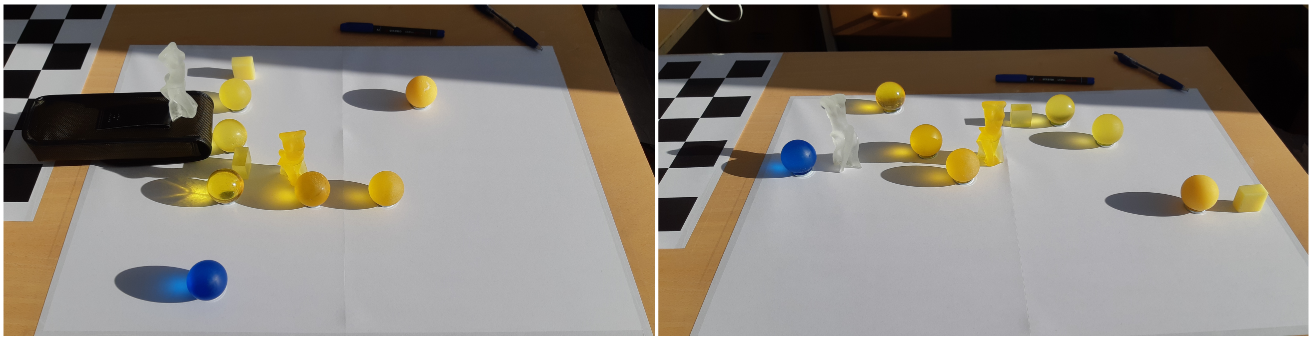

Figure 4: Examples of the two arrangements by one of the participants. **Left:** A 3D arrangement defined with five spheres. X axis - orange color. Y axis - yellow-blue colors. Center - maximum translucency. Further from the center - more opaque the object. Z axis - chromaticity. Box is used to hold the object in the 3^{rd} dimension. **Right:** 2D arrangement after accessing all objects. X axis - color axis: blue-achromatic-yellow. Y axis - translucency axis. Further from the center - more translucent the object.

nation incident from the window and artificial fluorescent illumination incident from the room. However, in contrast with real-life situations, observers were asked to wear gloves to avoid fingerprints and possible damage to the artwork. The data has been recorded with two video cameras, from frontal and side perspectives, and one microphone. A screenshot from a sample video can be seen in Figure 1. We measured light intensity at the table in the beginning and at the end of the experiment. However, analysis of the impact of illumination intensity is beyond the scope of this paper. Full experimental protocol is described in our previous work [7], while a description of the specific parts concerned by this work is provided below.

The experiment was composed of 11 tasks, while only first two tasks are in this paper. 17 observers with average age of 35.7 years participated in the experiment. 4 out of them were the authors of this paper. 14 observers were experts in the field and three of them were naïve to visual appearance studies. 2 observers were color deficient. 2 observers had artistic background, while 1 observer was a linguist. The experimental protocol was as follows:

Task 1:

- *Objects*: There are 48 rectangular parallelepipeds of different color, coarseness and translucency in the box.
- *Tasks:* 1) The first task is to cluster the objects into any number of groups the participant considers natural. 2) The experimenter asks the participant to discuss and explain the reasoning of clustering this way. 3) The experimenter asks the participant whether there could be any other ways of creating groups that look natural. 4) The experimenter selects one of the clusters and asks the participant to sub-cluster this group even further.

Task 2:

- *Objects:* There are five yellow spheres of different coarseness and translucency in the box. Besides, there are six more objects: two female busts, two spheres and two rectangular parallelepipeds. The objects are illustrated on Figure 2.
- *Tasks:* 1) The first task is to order the five spheres in any way the participant considers natural. They are encouraged to use any dimensions they think fit. 2) The participant is given six additional objects and is asked to locate the object in relation to the order he/she created with the first five spheres.

The observer is expected to fail placing all objects within the created order, and hence, to generate some questions on how to locate the new object. The goal is to identify potential cues to create an appearance ordering system. Besides, analysis of behavioral interaction when observers are trying to locate the objects within the scale would further help with understanding appearance assessment procedures.

Analysis

The results have been analyzed quantitatively as well as qualitatively. For quantitative representation a frequency analysis has been conducted. For clustering task occurrence of each cluster is represented as a histogram. The criteria used to define clusters are represented as percentages of the overall decisions taken. Likewise, different axis or dimensions and their criteria are presented by histograms and percentages for the ordering task.

The grounded theory analysis [9] has been used for qualitative explorations of the clustering task. The grounded theory analysis is a qualitative research methodology derived from the grounded theory approach [10]. The procedure is as follows: the very first step is codification when codes or labels are assigned to observations. On the second stage of the analysis, similar observations are grouped and described as categories. Category is a conceptual description of several observations. Afterwards links among categories are found and they are integrated into a single

Figure 5: An example of using combination of color- and translucency-based criteria for clustering. The observer has clustered objects into four groups: *"blue"*,*"white"*, *"yellow"*, and *"transparent"*.

system that is followed by modelling and theorization[1].

Finally, we will try to correlate quantitative and qualitative analyses, and explain numerical results with underlying behavioral patterns. Due to the complexity of the task, qualitative analysis of the ordering task is left beyond the scope of this paper and will be included in future communication.

Results

Clustering: Quantitative results

Several approaches have been revealed by the clustering task. The examples of clustering are illustrated on Figure 3.

The frequency analysis has shown that for initial clustering color or hue was the most widely used criterion with blue, yellow, and white clusters standing out at the histogram (Figure 6). However, in several cases above-mentioned hue-based groups coexisted with transparency-based groups as well. In 53% of the cases (9 out of 17) color was the sole criterion used for clustering, while combination of color and translucency has been used 5 times, i.e. in 30% of the cases (example is illustrated on Figure 5). Transparency, surface reflection properties and familiar object criteria were used once each. The histogram of sub-clustering of the first level clusters (Figure 7) illustrates that surface properties were secondary attribute for intra-group clustering. Group definitions varied a lot among subjects, and so was the number of groups. While 9 people felt comfortable with 3 simple groups, 4 people used 4 groups, 3 people used 5 groups, and one person used 6 groups for first level clustering. 3 groups remains most popular on the sub-clustering task too with 7 people opting for it. Groups of 2, 4, and 5 each were selected by 3 people. While one artist observer created 9 pairs out of 18 object group. The overall hierarchy of the criteria and corresponding labels are given on Figure 8.

Clustering: Qualitative results

Qualitative analysis of the clustering task revealed interesting process of individual decision making with many similarities with the existing models observed in the literature [11, 12, 13]. For example, *expected-value optimization* - opting for simple solution with least investment of time and effort; *satisficing* - looking for various options until satisfactory threshold is reached; *acquiesce to a person in authority* - like using references to the literature and CIE definitions in their judgments; *anti-autoritarianism* - opposite of the latter, deliberately taking non-conventional decisions, or *flipism* - some degree of randomness. The categories with the summary of the corresponding codes are illustrated in Table 1. *Situation* category covers the impact of the environment and interaction with the external world. *Structure expectation* category consists of the observations that subjects had anticipation of a hidden structure, i.e. assuming existence of the "right solution" of the task and tried to reach it by reverse engineering. Using *references* and comparison among objects, as well as with external stimuli has been observed a lot. *Clustering* category implies understanding of semantics, trial to cluster by several means: a) ma-

[1]Formulation of a theory.

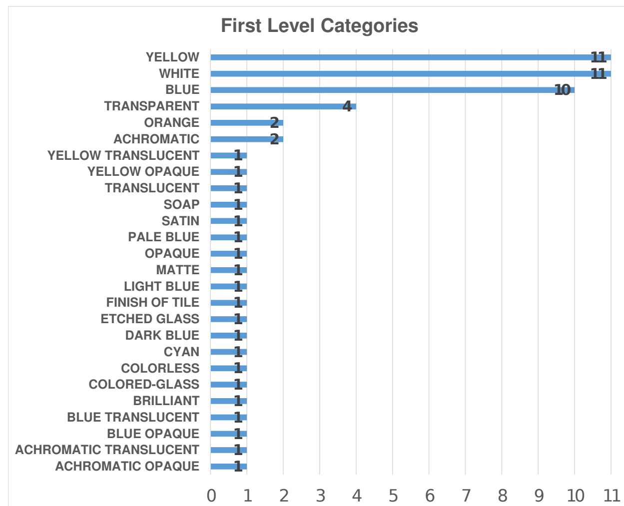

Figure 6: A histogram summarizing cluster labels used by subjects for the initial clustering. Hue-based labels: *"blue"*,*"white"*, and *"yellow"* stand out as most widely used ones.

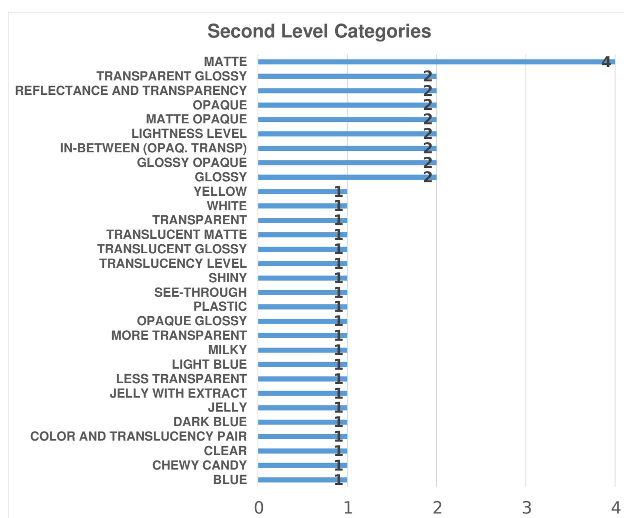

Figure 7: A histogram summarizing cluster labels used by subjects for the sub-clustering the original clusters. Surface properties along with transmission properties have been used as grouping criteria on this level.

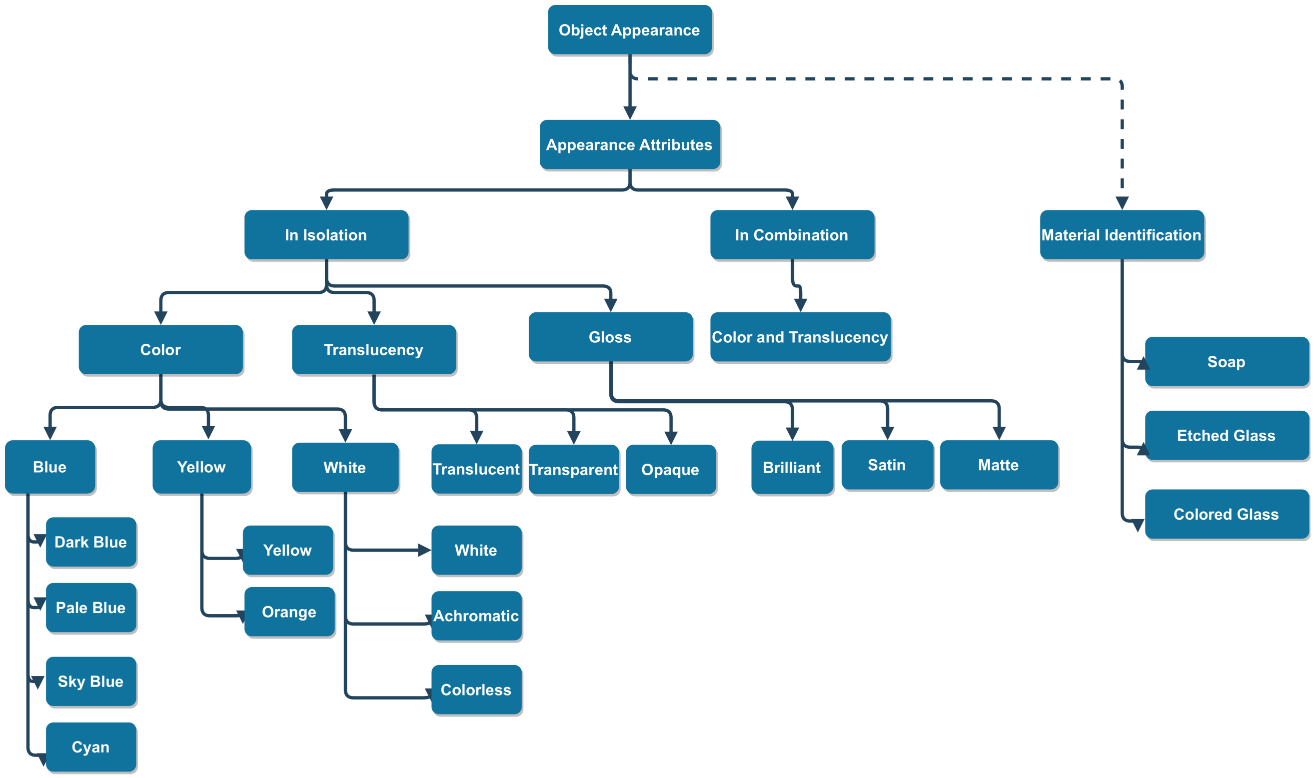

Figure 8: The diagram illustrates criteria usage to assess appearance similarity. While the majority of the observers used attributes with various labels, observers with artistic background used material identification for comparison. Proposed diagram is a simplified description of the major observations based on our data. It should be neither taken as a comprehensive map of appearance assessment process, nor generalized to other stimuli.

terial identification; b) ranking and searching for optimal thresholds to split; c) binary - *true/false* type check whether particular objects belongs to a particular group. *Strategy* category is mostly a collection of observations that people tried to select simplest approaches with least amount of time and effort investment needed. Finally, it is also worth mentioning that subjects were interrupted in the process after two levels of clustering, even if they wanted to continue. This aspect impacts the very final result, not the process itself. Hence it is not part of the theorization.

In total, the process looked as given on Figure 9. On the pre-analysis step subjects experienced stress, tried to find the physical structure ignoring popping out attributes. After some time, we observe acceptation and fatalism stage, when subjects fail finding a perfect layout and accept a simple solution. Finally, we have explicit signs of adaptation, when subjects start feeling more comfortable with the task (some of them explicitly mentioned that they had learnt throughout the process) and start scrupulous refinement and revision of their decisions. Overall, the observation are aligned and complementary to what was found in [8].

Ordering

Occurrence of the axes defined by 5 spheres, and with all objects altogether, are given on Figures 10 and 11 respectively. Transparency-translucency criteria was most widely used for the spaces defined with 5 spheres. We can draw parallel with the clustering task where color was most salient attribute due to its simplicity. In this case hue of the spheres is nearly identical and the color-based arrangement requires significant effort and judgment. Therefore, translucency as a more salient attribute prevails, as we have more apparent difference in translucency among spheres than that in color. Surprisingly, 16 out of 17 people used orthogonal axes and arrangement similar to Cartesian coordinate system, while 1 observer used polar coordinate system. Two observers mentioned circularity as a possible option for arrangement, although they were possibly biased either with the chromaticity diagram poster hanging in the experiment room, or with their colorimetric knowledge. Another interesting observation was that expert observers used caustics and shadows a lot to judge translucency. Interestingly, none of the observers used more than two axes for initial definition. While 12 people had one dimensional ranking, five people used two axes for arrangement of the spheres. Apart from that, introduction of new objects messed all their structure up and some of them faced impossibility to ”fit” new objects within pre-defined space. This makes us question possibility to generalize appearance ordering. We might always encounter a new object whose appearance cannot be adequately positioned in pre-defined appearance ordering space, in contrast with color spaces in existence.

Impossibility to describe whole new attributes of the six additional objects lead subjects to re-define their ordering systems and use more axes (5 people used 3 dimensional ordering, while 1 used 6 dimensions). 2 dimensions were used by 3 people, while 8 people simply put new objects into their one-attribute-based ranking line. However, it is obvious that 1 dimensional representation

Table 1: Categories and summary of the corresponding codes observed within the data. *Being cut in the process* could be additional category but it does not affect the clustering process.

Categories	**Situation**	**Structure Expectation**	**Reference or Absence of Reference**	**Semantics / Clustering**	**Strategy**
Summary of Codes	•Experience the conditions •Is adapted by the observer •Introduction to the object •Interaction	•On the set due to ones shape and its derivation in appearance	•Comparison •Between objects •To the light •To another shape •To memory •Guidance (What attribute should I cluster?) •Tentative •Doubt	•Description •Material identification •Attribute related •Related to order (along an axis) •Binary	•Simplicity •Pop-out (intense sensations and contrast) •Too many classes are tedious and do not pop-out

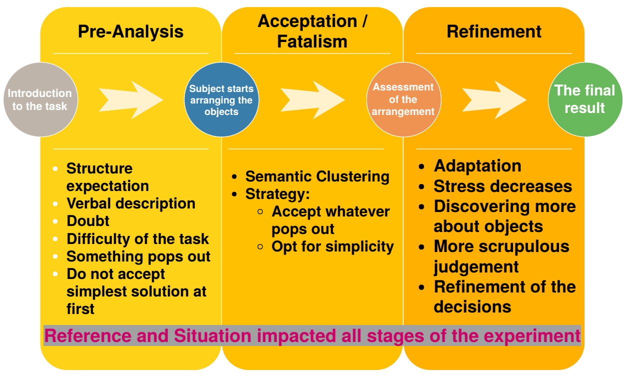

Figure 9: After linking the categories we can have a general picture of the visual clustering decision-making process. Stressful pre-analysis is followed by acceptation of the simplest structure that is scrupulously refined after adaptation.

did not describe the overall appearance and they simplified a task down to ordering by one attribute, rather than describing the entire appearance within the space. Even though those observers realized that their system failed to adequately represent overall appearance, 8 people refused to redefine the entire system using all objects at once, while 4 people were uncertain whether it was a good idea. Similar phenomenon has been observed in [8], and the fact can be explained either by simplicity, laziness, or general skepticism towards changes. Some of the examples of the ordering system created by the subjects is illustrated on Figure 12.

Discussion

Two groups of observers were identified by their overall appearance assessment. Refer to the diagram on Figure 8. While the vast majority of the observers conducted attribute-based comparison, the observers with artistic background opted for material identification, assigning group labels, like *"soap"*, or *"etched glass"*. On the other hand, people used not only separate attributes, like color, gloss, or translucency, but also their combinations too (for instance, clustering based color and transparency criteria into four groups: *"blue"*,*"yellow"*,*"white-achromatic"*, and *"transparent"*, as illustrated on Figure 5). While attribute usage was similar among observers, there was no consistency in thresholds used among clusters. For example, even though color was widely used attribute, some people used single *"blue"* cluster, while others had bluish objects allocated into several clusters,

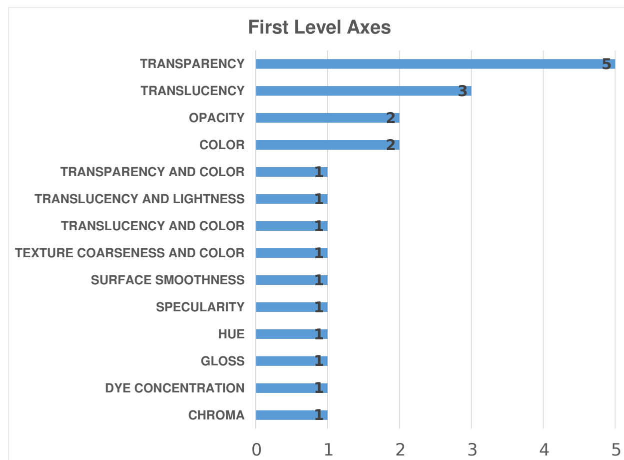

Figure 10: A histogram summarizing occurrence of the axes when space was defined using five spheres. As we see, transparency-translucency criterion prevails for the five spheres.

like *"dark blue"*, *"pale blue"*, or *"cyan"*. Composition of each of those groups varied a lot. On the other hand, hue-based uncertainty was almost non-existent and none of the bluish objects were classified under *"white-achromatic"*, or *"yellow"* labels.

Comparing with the behavior observed in other tasks [7], we can conclude that decision making is easy when a clear guideline exists (e.g. ranking by gloss). However, in absence of a clear guideline, decision making process becomes more cumbersome, requires more effort and subjects follow general decision making rules. We were not able to identify what is an attribute, a reference, or an étalon for appearance comparison (we can draw parallel with a kilogram as an attribute for mass comparison). While hard metrology of physical entities is possible, the potential for soft metrology capabilities for appearance remains unclear. Nevertheless, local comparison between close objects is a natural behavior, and thus, there might be a potential to develop metrology.

Frequency analysis in this experiment should be carefully taken. Although color stands out as a primary criterion for clustering, this might be the case due to the dataset and should not be generalized as a rule. Small number of easily separable hues lead to usage of expected-value optimization. However, whether this trend holds for high number of hues needs further investigation.

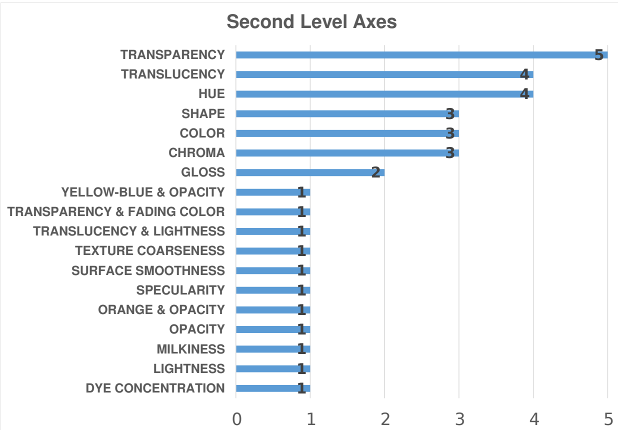

Figure 11: A histogram summarizing occurrence of the axes when additional six objects were introduced. The new objects create need for additional *shape, hue,* and *chroma,* dimensions.

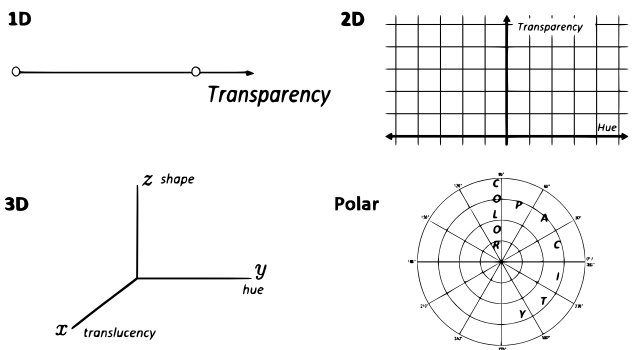

Figure 12: Examples of the ordering systems observed. From top-left clockwise: 1-dimensional system - ranking by *transparency*. 2-dimensional system with *hue* and *transparency* axes; A polar coordinate system with *color* as a radius and *opacity* as an angle; 3-dimensional system - with *hue, translucency,* and *shape* axes;

Summary and Future Work

First of all, it is worth mentioning that we have observed interesting cross-individual differences and similarities. The main concept is mostly similar among observers, while thresholds differ significantly (for instance, creating either one *"blue"* category, or two *"blue"* and *"cyan"* categories). Most of the observers agreed on the initial clustering, while on the detailed level cross-individual differences become apparent. However, the bias can exist due to randomness of the group left for sub-clustering. The most obvious cross-individual similarity for the ordering task was preference of orthogonal axes, and more specifically, Cartesian coordinate system. Apart from that, the results have also shown that professional background of the individual can have a significant impact on the results. Further topic of investigation should be whether cultural or other aspects play an important role as well.

Secondly, the qualitative analysis of the data has shown that even though some similarities have been observed, no obvious clustering or appearance ordering system popped out. The question for further investigation is whether it is possible to have an appearance ordering system at all based on our data, and if so, what was the reason we have not observed any? Our hypothesis is that sampling in our data was too sparse considering the number of dimensions. The question how to design an optimal experiment for natural appearance ordering system definition is still open. Multidimensional scaling can be an option for better visualization of the object similarities in the ordering tasks. As for the clustering task, the starting point for the future study can be using high number of hues, while keeping less diversity in other attributes. We hypothesize that color-based clustering will become challenging and color will not be a primary criterion anymore. In addition to that, the role of physical reality and interaction also remains unclear. One of the topics for future investigation should be replication of the similar tasks in Virtual Reality.

Finally, as the complexity of the observations cannot be fully described within the scope of this paper, comprehensive analysis of all eleven tasks (including [7]), as well as possible ways to validate our hypotheses will be discussed in further communications.

References

[1] Christian Eugène, "Measurement of "total visual appearance": a CIE challenge of soft metrology," in *12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement*, 2008, pp. 61–65.

[2] M Pointer, "A framework for the measurement of visual appearance," *CIE Publication. CIE 175:2006 ISBN: 978 3 901906 52 7*, 2006.

[3] Gaurav Sharma, Wencheng Wu, and Edul N Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application.*, vol. 30, no. 1, pp. 21–30, 2005.

[4] Gaurav Sharma and Raja Bala, *Digital color imaging handbook*, CRC press, 2002.

[5] Dorothy Nickerson, "History of the Munsell color system and its scientific application," *Journal of the Optical Society of America*, vol. 30, no. 12, pp. 575–586, 1940.

[6] International Color Consortium, "Specification ICC. 1: 2004-10," *Image technology colour management–Architecture, profile format, and data structure*, 2004.

[7] Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen, and Jon Yngve Hardeberg, "Behavioral investigation of visual appearance assessment," in *26th Color and Imaging Conference*, Vancouver, November 2018. 6 pages.

[8] Jean-Baptiste Thomas, Aurore Deniel, and Jon Y Hardeberg, "The plastique collection: A set of resin objects for material appearance research," *XIV Conferenza del Colore, Florence, Italy*, p. 12 pages, 2018.

[9] Pierre Paillé, "L'analyse par théorisation ancrée," *Cahiers de recherche sociologique*, , no. 23, pp. 147–181, 1994.

[10] Barney G Glaser, Anselm L Strauss, and Elizabeth Strutzel, "The discovery of grounded theory; strategies for qualitative research," *Nursing research*, vol. 17, no. 4, pp. 364, 1968.

[11] Kristina L Guo, "DECIDE: a decision-making model for more effective decision making by health care managers," *The health care manager*, vol. 27, no. 2, pp. 118–127, 2008.

[12] Leon Mann, Ros Harmoni, and Colin Power, "The GOFER course in decision making," *Teaching decision making to adolescents*, pp. 61–78, 1991.

[13] Pam Brown, "Career coach: Decision-making," *http://www.pulsetoday.co.uk/career-coach-decision-making/10967084.article (Retrieved: 27/01/2019)*, November 29, 2007.