



Delft University of Technology

A framework connecting vision and depiction

Wijntjes, M. W.A.; Middelkoop, C. van

DOI

[10.1080/13506285.2024.2320762](https://doi.org/10.1080/13506285.2024.2320762)

Publication date

2024

Document Version

Final published version

Published in

Visual Cognition

Citation (APA)

Wijntjes, M. W. A., & Middelkoop, C. V. (2024). A framework connecting vision and depiction. *Visual Cognition*, 32(6), 535-550. <https://doi.org/10.1080/13506285.2024.2320762>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.



Co-Editors: Benjamin Balaz & Jonathan Flombaum

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/pvis20

A framework connecting vision and depiction

M.W.A. Wijntjes & C. van Middelkoop

To cite this article: M.W.A. Wijntjes & C. van Middelkoop (22 Apr 2024): A framework connecting vision and depiction, Visual Cognition, DOI: [10.1080/13506285.2024.2320762](https://doi.org/10.1080/13506285.2024.2320762)

To link to this article: <https://doi.org/10.1080/13506285.2024.2320762>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 22 Apr 2024.



Submit your article to this journal [↗](#)



Article views: 61



View related articles [↗](#)



View Crossmark data [↗](#)

A framework connecting vision and depiction

M.W.A. Wijntjes ^a and C. van Middelkoop ^{a,b,c}

^aIndustrial Design Engineering, Delft University of Technology, Delft, The Netherlands; ^bWillem de Kooning Academy, Rotterdam University of Applied Sciences, Rotterdam, The Netherlands; ^cArt History and Material Culture, University of Groningen, Groningen, The Netherlands

ABSTRACT

We present a framework that connects ideas from the visual arts and visual perception. It adapts two existing frameworks for the analysis of form and content so that it can be used in an educational context for teaching perception through visual arts. The basis is the formal analysis of texture, colour, light, space, and material. This analysis can be conducted both on the medium and the motif, which adds a second level in addition to the formal level. Thirdly, a content level is discussed which combines a basic notion of semiotics and iconography. We share our experience of implementing pictorial analysis in design and perception education and discuss how the framework is used both in a quantitative and a qualitative fashion. Next to education, the framework provides a basis for further pictorial research.

ARTICLE HISTORY

Received 3 January 2023
Accepted 29 January 2024

KEYWORDS

Visual arts; visual perception;
formal analysis; art-based
education

Introduction

The visual arts have a close connection to the study of visual perception. Both artist and scientist have been interested in the relation between the 2D input (art-works or stimuli) and their perception/interpretation. Moreover, both artists and scientists are interested in the specific pictorial ingredients that signal specific perceptual phenomena. Art history can be regarded as a massive longitudinal perception experiment exploring the spectrum between the merely and clearly perceivable, and revealing the roles of convention and abstraction. Looking at art almost automatically results in learning something about perception when knowing what to look for. The current paper introduces a framework about vision and depiction that facilitates this process.

The framework we propose is a perception flavoured version of already existing frameworks of pictorial analysis used in art history and philosophy. In art history, the analysis of pictures is roughly divided in two categories: formal analysis and iconography. The word “formal” may elicit associations with something official, but in the context of the visual arts it means the analysis of *form* and refers to the visual elements that shape the work or art. It

can be contrasted with iconographic analysis, which is concerned with symbolism and reference. A well-known framework for formal analysis, *Principles of Art History* by Wölfflin (1915), proposes a model to distinguish artistic style on the basis of the five visual principles: linear–painterly, plane–recession, closed–open, multiplicity–unity and absolute–relative clarity. The aim of these principles was to discriminate between classical and baroque paintings. Recently, the framework has also been used in digital art history (Elgammal et al., 2018). However, the principles do not connect directly to the subject matter of vision science. Other formal frameworks are based on so-called “elements of art” or “elements of design,” and can be found in various places such as museum and art educational websites. A search on the internet reveals that the most common elements that are typically mentioned in relation to the “elements of art” are line, colour, shape, texture, and space. A more systematic overview and history of formal analysis is given by Munsterberg (2009), who writes that first Du Piles (1743) proposed composition, drawing, colour, and expression as formal elements, and then Fry (1920) proposed colour, line, light/dark, volume, mass, plane, composition, and

CONTACT M.W.A. Wijntjes  m.w.a.wijntjes@tudelft.nl  Industrial Design Engineering, Delft University of Technology, Landbergstraat 15, 2628CE Delft, The Netherlands

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

expressive content. Another famous formal framework is *Art and Visual Perception* by Arnheim (1974), which consists of the respective chapters balance, shape, form, growth, space, light, colour, movement, dynamics, and expression. These elements are summarized in Figure 1.

Many of these elements are topics in the study of psychophysics, where physical descriptions are related to perceptual judgements. Colour is a good example, but so are depth, light, and texture perception. As can be seen, all coloured elements in Figure 1 fall under these categories. One element that is omnipresent in formal frameworks but does not have a psychophysical counterpart is “line,” mostly because a line does not exist in nature. Certainly, we do not want to ignore the abundant usages of lines in art, but they are not a psychophysical element. Moreover, lines are typically part of the “medium,” i.e., the description of the flat 2D structure that makes up the image and are not part of our interpretation of the image, the “representation.” A number of other elements in the overview are also not psychophysical properties, such as style, expression, and balance. Style may not be an element but rather what a formal analysis leads up to: style can be described in terms of formal elements. Although movement can be perceived as a psychophysical element, our framework does not incorporate it as it is often associated with dynamic images, whereas our focus is primarily on static imagery.

Du Piles (1743) Composition Drawing Color Expression	Fry (1920) Color Line Light/dark Volume Mass Plane Composition	Arnheim (1974) Balance Shape Form Growth Space Light Color Movement Dynamics Expression	Getty (2023) Line Shape and form Space Color Texture
Kennedy Center (2023) Line Value Shapes Forms Space Color	Smarthistory (2023) Line Color Space Texture Scale	Agnesscott (2023) Line Medium Color Light Space Composition	Vision+Depiction Texture Colour/light Space Material

Figure 1. Overview of elements of various formal frameworks in art. The highlighted elements are the formal elements that are used in the framework shown in Figure 2. As can be seen, in our own framework we added “material” to the list. The colours refer to the colours used in Figure 4, where we analyse the usage of the Vision and Depiction Framework elements in practice.

We coloured the elements in three categories that form the initial basis for our framework: texture (red), colour and light (orange), and space (green). We added a fourth element that has rarely been used explicitly but has historical relevance and has been attracting attention from perception studies and 3D artists over the recent decades: the depiction of materials. These four formal elements constitute the first level of explanation, which is represented on the y-axis in the diagram shown in Figure 2. We also found “scale” and “medium” in our overview, which inspired a second level represented on the x-axis. Despite the fact that scale and medium are not often mentioned in the context of formal analysis, we are convinced that any formal analysis should be aware of the distinction between the description of the flat, physical medium and the imaginary, often non-flat representation. As any of the elements can be discussed in both contexts, we conceived that this is an extra dimension to the graphical representation of the framework that can be seen in Figure 2.

Next to form, there are content and context, which are represented on the z-axis in Figure 2. Interpreting the symbolism in artworks is called iconography. An iconographic analysis requires training and

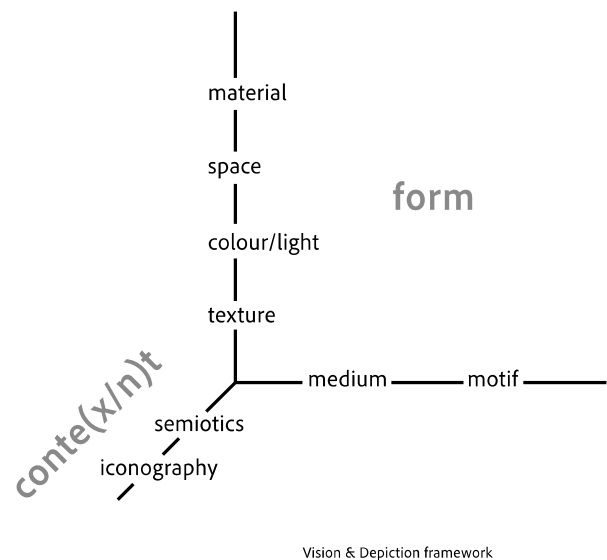


Figure 2. Framework to analyse the form and content of pictures. The y-axis refers to the traditional formal analysis elements from which we selected those that have been studied psychophysically. The x-axis implies that all formal elements can be studied both in the (physical) medium and the (perceived) motif. The z-axis relates the formal elements to the content and context, which are often discussed in the theory of semiotics and iconography.

knowledge, for example about the bible when interpreting paintings from western art history. Panofsky (1955) introduced iconological analysis, which is more concerned with the context of the painting. Furthermore, iconography is strongly related to the study of signs: semiotics. Especially, the distinction between icons and symbols is an interesting topic in the context of visual analysis. Although the content topics may not have a direct relation to perception, form and content are always related and to some extent inseparable (Sontag, 1966). It would be a missed opportunity if we did not discuss them in tandem. Hence, our complete framework consists of form and content and can be seen in Figure 2.

We have used this framework in education to teach students the relation between vision and depiction, we hence call it the Vision and Depiction Framework. The main idea is that analysing pictures using this framework reinforces and expands knowledge about perception. Yet there needs to be something to be reinforced, a theoretical foundation about the elements of the Vision and Depiction Framework. In the next section, we will outline examples related to the framework content. By reviewing common topics in perception and art history, we offer a foundation that is concrete although non-exhaustive: it should serve as a starting point.

The analysis of form

Medium versus motif

The French painter Maurice Denis famously argued that “It is well to remember that a picture before being a battle horse, a nude woman, or some anecdote, is essentially a flat surface covered with colours assembled in a certain order.” The first step of a formal analysis should be concerned with distinguishing these two stages of medium and what it represents (the motif). While “medium versus motif” alliterates well, it requires some clarification. Medium concerns the physical object of the picture, but we also use it to indicate the pictorial plane or the proximal stimulus. The latter term is used in perception science to indicate the raw signal that enters a sensory system. Motif is often used for recurring themes in artworks, but it can also be used to indicate the representation. In contrast to the pictorial plane, the motif relates to the pictorial *space*, i.e., the

space containing Maurice Denis’s battle horse. Unlike the proximal stimulus, the motif relates to the (interpretation of the) distal stimulus, the representation.

The dichotomy between medium and motif results in the “twofoldedness” of pictures (Wollheim & Eldridge, 2015). An observer can look at the physical surface of the picture (seeing-as) but also perceive the representation (seeing-in). According to Wollheim, the twofoldedness experience is simultaneous, which he opposes to Gombrich’s idea of an alternating awareness (Gombrich, 1960; Wollheim, 1998). In establishing what a picture is, Koenderink (2015) argues that a picture implies a double sided intentionality: it is intended by the artist to *be looked at as a picture*; and at the same time an observer actually *looks at the picture as a picture*, and not as a surface. Note that, in this definition, “picture” refers to what we call “motif.”

In vision science, the medium is often a digital image presented on a light emitting screen. In the arts, the medium can take many different forms, from stained glass to drawings on curved rocks and photos in lightboxes. Some media have interesting histories and relations to perception, such as paintings and prints. Paintings come with different paint (= binding material + pigment) and surfaces (e.g., panel, canvas, paper, wall). In the section about the depiction of materials, we will discuss how oil paint and tempera differently affect the rendering of stuff. Printmaking, as opposed to painting, aims to produce series of images. One of the oldest techniques is engraving, where the artist engraves paths on either a wood or metal surface where ink can accumulate. These line paths filled with ink are then transferred to paper when pressed. Given the limited means of lines, engravings can be rather impressive, sometimes reaching photographic appearance. Shading needed to create these convincing depictions is often achieved through hatching, but also by varying the line thickness, as is wonderfully illustrated in Claude Mellan’s “Face of Christ.”

Printmaking has often been used for creating reproductions of paintings, which is superseded by two other reproduction revolutions: that of *mechanical* reproduction (Benjamin, 2008; Berger, 2008) and that of the more recent *digital* reproduction. Most art historians emphasize the difference between the original and the reproduced in, for example, a book

or website. Not only is the medium different, also the scale is likely dissimilar, and the context is certainly distinct. There does not seem to be much perception research on the influence of medium, size, and context, but most art historians seem to agree that nothing beats the original. A good example is that Ernst Gombrich, in his popular *Story of Art* (Gombrich, 1995b), only discussed artworks he had seen with his own eyes (except some that were lost in fires). Perhaps surprisingly, art historians have never seemed worried about greyscale reproductions in books as there is nothing like the original; a reproduction merely serves as a reminder, an abstract shadow of the original.

Texture

An interesting transition from medium versus motif is the fact that both the picture surface and the representation can be described in terms of texture. Every medium builds the picture with different elements—grain in analogue photography, halftoning in printing, pixels in digital images, impasto textures in oil paint, thin lines in engraving, etc. Looking at the picture surface texture makes you realize that a picture is not something of infinite resolution but consists of discernable, finite elements. Before the elements form a recognizable shape or object, they form a texture. The transition from element to texture, to object is clearly related to the hierarchy of visual processing and perceptual organization. The integration of similarities and the segregation of dissimilarities are universal visual principles underlying texture perception (Julesz, 1981) and gestalt laws (Wagemans et al., 2012), and examples in the visual arts are abundant.

The picture surface textures arising from the techniques mentioned above do not automatically interact with the representation, e.g., the film grain is spread homogeneously over the picture area. But it is certainly possible to let the texture of the *medium* interact with the *motif*. A great example is the technique used by Bob Ross, but can also be found in some street artists: in order to produce artworks on the street, artists have discovered various rapid prototyping tricks to create stunningly convincing images. Their technique is based on applying paint using a mechanical process, resulting in textures that closely resemble real world elements like foliage, rocks,

clouds, etc. An example from Bob Ross would be spraying or dabbing paint with a brush. Street art showing fictional landscapes or cosmological scenes with planets and stars also makes use of textures resulting from spray paint but additionally uses sticky crumbled paper, creating more rough textures. It is somewhat surprising to the authors of this paper that we could not find empirical studies investigating the perceptual effects of various mechanical texture synthesis techniques. Yet the phenomenon does get attention in art history. The opposite also occurs, a total separation between the texture of the medium and texture within the representation, as can be seen in some of the works by Gerhard Richter, an artist worthy of adoration by vision scientists.

In *Art and Illusion*, Ernst Gombrich introduces the “etc. principle”, the “assumption we tend to make that to see a few members of a series is to see them all” which describes how vision extrapolates on the basis of local information. In the *Sense of Order*, this principle is reused but here in the context of texture perception: seeing a mass of things gives the impression that every detail is present. In the same book, Gombrich shows many examples of how artists used a technique that the perception scientist may call summary statistics. For example, zooming in on depictions of written texts or fine ornaments will reveal there is no legible text, nor detailed ornament, but rather a visual metamer similar to those studied in vision science (Balas et al., 2009; Freeman & Simoncelli, 2011). Indeed, tricks to elucidate the apparent richness of visual experience as described in vision science (Cohen et al., 2016) can be found in many paintings by zooming in and out, or better, moving back and forth to and from the original.

The subject of texture also involves a notion of blur and spatial frequency. It is a well-known photographic convention to sharpen the protagonist while blurring the context. Given its abundant use, it would be expected that sharp areas (high spatial frequency) capture visual attention more strongly than blurred areas (low spatial frequency). This was empirically confirmed with the addition that it only holds when inspecting the content and not the quality of the photo (Enns & MacDonald, 2013). This photo quality aspect is a second purpose behind using defocus: it adds aesthetic appeal. However, this common belief is less empirically robust (Zhang et al., 2014). While blur and sharpness are most often used juxtaposed,

it is also possible to use them overlaid, as exemplified by so-called hybrid images (Oliva et al., 2006). It is a relatively simple exercise to combine a blurred (low-pass filtered) image with another high-pass filtered image and thereby make the interpretation dependent on distance: when close by, the high-pass filtered image will be seen, while from far away, the blurred image will dominate. Although one may think that hybrid images are only artificial gimmicks, the concept is quite powerful as it explains many scale and (thus) distance dependent phenomena, such as being aware of the paint brush texture when being close to a painting. It may be noticed that we have arrived back at Wolheims' twofoldedness, discussed in the subsection "Medium versus motif," and offered a scale account of seeing-in and seeing-as.

The above examples display a rather wide definition of the term "texture" but do not contain texture as a material property, because that is a separate formal element. It should be noted that this goes against some of the art historical jargon where texture and material are sometimes used interchangeably (Gombrich, 1964). The ambiguity is often resolved by the context and, more importantly, the distinction itself does not affect the analysis of images.

Colour and light

Colour and light are inseparable, yet they are often discussed separated in formal analysis, as can be seen in Figure 1. By addressing them conjointly, we emphasize their entanglement, resulting in a rather wide spectrum of topics.

One of the most important topics in art history is the invention of pigment and binding material. We will have to omit comprehensive overviews—e.g., Ball (2003)—but it is interesting to consider for what reasons(s) the Egyptians started with a handful of pigments and why, in the seventeenth century, painters (e.g., Rembrandt and Vermeer) used a dozen pigments and how this increased to the usage of hundreds of pigments by the nineteenth century. The obvious question is "why would you need so many pigments if you can also just mix them?" While we are used to three primaries in computer screens creating a colour gamut specified by a CIE Chromaticity diagram, this is much less trivial for paint. The

difference is that paint gives a *reflective* colour and the screen an *emissive* colour. Colour mixing for these two is radically different and is called subtractive (paint) and additive (light). Although the process of paint mixing is complex, a simple model is to multiply the reflection spectra (it should be called multiplicative mixing). Doing this will make it clear that any colour mixing is making the colour duller than the original colours (Berns, 2016), which is the reason for the abundance of available pigments. Not only the history of pigments is interesting, also its binding material. The transition from egg yolk (tempera) to oil was especially pivotal, but we will discuss this in the section about material.

Many pigments merit individual stories, such as the precious ultramarine, which played such an important role from the Renaissance onwards, or white lead, which possessed unprecedented opacity but turned out to be poisonous and made artists go crazy. But one type of pigment deserves special attention, especially in the context of perception, which is fluorescent pigments. Famous artists like Frank Stella and Herber Aach (Aach, 1970) started to pioneer these paints in the 1960s, mesmerized by the illusion that the paint seemed to radiate light instead of reflecting it. One of the brands used by Stella was aptly called DayGlo. Fluorescent paint transforms invisible ultraviolet light to visible light. That is why "black light" (UV light) amplifies the effect of fluorescent lights so much, and that is also why in practice these pigments are used to increase the saliency of people like road workers, emergency crews, etc. The perceptual confusion caused by fluorescent paints can be a great starting point for further discussion about reflective and emissive colours. And again, the discussion could concern the medium or the motif. When it concerns the medium, it could be interesting to consider what art forms make use of emissive colours. Obviously, computer screens do, but only digital art is specially made for this medium and reproductions of paintings or photos originally made use of reflective colours. A pre-industrial medium that makes use of emissive colours (or better, transmissive colours) is stained glass. Emissive colours in the motif, i.e., the depiction of light sources, is another intriguing depiction challenge. The trick clearly is to increase the brightness beyond what could be expected on the basis of reflection, i.e., a brightness beyond white. The limited dynamic range of the

paint palette makes it particularly challenging and requires planning.

A subject connecting colour (this section) and space (the next section) is light and shadow. The interaction of light with the environment gives volume to shapes (shading) and space (cast shadows). Moreover, the treatment of light is a stylistic pattern that can be traced quite clearly throughout (western) art history. For example, in many paintings, light comes from the left side (Carbon & Pastukhov, 2018; Wijntjes, 2020), a bias also found in perception (Sun & Perona, 1998). The bias in depiction originates from practical considerations: when the window is on the left, the shadow of the pencil and hand are not obstructing the view. Indeed, windows are often seen on the left side (Van Zuijlen et al., 2021) in paintings. Therefore, it seems that the bias in depiction does not originate from the bias in vision, and hints towards the idea that vision is biased by depiction, a direction that seems to be quite unexplored.

When light hits a 3D surface, the reflected light that reaches the eye (or camera) depends on the orientation of the illumination, the orientation of the surface, and the reflectance properties (i.e., the optical material properties). The shading pattern resulting from a smoothly curved object, such as the human body, can result in a smooth gradient. Shading gradients can be seen abundantly throughout art history. There are famous old examples such as Roman mosaics, Pompeian murals, and Egyptian Fayum portraits (Gombrich, 1964). Even more astonishing is that the Greek painter Apelles in the 4th century BC was the inspiration for these mosaics and murals, and even some cave paintings seem to show shading gradients. When gradients are absent, as in Egyptian art, the appearance is immediately cartoon-like.

Cast shadows, on the other hand, are a different story. They are surprisingly absent (Gombrich, 1995a) and if present, they are often amorphous, bearing no resemblance with their caster, and function mainly to perceptually glue the object to a surface. The advantage of an amorphous shadow blob is that it reduces the risk of mistakes. Two typical “mistakes” tend to be made when the artist does render a detailed cast shadow: the projected shape and the perspective direction. Casati (2008) argued that the shadow primarily serves as a cue for relative position, for which a clear correspondence between caster and shadow is required. This

correspondence is optimal when the frontal outline shape of the caster is used (the “copy-cat” solution) instead of the physically correct projected shape (Casati, 2008). Indeed, human observers seem to be quite insensitive to the physical accuracy of cast shadow shapes (Jacobson & Werner, 2004; Mamassian, 2004), a phenomenon that was also quite successfully “gamified” (<https://en.wikipedia.org/wiki/Shadowmatic>). As for the perspective of shadows, humans are also insensitive (Pont et al., 2011) to the exact geometry. Cast shadows from sunlight are parallel in the world and should thus converge to a central vanishing point on the horizon (in case of a horizontal ground plane and orthogonally upright objects). This rarely occurs, and in cases like the surrealist painter Di Chirico this is actually on purpose. An interesting trick to avoid making perspective mistakes is to organize sunrays parallel to the projection plane. In that case, the cast shadows will also be parallel and do not need to converge, a convention that was often used by Canaletto (Wijntjes, 2020; Wijntjes & de Ridder, 2014).

Space

If one theme in art history is applicable to the science of perception, it is the invention of pictorial space and how this relates to depth perception. The various perspective systems that have been used throughout history and across various cultures have been neatly described by Willats (1997). One of the key concepts introduced in that book is the distinction between primary and secondary geometry. Primary geometry refers to the *projection* of the 3D world onto a 2D surface. While this geometry is of importance for understanding photography, and is obviously used in computer rendering systems, it is not directly used by a painting or drawing artist. While letting the pencil move over the paper, the artist makes use of *drawing rules*, for example by using one or multiple vanishing points in the case of linear perspective. These drawing rules are called secondary geometry. It is interesting to consider whether any primary geometry can be related to a secondary geometry and vice versa. A relatively trivial example is the relation between the primary geometry of linear perspective and the familiar drawing rules (secondary geometry). One famous example in the art history of perspectives is the use of “reverse” or “inverse” perspective where

parallels converge instead of diverge, for example in Rublev's Trinity. We leave it to the reader to consider what kind of primary geometry could be associated with this.

One secondary geometry for which it is surely impossible to find a primary geometry is that of the common "fold-out" or "intellectual" perspective that is so often seen in Egyptian art. "Fold-out" refers to the geometry of locally adjusting the viewpoint so that every individual object is seen from its most recognizable side, which results in an impression of folding out. "Intellectual" refers to depicting what is known, instead of what is seen. These accounts raise immediate associations with more contemporary notions of pictorial space and graphic design. The combination of multiple viewpoints in one picture is known from cubism, and the depiction of what is known, in an optimally recognizable fashion, is used in contemporary graphic design.

Willats (1997) describes many other perspective systems, such as oblique parallel perspective which occurs in Japanese paintings, horizontal parallel perspective in south Asian miniature art, linear perspective in western art, and curvilinear perspectives in some mirror reflections. Most of these systems "work," i.e., we can make spatial inferences although the result may depend on the paradigm used (Van Doorn et al., 2012). Furthermore, it can be interesting to apply a formal perspective analysis beyond art history, for example on the development of computer gaming environments.

The development of cinematographic techniques related to pictorial space deserves extra attention. Firstly, while most photo- and cinematography is confined to linear perspective, a number of interesting variations within these boundaries were developed. Secondly, the viewpoint became an important artistic variable since the pioneering of Orson Welles who drilled holes in the floor to achieve extremely low viewpoints. Hitchcock pioneered the "dolly zoom" (simultaneously zooming and moving) in *Vertigo*, resulting in an exciting non-rigid spatial illusion. Camera motion can be used to convey the three-dimensional shape of an object, based on the kinetic depth effect (Wallach & O'Connell, 1952). Motion parallax can be evoked by lateral camera movement, a visual phenomenon described by Helmholtz and later studied by Gibson (1950) and Rogers and Graham (1979), among others. It is interesting

to note that Disney's multiplane camera was specifically designed to evoke this type of three-dimensional perception, first used on *Snow White and the Seven Dwarfs* from 1937, and later used in *Bambi*, which was released in 1941. This effect in turn was used by the contemporary artists Persijn Broersen and Margit Lucács in their work *Mastering Bambi*, who put the background in the foreground by leaving out Bambi and making nature itself the protagonist.

Material

A thus-far rather neglected subject in both art history and perception is materials. There is no real equivalent of Alberti's treatise on perspective, while it certainly did not go unnoticed that the depiction of materials advanced substantially from van Eyck onwards. While the depiction of space can be attributed to the invention of linear perspective by Brunelleschi (and further explained by Alberti, 1435), the increase in convincing material depiction is often attributed to the invention of oil paint. The slow drying process, its ability to create smooth gradients and increased contrast, trumped the capacities offered by fast drying and lower contrast tempera. Van Eyck was one of the first adopting this new technique (Gombrich, 1964) and became a master in material depiction. For example, many of the material properties that received attention from perception research can be found in the Gent Altarpiece: transparency (the staff of the almighty), translucency (in many of the gems), glossiness/highlights (both on gems and pearls) more complex specular reflection (e.g., on the body armour) and the rendering of complex fabrics (such as the velvet brocade). We will shortly discuss the material properties below.

Transparency refers to the (partial) transmission of light without scattering, i.e., you can see clearly through a slab of transparent material. In the perception literature, a distinction is made between the transparency of thin sheets, like a plastic sheet, or thick 3D objects (like a cylinder or sphere). The perception of thin sheets was pioneered by Metelli (1974), who described a set of algebraic rules that determine whether two overlapping sheets will be seen as transparent or not (Metelli, 1970). Sayim and Cavanagh (2011) showed that these rules are sometimes obeyed by artists but also demonstrated cases where the artist violated them. They also

demonstrated violations of refraction patterns: the distortion of the background when looking through a thick transparent material (Fleming et al., 2011), such as the staff in the Gent Altarpiece. Violating the laws of physics without breaking perception seems a recurring theme in the history of art (Cavanagh, 2005).

Translucency on the representation, the motif, was sometimes accompanied by translucency in the paint itself, the medium (Bol, 2013): glaze layers were used to depict precious gems, which themselves are also observably translucent. Whether there is really a perceptual influence of translucency in the medium on translucency in the motif has not been empirically verified. A better known material interaction between medium and motif is the use of gold leaf, although that seems of a type different from our translucency example.

One of the most studied material properties is gloss (Chadwick & Kentridge, 2015). The dominant cue for glossiness is the highlight, in particular the contrast, sharpness, and coverage (Di Cicco et al., 2019; Marlow et al., 2012). The highlight is to material depiction what the vanishing point is to space depiction. Searching for highlights, or more general specular reflections, throughout art history will reveal that its usage goes back to at least the time of the Fayum portraits, where they were omnipresent in the eyes and sometimes skin of the sitters.

The analysis of content

Semiotics

The “content” reaches beyond the image, it refers to something, or as the semiotician would say: it signifies. Semiotics is a discipline originating from the language and philosophy that made its way into visual communication. It is the study of signs, and although this may evoke immediate association with traffic signs, semiotics can regard almost everything as a sign. A word (e.g., “apple”) can be a sign, signifying something (an apple). Interestingly, neither the sound of the word apple, nor the visual appearance of the five-letter pattern, resemble the signified in any way. Therefore, the signalling function must be based on convention, on a learned association between sign and signified. This type of sign is what the famous semiotician Peirce would call a “symbol”

(Atkin, 2022). A counterexample would be an onomatopoeia, which phonetically resembles what it signifies. While in language this seems relatively rare and primarily used for animal sounds (e.g., moo, meow, oink), in vision it happens all the time. For example, any photo resembles what it signifies (Goodman, 1976). These cases Peirce would call “icons”. There is a third type of sign, which is called an “index” and is used when there is a causal relation between sign and signified; a common example is smoke as an index for fire. An example from contemporary art is *Colour Studies* by Trevor Paglen, who photographed the night skies above California state prisons. The “bright orange and green hues of their always-on floodlights” are an index that signify the presence of prisons from large distances. While the index may not be directly relevant for those interested in art and perception, the continuum between symbol and icon is. The strength of the resemblance between sign and signified is called iconicity and it speaks for itself that, among the various visual signs, there exist many levels of iconicity. To some extent, iconicity may seem akin the abstract–figurative dimension, or how realistically something is depicted.

Although the scope of the current contribution is to describe a framework for the perceptual analysis of pictures, the scope of semiotics is wider, and the reader may find it interesting to contemplate the relation between semiotics and data visualization. While this area has attracted the attention of perception research (e.g., Ware, 2019) and an attempt to introduce “experimental semiotics” (Ware, 1993), the term “iconicity” is rarely used. The concept is commonly used in other visual research areas (e.g., anthropology, see Granito et al., 2022), and applying it to information visualization would result in a detailed analysis of how certain elements resemble what they signify. Primarily, the more creative visualizations make use of certain levels of iconicity, often in combination with metaphor. Then there are figures of speech that involve comparing two seemingly unrelated things or ideas to highlight a similarity between them.

Iconography

These semiotic considerations are quite theoretical and become more interesting when put in an art historical context, where we arrive at iconography. In a

broad sense, iconography is the interpretation of works of art (Müller, 2011). The most used framework for this interpretation is that of Panofsky (1955), who describes iconography as “that branch of the history of art that concerns itself with the subject matter or meaning of works of art, as opposed to their form.” Panofsky describes three levels of interpretation: (1) factual (e.g., a man with a lion at his feet), (2) symbolic (e.g., refers to Saint Jerome) and (3) contextual (e.g., represents the relation between humans and nature). These are our terms that we find instructive but are not standard; they are often referred to as primary, secondary and tertiary and also pre-iconographic, iconographic and iconological, respectively. It should be noted that these three different levels are used simultaneously in an actual iconographic analysis (Müller, 2011). This especially explains the relevance of the first level, because “factual” seems quite irrelevant in isolation, while the other two levels seem so relevant both in isolation and in relation.

We call the second level symbolic, as it mostly deals with symbolism related to religion and mythology. Especially in Renaissance art, for which Panofsky devised the method, religious and mythological reference is omnipresent. Therefore, this type of analysis requires prior knowledge that cannot be assumed in a general student population. While we believe it can be wiser to leave actual iconographic analysis to the art historians, the question “what does this picture refer to?” can perfectly well be asked and investigated with the means at hand.

It is interesting to note that there is a general classification system for iconography, called Iconclass. This system was introduced by the Dutch art historian Henri van der Waal (Couprie, 1983) and consists of nine main groups: religion, nature, human beings, society, abstract ideas, history, the Bible, literature, and mythology. As an example, there are many depictions of Saint Jerome, who can be identified by the lion at his feet. In Iconclass, his code is 11H(Jerome), with the key: 1-religion, 1-Christian religion, H-saints. This system can be seen as an iconographic equivalent of the many labelled image sets used in computer vision (e.g., Lin et al., 2014) but also art history (Strezoski & Worring, 2018; Van Zuijlen et al., 2021). Recently, the Iconclass system gained attention from the digital humanities (Milani & Fraternali, 2021).

The third level of Panofsky aims to contextualize the work in history, location, society, culture, etc.

Panofsky speaks of the “intrinsic meaning” of the work and calls this type of analysis “iconology”. While the first two layers are closely related to the intent of the painter, the third layer can lie beyond the artist’s influence as it is shaped by societal reception. To Panofsky, this third layer is the ultimate goal of analysing paintings. To some extent, this level is furthest away from visual perception. However, what makes it interesting to use pictures as research material, as starting points to learn about vision and depiction, is exactly this extra layer. For example, a famous press photo is not only well composed, uses optimal lighting and a subtle depth of field to create a clear narrative. Through its framing it also influences public opinion and becomes a symbol for historic events. Other pictures have a different relation to society. Certain Dutch still-lives from the seventeenth century are purely intended to be show-off works, for example the “Pronkstilleven” by Adriaan van Utrecht, which contained many different objects demonstrating he could depict anything in a virtuosically convincing fashion. The popularity of these types of work tells something about the public’s appreciation of the pictorial mastery of illusion. These are just two simple examples; the possibilities of doing this contextual, third level analysis are infinite and potentially very rewarding.

Lastly, let us briefly try to connect the formal analysis with the semiotic/iconographic analysis. Colours can play many different symbolic roles. For example, ultramarine, made from grinding Lapis Lazuli, was extremely expensive and often used to depict Holy Mary. It symbolizes both holiness but clearly also expensiveness and thus devotion of the maker (painter of funder). Shadows can also play symbolic roles. As a cast shadow is actually a picture, it can be used to symbolize the origin of painting itself (Gombrich, 1995a). While commenting on the various ways to depict space, Berger (2008) deemed linear perspective to be symbolic for European art as it “centres everything on the eye of the beholder” elevating the observer to a divine position as “the visible world is arranged for the spectator as the universe was once thought to be arranged for God.” Furthermore, Panofsky (2020) wrote a complete essay about the symbolic meaning of various perspective systems, although vision scientists may find some inaccuracies in Panofsky’s account of perception. Lastly, Bol and Lehman (2009) proposed a material

iconography, where they investigated skin and water depiction in relation to binding medium, establishing a clear relation between medium and motif. As with the previous paragraphs, these are only just some examples.

The framework in practice

After having outlined parallels in key concepts of vision and depiction, we will discuss learning activities. Over the past years, we have taught a course with approximately 100 design students where we made use of three types of learning activity: analysing, experimenting, and creating. The first activity aligns with the scope of our current contribution: analysing pictures within the vision and depiction framework discussed above. In the second activity, evaluate, students learn about empirical paradigms from the behavioural sciences, formulate a research question and conduct a (small) experiment. In the third activity, creating, the students communicate their findings in a visual manner, for example by recreating an artwork in a different medium, or adjusting it based on their empirical research. Depending on the type of curriculum, different weights can be given to these three facets. For example, for a psychology course one could consider skipping the “creation” activity.

Our starting point has been a curated picture collection, i.e., chosen by the educators. The advantage of curating a picture set, instead of letting students choose freely, is that educators have control over balancing, for example, gender and cultural backgrounds of makers, style, era, medium, materialization tools, size, purpose, etc. Also, it allows contemporary societal topics, cultural debate, and latest technologies to be involved. Next to this, by leaving out some of the decision-making processes and by focusing on final outcomes, we jumpstart the curiosity of our students and trigger a desire for experimentation. Students choose a fixed number of pictures from the curated selection and start their research through an inverse image search: to raise their awareness about pictures, we only provide pixels without metadata. Students are instructed to identify the formal elements they believe are most important, and to explain the underlying perceptual phenomena including whether they occur at the medium or motif level. Then they contextualize their findings

(using, for example, semiotics or iconography) and relate this to the background information they found through the reverse image search. In the appendix, more information can be found on how to curate the picture set. It should be noted that our framework, although clearly inspired by western art history and theory, is per definition not limited to this small subset of global culture. We strongly encourage the reader to look past the Western art history canon (Gil-Glazer, 2020). By deliberately labelling the picture collection as works of art, we create space for students to develop a more critical skillset and mindset.

For students with a background in the fundamental sciences (like our Masters students in design and engineering), an introduction to the arts through vision and depiction has the following advantages, even though “the idea that education has something to learn from the arts [might] cut across the grain of our traditional beliefs about how to improve educational practice” (Eisner, 2002). During the course, our students become more “qualitatively intelligent.” By having our students closely examine ways in which the forms of the artworks they choose from our picture collection are configured, their understanding of these qualitative relationships and the way the elements that constitute them are configured can also be applied to other things made, both practical and theoretical (from a visual image, a poem or a musical score, to an historical argument or a scientific theory). To let students learn from each other, we organize peer feedback sessions where they share and reflect on their analyses.

Next to learning from experiencing qualitative relationships and practising with making judgements in the absence of rules, students become familiar with “flexible purposing” and are more or less triggered into surrendering to “what the work in process suggests” (Dewey, 1934). In our case, this work can be seen as the three consecutive steps that constitute the course assignment. As uncertainty plays a pivotal role in creative learning (Beghetto & Jaeger, 2022), becoming aware of the fact that not everything knowable can be articulated in proportional form is another meaningful outcome of the course and which becomes most apparent in activity three, the creation phase of the course assignment.

Whether students are allowed to walk outside the territory naturally depends on the curriculum, and

our framework can equally well be used in a more focused context. Moreover, another version of the analysis activity can be one where students find the pictures themselves, instead of the curated picture set we use. It has some disadvantages, but if the learning goals purely focus on learning the formal elements that bind vision and depiction, it can be an interesting activity to collect pictures that show examples of all the different forms discussed in the framework.

Lastly, we hope to have raised some awareness of the connection between vision and depiction that may not only inspire education but also research. Many interesting questions seem to arise, for example, about the interaction between medium and motif, and there are likely many more unraised questions waiting.

Evaluation

In the academic year of submitting this paper (2022–2023), we let students analyse four images that they could choose from a curated set of 52. In total, we analysed 305 images (sometimes the student forgot to label them, and these data were omitted). First, we analysed image preference, i.e., whether some images were more popular than others. This was indeed the case, as can be seen in Figure 3. The distribution clearly shows preference, and 11 images accounted for 50% of the choices. Among these, Piet Mondrian, Felix Vallotton, Edmund Dulac, and Andy Warhol were probably relatively well known, while Karl Friedrich Schinkel, Herbert Bayer, Clement Hurd, Peter Saville, Toshio Saeki, and Georges Méliès were likely lesser known. Despite their lesser fame,

their images apparently appealed to the imagination and interest of the students. Furthermore, only two out of 52 images were not chosen by a student, showing that, despite certain preferences, almost all pictures received attention.

In addition to gathering statistics on the chosen images, we also collected vision and depiction keywords by asking the students to label their analyses. We specifically instructed the students not to discuss all elements of the framework for every picture, but rather to focus on specific elements. Yet, multiple labels per image were possible.

In Figure 4, the overall results are shown. Among the formal factors, colour (and light) were most popular, followed by space and texture, while material was chosen least. This resembles the dominance of colour/light and space in the formal analysis elements presented in Figure 1, where texture was least dominant and material was entirely absent. The formal element bias may not be completely attributable to the resemblance with previous formal frameworks, as we deliberately discussed each element with equal weight. One explanation could be that students were already familiar with formal analysis, and thus carried the bias already at the start of the course. A second explanation could be that pictures in general, or our selection in particular, carried this bias, i.e., that artists seem to experiment more with colour and space than with texture or material. This clearly depends on the viewer who notices and labels these.

A somewhat surprising result was that the majority of the formal analyses concerned the medium and not the motif. Most discussions about art and perception concern what is represented, the motif, so why

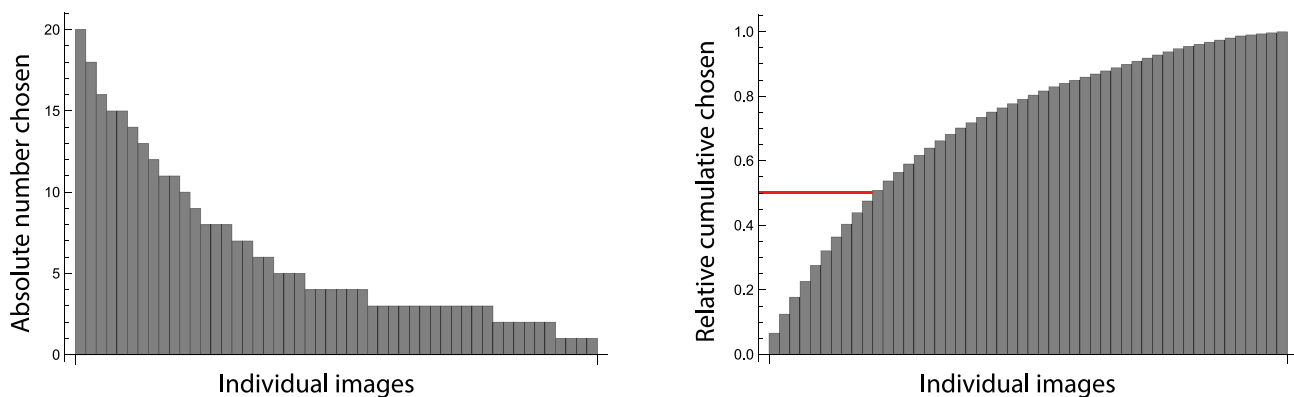


Figure 3. Statistics of chosen pictures. On the left, an ordered histogram showing a clear preference of some pictures over others. On the right, a cumulative histogram.

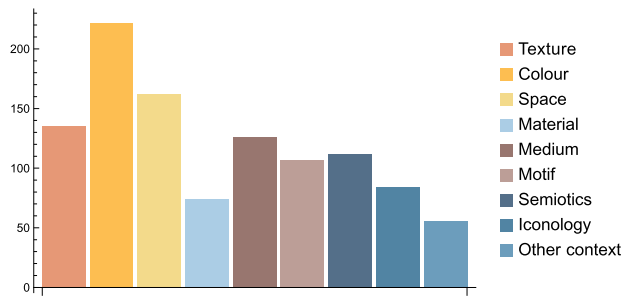


Figure 4. Histogram of V + D framework elements, the y-axis denotes absolute frequencies.

students were biased towards the medium was unexpected. Discussions about the medium often concern the technique of the making process. Perhaps the technical background of our students biased them in this direction. It should be noted that there is nothing “wrong” with a bias towards the medium over the motif; it was merely unexpected.

Thirdly, the context and content related analysis was dominated by semiotics. Again, this could be attributed to the design background of the students, as semiotics, the theory of signs, is often part of their

education. Furthermore, it should be noted that we only introduced both theories briefly and it can well be that iconography is a more complex topic to be put into practice.

Besides the overall average scores shown in **Figure 4**, we now discuss three example works, their labels and some of the students’ work.

The Islamic miniature from **Figure 5A** was made by Kamāl ud-Dīn Behzād, a famous Persian painter, depicting a scene from the Holy Quran where the prophet Yusef tries to escape Zulaikha. As we expected, students chose this work primarily to analyse pictorial space, although colour was also chosen relatively often. To gain an impression we copied two quotes: “*The flaming halo surrounding his head is the primary sign of his holiness. The absence of facial features further emphasizes the character’s sanctity, particularly within the context of the Islamic faith.*” Another student wrote: “*The lighting is even, without shadows or chiaroscuro, which further adds to the flatness in the painting. There are angled lines but little sense of axis, dimensionality, or depth.*”

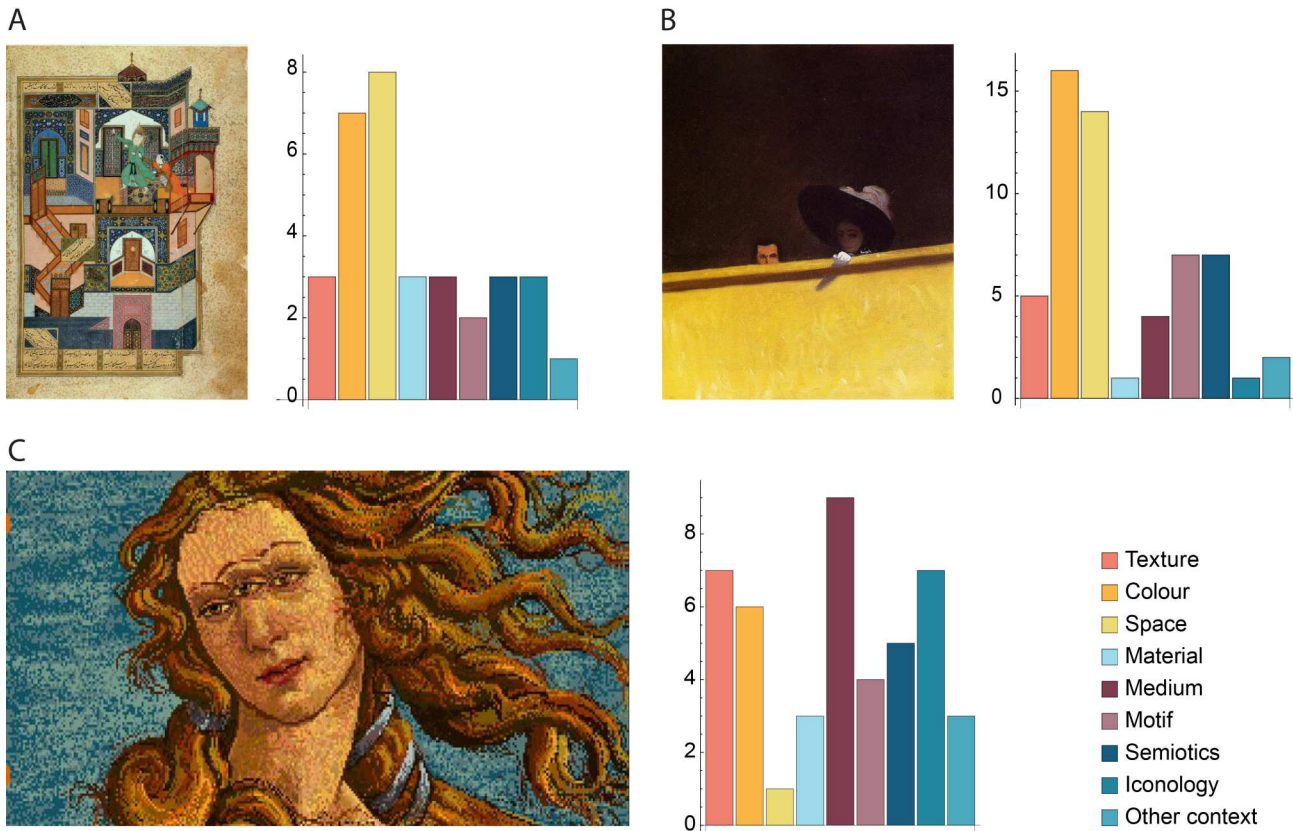


Figure 5. Three example works with histograms of chosen framework labels. The y-axis denotes absolute frequencies. (A) Kamāl ud-Dīn Behzād (1450–1535): *Yusuf and Zulaikha*. (B) Félix Vallotton (1865–1925): *La Loge de Théâtre, le Monsieur et la Dame*, 1909, private collection. (C) Andy Warhol: *Venus*, 1985, The Andy Warhol Museum, Pittsburgh, www.warhol.org.

In Figure 5B, the work *Box Seats at the Theater* by Felix Vallotton received much attention for colour/light and space, which were also the reasons we chose it. Furthermore, motif was chosen more often than medium, which makes sense as the representation was rather dominant. Lastly, the content was primarily analysed through semiotics in comparison to iconography, which also makes sense given that the scene did not originate from a story. Again, we quote:

As if shot through telelens, this painting only has two layers, which symbolically feel like different worlds. The yellow, illuminated balcony is visible to normal people in the audience, the contrasting dark purplish space behind is a mystery, only known to the upper-class.[...] With few depth cues, Vallotton paints a vivid story. As the lady comes forward, her white glove and hat break through the front layer, catching the spotlight from the back of the theatre. Considering the shadow angle, they are on the right side of the audience.

In Figure 5C, a relatively unknown work by Andy Warhol, is shown: a recently discovered digital work made on an amiga. The face is immediately recognizable as a reference to the Venus of Botticelli, but Warhol seems to have copy-pasted a third eye. Texture and colour dominate the formal elements while there is also a strong and expected dominance of medium over motif. This time, iconography was chosen more than semiotics which is to be expected of a work depicting a Roman goddess. We copied two quotes: *"The image texture is very pixelated, which is a direct result of the material that it is made of (computer pixels). This is a straight opposite of the original painting, which was made with very carefully and precise brush strokes."* Another student wrote: *"What is interesting to look into with this image is to see how the relation between depiction and resolution influence the idea of how futuristic an image is. Which raises a question about why the image looks retro rather than futuristic?"*

Discussion

The evaluation gave some quantitative and qualitative insights of the vision and depiction framework in practice. We found that there was some preferential bias, but that students also went less for popular pictures, as only two out of 52 pictures ended up not chosen at all. This wide spectrum of choices is

important because, while each student studied only four pictures, they could be confronted with many more: (1) during plenary formative feedback lectures where we discuss student work, and (2) during peer feedback meetings where students present their findings to each other.

Students seem to enjoy analysing a picture that is more than a stimulus. The observations they make often reach beyond the formal aspects of the picture; the framework seems to function as a steppingstone on which they can imagine their own narrative. In some cases, these analyses turn into personal perspectives. This can be encouraged as long as these accounts abide by the same rule that any other scientific contribution should follow: that of generalizability (i.e. not being applicable to only an individual case/student). Furthermore, the framework serves as a common ground to discuss a large variety of pictures with various cultural and societal backgrounds. The content of the images intrigues many students, sometimes because they identify with it, sometimes because it looks weird, and this curiosity carries over to the formal analysis and thinking about perception. Thus, although we cannot answer whether our framework increases the effectiveness of teaching about perception, the involvement of art, in all its versatility and variety, certainly seems to enhance engagement of the students (and the teachers).

Besides sharing our experience of how we have used the framework in practice, this paper has offered an alternative route to learning about perceptual phenomena: through artistic instead of scientific practice. We found that formal analysis offers common ground to vision and depiction, and we have briefly touched upon the art history and vision science of these formal elements. There is no doubt that this knowledge will grow and, besides education, will also find its way into science, where perceptual hypotheses will be inspired by artistic practice.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Maarten Wijntjes was supported by a Dutch Research Council (**Nederlandse Organisatie voor Wetenschappelijk** [NWO]) grant [Visual Communication of Material Properties" number 276.54.001].

ORCID

M.W.A. Wijntjes  <http://orcid.org/0000-0002-3657-735X>
C. van Middelkoop  <http://orcid.org/0000-0002-3425-7344>

References

- Aach, H. (1970). On the use and phenomena of fluorescent pigments in paintings. *Leonardo*, 3(2), 135–138. <https://doi.org/10.2307/1572077>
- Agnesscott.edu *Describing Art*. Retrieved July, 2023, from <https://www.agnesscott.edu/center-for-writing-and-speaking/handouts/describing-art.html>
- Alberti, L. B. (1435). *On painting*. Yale University Press.
- Arnheim, R. (1974). *Art and visual perception*. University of California Press.
- Atkin, A. (2022). Peirce's theory of signs. In E. N. Zalta, & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2022). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2022/entries/peirce-semiotics/>
- Balas, B. A., Nakano, L. B., & Rosenholtz, R. B. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, 9(12), 1–18. <https://doi.org/10.1167/9.12.13>
- Ball, P. (2003). *Bright earth: Art and the invention of color*. University of Chicago Press.
- Beghetto, R. A., & Jaeger, G. J. (2022). *Uncertainty: A catalyst for creativity, learning and development*. Springer.
- Benjamin, W. (2008). *The work of art in the age of mechanical reproduction*. Penguin Book.
- Berger, J. (2008). *Ways of seeing*. Penguin UK.
- Berns, R. S. (2016). *Color science and the visual arts: A guide for conservators, curators, and the curious*. The Getty Conservation Institute. <https://search.worldcat.org/title/927140853>
- Bol, M. (2013). Seeing through the paint. The dissemination of technical terminology between three Métiers: Pictura translucida, enameling and glass painting. In A. Speer (Ed.), *Zwischen Kunsthandwerk und Kunst: Die, Schemata diversarum artium* (pp. 145–162). DE GRUYTER. <https://doi.org/10.1515/9783110334821.145>
- Bol, M., & Lehman, A.-S. (2009). Towards a material iconography of translucent motifs in early Netherlandish painting. In *Papers Presented at the seventeenth symposium for the study of underdrawing and technology in painting held in Leuven, 22–24 October 2009*, 215–225.
- Carbon, C.-C., & Pastukhov, A. (2018). Reliable top-left light convention starts with early renaissance: An extensive approach comprising 10k artworks. *Frontiers in Psychology*, 9, 454. <https://doi.org/10.3389/fpsyg.2018.00454>
- Casati, R. (2008). The copycat solution to the shadow correspondence problem. *Perception*, 37(4), 495–503. <https://doi.org/10.1068/p5588>
- Cavanagh, P. (2005). The artist as neuroscientist. *Nature*, 434 (7031), 301–307. <https://doi.org/10.1038/434301a>
- Chadwick, A. C., & Kentridge, R. W. (2015). The perception of gloss: A review. *Vision Research*, 109(PB), 221–235. <https://doi.org/10.1016/j.visres.2014.10.026>
- Cohen, M. A., Dennett, D. C., & Kanwisher, N. (2016). What is the bandwidth of perceptual experience? *Trends in Cognitive Sciences*, 20(5), 324–335. <https://doi.org/10.1016/j.tics.2016.03.006>
- Couprie, L. D. (1983). Iconclass: An iconographic classification system. *Art Libraries Journal*, 8(2), 32–49. <https://doi.org/10.1017/S0307472200003436>
- Dewey, J. (1934). *Art as experience*. Minton Balch and Co.
- Di Cicco, F., Wijntjes, M. W., & Pont, S. C. (2019). Understanding gloss perception through the lens of art: Combining perception, image analysis, and painting recipes of 17th century painted grapes. *Journal of Vision*, 19(3), 7–7. <https://doi.org/10.1167/19.3.7>
- Du Piles, R. (1743). *The principles of painting*. J. Orborn.
- Eisner, E. W. (2002). *Elliott W. Eisner's John Dewey lecture*. Stanford University.
- Elgammal, A., Liu, B., Kim, D., Elhoseiny, M., & Mazzone, M. (2018). The shape of art history in the eyes of the machine. *32nd AAAI conference on artificial intelligence, AAAI 2018*, 2183–2191.
- Enns, J. T., & MacDonald, S. C. (2013). The role of clarity and blur in guiding visual attention in photographs. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 568–578. <https://doi.org/10.1037/a0029877>
- Fleming, R. W., Jakel, F., & Maloney, L. T. (2011). Visual perception of thick transparent materials. *Psychological Science*, 22 (6), 812–820. <https://doi.org/10.1177/0956797611408734>
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14(9), 1195–1201. <https://doi.org/10.1038/nn.2889>
- Fry, R. E. (1920). *Vision and design*. Dover Publications.
- Getty. (2023). The J. Paul Getty museum. *Elements of art*. Retrieved July, 2023, from https://www.getty.edu/education/teachers/building_lessons/formal_analysis.html
- Gibson, J. J. (1950). *The perception of the visual world*. Houghton Mifflin Company.
- Gil-Glazer, Y. (2020). Visual culture and critical pedagogy: From theory to practice. *Critical Studies in Education*, 61(1), 66–85. <https://doi.org/10.1080/17508487.2017.1292298>
- Gombrich, E. H. (1960). *Art and illusion—A study in the psychology of pictorial representation*. Phaidon Press Limited.
- Gombrich, E. H. (1964). Light, form and texture in the XVth century painting. *Journal of the Royal Society of Arts*, 112 (5099), 826–849.
- Gombrich, E. H. (1995a). *Shadows: The depiction of cast shadows in Western art*. Yale University Press.
- Gombrich, E. H. (1995b). *The story of art* (16th ed. rev., expanded, and redesigned). Prentice-Hall.
- Goodman, N. (1976). *Languages of Art*. Hackett Publishing Company.
- Granito, C., Tehrani, J. J., Kendal, J. R., & Scott-Phillips, T. C. (2022). Does group contact shape styles of pictorial representation? A case study of Australian rock art. *Human Nature*, 33(3), 237–260. <https://doi.org/10.1007/s12110-022-09430-2>

- Jacobson, J., & Werner, S. (2004). Why cast shadows are expendable: Insensitivity of human observers and the inherent ambiguity of cast shadows in pictorial art. *Perception*, 33(11), 1369–1383. <https://doi.org/10.1068/p5320>
- Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature*, 290(5802), 91–97. <https://doi.org/10.1038/290091a0>
- Kennedy Center. (2023). *Formal visual analysis: The elements & principles of composition*. Retrieved July, 2023, from <https://www.kennedy-center.org/education/resources-for-educators/classroom-resources/articles-and-how-tos/articles/educators/visual-arts/formal-visual-analysis-the-elements-and-principles-of-composition/>
- Koenderink, J. (2015). Perceptual organization in visual art. In J. Wagemans (Ed.), *The Oxford handbook of perceptual organization* (pp. 886–916). Oxford University Press.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., & Zitnick, C. L. (2014). Microsoft coco: Common objects in context. *European Conference on Computer Vision*, 740–755.
- Mamassian, P. (2004). Impossible shadows and the shadow correspondence problem. *Perception*, 33(11), 1279–1290. <https://doi.org/10.1068/p5280>
- Marlow, P. J., Kim, J., & Anderson, B. L. (2012). The perception and misperception of specular surface reflectance. *Current Biology*, 22(20), 1909–1913. <https://doi.org/10.1016/j.cub.2012.08.009>
- Metelli, F. (1970). An algebraic development of the theory of perceptual transparency. *Ergonomics*, 13(1), 59–66. <https://doi.org/10.1080/00140137008931118>
- Metelli, F. (1974). The perception of transparency. *Scientific American*, 230(4), 90–98. <https://doi.org/10.1038/scientificamerican0474-90>
- Milani, F., & Fraternali, P. (2021). A dataset and a convolutional model for iconography classification in paintings. *Journal on Computing and Cultural Heritage*, 14(4), 1–18. <https://doi.org/10.1145/3458885>
- Müller, M. G. (2011). Iconography and iconology as a visual method and approach. *The SAGE Handbook of Visual Research Methods*, 1, 283–297. <https://doi.org/10.4135/9781446268278.n15>
- Munsterberg, M. (2009). *Writing about art*. <https://writingaboutart.org>
- Oliva, A., Torralba, A., & Schyns, P. G. (2006). Hybrid images. *ACM Transactions on Graphics (TOG)*, 25(3), 527–532. <https://doi.org/10.1145/1141911.1141919>
- Panofsky, E. (1955). Iconography and iconology: An introduction to the study of renaissance art. In *Meaning in the Visual Arts*, 26–55.
- Panofsky, E. (2020). *Perspective as symbolic form*. Princeton University Press.
- Pont, S. C., Wijntjes, M. W. A., Oomes, A. H. J., van Doorn, A., van Nierop, O., de Ridder, H., & Koenderink, J. J. (2011). Cast shadows in wide perspective. *Perception*, 40(8), 938–948. <https://doi.org/10.1068/p6820>
- Rogers, B., & Graham, M. (1979). Motion parallax as an independent cue for depth perception. *Perception*, 8(2), 125–134. <https://doi.org/10.1068/p080125>
- Sayim, B., & Cavanagh, P. (2011). The art of transparency. *I-Perception*, 2(7), 679–696. <https://doi.org/10.1068/i0459aap>
- Smarthistory.org *Introduction: Close looking and approaches to art*. Kilroy-Ewbank, L. Retrieved July, 2023, from <https://smarthistory.org/reframing-art-history/introduction-close-looking-approaches/>
- Sontag, S. (1966). *Against interpretation*. Picador.
- Strezoski, G., & Worring, M. (2018). Omniart: A large-scale artistic benchmark. *ACM Transactions on Multimedia Computing, Communications and Applications*, 14(4), 1–21. <https://doi.org/10.1145/3273022>
- Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, 1(3), 183–184. <https://doi.org/10.1038/630>
- Van Doorn, A. J., Koenderink, J. J., Leyssen, M. H. R., & Wagemans, J. (2012). Interaction of depth probes and style of depiction. *I-Perception*, 3(8), 528–540. <https://doi.org/10.1068/i0500>
- Van Zuijlen, M. J. P., Lin, H., Bala, K., Pont, S. C., & Wijntjes, M. W. A. (2021). Materials in paintings (MIP): An interdisciplinary dataset for perception, art history, and computer vision. *PLoS One*, 16(8), e0255109. <https://doi.org/10.1371/journal.pone.0255109>
- Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin*, 138(6), 1172–1217. <https://doi.org/10.1037/a0029333>
- Wallach, H., & O'Connell, D. N. (1952). The kinetic depth effect. *Journal of Experimental Psychology*, 45(4), 205–217. <https://doi.org/10.1037/h0056880>
- Ware, C. (1993). The foundations of experimental semiotics: A theory of sensory and conventional representation. *Journal of Visual Languages & Computing*, 4(1), 91–100. <https://doi.org/10.1006/jvlc.1993.1006>
- Ware, C. (2019). *Information visualization: Perception for design*. Morgan Kaufmann.
- Wijntjes, M. W. A. (2020). Shadows, highlights and faces: The contribution of a “human in the loop” to digital art history. *Art and Perception*, <https://doi.org/10.1163/22134913-bja10022>
- Wijntjes, M. W. A., & de Ridder, H. (2014). Shading and shadowing on Canaletto's Piazza San Marco. *IS&T/SPIE Electronic Imaging*, 901415–901416.
- Willats, J. (1997). *Art and representation: New principles in the analysis of pictures*. Princeton University Press.
- Wölfflin, H. (1915). *Principles of art history*. Courier Corporation.
- Wollheim, R. (1998). On pictorial representation. *The Journal of Aesthetics and Art Criticism*, 56(3), 217. <https://doi.org/10.2307/432361>
- Wollheim, R., & Eldridge, R. T. (2015). *Art and its objects: With six supplementary essays* (2nd ed., Cambridge Philosophy Classics edition). Cambridge University Press.
- Zhang, T., Nefs, H. T., Redi, J., & Heynderickx, I. (2014). *The aesthetic appeal of depth of field in photographs*.

Appendix: Picture set curation

At the time of writing, we have used the same curated image set for four years. The images were normally offered on a webpage where they were directly linked to a reverse images search service (such as Google or Bing). As these services function unpredictably over a long period (e.g., Google may cancel automatic accessibility), we only offer the images here:

<https://visionanddepiction.s3.eu-central-1.amazonaws.com/grandgallery.html>

The number of chosen images has to be large enough (around 40/50 pictures at least) to suggest a wealth of options, and trigger a multitude in different points of view. The pictures should be ordered in such a way that no additional meaning (beyond the meaning captured within the picture frame/image borders) is projected through hierarchy. In a table this could be done by listing words alphabetically, in the picture overview we use to present our image selection to the students the image placement is based on image proportions (ranking from widest to tallest). This natural flow throughout the representation of the image collection allows students to focus on what happens within the picture frame.

Ideally all elements from our framework are equally represented in the images: around 12 pictures representing “material,” 12 being illustrative of “space,” 12 exemplifying “colour/light,” etc.

This works well if an artist/maker creates works that fit into different categories. In this way, the framework can be put to use in

a more holistic manner; students start to see how the different elements relate.

It is recommendable not to use the same artist/maker twice within the cluster of one element. Similarities in style could lead to visual dominance of one maker, potentially shifting the focus away from the picture and towards an interesting yet invisible context.

Per element, a mixture of production techniques used to create the pictures is desirable. The broader the range and the more extreme the examples, the more curious the students become: for example, some pictures that are made of egg tempera and others that are constructed out of pixels in the cluster of colour/light.

As students are completely free in making their own selection from the “given picture set,” they will often come up with completely different groupings from what the curator of the picture set might have intended or anticipated. Still, it is important to curate the entire set with the intention to be as “complete”/“visually saturated” as possible, and with the interconnectedness of the individual pictures in mind.

Their interconnectedness could be based on, for example, recurring visual elements throughout the picture set (such as a centred circle, an horizon, a specific colour), but also occurring on a not immediately visual level (age of image, size of the image in real life, or the human/non-human nature of the maker).

Further information on the course can be found at <https://whenimagesremain.github.io>.