On the appearance of objects and materials: Qualitative analysis of experimental observations

Davit Gigilashvili, Jean-Baptiste Thomas, Marius Pedersen and Jon Yngve Hardeberg

Department of Computer Science, Norwegian University of Science and Technology, Norway
Email: davit.gigilashvili@ntnu.no

Perception of appearance of different materials and objects is a complex psychophysical phenomenon and its neurophysiological and behavioral mechanisms are far from being fully understood. The various appearance attributes are usually studied separately. In addition, no comprehensive and functional total appearance modelling has been done up-to date. We have conducted experiments using physical objects asking observers to describe the objects and carry out visual tasks. The process has been videotaped and analysed qualitatively using the Grounded Theory Analysis, a qualitative research methodology from social science. In this work, we construct a qualitative model of this data and compare it to material appearance models. The model highlights the impact of the conditions of observation, and the necessity of a reference and comparison for adequate assessment of material appearance. Then we formulate a set of research hypotheses. While our model only describes our data, the hypotheses could be general if they are verified by quantitative studies. In order to assess the potential generalisation of the model, the hypotheses are discussed in context of different quantitative state-of-the-art works.

Received 14 December 2020; revised 30 March 2021; accepted 04 May 2021
Published online: 12 June 2021

Introduction

We observe the emergence of new ways to fabricate objects and materials, such as 3D printing [1] and advanced surface processing [2-3]. Object manufacturing is also related to digital edition and design [4]. Both need to be supported by an adequate description of material appearance. This description may be produced with a physical measurement and its correlation with human perception but could also be related to semantic communication. A further challenge comes with the development of programmable matter [5-7]. We foresee that an object's appearance will not be limited to the natural appearance of the material it is made of, but also an object may have an evolving shape, that impacts its appearance. Therefore, description, quantification, and communication of appearance are important.

According to the ASTM E284-17, Standard Terminology of Appearance [8], the **appearance of an object** is "*the collected visual aspects of an object or a scene*"; while **perceived appearance** is defined as "*the visual perception of an object, including size, shape, color, texture, gloss, transparency, opacity, etc., separately or integrated.*" The same dictionary highlights that "*appearance, including the appearance of objects, materials, and light sources, is of importance in many arts, industries, and scientific disciplines.*" Appearance is a complex phenomenon that is far from being comprehensively understood. Considering its complex nature, it is usually broken down into various attributes that entail only particular dimensions of appearance. The CIE (Commission Internationale de l'Eclairage, International Commission on Illumination) defines colour, gloss, translucency and texture as four major appearance attributes [9].

Appearance has long been a point of scholarly interest from physical [10-11] (e.g. solving radiative transfer equation [12]), psychological [13], and philosophical [14-15] points of view. Hunter and Harold [10] provided the first significant summary of appearance measurement techniques, which aim "*to obtain numbers that are representative of the way objects and materials look*". However, they consider that comprehensive analyses of total appearance is impossible and impractical and argue that, at least, "*measurements of specific attributes of appearance can be exceedingly useful and economically important*". Their work is far from modelling total appearance and provides little guidance on the correlation between metrology and perception.

Practical aspects of total appearance by Hutchings [14-15] focused on unifying knowledge of appearance from science disciplines and arts, which "*can be based on a quantitative understanding of the basic perceptions of form, colour, translucency, gloss, and movement.*" He describes and structures seven factors that influence total appearance [14, 16]: appearance images; immediate environment factors; inherited and learned responses to specifics; receptor mechanisms; design; object properties, and light source properties and defines it as: "*total appearance combines a description of the appearance of each element of a scene... with a personal interpretation of the total scene in term of its recognition and expectation*". Eugène [13] highlights the definition recommended by the CIE "*the total appearance points out the visual aspects of objects and scenes*" [9]. On a semantic level, Eugène considers appearance measurement challenging, because it involves subjective judgment and argues that "*a goal of making measurements that ensures appropriate quality control in the manufacturing process is probably achievable, but the measurement process will be multidimensional, product specific and probably application specific*". Choudhury [11] also reviewed total appearance as a concept and described a four-step flow of total appearance from molecular composition of an object to the high level cognitive interpretation of appearance by a human observer.

Despite those attempts, the objects' total appearance is so difficult that most research focuses on the total appearance of a material. Most recent quantitative studies aim to provide a correlation model between optical properties and perception of a single appearance attribute (e.g. [17]). Works in computer graphics, vision, and metrology focus on very narrow specific cases and provide a quantitative analysis of particular appearance attributes [18-25], or investigate the role of image attributes on appearance, e.g. [26]. Many are based on psychophysical studies with human subject involvement. However, the constraints imposed on the experimental conditions of those works limit, in general, their relevance in real life, such as, the viewing condition in colorimetry. The majority of these studies are based on images, either synthetic [23, 25] or real [27-29], shown on displays with no possibility for physical interaction. Wherever physical samples are used [30-31], interaction and possible observation geometries are still strictly constrained. While the attributes are studied separately, it is unlikely that individual attributes of appearance are independent, e.g. transparency may impact gloss perception [32]. Furthermore, there is inconsistency in terminology. On the one hand, terminology differs across communities, e.g. *texture* in computer graphics refers to the image mapped on a mesh, while in the context of textiles, *texture* is primarily a tactile attribute describing surface geometry. On the other hand, terminology can also be ambiguous within the field of appearance, e.g. translucency, transparency, perceived translucency or opacity are sometimes used interchangeably, as in [25], which can impact the experimental observations. Further work is needed to develop a quantitative model.

In parallel to the many quantitative studies, we propose building a qualitative model of material appearance outlining general processes to formulate relevant research hypotheses. Analysing and testing those hypotheses reveals more details of total appearance mechanisms, including people's behaviour to assess appearance, the way they perceive and communicate appearance. We hypothesise that appearance is a social interaction, between an object in a scene and a person, or between two persons communicating about one object in a scene. Therefore, we approach the problem from a social science perspective and investigate how subjects interact with objects and communicate with other people. For this purpose, we conducted an experiment and applied the Grounded Theory Analysis [33], derived from the Grounded Theory Approach [34-35], to the data collected. This method belongs to the class of inductive research methods[1]. We conducted the experiment using physical objects from the *Plastique* artwork [39] comprising resin spheres, cuboids, and complex female bust sculptures with different mixes of colorants and surface roughness properties. The process and the results were videotaped and then analysed.

In the next section, we introduce the experiment. Then, we develop the qualitative model of our data. From this observation, we formulate research hypotheses and discuss them. We conclude by highlighting the potential limitations of this work.

Materials and methods: the social experiment

We conducted an experiment based on an interview format, which consisted of 11 visual tasks where the observer was asked to interact with physical objects, describe them and explain their choices (both rationales and actions). The experimenter asked additional questions to clarify the motives of particular actions, and to disambiguate the interpretation of the concepts by the participant. The study was reported to and approved by the NSD - Norwegian Centre for Research Data (project number 59754).

Stimuli

Generating the proper visual stimuli for the social interaction was one of the fundamental challenges in the preparation process. This study is based on real physical objects and this choice is discussed in **Appendix 1**. The objects belong to the artwork collection *Plastique* that was commissioned to the independent artist Aurore Deniel from "Aden Keramikk"[2]. Technical details of production, and a description of the collection and subsequent analysis of the creation process are reported in [39]. The objects in the artwork are made of resin and come in three different shapes (cuboid, spherical, and complex female bust), various colorant mixtures (from achromatic to blue and yellow), and three levels of surface coarseness (also referred to as roughness).

[1] An example and method description in English can be found in e.g. [36], many other examples of studies can be found in the literature, focusing on diverse social aspects, such as [37-38].

[2] Aden Keramikk website, https://auroredeniel.wixsite.com/adenraku – last accessed 21 November 2019.

 ISSN 2227-1309

Experimental protocol

The interviews were held in two rooms with different mixed illuminations from direct sunlight (subject to weather conditions) and artificial fluorescent lighting systems. The illumination was measured with a photometer at the beginning and at the end of the interview to record changes of viewing conditions. The desk, where the objects were introduced to the participant, contained some potential visual references: a white sheet of paper, a checkerboard and a pen with text on it. We expected the observer to use them as a background of reference for appearance assessment. The observers were not explicitly instructed to use these objects to preserve their natural behaviour. Additionally, the checkerboard could serve geometric calibration for the camera positions.

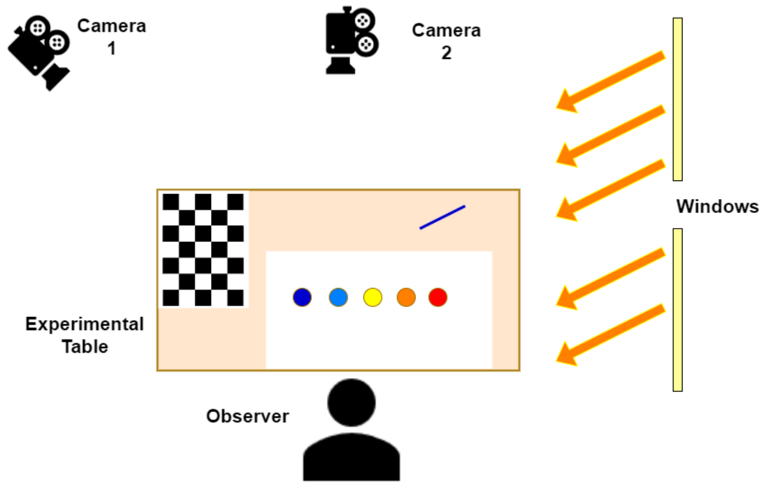

Figure 1: ***A Bird's-Eye Representation of the Experimental Setup.*** *The natural illumination incident from the windows is mixed with the artificial light incident from the ceiling (not shown). The different angles of the two cameras helped us analyse the behaviour of the observers.*

People had complete freedom to interact with the objects, to touch and move them. The entire process was videotaped by two cameras (Figure 1), from front and side, to detect all potentially interesting movements and facial expressions. 17 observers, 11 males and 6 females, participated in the experiment. All of them were proficient in English. 12 of them had a scientific background related to colour, vision, and appearance studies; 2 participants had an artistic background, while 3 observers were considered naïve. Their age ranged between 24 and 60, with 34 being the median age. One participant was colour deficient, the others performed the interview with corrected-to-normal vision, when needed. The experiment was conducted between March and May 2018. The experiment was arranged during the day, in order to have direct sunlight in the room. On average, illuminance at the table in the beginning of the experiment was 1512 lux and colour temperature was 5306 K, the standard deviation among all experiments was 766 lux and 615 K, respectively. In addition, illuminance difference and colour temperature difference between starting and ending point of each interview was on average 683 lux and 497 K, respectively. We assume that some changes in participants' behaviour might be related to the amount or quality of incoming light (e.g. using artificial light source for translucency assessment rather than sunlight or vice versa).

12 observers were interviewed by one interviewer and the other 5 by another one. Although the social interaction, particularly the conversation between the participant and the experimenter, was subject to improvisation and individual development, the experiment followed a well-defined routine. The observers went through 11 tasks involving set of objects grouped in 9 boxes (Figure 2). Two boxes were used twice, although this was not revealed to the participants. In the first task (box Q), observers were

asked to cluster 48 cuboid objects in any way they considered natural. We wanted to observe whether one particular appearance attribute was predominant in a grouping task. In the second task (box C), observers were asked to arrange five different yellow spheres in a meaningful way, i.e. creating some ordering system for them. Afterwards, they were given additional objects with different shape, colour, and other attributes, to be placed into their ordering system. With this experiment, we tried to explore potential appearance ordering systems. Tasks 3 through 10 were composed of two parts. First, observers were asked for a semantic description of the objects without touching them. The second implied ranking them by either glossiness (boxes X, M, P, A) or translucency (boxes F, X, A, Z). It is worth mentioning that the phrase "how light is going through" was used instead of "translucency", to avoid potential confusion by the term. The experiment was concluded with a binary opaque/non-opaque classification of six spherical objects (box T) with and without high intensity directional flashlight.

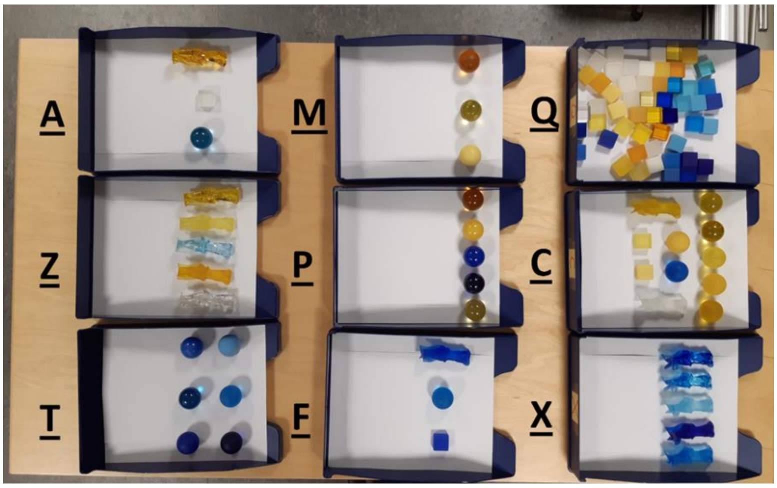

Figure 2: ***The Nine Sets of Objects.*** *The nine sets of objects have been used for eleven tasks throughout the experiment. The single letter identifiers of the boxes are completely arbitrary. The figure has been reproduced from [40]. Reprinted with permission of IS&T: The Society for Imaging Science and Technology sole copyright owners of, "CIC26: Twenty-sixth Color and Imaging Conference 2018".*

Data analysis

The data collection process was followed by a thorough data analysis that consisted of three stages:

1. Two independent manual transcriptions of the collected data, i.e. more than 20 hours of video materials, were performed. This includes transcribing speech, as well as taking notes on behaviour and movements.
2. We performed a quantitative study on the results of the tasks by frequency analysis. This analysis was independent from transcription and was based on the task results recorded throughout the experiment. The quantitative data were presented and discussed at conferences [32, 40-41].
3. The qualitative analysis was based on the transcribed material using the Grounded Theory Analysis. Those observations were augmented and strengthened by the results of the quantitative analysis.

Qualitative model of material appearance assessment

We used the Grounded Theory Analysis [33], derived from the Grounded Theory Approach [34-35], to analyse the data. The method includes a comprehensive description of the observations and labelling them with codes (*coding* step). We watched the videotaped experiments (around 20 hours of video), manually extracted all observations, and labelled them accordingly. Later, conceptually similar observations are grouped into categories (*categorisation* step). For instance, we observed that if the object is lit from behind or if it is placed on a textured background, it can look more translucent. These observations are grouped together into the "**Conditions of Observation**" category. Those categories were carefully designed, defined and consolidated - in particular, they were consolidated with the quantification of some of the observations. Afterwards, we identified how different categories interact with each other (*co-linking step*) that eventually leads to *modelling* through the *integration*, where we redefined and refined what we observed. The process led to *theorisation*. According to the Grounded Theory Analysis as described in [33], theorisation is a process that is more advanced than a mere description of observation (more conceptual and better structured), yet still anchored in the observation, but far from a general theory. The potential of generalisation towards a theory of our *theorisation* is discussed in the next sections. The coding part was performed two times independently by two persons. The categories were consolidated and revised, and the subsequent steps were conducted jointly.

The main reason for choosing this method is that the result, while qualitative, should guarantee to be strongly rooted in the data, and there are security mechanisms that avoid falling into an individual interpretation, e.g. the verification that all the codes are belonging to at least one category. Another reason is that this method is known to allow the experimenter to improve his or her understanding of the phenomenon to be studied, and the authors of this article benefited greatly from this collateral effect.

Definition of categories

We have identified the following categories that encapsulate all the codes observed in the codification step:

1. **Object** is a given sample to be considered for a particular task. It is very stable because its intrinsic parameters are static (e.g. shape, surface, size, but also specific light effect). However, it is dynamic at the same time because its appearance may vary depending on the conditions of observation.
2. **Conditions of Observation** is a set of extrinsic factors that permit the observation, contribute to the appearance of a given object and the communication of it. Conditions of observation is the place and an individual observer (illumination geometry and spectral power distribution, experimental room interior, viewing angle, personal vision, physiological condition and mood, background, vocabulary pool, etc.) - We want to highlight that observer is not a separate category but part of the conditions of observation. We are presenting an objective cross-observer generic model representing a task-motivated material assessment process. The way a subjective psychological or physiological condition of the observer contributes to the overall process is by nature no different from illumination geometry or other external conditions of observation
3. **Methodology** is a stable systematic way to act and make decisions towards completion of a task. Methodology can be based on intuition or experience, and it could converge and be revised after trial and failure (calls **Learning and Adaptation**).
4. **Comparison** is an action that permits judgement of the objects by referring to something else, making assessment relative to a **Reference**. Similarities and differences are judged either with an arbitrarily chosen reference or among different states of the object itself, that becomes the reference.

5. **Reference** is the observation, memory, concept, etc. an object or a set of objects are compared with. This is one of the most important categories when we want to discuss measurement of appearance.
6. **Vocabulary Search** is the process to identify and select the right **Vocabulary** in order to communicate and express the perceived appearance of a given object or set of objects. In the process of **Vocabulary Search**, different methodologies might be applied, including, but not limited to, citing standard definitions from the literature, recalling familiar objects from memory in order to draw parallels, or looking up for proper words on the Internet.
7. **Vocabulary** is a selected set of words, like adjectives, nouns, phrases (e.g. "blown-up glass") - all attributes and labels used to describe the appearance of a given object or set of objects. The selection of this set is derived from the **Vocabulary Search** and serves as a basis for the **Semantic Description**.
8. **Semantic Description** consists of tentatives to name, or to describe the appearance of one given object or a given set of objects.
9. **Completion of a Visual Task** is a process to successfully perform a given mission that relies on the analysis of the visual appearance of a given set of objects but also on the **Task Interpretation**.
10. **Task** is a given mission an observer is instructed to accomplish by an **Experimenter**. We used those tasks to lead the interviews.
11. **Experimenter** is a person, in our case one of the authors of the paper, who introduces tasks to the observers and guides the entire process by oral communication with an observer. The communication and interaction with an observer were subject to individual improvisation by the experimenter. Thus, this impacted the data and made all experiments unique.
12. **Structure Expectation** is an assumption by an observer that there exists a structure in the data. This structure, that may or may not exist, will be used as a cue to perform the task, instead of, or in addition to, relying on visual qualities. This implies that the participant assumes that there is an expectation or a solution known by the experimenter, which was not the case.
13. **Task Interpretation** is a decoding process of the oral description of the task conveyed by the **Experimenter**. The observer tries to understand what they are expected to do and selects a **Methodology** to reach the goal.
14. **Decision-making** is a general approach that leads the observer to the strategy on how to perform a **Task** that involves freedom of interpretation. This was not observed in all experiments, because some tasks were less prone to interpretation.
15. **Learning and Adaptation** is a function of time affecting actions of the observer. It impacts the processes we have observed. As the observer interacts with the corpus of data, their understanding of the data is refined based on the recently acquired experience. Secondary visual attributes, like scratches and imperfections start to be taken into account, leading potentially to refinement in **Methodology**. Observers start recognising similarities with the part of the corpus already studied and behave accordingly. It can have a positive impact and facilitate the task completion or a negative impact related to exhaustion, shortcut or overconfidence.

Definition of the qualitative model

The resulting model of the data is illustrated in Figure 3. The model consists of two blocks. The pivotal visual part unfolds the flow of the process from introduction of the object towards the completion of a particular mission. An auxiliary decision-making part describes all the factors that could impact a methodology selection in the process of task performance. It is worth mentioning that the decision-making part only impacts the result of the experiment, i.e. what we observe by the frequency analysis, but does not change the model and the flow of the processes itself. The structure of the model is independent of the observer and the task.

The **Object** is observed in certain **Conditions of Observation**. The combination of both categories creates in fact the core of the sensory perception of the object by a person. While the **Object** has some absolute properties, total appearance is impacted by the various **Conditions of Observation**. Anything that can impact the perception of the appearance of an object is considered a **Condition of Observation**. While usually conditions impact the object appearance, the interaction is both-ways, as an object could also impact the conditions (e.g. produce caustics, evoke particular memories). The category **Methodology** is at the heart of the observation. In fact, we observed how the participants perform the task and describe their actions and decisions. Indeed, the **Object** and **Conditions of Observation** constrain the **Methodology**. However, we observed that there are major contributions from **Comparison** and the **Decision-Making** which define or constrain the **Methodology**, and in our data, they might be as important as the perception part because they are very general. Both of them are induced by the **Task** given to the observer. The **Comparison** is required to analyse the samples, and this is done by **Reference** to *something*. As we shall see, the observation that a reference is systematically used is a crucial piece of information, which is both very positive from a perspective of metrology, but also a great challenge when it comes to selection of an appropriate reference. **Decision-making** is required when a **Task** leaves room for interpretation, and is based on the **Task Interpretation**. It is closely related to the **Task** itself, the way it is conveyed by the **Experimenter**, and constrained by the **Structure Expectation** on the data. The latter was observed in our experiment, but it is hard to anticipate whether this will be observed in a more free context. Observers applied various decision-making models to come up with an efficient strategy and select a particular **Methodology** to complete a mission [41]. Based on the **Methodology**, the visual task is solved and the observer reaches the **Completion of a Visual Task**. We also observe that the **Methodology** is used to structure the **Vocabulary Search**, that led to a selection of **Vocabulary** used to come up with a **Semantic Description**. Several methodologies were observed to be pre-selected, in order to find, choose, and convey the **Vocabulary** necessary for **Semantic Description**. **Semantic Description** can be a substantial prerequisite for the **Completion of a Visual Task**. We observed that subjects tend to describe objects in the process of **Completion of a Visual Task** even if they are not explicitly instructed to do so. In order to assess appearance, they seem to construct a semantic image of the target in their mind with or without explicit oral expression. In addition, the description of the objects might already include the draft solution of the visual task (for instance, object A is described as glossier than B and as less glossy than C, while the visual task is to rank the three by glossiness). Finally, we should highlight that a significant impact of **Learning and Adaptation** was observed throughout the experiment and it impacts all other categories.

Verification and analysis

In order to demonstrate how the model is rooted in the data, we describe an example case in **Appendix 2**, where the observer is asked to rank five spheres by their glossiness. We recall that this model is a model of our data. However, it is interesting to study how those data compares to general models of material or object appearance by Hutchings [16], Choudhury [11], and Eugène [13]. They all referred to the scene context, supported by the CIE definition that also includes scene concept into the total appearance [9, 13]. In our data we can observe how this context is verbalised by the observers. The context is summarised in the **Conditions of Observation**. These conditions were experienced by the observer, but explicitly mentioned only when these conditions constrained successful completion of the task. Otherwise, the impact of the scene was encapsulated in the **Semantic Description** and in the

Completion of a Visual Task. For example, observers ranked an object by gloss, using distinctness-of-image gloss when the light was low enough, without further discussing the environment. However, when intense direct sunlight made it impossible to observe distinctness-of-image gloss, the observers discussed the scene and mentioned that the sunlight in the scene made task completion difficult.

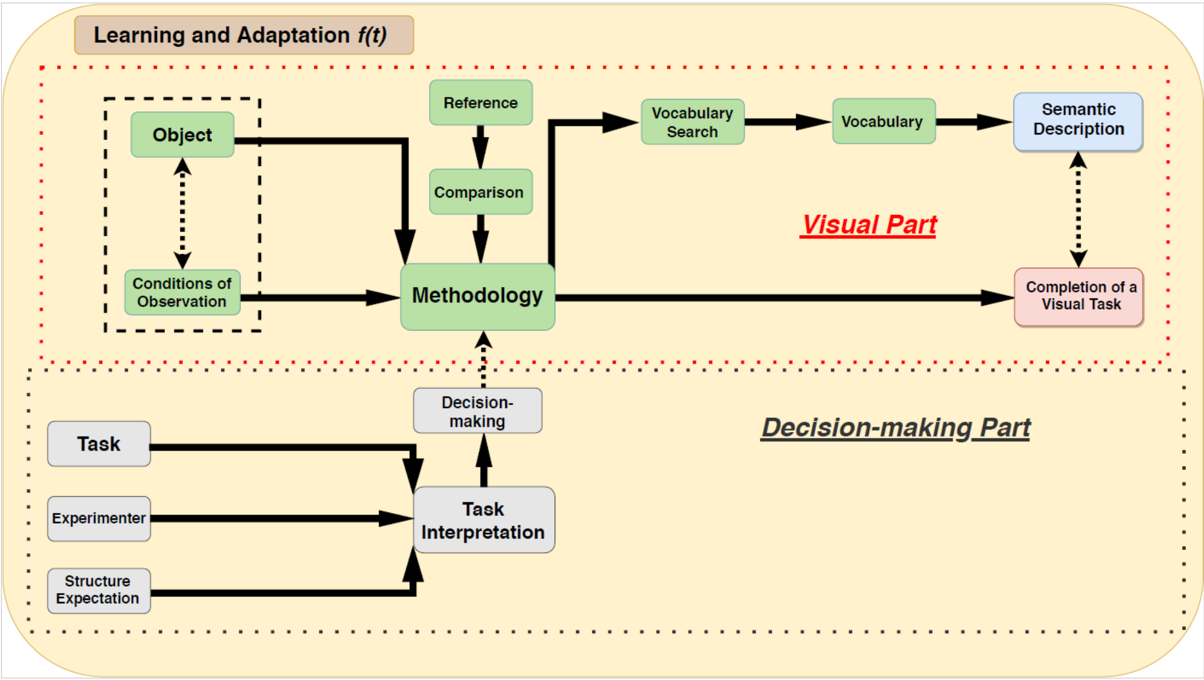

Figure 3: ***Qualitative Model of Material Appearance Assessment.*** *The primary Visual Part of the model details the flow of the process from introduction of an object in particular conditions to semantic description of its appearance and completion of a visual task using this object. Auxiliary Decision-making part illustrates categories impacting methodology selection in the Visual Part, while Learning and Adaptation impacts the entire process as a function of time (f(t)).*

Eugène [13] supports the idea of total appearance implying higher level semantics, for instance concepts like, "visually assessed safety", "visual identification of the scene", "visually assessed usefulness of the scene" etc. in addition to Hunter's attributes. In our data, this appears in the **Semantic Description** when observers describe the objects as "like food", "fragile", "pricy". In addition to appearance attributes, they also referred to high level semantics, like usefulness ("decoration", "soap"), safety ("fragile"), in order to express and communicate the appearance of the objects and materials.

Apart from that, Hutchings considers that "there are two classes of appearance images: the impact (or Gestalt) image, and the sensory image. The impact image is the initial perception of the object plus an initial opinion or judgment." [16] This is also present in our model, where the sensory image is limited by the **Object** and the **Conditions of Observation**. This is also the case for Choudhury's model [11], where the three first stages correspond to the sensory image of Hutchings and the fourth one is related to higher cognitive interpretation. Choudhury also emphasises the physiological phenomena as an explanation of the process, which we do not consider.

To conclude on those comparisons, it appears that the works discussed above focus much more on the sensory analysis, while we observe more on the human behaviour, semantic description, decision-making and task-solving than them. Compared to their works on those aspects, which are a formulation of opinions, what we observe is rooted in our data. Our model is centred around the completion of a visual task, while there is no motive of appearance interpretation introduced in those other works. We, however, all agree on the idea that conditions of observation (including environmental or individual background aspects of a human subject) have a tremendous impact on perceived appearance.

Three key behavioural observations

The omnipresence of a reference

Comparison with a reference turned out to be a pivotal point of all methodologies applied for visual task performance, as well as for semantic description. The reference varied and was any of the following, but perhaps not limited to:

a) Comparison to the appearance of another object (e.g. comparing two objects to decide which one is glossier).
b) Comparison of the appearance of the same object under different conditions of observation (e.g. move an object from shadow to direct sunlight to assess its translucency).
c) Comparison of the perception of the background through the object or by direct view (e.g. try to read a text through the object and see how much is it distorted to assess transparency).
d) Comparison to memory of familiar objects (e.g. comparison with an appearance of a favourite childhood candy).
e) Comparison to a hypothetical idealistic object or material (e.g. comparison of a glossy object to a perfect mirror).
f) Comparison to a definition (e.g. "gloss, n. — angular selectivity of reflectance, involving surface-reflected light, responsible for the degree to which reflected highlights or images of objects may be seen as superimposed on a surface" [8] - thus, only the surface is analysed, rather than the actual sensation of gloss).

Comparison with a reference is a measurement process. The standardisation of this reference as a unit of measurement is the fundamental aspect of metrology. In order to quantify and communicate visual appearance, subjects need such a reference that will be used for quantification of the appearance. If one does not exist, we have observed that they try to create one themselves. However, the process to come up with a standard is difficult. For instance, a standard for length implies the usage of one unit, and a standard for speed is based on two units (distance and time), while the standard for appearance should regard many components considering the complicated nature of appearance as a phenomenon. Even though the selection of references is very subjective by nature, the process is still conditioned by the physical world. We have observed that people without much training perform surprisingly well on complex tasks that are impossible nowadays for machines and tools [27, 29, 42]. We believe that in case appropriate physical measures and references are used, we should be able to mimic this ability. Even though Eugène [13] argues that "*it is unlikely that any physical scale called "appearance" will be possible*", he admits that "*it is necessary to find physical parameters that can be measured and the most obvious area for exploitation is that described in terms of the optical properties*". References vary depending on the context: comparison can be with a local reference (e.g. with another object), or with a global reference (e.g. the appearance of marble according to the subject's memory); comparison can be with objective things (e.g. definition of blue), as well as subjective ones (e.g. a gummy bear that tastes very good). However, communication of appearance requires generalisation and some objectivity - in most cases, we have a common understanding and agreement on the definition of the words we use to communicate appearance (e.g. "green" refers to a set of colours most of the general populace agree upon with some marginal exceptions, e.g. [43]).

When **global references** are not enough for a given visual task, the Human Visual System (HVS) might use a **local reference**. Simultaneous contrast and dynamic range adaptation are a good demonstration of this. We have observed in our data that the reference is floating, i.e. varying across situations. We believe that this can be a general pattern for material appearance assessment. In other words, the reference could be application-, material-, or situation-specific. We have observed that references have been selected based on the peculiarities of a given scene. When observers were asked to assess the translucency of an object, they usually looked through the object towards the brightest light source (usually the sun), comparing the original appearance with the appearance of the same object under back-lit illumination geometry (back-lit geometry is typically used for measuring "through translucency" [44] or transmission of translucent materials [45]). When the sunlight was not visible observers tended to use an artificial light source of the room instead. Change of reference depending on the illuminance of the artificial light sources has also been observed in [46]. As this was subject to presence of the bright light source, some observers also moved their fingers behind the object comparing the cues between blocked and non-blocked light source conditions. This supports the notion that illumination and room interior, i.e. *Conditions of Observation*, impact *Methodology*, thus reference selection. Back-lit illumination geometry has been already demonstrated to increase the perceived translucency of the materials [23, 25].

Although the HVS is very sensitive, it is not capable of standalone quantitative measurements. Humans can discriminate perhaps 5 to 10 million colours when seen side-by-side [47]. However, when the stimuli are seen with long time intervals, it is difficult to tell the difference, unless the difference is very large - proposedly, our memory stores only around 300 colours [10]. While memory as a global reference has limited capacity, presence of a local reference in a particular point of time, could dramatically enhance the discriminative capabilities of the HVS.

For such a high dimensional problem, probably the reference should not be very different from the target. Deborah [48] addresses the importance of reference selection in the context of spectral differences, considering it an important aspect for a metrological hyperspectral image analysis. The author represents an image as spectral sets falling within a convex hull and argues that if the reference is far outside of the convex hull, the distance to all cluster centres will be nearly identical and discrimination will be poor. Drawing a parallel with appearance, we have observed that a transparent reference medium is a poor measure of apparent translucency differences [49].

Fleming discusses "statistical appearance models" as a potential mechanism for material appearance perception [50]. The author argues that instead of estimating physical properties of materials, our visual system identifies salient features of a given material and creates an internal generative model to estimate how these features behave (i.e. vary across conditions), in order to identify a material in different contexts. The model "*seeks to discover in what ways different material samples look different from one another*", where comparison process and need for a reference seems inevitable. He further argues that our brain tries to characterise systematic changes in the look of materials and the model is "*refined and corrected through experience with other samples*". This process highlights the importance of reference in material perception, and resembles searching for the optimal reference in our data. The author also describes two pivotal forms of material perception: estimation - assessment of potential characteristics, and categorisation – assigning a particular label or material name. Considering his explanation that "*material estimation is the process of establishing the true position of a given sample within the feature space, and material categorization is the process of identifying the boundaries separating different classes of material*", it becomes obvious that neither process is possible without comparison with a reference. Furthermore, material perception as a categorisation process has another interesting aspect - it implies "*access to stored knowledge about other members of the same class*". This phenomenon has been observed in our data and we describe it as a *reference to memory*.

Multisensory impact on appearance

While reference selection and change might imply direct interaction with the object, the interaction can itself provide additional information for appearance assessment, because relying on visual stimuli might still not be enough for material identification, as demonstrated in [51]. We noticed that observers frequently failed to guess the material without touching the object, even though they could move themselves and inspect fixed objects from various viewpoints. Multisensory information, like auditory (knocking objects on the table), tactile information (examining the surface with a finger), or weighting them by hand, have been used to identify material and to describe it [52]. However, it is worth mentioning that after some time, observers demonstrated adaptation, as they got familiarised with the dataset and concluded that the collection is composed of resin materials only.

Choudhury notes that "*although visual perception apparently seems to be independent of human sensation, some properties are perceived in different ways by more than one sense. Individual visual attributes may arise from combination of signals from different senses*" [11]. Limited multisensory interaction in computer graphics might lead to material metamerism and unrealistically large constancy of appearance attributes [52]. This supports our idea that physical objects are important for studying appearance. While we have observed in our data that multisensory information facilitates material identification, neither of the following is clear: whether material identification impacts the perception of the appearance, or whether auditory or tactile information impacts visual appearance. For instance, does the object identified as glass look glossier because this is a typical look for glassy objects? Or if we feel with our finger that the surface of a material is smooth, will it look glossier? It has been shown that priors and expectations regarding familiar-looking materials might actually impact the perception of various mechanical and optical properties of materials [53]. To what extent this applies to visual appearance attributes definitely deserves further study.

Semantic aspects

Analysis of the semantic description has also revealed interesting trends. In [41] we have introduced a hierarchy of the criteria used to assess appearance similarity. Interestingly, it resembles to the vocabulary used for semantic description of the appearance of the objects. The observers have taken different approaches for semantic description that could be diversified into several categories either by tactics, scale, or semantics of the description.

Tactics: 1. Material identification (e.g. amber, ice, silicate, glass, plastic) 2. Attribute-based (glossy, blue, transparent) 3. Familiar object and function identification (e.g. soap, fortune-telling crystal ball, souvenir sold in shops, eraser) 4. Any combination of the previous.

Scale: 1. Absolute (describe just the object) 2. Relative (glossier than this; rougher than that surface).

Semantically: 1. Description as quantification of appearance attributes - the same routine for all objects, e.g. "this object is blue and somewhat glossy". 2. Description as a creative process (comparison with unusual stuff like sorcery; analysing and describing impact of artefacts on caustic formation; conveying appearance with emotions, like "this looks boring").

All these approaches to semantic description involve comparison with various references. It is worth noting that selecting the attributes to communicate the appearance might be dependent on the similarity or dissimilarity within the corpus. For example, when the shape of all objects under question was identical, shape was mentioned less frequently in semantic description than in the cases, where observers had to describe objects with different shapes.

Formulation of the research hypotheses

While the above discussion refers to our data only, the model and the observations might be general to some extent. We formulate 20 research hypotheses (**H1-H20** in the rest of the paper), which, if validated quantitatively, can help us to understand the generality and the limits of our model. The verification of the hypotheses is usually based on quantitative experiments. Some related experiments are already reported in the literature and we use this literature to have a critical reading on those hypotheses. We want to make clear that the verification of the hypotheses do not challenge the existence of the qualitative model, since this is a model of the collected data.

Reference

H1: It is possible to measure and predict perceived appearance. There should be reference(s) and comparison protocol(s), presumably specific to a given material and conditions, that permit objective instrumental measurement of perceived appearance. The critical challenge is to discover these references and comparison protocols.

H2: Human subjects limit one comparison to a single reference at a discrete point of time in appearance assessment process. We have observed that oftentimes, ranking, clustering and ordering visual tasks were broken down into several pair-comparison tasks. For instance, when a subject was asked to rank objects by glossiness, they compared a given object with other objects individually, one by one.

H3: A general appearance ordering system (empirical) cannot exist in sensibly low dimensions. It should be either application specific, local, or most probably unintelligibly high dimensional. If such system would ever exist, it will be strongly non-uniform by nature. There have been several studies in context of material appearance, where *n* manually selected attributes, i.e. features, have been quantified psychophysically to learn how materials relate with one another in a given *n-dimensional* feature space [28, 52, 54]. However, it is observed in [41] that a manually defined system often fails to accommodate new out-of-the-corpus objects

Conditions of observation

H4: Multisensory information and interaction level impact the robustness of appearance constancy. On multiple occasions we observed multisensory impact on visual assessment. Although visual information is unarguably essential to visual appearance, the role of other senses is yet to be understood. It has been shown that different senses, such as visual, tactile and olfactory impact each other in aesthetics impression [55], object recognition [56], material identification [57-58] and material perception [59]. However, the exact way multisensory information contributes to visual appearance is not understood yet.

Object

H5: Shape difference can dramatically impact appearance difference even for identical materials. This observation is consistent with the state-of-the-art. Vangorp *et al.* [60] illustrated that difference in shape, particularly tesselated geometry, diminishes material matching accuracy and comparison is easier between identical shapes. It also impacts perceived translucency differences [49]. As perceptual attributes, such as gloss [61-63] or lightness [64] vary across shapes, it is no surprise that total appearance is also impacted.

H6: Confusion between subsurface and surface scattering might lead to equivalent appearance through different physical material properties. We believe this point boils down to the question whether the HVS can separate contributions of surface and subsurface scattering to the image information. If this is not the case, it could support our proposal that **translucency impacts gloss perception**. We think the confusion can be minimal for gloss if a sharp image of the environment is reflected from the surface, which is subject to presence of well-structured real-world illumination [65]. However, the orientation of the reflected image can also cause confusion between transmission and reflection phenomena [66].

Translucency perception

H7: The amount of transmitted light and preservation of the light structure after transmission are independent, but core dimensions for translucency assessment. From the perspective of hard metrology, this observation can be related to concepts such as, *direct*, *diffuse* and *total transmittance*, as well as *clarity* and *haze* [9, 45]. However, perceptual dimensions of translucency are yet to be understood. In a translucency classification system proposed by Gerardin et al. [67] independent orthogonal dimensions of diffusion and absorption are roughly equivalent to these quantities. However, the authors argue that increasing scattering (i.e. diminishing light structure preservation) makes transparent material to some extent translucent and finally opaque; while increasing absorption (i.e. amount of light) does not cause translucency and ranges from transparency to opacity without translucency in between. This is contradictory to some of our observations that people consider absorbing objects less translucent, even in case of identical scattering properties. We have observed that **the assessment procedure of perceptual translucency difference depends on the subjective interpretation of the term and needs to be standardised**.

H8: A given material looks more translucent when an object made of it has thin parts. This phenomenon is illustrated in Figure 4. The observers considering objects with thin-parts more translucent, instead of referring to low level image cues, explicitly mention that they understand and see that the light is being transmitted through the object. This can be an indication that Fleming and Bülthoff's [25] conclusion that the HVS does not invert optics to assess translucency might not hold for thin objects. In general, shorter the distance a photon needs to travel through a medium, easier to detect light transmission. Scale and thickness of the object impact perceived translucency and thin parts, such as edges, are usually informative translucency cues [17, 25]. In addition, thin parts, such as fine surface details and bumps, might blur the background image and make transparent materials appear translucent (Figure 5). Therefore, this hypothesis can be reformulated as a more general statement that **object shape and size impact perceived translucency of the material**.

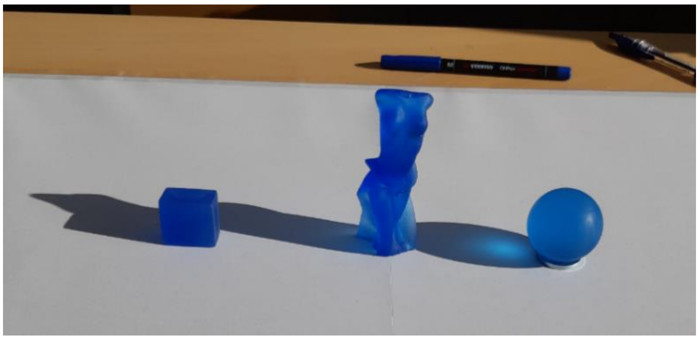

Figure 4: ***Three Blue Objects Used in the Experiment.*** *The cuboid and the female sculpture have equal density of the blue colorants, while the sphere has less blue colorants in the volume. On the other hand, the surface coarseness of the sphere and the sculpture is identical, while the cuboid has rougher surface than the other two. Combination of the two factors, led the vast majority of the observers to consider the cuboid least translucent. On the other hand, there was no statistically significant difference in apparent translucency of the sphere and the female sculpture, despite higher density of the colorants in the latter. This can be explained with the fact that a sphere has a dense shape, while the sculpture has thin parts letting the light through.*

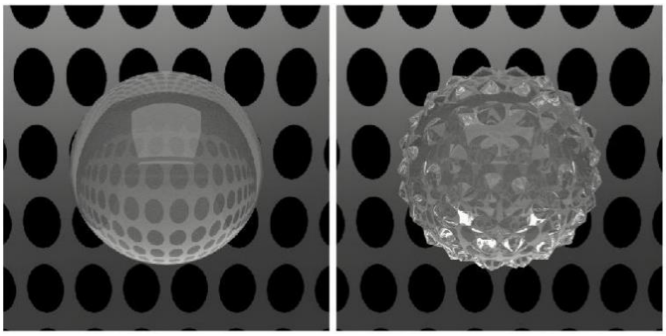

Figure 5: ***Same Material, Different Transparency.*** *Although the material is identical in both objects, meso-scale geometry of the right objects removes see-through cues impacting perceived transparency and translucency of the material and object. The images have been reproduced from [49]. Reprinted with permission of IS&T: The Society for Imaging Science and Technology sole copyright owners of, "CIC27: Twenty-seventh Color and Imaging Conference 2019".*

H9: Back-lit is a preferred lighting geometry for translucency assessment. We have observed that observers tend to locate the illumination source in the scene (typically the sun in our context) and look towards it through the object to assess translucency. One interpretation of this behaviour can be a potential attempt to invert optics and observe transmission. Xiao *et al.* [23] have shown that materials typically look more translucent when they are back-lit. The magnitude of difference between translucent and opaque objects is expected to be larger in this condition and moving them from front- to backlight has stronger impact on translucent objects' appearance, as translucent objects, unlike opaque ones, start to shine or glow on the backlight. This is related to the above-discussed notion of comparison with a reference. A typical reference can be the appearance of the same object under different illumination conditions. On the other hand, it is worth mentioning that transparent objects might look less transparent on a high-illuminance backlight, as observers do not see the scene through the object due to the limited dynamic range of the HVS [46].

H10: Dynamic and heterogeneous backgrounds enhance perceived translucency or transparency. We have observed that human observers frequently use object and background relative motion to estimate light transmission properties of a material. This implies both - moving an object over a heterogeneous background, e.g. checkerboard, as well as moving background objects behind a static object, e.g. moving one's own fingers or a pen behind the object. While in a static scene the HVS has a reduced ability to separate reflection and transmission components of the visual stimulus, human subjects try to observe and estimate the magnitude of the changes induced by the background change. Commercial measurement systems measure transmission from a static point perspective (e.g. ISO 13468 for plastics [68]) limiting the capability of measured quantities to adequately describe visual sensation in real life encounters.

H11: Lightness impacts perceived translucency (lighter objects look more translucent). Many translucent materials, such as snow, cream, milk, wax and soap, are typically light-coloured and have diffusive, hazy appearance usually described by observers as "milky". Therefore, "milkiness" of light-coloured objects might be the cause for perceived translucency (refer to Figure 6). Lightness has been shown to be correlated with luminance [69,70]. Subsurface scattering can contribute to luminance and highly scattering media usually look lighter. However, lightness information alone cannot be discriminative enough for assessing translucency. Marlow *et al.* [71] demonstrated that if luminance gradients co-vary with surface geometry, surface looks opaque, while if luminance information seems independent from surface geometry, perception of subsurface scattering is evoked. This indicates that in addition to lightness, interpretation of the 3D shape is also involved.

*Figure 6: **"Milky" Translucent-looking Objects.** With their light and "milky" appearance, the objects evoke perception of translucency in some human observers.*

H12: Glossiness impacts translucency perception. Some of our observers considered glossy objects more translucent. It has been shown that gloss enhances perception of translucency [72] and realism of translucency appearance (refer to Figure 8 in [25]), proposedly because many translucent materials we interact with on a daily basis are glossy and "*the human visual system may "expect" translucent materials to exhibit specular reflections*"[25]. Hence, contribution of gloss to translucency perception might come down to the material identification problem. Schmid *et al.* [73] propose that neural aspects of gloss perception should be addressed in the context of material identification. However, the role of material association should be taken with care. Some materials (e.g. glass) appear glossy and translucent, but others (e.g. metals) can be glossy and opaque [28, 54].

H13: Presence of caustics is a cue to assess translucency and may increase perceived degree of translucency. We noticed that caustics were often used as a cue for translucency and transparency assessment by the observers, and in some scenes, might be the sole cue to translucency of the material, as illustrated in Figure 7. Caustic pattern projected by an object onto a different surface contains interesting information regarding its properties (refer to the top image in Figure 8). It was shown that when the floor and the caustic pattern projected onto it are removed, the material is judged less translucent [74].

Gloss perception

H14: Translucency impacts the perceived glossiness of an object. We observed that gloss-based ranking has been possible for the objects with identical surface reflectance but different translucency. It has been demonstrated that translucency can impact gloss and the magnitude of this impact depends on the shape and surface roughness of the object [75]. Translucent objects with complex shape might produce highlights that originate from inside the medium - like, internal reflections, scattering and caustics. Considering the limit of the dynamic range perceived by the HVS, these highlights might be mistaken for specular reflections evoking glossiness perception [32], as shown in Figure 8. Objects can look very glassy and glossy due to internal reflections and caustics even if specular reflections are negligible (refer to Figure 8 in [51]). Additionally, Pellacini *et al.* [76] have shown that contrast between specular and non-specular regions is an important factor for gloss "*light colored surfaces appearing less glossy than dark ones having the same finish*". The amount of subsurface scattering can affect lightness of the non-specular regions, while having little impact on specular ones. Hence, for some shapes, they can modulate contrast gloss of translucent objects [75].

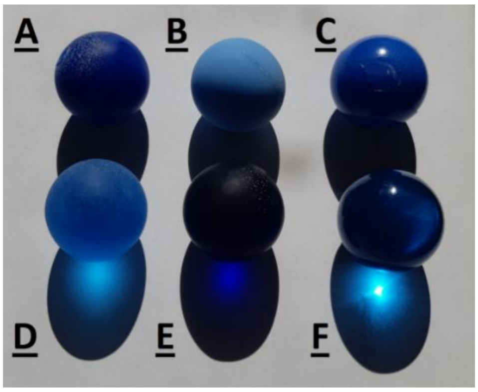

Figure 7: ***Translucency and Caustics.*** *Caustic pattern might provide information regarding color and light transmission properties of the material. For object E, it is the sole cue that makes us deduce the material is translucent. The figure has been reproduced from [46]. Reprinted with permission of IS&T: The Society for Imaging Science and Technology sole copyright owners of, "CIC27: Twenty-seventh Color and Imaging Conference 2019".*

Figure 8: ***Objects Used in Gloss Ranking Experiments.*** *We identified three groups of people: those who tied all spheres (top image) due to similarity in surface coarseness (35.29% of the observers); those who considered translucent objects more glossy, because of higher luminance and "shininess" (35.29%); and those who considered opaque ones glossier due to higher contrast and more visible distinctness-of-image gloss on them (29.42%). In the follow-up experiment with female sculptures (bottom image) the majority of the observers (78.50%) stated that the transparent ones were glossier. [32] The complex macro-geometry of the surface made it impossible to observe distinctness-of-image gloss, while these objects produced complex caustic patterns that could be mistaken for specular reflections. The top image has been reproduced from [40]. Reprinted with permission of IS&T: The Society for Imaging Science and Technology sole copyright owners of, "CIC26: Twenty-sixth Color and Imaging Conference 2018".*

H15: Complex shape makes materials look glossier. Some observers noted that a complex bust figure looked glossier than a sphere and a cube, because it shines more and has more specular regions. The state-of-the-art shows that shape can considerably impact gloss perception, even if surface reflectance is identical. It has been shown that surface reflectance constancy of the HVS fails across shapes [22] and perceived gloss is correlated with perceived surface bumpiness [62-63, 77]. However, we see two challenges that need to be addressed:

- What is the threshold between shape change and surface change? What scale do we mean with the hypothesis mentioned earlier? Can we really change a shape without changing a surface, and if so, to what extent can we change shape not to impact the surface?
- All shape changes are due to a manipulation of a controlled parameter (e.g. RMS height deviation). Can we have a shape descriptor statistic that could predict the glossiness of a given material for any random shape?

H16: Motion facilitates gloss perception. We have observed that motion was widely used for glossiness estimation by the observers. They either moved their head or moved the objects to monitor the motion of the highlights. This is consistent with the state-of-the-art. Impact of head motion has been already observed to be important for gloss, as "temporal changes of the retinal image caused by the observer's head motion" and "image differences between the two eyes in stereo viewing" both significantly increase perceived gloss [78]. Motion seemingly helps the HVS distinguish specular reflections and surface texture. Unlike texture, specular reflections remain static relative to the observer on rotating spheres [79] and "objects with normal specular motion to appear shinier than those with sticky reflections" [80]. Motion improves gloss constancy [80] and can even increase the magnitude of perceived gloss [81].

Opacity perception

H17: Opacity does not imply a complete absence of transmission. We have observed that some objects manifesting translucency cues when exposed to high illuminance directional backlight were considered opaque under diffuse and low intensity illumination. While perceived opacity is proposedly impacted by the amount of transmitted light, the latter itself depends on the amount of light incident on the back side of the object. The amount of transmission tolerated for classifying the object opaque varied across observers. We concluded that opacity perception or more likely the interpretation of the concept depends on the thresholds that are floating and subjective by nature. The same trend was observed in [46]. Moreover, Marlow *et al.* [71] argue that the HVS relies on the co-variance between shading and surface orientation for distinction between translucent and opaque objects. They demonstrated that optically translucent object might look opaque "*if the light transported through the material accidentally preserves the co-variation of intensity and surface orientation*", as if it was a result of reflection rather than transmission which again supports our hypothesis that opacity can be perceived even if subsurface scattering event occurs.

Appearance attributes and subjective material properties

H18: Glossy objects look more fragile and precious. Glossy objects with the complex shape have been described as fragile, expensive and precious. Our observations are partially consistent with the state-of-the-art. Fujisaki *et al.* [82] found that for wooden materials gloss and expensiveness are positively correlated. Contrasting results have been reported on the correlation between gloss and fragility, which was either positive [28] or negative [54] on different occasions. Additional role can be played with the positive correlation between glossiness and prettiness [52, 54], although some authors found no significant correlation between the two [28, 82]. We believe material identification is also an important factor, as metal, glass, and plastic can all be very glossy, they are not necessarily perceived equally fragile, neither equally precious. Material recognition and semantic interpretation of objects' function have been major contributing factors to subjective perceptual qualities in our experiment. Although observers, by visual inspection, described glossy bust figures as glass or precious stone decorations "found in a fancy store" (per contra, spheres have been described as an "ice ball", "candy", or a "billiard ball"), the auditory and tactile information made them revise their descriptions ("ah, this sounds like a cheap plastic" noted an observer after knocking the figure on the table).

H19: Darker objects look heavier. This phenomenon is correlated with *brightness-weight illusion* meaning that when lifted, a light-coloured object feels heavier than a darker object of the same mass, because of the anticipation that darker objects are generally heavier [83]. Bullough [84] demonstrated that darker-coloured objects are perceived heavier, proposing an explanation that darker colours evoke a perception of "*more of it*", potentially referring to "more pigments". Interestingly, our observers provided similar justification. This finding has been supported by numerous studies [85-87]. Another intriguing explanation is that in English the same adjective *light* is used to describe both properties - low weight and high brightness [85].

Artefacts

H20: Complex surface geometry can mask imperfections and artefacts. We have observed that scratches, bubbles and other imperfections were mentioned more often when describing spheres and cuboids, and rarely for a complex bust shape. Considering that the retinal image is actually a 2D projection of the 3D object, we believe this phenomenon is related to the concept of visual masking in image quality, when noise is more apparent in homogeneous parts of the image, while it gets masked in high frequency areas [88].

Conclusions

While the vast majority of appearance studies focus on either instrumental measurement or psychophysics, we analysed material appearance from a social science perspective. We propose that appearance is a social interaction that implies communication. We have conducted interviews where people were asked to perform visual tasks on objects of different appearances, describe the objects, explain their actions and interact with the interviewer and the objects. Those interviews were videotaped. This large collection of data was analysed with the Grounded Theory Analysis and we constructed a model to have a structured representation of the observations. This qualitative model and its implications were described in the corresponding section. We conducted an analytical survey of the literature in the perspective of this model, and formalised future research hypotheses. In particular, we found that selecting a reference and the comparison with this reference have been the essential instruments for appearance assessment and communication in our scenario. In this work we addressed the appearance of objects, which have context, rather than the appearance of abstract materials.

Our results are to be taken with care because no level of generalisation can be assumed or stated from the specific research methodology we used. Indeed, we used an inductive research method, while deductive research methods are more common in the study of appearance. The observations are limited to the conducted experiment, but when we compared our work with the state of the art, we found encouraging echoes.

Further quantitative verification of the hypotheses is a straightforward follow up of this work. Psychophysical experimental design might also benefit from our behavioural observations on natural ways of object appearance assessment. For instance, the use of extended reality technologies might permit more freedom in future experimental processes.

References

1. Brunton A, Arikan CA, Tanksale TM and Urban P (2018), 3D printing spatially varying color and translucency, *ACM Transactions on Graphics (TOG)*, **37** (4), 157:1-13.
2. Totten GE and Liang H (2004), *Surface Modification and Mechanisms: Friction, Stress, and Reaction Engineering*, CRC Press.
3. Rong MZ, Zhang M and Ruan W (2006). Surface modification of nanoscale fillers for improving properties of polymer nanocomposites: a review, *Materials Science and Technology*, **22** (7), 787-796.
4. Dorsey J, Rushmeier H and Sillion F (2010), *Digital Modeling of Material Appearance*, Elsevier.
5. McCarthy W (2000), Programmable matter: How quantum wellstone ushered in our modern age, *Nature*, **407** (6804), 569.
6. Campbell TA, Tibbits S and Garrett B (2014), The programmable world, *Scientific American*, **311** (5), 60-65.
7. Campbell TA, Tibbits S and Garrett B (2014), The next wave: 4D printing programming the material world, *Technical Report*, Atlantic Council, Washington DC (USA).
8. ASTM (2017), *ASTM E284-17 Standard Terminology of Appearance*, ASTM International, West Conshohocken PA (USA) – https://doi.org/10.1520/E0284-17 – last accessed 11 June 2021.
9. CIE (2006), *CIE 175:2006 A framework for the measurement of visual appearance*, International Commission on Illumination.
10. Hunter RS and Harold RW (1987), *The Measurement of Appearance*, John Wiley & Sons.
11. Choudhury AKR (2014), *Principles of Colour and Appearance Measurement: Object Appearance, Colour Perception and Instrumental Measurement*, Elsevier.
12. Ishimaru A (1991), Wave propagation and scattering in random media and rough surfaces, *Proceedings of the IEEE*, **79** (10), 1359-1366.
13. Eugène C (2008), Measurement of "total visual appearance": a CIE challenge of soft metrology, *Proceedings of the 12th IMEKO TC1 TC7 Joint Symposium on Man, Science Measurement*, 61-65, Annecy (France).
14. Hutchings J (1995), The continuity of colour, design, art, and science. I. The philosophy of the total appearance concept and image measurement, *Color Research and Application*, **20** (5), 296-306.
15. Hutchings J (1995), The continuity of colour, design, art, and science. II. Application of the total appearance concept to image creation, *Color Research and Application*, **20** (5), 307-312.
16. Hutchings JB (1999), *Food Color and Appearance*, Second Edition, New York: Aspen Publishers.
17. Urban P, Tanksale TM, Brunton A, Vu BM and Nakauchi S (2019), Redefining A in RGBA: Towards a standard for graphical 3D printing, *ACM Transactions on Graphics (TOG)*, **38** (3), 21:1-14.
18. Hunter RS (1937), Methods of determining gloss, *NBS Research Paper RP*, **958**, 19-39.
19. Motoyoshi I (2010), Highlight–shading relationship as a cue for the perception of translucent and transparent materials, *Journal of Vision*, **10** (9), 6:1-11.
20. Motoyoshi I, Nishida S, Sharan L and Adelson EH (2007) Image statistics and the perception of surface qualities, *Nature*, **447** (7141), 206-209.
21. Nicodemus FE (1965), Directional reflectance and emissivity of an opaque surface, *Applied Optics*, **4** (7), 767-775.
22. Nishida S and Shinya M (1998), Use of image-based information in judgments of surface-reflectance properties, *Journal of the Optical Society of America A*, **15** (12), 2951-2965.
23. Xiao B, Walter B, Gkioulekas I, Zickler T, Adelson E and Bala K (2014), Looking against the light: How perception of translucency depends on lighting direction, *Journal of Vision*, **14** (3), 17:1-22.

24. Chowdhury NS, Marlow PJ and Kim J (2017), Translucency and the perception of shape, *Journal of Vision*, **17** (3), 17:1-14.
25. Fleming RW and Bulthoff HH (2005), Low-level image cues in the perception of translucent materials, *ACM Transactions on Applied Perception (TAP)*, **2** (3), 346-382.
26. Tanaka M, Nakayama D and Horiuchi T (2020), Analysis of factors affecting the contrast effect for total appearance, *Journal of the International Colour Association*, **25**, 1-11.
27. Sharan L, Rosenholtz R and Adelson EH (2014), Accuracy and speed of material categorization in real-world images, *Journal of Vision*, **14** (9), 12:1-24.
28. Fleming RW, Wiebel C and Gegenfurtner K (2013), Perceptual qualities and material classes, *Journal of Vision*, **13** (8), 9:1-20.
29. Sharan L, Liu C, Rosenholtz R and Adelson EH (2013), Recognizing materials using perceptually inspired features, *International Journal of Computer Vision*, **103** (3), 348-371.
30. Vu BM, Urban P, Tanksale TM and Nakauchi S (2016), Visual perception of 3D printed translucent objects, *Proceedings of the Twenty-fourth Color and Imaging Conference*, 94-99, San Diego (USA).
31. Smet K, Ryckaert WR, Pointer MR, Deconinck G and Hanselaer P (2011), Colour appearance rating of familiar real objects, *Color Research and Application*, **36** (3), 192-200.
32. Gigilashvili D, Thomas JB, Pedersen M and Hardeberg JY (2019), Perceived glossiness: Beyond surface properties, *Proceedings of the Twenty-seventh Color and Imaging Conference*, 37-42, Paris (France).
33. Paille P (1994), L'analyse par théorisation ancrée, *Cahiers de recherche sociologique*, **1** (23), 147-181.
34. Glaser B and Strauss A (1967), *The Discovery of Grounded Theory: Strategies for Qualitative Research*, Chicago: Aldine Publishing.
35. Corbin J and Strauss A (1998), *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, Thousand Oaks, CA: Sage publications.
36. Jacob JD and Holmes D (2011), Working under threat: Fear and nurse–patient interactions in a forensic psychiatric setting, *Journal of Forensic Nursing*, **7** (2), 68-77.
37. Gaucher N and Payot A (2011), From powerlessness to empowerment: Mothers expect more than information from the prenatal consultation for preterm labour, *Paediatrics Child Health*, **12** (16), 10:638-642.
38. Rippon D, McDonnell A, Smith M, McCreadie M and Wetherell M (2020), A grounded theory study on work related stress in professionals who provide health social care for people who exhibit behaviours that challenge, *PLOS ONE*, **15** (2), e0229706.
39. Thomas JB, Deniel A and Hardeberg JY, The Plastique collection: A set of resin objects for material appearance research, *Proceedings of the XIV Conferenza del Colore*, 13-24, Florence (Italy).
40. Gigilashvili D, Thomas JB, Hardeberg JY and Pedersen M (2018), Behavioral investigation of visual appearance assessment, *Proceedings of the Twenty-sixth Color and Imaging Conference*, 294-299, Vancouver (Canada).
41. Gigilashvili D, Thomas JB, Pedersen M and Hardeberg JY (2019), Material appearance: ordering and clustering, *Proceedings of the IS&T International Symposium on Electronic Imaging: Material Appearance*, 202-1-202-7, Burlingame (USA).
42. Sharan L, Rosenholtz R and Adelson E (2009), Material perception: What can you see in a brief glance? [Abstract], *Journal of Vision*, **9** (8), 784.
43. Koren M (2018), What color Is a tennis ball? An investigation into a surprisingly divisive question, *The Atlantic* – https://www.theatlantic.com/science/archive/2018/02/what-color-tennis-ball-green-yellow/523521/ – last accessed 23 February 2020.
44. Bucknell SP, Winsey NJP and Gale DR (2000), Translucency measurement, *Google Patents*, US Patent 6,111,653.
45. Tirpak A and Young R (2008), Accurate transmission measurements of translucent materials - The light scattering of translucent materials means that two sources are needed for the best transmission measurements, *Photonics Spectra*, **42** (2), 1-6.
46. Gigilashvili D, Mirjalili F and Hardeberg JY (2019), Illuminance Impacts Opacity Perception of Textile Materials, *Proceedings of the Twenty-seventh Color and Imaging Conference*, 126-131, Paris (France).

47. Thomas JB, Colantoni P and Trémeau A (2013), On the uniform sampling of CIELAB color space and the number of discernible colors, in *Computational Color Imaging*, Tominaga S, Schettini R and Trémeau A (eds), 53-67, Heidelberg: Springer Berlin Heidelberg.

48. Deborah, H (2016), Towards spectral mathematical morphology, *PhD Thesis*, NTNU - Université de Poitiers.

49. Gigilashvili D, Urban P, Thomas JB, Hardeberg JY and Pedersen M (2019), Impact of shape on apparent translucency differences, *Proceedings of the Twenty-seventh Color and Imaging Conference*, 132-137, Paris (France).

50. Fleming RW (2014), Visual perception of materials and their properties, *Vision Research*, **94**, 62-75.

51. Todd JT and Norman JF (2019), Reflections on glass, *Journal of Vision*, **19** (4), 26:1-21.

52. Filip J, Kolafová M, Havlíček M, Vávra R, Haindl M and Rushmeier H (2018), Evaluating physical and rendered material appearance, *The Visual* Computer, **34**, 805-816.

53. Alley LM, Schmid AC and Doerschner K (2019), Expectations affect the perception of material properties, *Journal of Vision*, **20** (12), 1:1-20.

54. Tanaka M and Horiuchi T (2015), Investigating perceptual qualities of static surface appearance using real materials and displayed images, *Vision Research*, **115**, 246-258.

55. Fiore AM (1993), Multisensory integration of visual, tactile, and olfactory aesthetic cues of appearance, *Clothing and Textiles Research Journal*, **11** (2), 45-52.

56. Liu H, Yu Y, Sun F and Gu J (2016), Visual–tactile fusion for object recognition, *IEEE Transactions on Automation Science and Engineering*, **14** (2), 996-1008.

57. Liu H and Sun F (2017), Material identification using tactile perception: A semantics-regularized dictionary learning method, *IEEE/ASME Transactions on Mechatronics*, **23** (3), 1050-1058.

58. Fujisaki W, Goda N, Motoyoshi I, Komatsu H and Nishida S (2014), Audiovisual integration in the human perception of materials, *Journal of Vision*, **14** (4), 12:1-20.

59. Wongsriruksa S, Howes P, Conreen M and Miodownik M (2012), The use of physical property data to predict the touch perception of materials, *Materials and Design*, **42**, 238-244.

60. Vangorp P, Laurijssen J and Dutre P (2007), The influence of shape on the perception of material reflectance, *ACM Transactions on Graphics (TOG)*, **26**, 77:1-77:10.

61. Olkkonen M and Brainard DH (2011), Joint effects of illumination geometry and object shape in the perception of surface reflectance, *i-Perception*, **2** (9), 1014-1034.

62. Ho YX, Landy MS and Maloney LT (2008), Conjoint measurement of gloss and surface texture. Psychological Science, **19** (2), 196-204.

63. Marlow PJ, Kim J and Anderson BL (2012), The perception and misperception of specular surface reflectance, *Current Biology*, **22** (20), 1909-1913.

64. Schmid AC and Anderson BL (2014), Do surface reflectance properties and 3-D mesostructure influence the perception of lightness?, *Journal of Vision*, **14** (8), 24:1-24.

65. Fleming RW, Dror RO and Adelson EH (2003), Real-world illumination and the perception of surface reflectance properties, *Journal of Vision*, **3**, 347-368.

66. Kim J and Marlow PJ (2016), Turning the world upside down to understand perceived transparency, *i-Perception*, **7** (5), 1-5.

67. Gerardin M, Simonot L, Farrugia JP, Iehl JC, Fournel T and Hébert M (2019), A translucency classification for computer graphics, *Proceedings of the IS&T International Symposium on Electronic Imaging: Material Appearance*, 203-1-203-6, Burlingame (USA).

68. ISO (2019), *Plastics — Determination of the Total Luminous Transmittance of Transparent Materials — Part 1: Single-beam Instrument*. Geneva, CH: International Organization for Standardization.

69. Gilchrist A, Kossyfidis C, Bonato F, Agostini T, Cataliotti J, Li X, Spehar B, Annan V and Economou E (1999), An anchoring theory of lightness perception, *Psychological Review*, **106** (4), 795-834.

70. Land EH and McCann JJ (1971), Lightness and retinex theory, *Journal of the Optical Society of America*, **61** (1), 1-11.

71. Marlow PJ, Kim J and Anderson BL (2017), Perception and misperception of surface opacity, *Proceedings of the National Academy of Sciences*, **114** (52), 13840-13845.

72. Yu H, Liu PX and Hu L (2019), A highlight-generation method for rendering translucent objects, *Sensors*, **19** (4), 860:1–860:15.

73. Schmid AC, Barla P and Doerschner K (2020), Material category determined by specular reflection structure mediates the processing of image features for perceived gloss, Preprint Version, *bioRxiv*, 1-45.

74. Gigilashvili D, Dubouchet L, Pedersen M and Hardeberg JY (2020), Caustics and translucency perception, *Proceedings of the IS&T International Symposium on Electronic Imaging: Material Appearance*, 003-1-003-6, Burlingame (USA).

75. Gigilashvili D, Shi W, Wang Z, Pedersen M, Hardeberg JY and Rushmeier H (2021), The role of subsurface scattering in glossiness perception, *ACM Transactions on Applied Perception*, **18** (3), 10:1-26.

76. Pellacini F, Ferwerda JA and Greenberg DP (2000), Toward a psychophysically-based light reflection model for image synthesis, *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, 55-64, New Orleans (USA).

77. Qi L, Chantler MJ, Siebert JP and Dong J (2014), Why do rough surfaces appear glossy?, *Journal of the Optical Society of America A*, **31** (5), 935-943.

78. Sakano Y and Ando H (2010), Effects of head motion and stereo viewing on perceived glossiness, *Journal of Vision*, **10** (9), 15:1-14.

79. Chadwick AC and Kentridge R (2015), The perception of gloss: A review, *Vision Research*, **109**, 221-235.

80. Doerschner K, Fleming RW, Yilmaz O, Schrater PR, Hartung B and Kersten D (2011), Visual motion and the perception of surface material, *Current Biology*, **21** (23), 2010-2016.

81. Wendt G and Faul F (2018), Can color and motion information be used to disentangle the influence of multiple light sources on gloss perception?, *i-Perception*, **9** (5), 1-26.

82. Fujisaki W, Tokita M and Kariya K (2015), Perception of the material properties of wood based on vision, audition, and touch, *Vision Research*, **109**, 185-200.

83. De Camp JE (1917), The influence of color on apparent weight: A preliminary study, *Journal of Experimental Psychology*, **2** (5), 347- 370.

84. Bullough E (1907), On the apparent heaviness of colours: A contribution to the aesthetics of colour, *British Journal of Psychology*, **2** (2), 111-152.

85. Walker P, Francis BJ and Walker L (2010), The brightness-weight illusion, *Experimental Psychology*, **57** (6), 462-469.

86. Walker P (2012), Cross-sensory correspondences and naive conceptions of natural phenomena, *Perception*, **41** (5), 620-622.

87. Hagtvedt H (2014), Dark is durable, light is convenient: Color value influences perceived product attributes, *ACR North American Advances*, **42**, 27-31.

88. Chandler DM (2013), Seven challenges in image quality assessment: past, present, and future research, *ISRN Signal Processing*, Article ID 905685, 1-53.

Appendix 1

Different ways to display stimuli in appearance research

There are three ways to generate the visual stimuli: direct view to the real physical objects, photographing the real objects, and using computer graphics to generate synthetic images. However, the ways to present them to the observer are two: either present the object directly, or to display it through an intermediate medium - e.g. computer display or VR headset. By presenting the stimulus on an intermediary display the dimensionality of the stimulus reduces (e.g. from infinite dimensions in a natural scene to 5D in 2D displayed colour image). Therefore, the way of stimuli introduction should be carefully chosen. The advantages and disadvantages of different methods for displaying the stimuli are summarised in Table A1.

Table A1: Advantages and disadvantages of using tangible and displayed stimuli.

	Advantages	**Disadvantages**
Physical Objects	• Subjects can freely interact with the physical objects - i.e. possibility to apply all behavioural patterns we use in our daily lives for appearance assessment (move head, move object). • Multisensory information is present (e.g. tactile, auditory). • Binocular vision. • Realistic environment. • Artefacts make objects realistic. • In the real world we have access to full scene context that is often not possible in graphics.	• Difficult to model, measure, and replicate. • High cost of manufacturing. • Unpredictable effects of aging. • Unwanted artefacts. • Risk of damaging. • Limited access across the scientific community. • Limited reproducibility of the experiments (due to access, aging).
Displayed Images	• Full control of the material parameters (e.g. phase function, absorption and scattering) and scene (illumination, background). • Simplicity of manipulation of any material or scene parameters. • Relatively low cost of production/generation. • Better reproducibility. • Easier to share the data across the scientific community. • Realistic photographs can be used. • Free from aging effects.	• Graphic rendering is based on a model that might be limited and might significantly impact result of the experiment. Physically based rendering is extremely time-consuming. • It is very difficult to relate a radiate image and stimuli to the optical model due to digitisation of the information and calibration of the display. If it is relative to display (and full calibration, even though might be reproducible), it is still not correlated to the optical model. • Many factors, like resolution, colour gamut or heterogeneity of the display might impact the results. • Dynamic range of the displays are lower. • Interactivity is limited in computer graphics. • Multisensory information is absent, or extremely limited. • Often no stereo vision is possible. • The environment is often unrealistic in computer graphics (e.g. neutral grey background). • No virtual system replicates fully the complex lighting environments we
Virtual Reality	• All display-related advantages apply to VR as well. • VR might enable binocularity and motion. • More realistic interactivity than in case of displays. • Not affected by the ambient illumination. • Less distraction from the ambience.	

Advantages	Disadvantages
	encounter in real lives, especially characterising directional spectral variation in natural environments. • Lack of imperfections in computer graphics not only reduce naturalness of the stimuli, also undermines robustness of the models built based on them. • While photographs are realistic and superior to synthetic stimuli in several above-mentioned aspects, they do not contain the information regarding the physical material properties, and we are limited to image statistics extraction.

Appendix 2

An example of the observations, with the transcript, the action performed and their interpretation within the model

We introduced 15 categories that unify conceptually similar observations. Afterwards, we also presented the qualitative model that not only shows how the categories relate with one another, but also explains the entire pipeline of the material appearance assessment in context of our tasks. At first glance, it might be ambiguous in what way the videotaped experiment is processed using the Grounded Theory Analysis. In order to illustrate exactly how the model is rooted in the data, below we present a detailed transcript of the 6.5-minute excerpt from the actual experiment where the observer tries to rank five spheres by glossiness (refer to Table A2). The first column shows the time frame (in mm:ss format) the comments in the corresponding row are referring to. The second column contains the speech from a given time frame - either quoted, or paraphrased. The third column describes the actions happening within a given time frame. The fourth column comments the content and explains the process in context of our model.

Table A2: An example task transcript illustrating how the model describes the data.

Time	Speech	Action	Comment
00:00 to 00:20	The experimenter introduces objects to the observer.	The experimenter puts objects in front of the observer.	***Object*** enters the scene under given ***Conditions of Observation***. The ***Experimenter*** starts impacting the process.
00:20 to 00:30		The observer starts inspecting the objects.	The appearance perception is evoked by the combination of two factors: characteristics of the ***Object***, and the ***Conditions of Observation***, like illumination conditions.
00:30 to 00:45	The experimenter says that as the observer has got used to this dataset, he can again describe them by appearance.		The ***Experimenter*** contributes to ***Task Interpretation***. The experimenter means that the observer has already seen similar objects in previous tasks, and ***Learning and Adaptation*** facilitates the process.
00:45 to 01:20	Observer describes: "even without taking them and looking through them towards the sun, which is an usual way for translucency, even without that, I see that this is yellowish and very translucent, these are opaque, opaque I do not know color, bluish and somewhat translucent, orange and very translucent".	Observer moves his head to the sides while examining objects. Points one by one to the caustics of the objects with an index finger, while describing the appearance. The judgement is based solely on the caustic pattern projected onto the table.	The observer has come up with a particular ***Methodology*** (that involves assessment of the caustic pattern). He needs a ***Reference*** for ***Comparison***. In this case, he compares appearance of the two objects between the two observation geometries (when moving the head), where the ***Reference*** is the appearance in normal sitting condition that is compared with the appearance of the same object seen with a head tilted to the side. For ***Semantic Description***, the observer needs ***Vocabulary Search***. His

			professional background in material appearance is a ***Condition of Observation*** that contributes to his ***Methodology*** and ***Vocabulary Search***, coming up with a particular ***Vocabulary*** that is composed of appearance attribute terminology related to colour, and light transmittance properties. When exact word was not found with ***Vocabulary Search***, the ***Comparison*** with the nearest ***Reference*** is used to express uncertainty, like words ”yellowish”, ”bluish”, and ”somewhat translucent”.
01:25 to 01:35	The experimenter asks: ”so, you put them against light, so you can see the shadow in front of you as a colour palette?”. The observer confirms.		The ***Experimenter*** clarifies the ***Task Interpretation*** and selected ***Methodology***.
01:35 to 01:47	The observer continues description: ”well, they are pretty glossy. No texture, they have all spherical shape”.	The observer moves his head to the sides, looks from the top to observe the image in the reflections.	The ***Vocabulary*** is still strongly impacted by the ***Conditions of Observation*** - the background of the observer, and the illumination conditions in the room. The observer continues using ***Comparison*** between two observation geometries.
01:47 to 02:28	The observer continues description: ”I see some kind of artifacts. Here the scratches are deeper. This one has more severe artifacts. Apart from artifacts, they are all glossy, those three are translucent, those two are opaque. They differ in colour, yellowish, this is kind of yellow too, orange, dark blue, light blue. ”	Observer picks one object and looks closely. Then picks the next one.	As the time passes, ***Learning and Adaptation*** helps the observer to include more details in the ***Semantic Description***.
02:28 to 02:45	The experimenter introduces the visual task: ”now I will ask you a very specific task. Rank them by glossiness again. As you said, they are very glossy, so it might be more difficult. ”		The ***Task*** is presented. ***Experimenter*** conveys the message and the observer starts ***Task Interpretation***.
02:45 to 02:55	”That’s true” - the observer admits the task is difficult.	The observer picks two objects up, and looks at them from the side, holding them next to each other.	The observer has ***Structure Expectation***. The task is considered ”difficult”, because the observer assumes the ranking should be possible and there is the ”right answer”, even though

			all objects look "very glossy". This impacts the rest of ***Task Interpretation***. After ***Task Interpretation***, the observer has taken his time for ***Decision-Making*** and came up with a ***Methodology*** (that will be refined over time due to ***Learning and Adaptation***). The observer clearly needs a ***Reference*** for ***Comparison*** to quantify appearance of a particular object. So, he picks two objects and compares them against each other.
02:55 to 03:16	Experimenter gives further instructions: "one thing you could consider is artifacts, if you can't find any other difference; but, first of all, I want to ask you to classify without taking them into account."		This is a pure improvisation by the ***Experimenter*** that impacts ***Task Interpretation*** and further ***Decision-making***.
03:16 to 03:21		The observer continues picking pairs of objects and inspecting them. Comparing each other.	***Comparison*** with a ***Reference***.
03:26 to 03:31		The observer puts two spheres next to the third one, and compares the three.	***Comparison*** with a ***Reference***.
03:31 to 03:41		The observer moves his hand atop the objects, and looks at the reflections.	New details appear in selected ***Methodology***. In addition to picking objects up and comparing them, the observer starts a different kind of ***Comparison*** with a different ***Reference*** - he compares reflection image on the same sphere among several conditions - among several positions of his hand. According to the selected ***Methodology***, better the hand movement is depicted in the surface reflection image, glossier the object.
	Observer: "this is I think the most glossy one, without considering the artifacts."	The observer picks the dark blue object and examines from close. Then puts it on the right hand side of the table, as being ranked glossiest.	The ***Semantic Description*** is regularly used for ***Completion of a Visual Task***.
03:41 to 03:46		Puts his hand close to the sphere surface and observes closely. Then puts the blue one next to the one ranked	

		first. Then chooses the third one.	
03:46 to 04:02	The observer explains his decisions: ”these specular reflections look the same on all of them. Except for the damaged areas. The way I am going to classify them is whether I see myself on them. Whether it has a mirror effect or not. ”		The observer explains the ***Methodology***, and the ***Decision-making*** process that lead him to this particular ***Methodology***.
04:02 to 04:08	Experimenter: ”so you are not using specular effect, but how you can use them as a mirror.”		The ***Experimenter*** clarifies the ***Task Interpretation*** and selected ***Methodology***.
04:08 to 04:18	Observer: ”yes, I tried to use specular reflections, but they all look the same.”		The combination of ***Object*** and ***Conditions of Observation*** have impacted ***Methodology*** selection.
04:18 to 04:28		The observer blocks direct sunlight with his hands towards two translucent spheres, and looks at them in the shadow. Then picks them up and inspects closely.	Again, ***Comparison*** with a ***Reference*** in several conditions.
04:28 to 04:35		The observer takes decision one of them is glossier. Puts it on the fourth place, while the last one is put on the fifth place.	The ***Comparison*** with a ***Reference*** using particular ***Methodology*** leads to ***Visual Task Completion***.
04:35 to 04:40	Experimenter: artifacts would have changed this order, or not?		
04:40 to 06:14	The observer explains the process: ”it depends how you look at it. At first, I did not pay attention to them, because I know they are not intended to be there. So, I judged just the normal part. But between this two”, - points to the last two ones - ”when I did not have any other choice, because I couldn’t use them as a mirror, and specular reflections are same, so I look at them and decided which one has more damaged areas that reflects less light. It’s very very last cue, I looked specular reflections first of all, but they are the same. Then I saw my gloves on this one [glossiest one], here it’s a bit blurry [second and third	Picks the two objects again and shows the areas, which do not reflect in a specular direction due to scratches.	The observer explains the ***Methodology***, and the ***Decision-making*** process that lead him to this particular ***Methodology***. Also names particular **References** used.

	ones]", - moves his hand atop the object. "And here [two least glossy ones] very little bit. Here (first two ones), I even see my face, while here [last two ones], I just see my gloves when I bring it very close to the surface."	
06:14 to 06:30		The experimenter thanks the observer, the result is recorded (photographed), and they switch to a new task.