Non-Photorealistic Rendering in Context: An Observational Study

Tobias Isenberg* Petra Neumann* Sheelagh Carpendale* Mario Costa Sousa* Joaquim A. Jorge†
Department of Computer Science
University of Calgary, Canada
Instituto Superior Técnico
Lisboa, Portugal

Abstract

Pen-and-ink line drawing techniques are frequently used to depict form, tone, and texture in artistic, technical, and scientific illustration. In non-photorealistic rendering (NPR), considerable progress has been made towards reproducing traditional pen-and-ink techniques for rendering 3D objects. However, formal evaluation and validation of these NPR images remain an important open research problem. In this paper we present an observational study with three groups of users to examine their understanding and assessment of hand-drawn pen-and-ink illustrations of objects in comparison with NPR renditions of the same 3D objects. The results show that people perceive differences between those two types of illustration but that those that look computer-generated are still highly valued as scientific illustrations.

CR Categories: I.3.0 [Computer Graphics]: General

Keywords: Non-photorealistic rendering (NPR), evaluation of NPR and traditional scientific illustration, observational study, pen-and-ink illustration.

1 Introduction and Motivation

After more than two decades of intensive research, non-photorealistic rendering (NPR) is an established and important field within computer graphics [Gooch and Gooch 2001; Strothotte and Schlechtweg 2002]. NPR breaks free from the constraint of (photo-)realism that many other rendering techniques strive for. It generates images and animations that at least in some parts appear to be made by hand [Strothotte and Schlechtweg 2002]. Within NPR, one exciting direction deals with how computers can be used to generate line drawing illustrations, often with the goal to depict scientific subject matter. For achieving "non-photorealism" in these types of images, NPR takes inspiration from a long tradition of artistic and illustrative depiction. Over hundreds of years, hand-drawn scientific illustrations have achieved a high level of sophistication. NPR often tries to imitate long established illustration techniques but we strongly feel that the NPR research has reached a point of sophistication at which it is time to halt and investigate where our research stands compared to hand-drawn illustrations.

We conducted an observational study to examine how people understand and assess both traditionally created hand-drawn illustrations and computer-generated non-photorealistic illustrations. We chose pen-and-ink illustrations as our study domain since the techniques have reached a mature state, having been a topic of NPR research for more than ten years. Pen-and-ink illustration is inspired by illustration techniques frequently found in hand-drawn scientific illustrations and provides a relatively cohesive domain. The purpose of this study was to improve our understanding of the differences between hand-drawn and computer-generated images at the current stage of NPR research and open up new or validate current research directions for the NPR community. We studied how people view both types of images, asked about their imaginable contexts of usage, assessed participant's likes and dislikes, and asked about directions for improvement of images that were found to be less appealing. To some degree this study is also an attempt of a Turing test for pen-and-ink line drawings since we determined which images were described as having a computer-generated feel to them. Our findings reveal that there are still obvious differences between computer-generated and hand-drawn illustrations.

We start by discussing related work in Section 2. Then we explain the design of the observational study in Section 3. Afterwards, Section 4 presents the results and general observations. In the next Section 5 we discuss and interpret these results. Finally, we conclude with a summary and suggestions for future work in Section 6.

2 Related Work

Relatively few papers have been devoted specifically to evaluations of NPR methods, systems, and images. We consider these studies from six categories of evaluation goals and applications as relevant.

Communication in architecture and design: The first NPR evaluation [Schumann et al. 1996] studied the usability of computer-generated images with respect to communicative goals during design concept development with 54 architects as participants. These participants were shown three different images portraying the same architectural object: a CAD plot image, a constant shading image and a NPR image generated by a sketch-renderer developed by the authors [Strothotte et al. 1994]. Their results show that these three image types have different effects on viewers. For example, NPR images appeared to invite more interaction than the other two types.

Space perception in immersive environments: The first experiments to examine and evaluate space perception in a functional, non-photorealistically rendered immersive environment [Gooch and Willemsen 2002], involved direct walking tasks in a physical hallway and NPR renditions (silhouettes, boundaries, and creases) of a 3D model of the same hallway, visualized through a head-mounted display. Their study provided important indications of the degree to which NPR images are capable of conveying a veridical sense of spatial organization.

Psychology of NPR: Duke et al. [2003] explored the affective qualities of images in a series of experiments. Their main conclusion was that understanding of rendered images requires models that go beyond perception to harness the dynamics of semantic processing in the context of specific tasks. Indeed, they showed through experimental evidence that rendering styles can convey meaning and influence judgment in non-trivial ways.

---

*e-mail: {isenberg | pneumann | sheelagh | mario}@cpsc.ucalgary.ca
†e-mail: jorgej@acm.org

In a similar vein, Halper et al. [2003a; 2003b] suggested applying psychological theory to NPR in terms of biological, social, and environmental paradigms emerging from studies to explore the relationship between rendering style and affect. The authors assessed NPR styles by psychological measures ranging from statistical analysis of user selections to direct analysis of brain activity. One interesting result identified elements and patterns used in rendering styles which directly affect social perceptions such as danger, safety, strength, or weakness. This suggests that insights into psychological dimensions of rendering can be successfully used to select effective presentation styles in the rendering pipeline, encouraging the use of NPR tools by non-experts towards intent-driven illustration.

Santella and DeCarlo [2004] conducted a study in which they employ eye tracking to determine where viewers look in regular photographs and NPR images, the latter being created with high detail, low detail, and selective preserved detail according to eye tracking data as well as salience maps. Their results show that viewers' gazes are drawn to more detailed regions in partially abstracted images, supporting the empirical findings of Strothotte et al. [1994]. From that the authors conclude that meaningful abstraction is essential for NPR styles to be effective and that more emphasis should be put on creating meaningful instead of uniform abstractions.

Recognition and learning from facial illustration and caricatures: Gooch et al. [2004] conducted a psychophysical study to observe the influence of facial illustration and caricature algorithms in humans. Specifically, they assessed the effect on speed and accuracy of recognition and learning faces by humans. To this end, subjects were presented with sequences of familiar and unfamiliar faces. Results showed that computer-generated illustrations and caricatures were as effective as photographs in recognition tasks. For the learning tasks, illustrations were two times faster and caricatures were one and a half times faster than photographs.

Simulation and training in medical visualization: Tietjen et al. [2005] published the first evaluation study of NPR techniques applied to medical data for surgery planning and training. In their study, various renditions were presented along with questionnaires to different groups of users including medical doctors and medical laypersons such as patients who usually receive pre-surgery consultation. Their study indicated the importance of silhouette in combination with transparent surfaces and hybrid visualizations (i. e., combinations with iso-surfaces and direct volume rendering).

Shape and data depiction from textures: Kim et al. [2004] describe two comprehensive experiments that assess the effectiveness of texture patterns in conveying 3D shape information. They investigated how particular texture components influenced shape perception. To this end, they assessed the ability of observers to identify intrinsic shape features and surface orientation using different viewing conditions. They found that the "principal direction grid" pattern was more effective at promoting shape perception and that oblique viewpoints seem to favor shape classification.

In a study on a similar subject, Jackson et al. [2003] examined how well different visualization techniques allow viewers to comprehend flow fields. To back up a previous study based on numeric measurements, they asked a professional graphic designer to comment on the visualization. They found that the subjective results of the designer's critique corresponded well to the previously measured numeric values for performance. In a more recent study, Acevedo et al. [2005] evaluated the effectiveness of different 2D visualization methods by asking university design educators to critique them, an approach that is very similar to ours.

In contrast to these previous approaches, we are the first to compare non-photorealistic with hand-made pen-and-ink illustrations in an observational study.

3 Observational Study

The goal of our study was to gain understanding in the way viewers evaluate traditional hand-drawn illustrations compared to computer-generated techniques. Since this assessment of images is a complex, interwoven, and difficult variable to measure, our intention was to look for larger trends, patterns, or assessment styles. With these types of qualitative information goals, we did not set out with a predetermined hypothesis of how our participants would assess the provided images. Hence, our chosen study approach is taken from the field of *qualitative research* [Denzin and Lincoln 2000]. This type of research involves the gathering and use of qualitative data, such as from interviews and participant observation data and is often conducted in situations when appropriate measurable variables are not known. This is distinct from quantitative research that measures pre-defined variables according to quantity, amount, intensity, or frequency and then uses statistical models to explain the results. While quantitative data allows the testing of hypotheses through statistics, qualitative data is descriptive and, thus, allows insight into process, meaning, and understanding from which researchers can build abstractions, concepts, or hypotheses about human experience [Creswell 2000]. Since we wanted to discover the types and varieties of classifications and categories that people used when thinking about illustration, the qualitative approach was most appropriate.

Our qualitative analysis involves a combination of user observation with systematic data collection and semi-structured interviewing. During the observation period, we gave the participants an *unconstrained pile-sort task* [Weller and Romney 1988], a task suitable to relate data items, in which the participant rather than the researcher determines the salient criteria for distinguishing between the items.The unconstrained pile-sorting enabled the participants to familiarize themselves with the images without specific assessment criteria in mind and also allowed us to better extract their unbiased opinions about the images. In the following second part, the semi-structured interview, we asked the participants for their assessment of the shown images, discussing their categorizations and further questions in more detail.

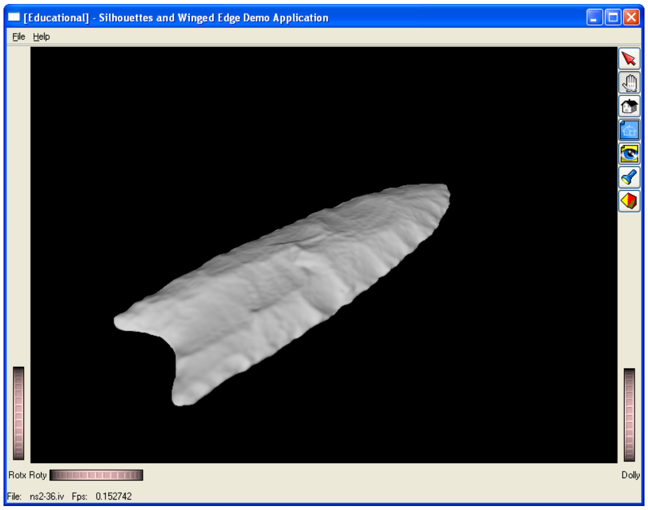
3.1 Participants

For the initial setup of this study we identified four main groups of people who work with illustrations. The first group, *domain experts* in a field, often use illustrations to teach or learn, and have a very good understanding of the objects or processes to be illustrated. They typically either use tools to generate the illustrations themselves or get *professional illustrators*, the second group, to create them. *Illustration end users* take illustrations to learn about the depicted subjects without having prior expert knowledge. This group includes students ranging from kindergarten to university students and beyond, as well as the general public. The last group we were interested in consists of *NPR researchers* whose work involves developing tools to create computer-generated illustrations.

For this observational study, we restricted ourselves to study three of the above four groups. Eight *illustration end users* (general university population; 5 male, 3 female), eight *professional illustrators* (either advanced art students or professional artists and illustrators, both with a background in drawing; 3 male, 5 female), and eight *NPR researchers* (graduate computer graphics students with an NPR background; 7 male, 1 female) participated in our study (24 participants in total).

3.2 Apparatus

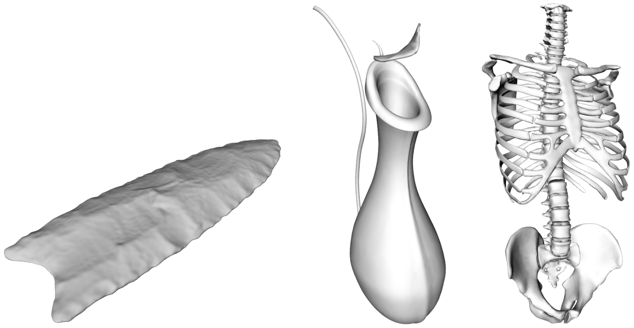
The experiment was conducted in a small, quiet, naturally-lit room on a large table with 30 different study images. These comprised scientific illustrations of three models, an *archaeological model* consisting of an arrowhead with a number of bumps and ridges on an otherwise flat surface, a *botanical model* consisting of the trap of a tropical pitcher plant with a smooth surface and mostly round shape, and a spatially more complex *medical model* of a human skeleton's torso. For each object we acquired ten images, five computer-generated and five hand-drawn by professional illustrators. To allow a balanced comparison between hand-made and computer-generated illustrations, all images showed approximately the same view of the objects. While this is relatively easy to achieve using NPR tools, we created a simple 3D viewer for the illustrators that showed the Gouraud-shaded model in gray-scale (see Figure 1).



File: ns2-36.iv Fps: 0.152742

Figure 1: Screenshot of simple viewer provided to the illustrators.

It consisted of a separate Win32 application for each of the three models. Each model could be rotated using the mouse and a virtual trackball to give the illustrators the chance to examine the objects from all sides. However, the viewer applications included a default view (see Figure 2) from which the objects were to be drawn.



(a) Archaeological model. (b) Botanical model. (c) Medical model.

Figure 2: The default views of the archaeological, botanical, and medical models as presented to the illustrators.

For each of the three models, five professional illustrators were asked to draw one image using their preferred pen-and-ink technique. We received both completely hand-made drawings, drawings created with computer support in printed form, and scanned-in hand-made drawings. To generate images using the computer we used five fairly recent NPR techniques matching the five illustrators to have an equal number of hand-made and computer-generated images. We chose NPR methods which emulate pen-and-ink techniques such as stippling, hatching, and cross-hatching to match the ones the illustrators were also asked to use. In addition, we chose techniques that worked on triangular models or gray-scale images and produced high quality and high resolution output, adequate for print reproduction (vector graphics or high resolution pixel images, all in black-and-white). The default views for the computer-generated images were slightly altered (rotated by a few degrees) to match the slight variations in the images received from the professional illustrators. Figures 3 and 4 give an overview of the hand-drawn and computer-generated illustrations and Table 1 provides the study images' authors as well as the NPR techniques used to create them.

To ensure that images could not be told apart by paper or ink usage we scanned all images received from illustrators in paper form as 1200 dpi black-and-white images and printed them in their original size (using a HP LaserJet 4100 PS at 1200 dpi). We also printed all computer-generated images, their sizes were slightly varied but kept similar to those used by the illustrators. This scan-and-print step also simulates the usage of illustrations as they appear in printed books. The paper used for all printing was 216 mm by 279 mm (Letter sized) color laser paper with $105\,\frac{\mathrm{g}}{\mathrm{m}^2}$ weight.

3.3 Procedure

Each study session involved a single participant working through three stages: pile-sort, interview, and questionnaire as follows, with data being recorded using video, audio, and field notes:

Stage 1: *Pile-sorting task.* Pile-sorting is a method for systematic data collection in qualitative research that is easily understood by participants and facilitates conversation [Weller and Romney 1988] which was important for Stage 2 of our study. Participants were given the 30 illustrations with a different random order for each participant. They were then asked to sort them into piles according to their own criteria. As an example, the participants were shown how the images could be sorted by object or by illustration size and asked not to use these two sample criteria for their sorting. We chose an *unconstrained* pile-sort approach in which participants could make as many piles as they wanted and take as much time as they wanted. We did not restrict the number of piles participants were allowed to make in order to maintain the open character of this task. Participants were also asked to think-aloud during the pile-sort in order to be able to capture their thoughts on video.

Stage 2: *Semi-structured interview.* After the sorting was finished, the images in the piles were, while keeping piles together, spread on the table so that each image was visible and could be discussed by itself or in connection to its category (see Figure 5). The participants were first asked to identify the criteria that lead to each pile. In the remainder of the interview, the participants were asked the following questions:

- Which images do you like best? Why?
- Which images do you like least? Why?
- In what context would you use any of these images? Where would you like to see them? Why?
- Which images would you use in university textbooks? Why?
- Which images would you use in textbooks for children (late kindergarten to early elementary school)? Why?

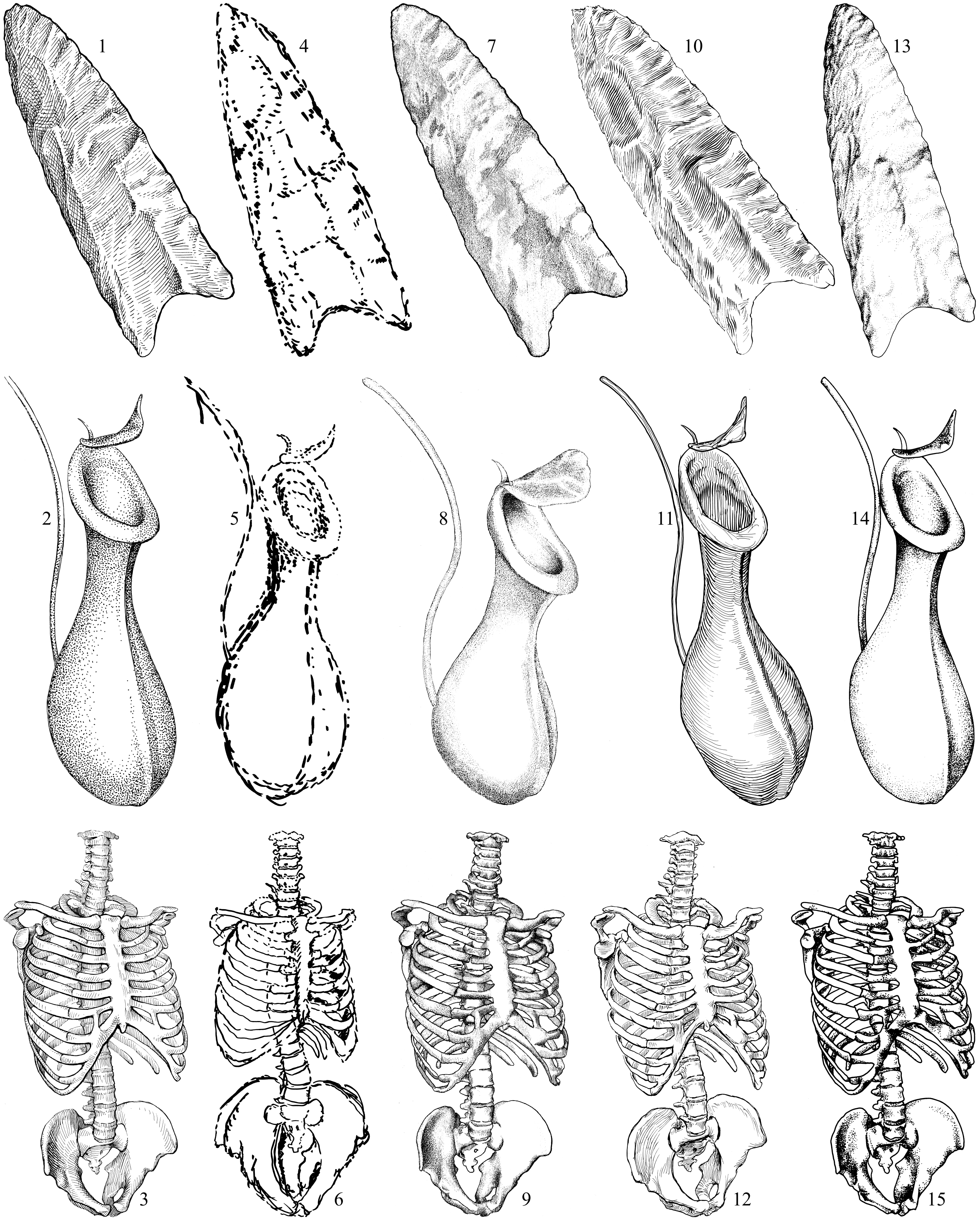


Figure 3: All hand-drawn study images with their numbers used throughout the paper. All images are copyright of their respective authors (refer to Table 1), used with permission.

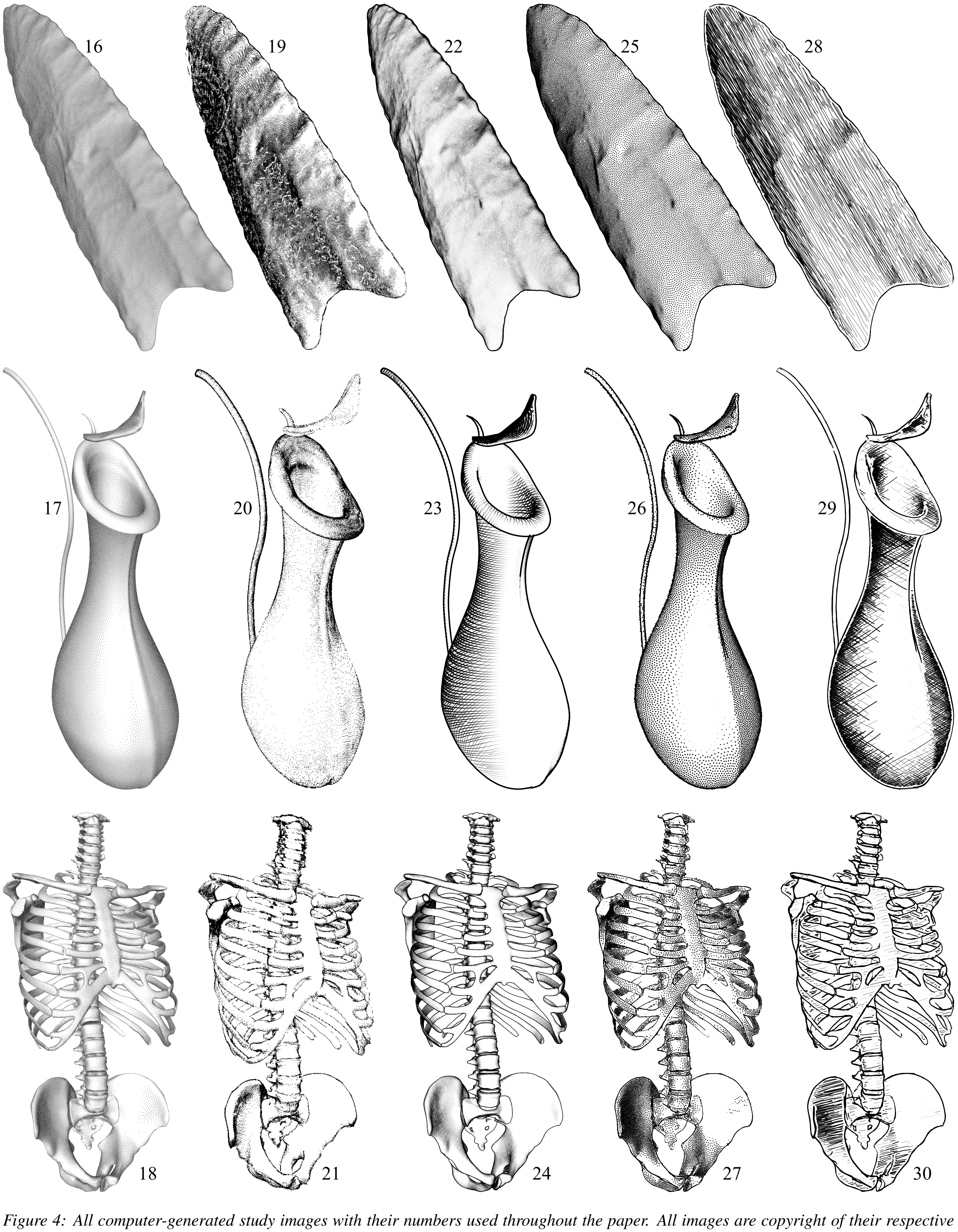


Figure 4: All computer-generated study images with their numbers used throughout the paper. All images are copyright of their respective authors (refer to Table 1), used with permission.

images	illustration author	drawing	drawing or NPR technique	file type	resolution
1–3	William M. Andrews	hand	cross-hatching or stippling w/ silhouettes	pixel	1200 dpi
4–6	Davide Brunelli	hand	sketchy hatched outlines	pixel	1200 dpi
7–9	Humberto Costa Sousa Filho	hand	silhouette, scratchboard w/ hatching and some stippling	pixel	1200 dpi
10–12	Andrew E. B. Swift	hand	hatching or cross-hatching w/ silhouettes	pixel	1200 dpi
13–15	Lynda Smith Touart	hand	stippling w/ outlines	pixel	1200 dpi
16–18	Tobias Isenberg	NPR	[Secord 2002], stippling w/o silhouettes	vector	n/a
19–21	Mario Costa Sousa	NPR	[Sousa et al. 2003; Sousa et al. 2004], precise ink marking	pixel	128 dpi
22–24	Tobias Isenberg	NPR	[Zander et al. 2004], cross-hatching w/ silhouettes	vector	n/a
25–27	Tobias Germer	NPR	[Schlechtweg et al. 2005], stippling w/ silhouettes	vector	n/a
28–30	Tobias Germer	NPR	[Schlechtweg et al. 2005], cross-hatching w/ silhouettes	vector	n/a

Table 1: Data about the illustration images used in the study. The first image of each group is always the archaeological model (e. g., 1, 4, 7, etc.), the second one the botanical model (e. g., 2, 5, 8, etc.), and the third one the medical model (e. g., 3, 6, 9, etc.). Also see Figures 3 and 4.

Figure 5: Discussion session after the pile-sorting.

- Which images have the most computer-generated feel? Why?
- Which images look most hand-drawn? Why?
- Is there anything else that you noticed about these images?

Stage 3: *Post-session questionnaire.* Participants completed a questionnaire asking about their experience creating illustrations by hand or with the computer and their experience in viewing and working with illustrations.

4 Results

We present results of the study according to the stages of the experiment, discussing the pile-sort first and the interview afterwards.

4.1 Pile-Sorting Task

The pile-sorting stage took 33:38 minutes on average; 29:54 minutes for general students, 34:02 minutes for NPR students, and 36:57 minutes for artists. Participants used five different categorization types to sort images in piles and four participants used more than one category to create piles. Table 2 gives an overview of the category types used by our three groups.

Most people categorized with respect to drawing / rendering style based on the type of mark being used (lines vs. dots), line style, or tone. Other criteria were the amount of realism and detail, information content, object orientation, or the overall look and feel of an image in terms of aesthetics. We performed a Fisher's exact test on Table 2 to determine if the categories depend on the three groups. The results showed no statistically significant difference ($p = 0.981$, two-sided Fisher's exact test) which indicated that the two variables, categorization and group, are independent. We, therefore, report combined results of all groups below.

category	general students	artists	NPR students	sum
drawing style	5	5	7	17
realism / detail	2	2	2	6
aesthetics	1	1	1	3
other	1	1	-	2

Table 2: Categories used for sorting the images by our participant groups. Some participants used more than one category for sorting. 'Other' categories are information content and orientation / view.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1	24	2	11	6	2	3	2	3	9	14	12	15	3	2	2	0	0	0	2	2	1	5	12	6	2	3	2	11	10	8
2	2	24	2	3	3	3	6	9	3	2	3	2	13	15	11	5	5	3	7	14	7	1	2	2	8	15	10	3	2	3
3	11	2	24	1	0	3	2	2	9	14	15	16	2	2	6	1	0	1	2	2	3	2	13	6	3	4	5	11	9	13
4	6	3	1	24	20	20	3	2	1	5	2	1	4	3	3	1	1	1	1	3	3	0	1	1	2	2	2	5	2	4
5	2	3	0	20	24	19	2	3	0	2	3	0	2	4	2	1	1	1	1	3	3	0	1	0	1	2	1	2	3	3
6	3	3	3	20	19	24	2	2	2	3	2	3	3	3	4	1	1	1	1	3	4	0	1	1	2	2	3	3	2	6
7	2	6	2	3	2	2	24	12	7	2	2	1	10	6	6	4	4	3	9	9	11	4	3	0	6	4	6	4	2	3
8	3	9	2	2	3	2	12	24	8	2	2	3	9	8	6	3	4	2	8	11	10	3	3	2	4	7	4	3	2	3
9	9	3	9	1	0	2	7	8	24	7	7	11	4	3	4	1	3	4	6	5	6	5	9	8	4	2	6	6	9	8
10	14	2	14	5	2	3	2	2	7	24	19	14	3	2	2	0	0	0	1	2	0	2	8	4	2	3	2	13	9	11
11	12	3	15	2	3	2	2	2	7	19	24	16	2	4	3	0	0	0	1	3	1	2	10	4	2	4	3	11	11	11
12	15	2	16	1	0	3	1	3	11	14	16	24	2	2	6	0	0	1	2	2	4	3	11	7	3	4	5	9	9	11
13	3	13	2	4	2	3	10	9	4	3	2	2	24	16	14	6	4	3	6	16	9	2	1	1	8	10	9	4	1	3
14	2	15	2	3	4	3	6	8	3	2	4	2	16	24	17	5	5	3	5	14	6	2	2	1	7	12	10	3	3	3
15	2	11	6	3	2	4	6	6	4	2	3	6	14	17	24	3	3	4	4	10	9	1	3	4	8	11	10	4	2	6
16	0	5	1	1	1	1	4	3	1	0	0	0	6	5	3	24	16	15	8	5	3	11	1	4	7	4	4	0	1	0
17	0	5	0	1	1	1	4	4	3	0	0	0	4	5	3	16	24	18	7	5	4	9	2	4	6	4	6	0	1	0
18	0	3	1	1	1	1	3	2	4	0	0	1	3	3	4	15	18	24	6	3	5	9	0	8	5	2	6	0	0	1
19	2	7	2	1	1	1	9	8	6	1	1	2	6	5	4	8	7	6	24	5	11	7	4	4	11	5	7	0	4	0
20	2	14	2	3	3	3	9	11	5	2	3	2	16	14	10	5	5	3	5	24	9	2	1	2	7	12	10	3	3	3
21	1	7	3	3	3	4	11	10	6	0	1	4	9	6	9	3	4	5	11	9	24	3	3	3	4	4	7	2	3	4
22	5	1	2	0	0	0	4	3	5	2	2	3	2	2	1	11	9	9	7	2	3	24	6	13	4	0	1	2	4	2
23	12	2	13	1	1	1	3	3	9	8	10	11	1	2	3	1	2	0	4	1	3	6	24	9	4	4	4	8	14	8
24	6	2	6	1	0	1	0	2	8	4	4	7	1	1	4	4	4	8	4	2	3	13	9	24	3	2	4	4	7	5
25	2	8	3	2	1	2	6	4	4	2	2	3	8	7	8	7	6	5	11	7	4	4	4	3	24	17	12	1	3	1
26	3	15	4	2	2	2	4	7	2	3	4	4	10	12	11	4	4	2	5	12	4	0	4	2	17	24	13	2	3	2
27	2	10	5	2	1	3	6	4	6	2	3	5	9	10	10	4	6	6	7	10	7	1	4	4	12	13	24	2	4	4
28	11	3	11	5	2	3	4	3	6	13	11	9	4	3	4	0	0	0	0	3	2	2	8	4	1	2	2	24	13	18
29	10	2	9	2	3	2	2	2	9	9	11	9	1	3	2	1	1	0	4	3	3	4	14	7	3	3	4	13	24	13
30	8	3	13	4	3	6	3	3	8	11	11	11	3	3	6	0	0	1	0	3	4	2	8	5	1	2	4	18	13	24

Table 3: Correlation table for the pile-sort task (all participants).

An image-by-image similarity matrix (see Table 3) was created from each individual's pile-sort by tabulating the co-occurrence of images in piles so that those that were grouped together were

image	4	6	8	14	21	5	13	20	7	9	15	19	25	26	28	2	23	27	30	29	1	3	10	16	18	22	12	11	24	17
hd vs. cg	hd	hd	hd	hd	cg	hd	hd	cg	hd	hd	hd	cg	cg	cg	cg	hd	cg	cg	cg	cg	hd	hd	hd	cg	cg	cg	hd	hd	cg	cg
# liked	0	1	1	1	1	2	2	2	3	3	3	3	3	3	3	4	4	4	4	5	7	7	8	8	9	9	10	11	11	14
%	0	4	4	4	4	8	8	8	13	13	13	13	13	13	13	17	17	17	17	21	29	29	33	33	38	38	42	46	46	58

image	3	9	17	18	24	27	1	2	7	14	15	22	8	10	11	12	23	25	30	13	20	21	16	26	19	29	28	5	6	4
hd vs. cg	hd	hd	cg	cg	cg	cg	hd	hd	hd	hd	hd	cg	hd	hd	hd	hd	hd	cg	cg	hd	cg	cg	cg	cg	cg	cg	cg	hd	hd	hd
# disliked	0	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	4	4	5	6	14	15	15	17
%	0	0	0	0	0	0	4	4	4	4	4	4	8	8	8	8	8	8	8	13	13	13	17	17	21	25	58	63	63	71

Table 4: Answers on whether a certain image is particularly liked or particularly disliked (all participants, increasing order).

counted as being similar. We combined these similarity matrices and analyzed them with hierarchical clustering using average linkage (between-groups) with chi-square as the dissimilarity measure.

Figure 6 gives an overview of the hierarchical cluster analysis represented by a 2D dendrogram that illustrates the fusions or divisions of clusters made at each successive stage of analysis. Here, the different cluster steps are displayed on the y-axis and the data items on the x-axis. The Step 1 links show that the two linked images were frequently placed in one pile. For example, on the left hand side Images 4 & 6 were in the same pile for 20 of the 24 participants. A small separation in step number for subsequent links indicates a similar frequency of co-placement in a pile, a high level of consistency between images, or a lack of distinction. Following the same example Images 4, 5, & 6 are all linked at Step 3 (Images 4 & 5 were also piled together by 20 participants and Images 5 & 6 were in the same pile for 19 participants). Conversely, the greater differences in clustering step, the lower the degree of consistency or similarity between linked clusters. These inconsistent links can represent a natural division in the data set. Applying these criteria indicates four clusters as highlighted in Figure 6. A discriminant function analysis showed a statistically significant difference between the clusters (Test of Functions: 1 through 3, Wilk's $\lambda < 0.001, \chi^2(78) = 283.449, p < 0.001$). The four clusters are discussed next.

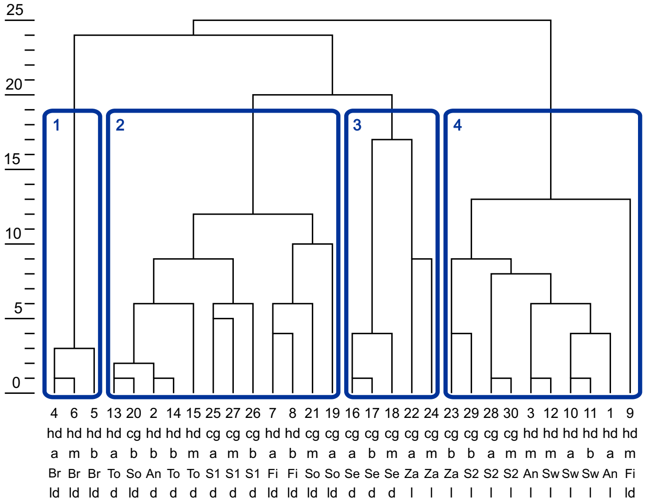


Figure 6: Cluster graph for all participants. The first data line contains the image number (refer to Table 1 and Figures 3 and 4), the second line describes whether the image was hand-drawn (hd) or computer-generated (cg), the third line the model (archaeological, biological, or medical), the fourth line the first two letters of the illustrator's or NPR algorithm's first author's last name, and the last line whether it contains mainly dots (d), lines (l), or a mixture (ld). Four clusters are emphasized (blue boxes) and discussed in the text.

- Cluster 1 includes three hand-drawn images by the same artist. The images in this cluster are clearly distinct from the other images. As detailed in the above explanation these images were grouped in piles by approximately 80% of the participants. Piles that included these images were titled *"less clear, simple drafts, sketchy, incomplete drawing, artistic, not very detailed, general shape, little tone."*

- Cluster 2 consists of illustrations that use stippling or a combination of dots and lines. It contains both computer-generated and hand-drawn illustrations. The illustrations in this cluster were drawn by two artists and two NPR algorithms. Smaller sub-clusters appear that also show a combination of hand-drawn and computer-generated images, in particular subclusters (13, 20, 2, 14) and (7, 8, 21).

- Cluster 3 contains highly-detailed computer-generated images with stippling or hatching marks. The images were generated using the methods by Secord [2002] and Zander et al. [2004].

- Cluster 4 contains hatched illustrations and one image that uses a mixture of lines and dots (Image 9). The illustrations are both hand-drawn and computer-generated and were drawn by two algorithms and three artists.

4.2 Semi-Structured Interview

During the interview part we gathered information about the participants' criteria for establishing piles and then asked more specific questions about the images that were spread out on the table. In general, the three user groups had similar opinions about the images. However, the artists were more outspoken and articulate in discussing their thoughts and clearly had previous experience and their own vocabulary for describing the illustrations. NPR researchers investigating line rendering or stippling usually looked at the criteria important for their own research to judge the images. Participants in the general category described the images in broader terms and mostly from their experience with viewing or teaching using illustrations and diagrams. They also talked about their experience with using graphical editors, GIS systems, or printers to compare the images. In the following we will discuss the results according to the questions asked.

Which images do you particularly like or appeal to you? This question was asked in an open way. Participants could decide by themselves what criteria to set for determining what for them made a good illustration. Many participants answered this question in the context of scientific illustration by how clearly rendered images were, how well they depicted shape, or their amount of 'realism'; others went by the aesthetic appeal of the images; or a mixture of both. There was no absolute favorite illustration (see Table 4). Of the top 3, top 6, and top 10 ranked images, 2 (66.6%), 4 (66.6%), and 5 (50.0%) were computer-generated, respectively.

Most participants answered this question in the context of scientific illustration. This is the reason why Image 4, 5, 6, and 28 were often named to not be liked as much. These images were described as too sketchy and defining shape less well in the context of learning and teaching. On the other hand, outside of the context of scientific illustrations, participants found these images to be valuable in an arts context and as having an interesting drawing style to look at.

In what context would you like to see or would you use these images? This question was mostly answered in context to specific images or piles. 22 participants mentioned that most images were useful for instructive purposes, as in scientific or artistic textbooks, as diagrams in classes, for museum displays, magazines, periodicals, journals, papers, or encyclopedias and dictionaries. Others suggested usage as art displays (8), comics or graphic novels (3), computer display for games, webpages, or in interactive software (3), bathroom tiles (1), restaurant signs (1), or advertisement (1).

Which images would you use in university textbooks? Table 5 shows which images were named most suitable for inclusion in a university level textbook. Except for the hand-drawn Image 12, these are all computer-generated images that stood out as being computer-generated. They were described as *"highly detailed, more realistic, better shading, good texture display, traditional illustration style, clear, good sense for 3D"* etc. However, not all computer-generated images that stood out as being computer-generated were also named to be good for university textbooks (Images 24, 25, and 26). Except for Image 19, the images that participants named in this question are in the top 6 most 'liked' category (Table 4). The computer-generated Image 19 was only particularly liked by 3 participants, while the hand-drawn Image 11, the third most liked image, was not under the top 6 named to be suitable for a university level textbook.

image	24	12	17	22	19	18
#	8	9	10	10	13	15
%	33%	38%	42%	42%	54%	63%

Table 5: Images named most suitable for a university level textbook.

Which images would you use for children's textbooks? Images most often named as suitable for children's textbooks (kindergarten to early elementary school level) were Images 11, 12, 17, 1, and 18 which is obviously different from what was recommended for the university level (see Table 6). In general, we received diverging answers, some participants suggested to use simpler images for children, others thought children should see the same or even more detailed and realistic images compared to adults.

image	1	18	17	11	12
#	7	7	8	9	9
%	29%	29%	33%	38%	38%

Table 6: Images named most suitable for a children's textbook.

Which images look most computer-generated and which look most hand-drawn? In the pile-sorting task, none of the participants set out to make piles that clearly distinguished computer-generated from hand-drawn images. With this question we tried to determine which images had the most obvious computer-generated or hand-drawn look and why. Participants did not have to classify every image but rather could pick a few that clearly stood out in their opinion (see Table 7). We observed that several categorizations included complete piles that participants described as looking computer-generated or hand-drawn, some participants even described all of their piles as looking one way or the other when asked, but this criterion was not consciously used to group them.

The results show that many hand-drawn images often stood out as such, most notably Images 4, 11, 5, 6, 10, 3, 12, and 1. All these images used a hatching drawing style. The hand-drawn images least often thought to stand out as hand-drawn were Images 2, 7, 13, 14, 9, 8, and 15. These images used stippling or a mixture of lines and dots as marks. Similarly, some computer-generated images were often named as standing out as such, most notably Images 18, 16, 17, 25, 22, 24, 26, and 19. These images used stippling or high density lines. Images least often named in this respect were Images 28, 30, 21, 29, 20, 23, and 27.

On the other hand, hand-drawn images were rarely thought to stand out as computer-generated with Image 7 being named most often (3 ×). Computer-generated images were a bit more often thought to stand out as hand-drawn, particularly Images 28 (16 ×), 30 (15 ×), 29 (11 ×), and 23 (7 ×). These images used longer lines with a lower density than images using lines that were seen to stand out as computer-generated. Figure 6 shows that these are also those computer-generated ones grouped in Cluster 3 with only hand-made images.

Participants in the general and NPR category overall had more difficulty describing what made an image look computer-generated or hand-drawn. Many mentioned that images looking obviously computer-generated had more of a 3D feel to it, were much more detailed, complex, and uniform looking. Images looking most hand-drawn to them looked more abstract, sketchy, rough, or free-form. In discussing this, artists paid a lot more attention to line quality. In addition to the criteria mentioned by general and NPR students for both types of images, many artists noted that images that look computer-generated used mechanical lines, that the line and mark making was not as apparent, or that patterns were emerging as stippling artifacts. They described hand-drawn images as being more organic, tentative, broken up, having different line qualities or line weights, and having little inaccuracies as well as inconsistencies.

5 General Discussion

Through the study we gained a lot of insight into both traditional hand-drawn and computer-generated scientific illustration. The study results show that participants can mostly distinguish between hand-drawn and computer-generated images but that those found to stand out as computer-generated were still often thought to be very successful in the context of scientific illustration. We also found differences in the assessment of illustrations using stippling vs. hatching styles. In the following section we discuss our results in more detail.

5.1 Pile-Sorting Task

During the pile-sort we observed very similar piling strategies by all participants. The initial piling task lead the participants to engage with the shown images and to form opinions about the illustrations. After the pile-sorting the participants discussed their thoughts and interpretations of the images freely with us, so the pile-sort formed a valuable tool to facilitate later discussion. From the cluster analysis of participants' piles we found three main clusters that each used a particular drawing style (hatching vs. stippling). This supports our observation that 17 participants used the drawing style as a sorting criterion but also suggests that the drawing style might play an important role for the other sorting criteria that were used (see Table 2). We also found that while there are, with the exception

image	3	5	6	10	11	4	12	13	15	28	30	1	2	8	9	14	21	7	29	20	23	27	19	24	26	22	25	16	17	18
hd vs. cg	hd	hd	hd	hd	hd	hd	hd	hd	hd	cg	cg	hd	hd	hd	hd	hd	cg	hd	cg	cg	cg	cg	cg	cg	cg	cg	cg	cg	cg	cg
# named cg	0	0	0	0	0	1	1	1	1	1	1	2	2	2	2	2	2	3	3	4	5	6	10	11	11	13	14	19	19	21
%	0	0	0	0	0	4	4	4	4	4	4	8	8	8	8	8	8	13	13	17	21	25	42	46	46	54	58	79	79	88

image	16	17	18	19	22	2	25	27	7	13	20	24	26	14	21	9	23	8	15	29	1	12	30	3	28	10	5	6	4	11
hd vs. cg	cg	cg	cg	cg	cg	hd	cg	cg	hd	hd	cg	cg	cg	hd	cg	hd	cg	hd	hd	cg	hd	hd	cg	hd	cg	hd	hd	dh	hd	hd
# named hd	0	0	0	1	1	2	2	2	3	3	3	3	3	5	6	7	7	8	8	11	12	14	15	16	16	17	19	19	20	20
%	0	0	0	4	4	8	8	8	13	13	13	13	13	21	25	29	29	33	33	46	50	58	63	67	67	71	79	79	83	83

Table 7: Answers on whether a certain image stands out as hand-drawn (hd) or computer-generated (cg) (all participants, increasing order).

of Cluster 1, no other main clusters that contain only images by one artist or algorithm, there were some smaller subclusters that contained illustrations by a unique illustrator. Since the drawing style is often very similar across authors (with a few exceptions in our dataset), these subclusters might have formed through the main sorting criteria. In the same way, the drawing style sorting also influenced how hand-drawn and computer-generated images were clustered. The one cluster that does not have a homogeneous drawing style (Cluster 3) contains highly detailed computer-generated images. This suggests that the high amount of detail was seen as a drawing style itself, in contrast to the usual sorting in lines vs. dots. Also, the three images in Cluster 1 were seen as very different. While they use a combination of lines and dots, their illustration type might have also been characterized as a drawing style due to its looseness and sketchiness. We did not see apparent clustering according to model type. This was due to the fact, that we specifically asked participants not to sort on this criteria.

The pile-sort task lead the participants to focus on their own thoughts and ideas about these images. The fact that these thoughts were upper most in their minds lead to rich and informative responses during our semi-structured interviews in which we talked with the participants about their pile choices and asked them specific questions. These will be discussed below.

5.2 Semi-Structured Interview

In general, the non-photorealistic pen-and-ink illustrations we tested are not yet able to pass the Turing test. With three exceptions, all of the computer-generated images were named to stand out as hand-drawings by 7 participants or less. Some images were never named and were, in contrast, almost always recognized as computer-generated: the very detailed stippling illustrations created with the technique by Secord [2002]. Also, images created with the techniques by Zander et al. [2004] and Sousa et al. [2003; 2004] often stood out as computer-generated illustrations. However, we cannot deduct that the algorithm is solely responsible for the decisions participants made. How an image appears to viewers also depends on the display parameters such as lighting or the number of lines or dots chosen and how the algorithm applies to different models. For example, the two highly detailed images created with Zander et al. [2004]'s technique (Images 22 and 24) were judged differently than Image 23 that was created with the same method using fewer lines. They were also often piled differently from each other as can be seen in Figure 6.

However, it was notable that the three hatched RENDERBOTS images [Schlechtweg et al. 2005] frequently stood out as hand-drawings, Image 30 by 63% and Image 28 by 67% of the participants. On the other hand, these images were also described as being *"sketchy, simplified, insufficiently detailed, not portraying the shape of the objects well, and ill suited for the context of scientific illustration."* This is not surprising since this technique depended on bitmap data as input and we did not provide a normal map for surface information which would have improved the major points criticized by participants. These images may have appeared hand-drawn because they had characteristics that the participants named as properties of hand-drawn images (varying line direction, squigglier silhouettes, and less realism in that lines did not follow the surface of the objects).

From the comments made by participants we conclude that computer-generated images are recognized because they are usually cleaner and tighter than hand-drawings, are too perfect (round dots, long lines), too sterile, and do not have much variation in their lines or stipples. They are also often characterized as having too much detail and being too complex which would result in a tedious illustration production process for a human. Computer-generated images look three-dimensional, *"close to 3D objects"*, use a lot of shading and lighting, and seem to apply much attention to these. An interesting finding was that images using lines often implied a hand-drawing while stippling images often had a computer-generated feel to them. This is also partially reflected in the clusters highlighted in Figure 6. Several of the general participants mentioned their disbelief that humans could place many regular dots to create shading or would have the patience to do it. Characteristics named for hand-drawn images included: line variability, a *"rougher"* look, organic feeling, weighted lines, not too much detail, looseness, *"odd"* shape. Variability in the mark, mark direction, or mark placement was named most often (by 20 of 24 participants).

However, the fact that some of the computer-generated NPR images are readily recognized as such does not mean that people do not like them or that they are considered to be bad illustrations. In fact, the highly detailed stippled and cross-hatched images, characterized as realistic and three-dimensional looking, were often named as well suited for university level textbook illustrations and were also liked more often than less precise illustrations in this context.

It seems that there are two possible reasons for many people to find illustrations attractive. On the one hand, they like very detailed and very realistic images because these convey much information. In this case the individual mark that creates the image is not as important but the overall appearance is. On the other hand, people enjoy illustrations with character because in these the artistic appearance is appreciated. Here, the individual lines are very important because they convey information such as shape or material but also add the mentioned *"character"* and *"life"* to the images.

5.3 Impact for NPR Research

The findings from our study point to several recommendations and directions for NPR research in the area of illustration.

Know your goal. When conceiving illustration techniques, it is important to be clear about the goals or application areas for our renditions. In terms of good scientific illustration, some of the tested

NPR images were quite successful in depicting shape and surface as well as in giving a sense of three-dimensionality. Participants appreciated high detail, clear depiction of shape, and a clean, less sketchy look of these images. On the other hand, the goal could be to create images that look hand-drawn (for suggestions see discussion about avoiding patterns and regularities). However, in terms of scientific illustration, a hand-drawn look might not necessarily be the main goal. A few of the computer-generated images in our set looked hand-drawn but these few were not labeled *"good for use in a university level textbook"* due to their lack of shape and surface depiction. The hand-drawn image named most often to be suitable for a university textbook was also described as having a clean and simpler look that was good for an overview of the shape. According to our participants, being clear about shape depiction was most important for communicating information about the objects.

Know your audience. It became clear in the study that participants thought images would work better or worse in particular contexts. This has practical impact. We need to be informed about the audience for our illustration techniques, who they are, what they want, and how they will use the images. Questions include:

What is the purpose of the illustration? What should the viewer of the illustration gain from viewing? Several participants suggested different illustrations for giving a good overview of the shape of an object than for learning about the exact shape and surface details. For these and other purposes, different illustration styles and parameters will be more or less successful. Several participants also mentioned that they were used to seeing certain illustration styles more than others. If illustration conventions are important in a certain discipline these should also be provided and adhered by the illustration rendering tool.

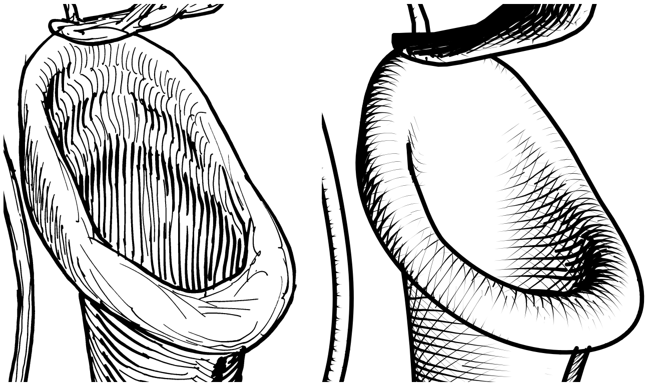
What will be the viewing context? Many participants pointed out that viewing the images from far away was notably different than viewing them close up. For making an illustration it is important whether it will be printed in small size in a book where it will be closely examined or put on a poster, overhead, or a conference presentation at a larger size where it will most likely be seen from far away. Different aspects of the illustration technique will become apparent to viewers and might lead them to different conclusions about the illustrations and the portrayed information. A good rendering technique could, for example, suggest different parameters for changing viewing contexts (e. g., [Salisbury et al. 1996]).

Strive for high-quality output. In some cases, computer-generated images were recognized as such because participants noticed larger pixels when the NPR images were created based on pixels and used a lower resolution. For high-quality print reproduction it is essential that NPR techniques do not rely on pixel primitives as marks because at the required high resolutions individual pixels will be very small (i. e., one pixel typically becomes one printer dot) so that the marks that rely on these pixels will hardly be visible. Illustrations should either be created as high resolution pixel images with resolution-independent marks or, preferably, as vector graphics (see also [Isenberg et al. 2005]).

Develop NPR techniques to portray material properties. One of the still largely unsolved problems in pen-and-ink NPR is how to portray object materials using rendering techniques. Thus far, we have concentrated on conveying tone and shading through mark properties and placement but we are generally lacking ways to portray materials. Participants criticized that some of the NPR images *"look like plastic"*, that the ribcage of the medical model *"does not look like bone"*, or that the material of the arrowhead in some images *"looks like wood but not like rock"*. However, in traditional hand-made pen-and-ink illustrations the ink marks portray shading and surface material at the same time. We need to find ways to incorporate this into pen-and-ink rendering so that people would be able to choose from material libraries in NPR tools as started by Winkenbach and Salesin [1994], much like in photorealistic tools.

Know your models. We have noticed that some participants found that certain rendering styles did not work well with certain types of objects. For example, illustrator William M. Andrews chose hatching for the archaeological and medical models and stippling for the botanical illustration. Similarly, several participants noted that certain illustration styles did not work well with certain models and that they would like them better on other models. Interestingly enough, this was not only mentioned about computer-generated but also about hand-drawn illustrations. This may be due to the fact that our instructions to the illustrators in terms of objects, techniques, and time to complete the illustration restricted them to some degree.

Work with the models. One comment that we heard from many participants when asked what makes computer-generated images stand out as such was that these images were lacking character and followed a 3D model too closely. In fact, when directly comparing hand-drawn images with their computer-generated counterparts, one can notice that hand-drawn images are more expressive and diverge from the 3D models to better illustrate certain features of the shapes. For example, see Figure 7 where the details about how the inner surface approaches the lip are much clearer in the hand-drawn image. Participants described this effect by pointing out that edges are very important in the illustrations and that some computer-generated images looked flat, in particular, some of the archaeological arrowhead illustrations. Thus, one possible direction in NPR research is to not only work on non-photorealistic rendering but also on non-photorealistic models. For example, NPR models may need more attention to details. This would not only give non-photorealistic renditions more illustrative power but also give them the *"character"* that many participants found missing in them. Some ideas and initial work to achieve this were discussed by Willats and Durand [2005] but there appears to be much more to explore in this area.



(a) Detail from Image 11. (b) Detail from Image 23.

Figure 7: Details from both a hand-drawn (a) and a computer-generated (b) image illustrate how the shape of the object is interpreted in the hand-drawn image to show shape features better.

Avoid patterns and regularities. If the goal for a computer-generated scientific illustration is to look more hand-drawn, there are several possible areas for improvement. In general, some participants noted regularities and repeated patterns right away. Stippled images were recognized as being computer-generated due to stippling artifacts (see Figure 8). Participants called those *"snowflakes"*, *"worm holes"*, or *"bands"*. They also noted the regular placement, distance, size, and shape of dots. One possible improvement would be to define different dot shapes that are applied with changing orientations and using a more random placement for the dots. Hatched images were often recognized as being computer-generated due to parallel line placement, mathematical appearance of curves and line intersections, and less variation in line weight. Interestingly the RenderBots algorithm [Schlechtweg et al. 2005] uses a more random approach to line placement and created images with the most hand-drawn appearance for a computer-generated image.

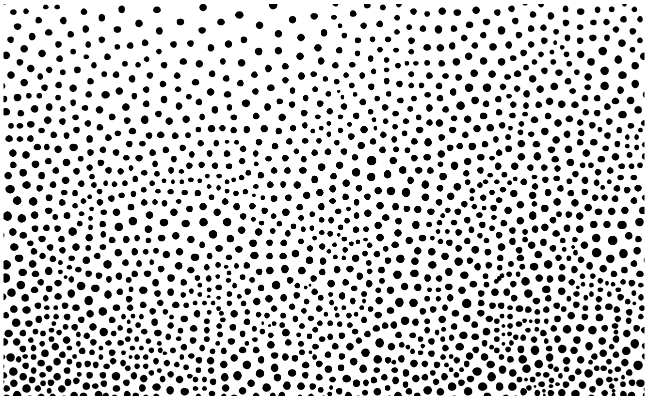
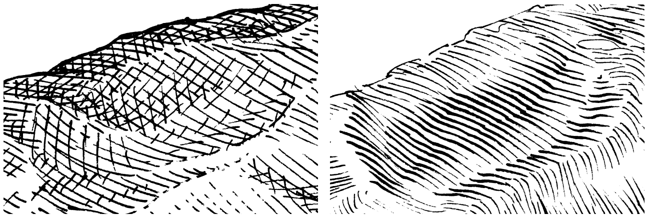


Figure 8: Detail from Image 16 showing the linear and circular patterns generated by a Voronoi relaxation based stippling as well as the non-circular but algorithmically rounded dots.

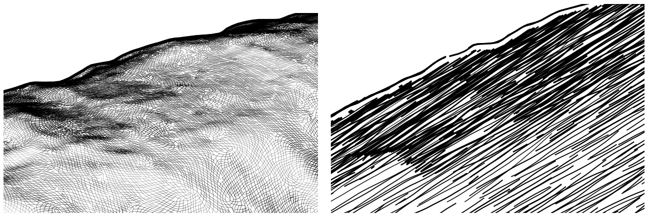
Pay more attention to the lines and dots and follow established rules of the traditional techniques. One aspect of computer-generated images that was criticized fairly frequently by artists and sometimes by other participants was the lack of mark making evidence and consequently the lack of human element. They described the line weights and line placement as mechanical or obviously according to an illumination model, as some NPR students put it. Thus, to make illustrations appear more expressive and interesting factors to address include: paying more attention to how lines or dots are placed; working more with indicating curvature through line weights as done in traditional illustrations (see comparison of details from hand-drawn and computer-generated illustrations in Figure 9). This could help to reduce the amount of marks necessary to reach a certain level of detail. Initial work in this field has been done by Sousa and Prusinkiewicz [2003] and DeCarlo et al. [2003] but there is plenty of scope for extension.

In another example, one artist mentioned that Image 26 (RenderBots-stippled botanic image) looked particularly computer-generated, in that the dots seemed to get bigger and were more spaced out in brighter regions as opposed to being smaller and closer together in darker areas. On close examination one can find that this is not really the case, the dot size remains constant except in very bright areas where it rapidly decreases. However, when looked at from farther away, the dots indeed seem to get bigger in brighter regions. It is probable that some perceptual reason causes this effect but was not apparent in the hand-stippled examples in our image set although their dot sizes also seem to be fairly constant before decreasing when getting closer to bright regions.

Pay attention to the tools. From creating the images for this study we learned that making good illustrations is not only dependent on the type of illustration style but also often on a number of parameters to set, the chosen lighting conditions or the scene compositions. Very few tools give appropriate assistance to the user of NPR illustration programs. Even experienced computer graphics researchers often have difficulties using other researchers' programs since they simply have not been designed with the end-user in mind. People who will create illustrations with our tools have to know about how to create good illustrations or have to be assisted in the process. One possible direction would be to provide templates that work for different illustration goals or application domains. Another would be to conduct further research in developing appropriate interaction techniques, developing different balances between algorithmic assistance and hand drawing.



(a) Detail from Image 1. (b) Detail from Image 10.



(c) Detail from Image 22. (d) Detail from Image 28.

Figure 9: Details from hand-drawn images (top) compared with computer-generated details (bottom) of the same region.

5.4 Critical Reflection

Although the study yielded extremely interesting and promising results, there are still some issues that can be improved and some issues to mention about the interpretation of our results. One of these issues is that we can, of course, only evaluate and talk about the images that we received or created and that we showed to our participants. There are certainly a lot more pen-and-ink techniques out there that we did not or could not include due to limited availability of the rendering tools, the required high quality for print reproduction, or the constraint that we wanted to have the same number of images as we had hand-drawn illustrations. We cannot say whether including more and/or other images and techniques would have produced the same or completely different results. Also, we have to be careful in generalizing the results we received for the images from one technique to be true for all images generated with this method. There are many parameters with each of the employed tools that can be modified so that other parameter values might have lead to other results. A glimpse of these issues can be seen with Images 22 and 24 as opposed to Image 23 as discussed in Section 5.2. Related to these two issues is that all computer-generated images were created by NPR researchers rather than professional illustrators. If we would have had the time and resources to teach illustrators to use NPR tools and have them create the images, they would most probably have been able to come up with much better illustrations. However, as mentioned in the previous section we also realized during the study preparation that some of the tools that we have written to demonstrate non-photorealistic rendering techniques pose quite a challenge when it comes to creating good illustrations. Good geometric models that our tools can handle are sometimes hard to obtain or to create and the usability and intuitiveness of parameter adjustments needs to be improved. In addition, professional illustrators pointed out in separate conversations that they would like semi-interactive tools best, i. e., tools that provide an automated way to place marks but that also allows them to interactively make changes afterwards.

6 Conclusions and Future Work

In conclusion we can say that we still have to learn a lot in terms of what makes a good scientific illustration. We found that participants liked the realism and clear depiction of shape in computer-generated images created using a high number of lines or dots. However, as one participant put it: *"it's a fine line between looking real and getting across what the image is supposed to be"*. Hand-drawn images clearly still seemed different from computer-generated images. Not all NPR algorithms used to render the images in this paper were equally successful in creating good scientific illustrations. We believe that what makes a good scientific illustration has to be determined per algorithm or technique and we encourage researchers to evaluate their algorithms by showing them to participants using different models and parameters. We still have a lot to learn from professional illustrators.

There are several avenues for future work after this study that we plan to pursue. We plan to talk to and evaluate illustrations with the domain experts, a group that was left out in this study. Also, we only had general art students and general illustrators in the illustrators group. It would be very interesting to repeat the study specifically with professional scientific and medical illustrators. Finally, we would like to generalize the findings of this study by using different models and NPR algorithms.

Acknowledgments

We would like to thank all artists who provided their time and artistry and created wonderful illustrations, Adrian Secord for his stippling tool, Tobias Germer for generating images with his RENDERBOTS on short notice, all study participants for discussing their thoughts about the illustrations, all people in the Interactions Lab, Dr. Tak Shing Fung for his advice with the statistical analysis, and our funding agencies Alberta Ingenuity (AI), the Informatics Circle of Research Excellence (iCore), and the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

ACEVEDO, D., LAIDLAW, D. H., AND DRURY, F. 2005. Using Visual Design Expertise to Characterize the Effectiveness of 2D Scientific Visualization Methods. In *Poster Comp. of IEEE VIS & InfoVis 2005*, 111–112.

CRESWELL, J. W. 2000. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 2^{nd} ed. Sage Publications Inc., Thousand Oaks, CA.

DECARLO, D., FINKELSTEIN, A., RUSINKIEWICZ, S., AND SANTELLA, A. 2003. Suggestive Contours for Conveying Shape. *ACM Transactions on Graphics 22*, 3 (July), 848–855.

DENZIN, N. K., AND LINCOLN, Y. S., Eds. 2000. *Handbook of Qualitative Research*, 2^{nd} ed. Sage Publications Inc., Thousand Oaks, CA.

DUKE, D. J., BARNARD, P. J., HALPER, N., AND MELLIN, M. 2003. Rendering and Affect. *Computer Graphics Forum 22*, 3 (Sept.), 359–368.

GOOCH, B., AND GOOCH, A. 2001. *Non-Photorealistic Rendering*. AK Peters Ltd., Natick.

GOOCH, A. A., AND WILLEMSEN, P. 2002. Evaluating Space Perception in NPR Immersive Environments. In *Proc. of NPAR 2002*, ACM Press, New York, 105–110.

GOOCH, B., REINHARD, E., AND GOOCH, A. A. 2004. Human Facial Illustrations: Creation and Psychophysical Evaluation. *ACM Transactions on Graphics 23*, 1 (Jan.), 27–44.

HALPER, N., MELLIN, M., HERRMANN, C. S., LINNEWEBER, V., AND STROTHOTTE, T. 2003. Psychology and Non-Photorealistic Rendering: The Beginning of a Beautiful Relationship. In *Mensch & Computer 2003*, Teubner Verlag, Stuttgart, 277–286.

HALPER, N., MELLIN, M., HERRMANN, C. S., LINNEWEBER, V., AND STROTHOTTE, T. 2003. Towards an Understanding of the Psychology of Non-Photorealistic Rendering. In *Proc. Workshop Computational Visualistics, Media Informatics and Virtual Communities*, Deutscher Universitäts-Verlag, Wiesbaden, 67–78.

ISENBERG, T., CARPENDALE, M. S. T., AND SOUSA, M. C. 2005. Breaking the Pixel Barrier. In *Proc. of Computational Aesthetics 2005*, Eurographics Association, Aire-la-Ville, Switzerland, 41–48.

JACKSON, C. D., ACEVEDO, D., LAIDLAW, D. H., DRURY, F., VOTE, E., AND KEEFE, D. 2003. Designer-Critiqued Comparison of 2D Vector Visualization Methods: A Pilot Study. In *ACM SIGGRAPH 2003 Conference Abstracts and Applications*, ACM Press, New York.

KIM, S., HAGH-SHENAS, H., AND INTERRANTE, V. 2004. Conveying Shape with Texture: Experimental Investigation of Texture's Effects on Shape Categorization Judgments. *IEEE Transactions on Visualization and Computer Graphics 10*, 4 (July), 471–483.

SALISBURY, M. P., ANDERSON, C., LISCHINSKI, D., AND SALESIN, D. H. 1996. Scale-Dependent Reproduction of Pen-and-Ink Illustration. In *Proc. of SIGGRAPH 96*, ACM Press, New York, 461–468.

SANTELLA, A., AND DECARLO, D. 2004. Visual Interest and NPR: an Evaluation and Manifesto. In *Proc. of NPAR 2004*, ACM Press, New York, 71–78.

SCHLECHTWEG, S., GERMER, T., AND STROTHOTTE, T. 2005. RenderBots—Multi Agent Systems for Direct Image Generation. *Computer Graphics Forum 24*, 2 (June), 137–148.

SCHUMANN, J., STROTHOTTE, T., RAAB, A., AND LASER, S. 1996. Assessing the Effect of Non-photorealistic Rendered Images in CAD. In *Proc. of CHI'96*, ACM Press, New York, 35–42.

SECORD, A. 2002. Weighted Voronoi Stippling. In *Proc. of NPAR 2002*, ACM Press, New York, 37–44.

SOUSA, M. C., AND PRUSINKIEWICZ, P. 2003. A Few Good Lines: Suggestive Drawing of 3D Models. *Computer Graphics Forum 22*, 3 (Sept.), 381–390.

SOUSA, M. C., FOSTER, K., WYVILL, B., AND SAMAVATI, F. 2003. Precise Ink Drawing of 3D Models. *Computer Graphics Forum 22*, 3 (Sept.), 369–379.

SOUSA, M. C., SAMAVATI, F. F., AND BRUNN, M. 2004. Depicting Shape Features with Directional Strokes and Spotlighting. In *Proc. of CGI 2004*, IEEE, Los Alamitos, CA, 214–221.

STROTHOTTE, T., AND SCHLECHTWEG, S. 2002. *Non-Photorealistic Computer Graphics. Modelling, Animation, and Rendering*. Morgan Kaufmann Publishers, San Francisco.

STROTHOTTE, T., PREIM, B., RAAB, A., SCHUMANN, J., AND FORSEY, D. R. 1994. How to Render Frames and Influence People. *Computer Graphics Forum 13*, 3 (Aug.), 455–466.

TIETJEN, C., ISENBERG, T., AND PREIM, B. 2005. Combining Silhouettes, Shading, and Volume Rendering for Surgery Education and Planning. In *Proc. of EuroVis 2005*, Eurographics Association, Aire-la-Ville, Switzerland, 303–310, 335.

WELLER, S. C., AND ROMNEY, A. K. 1988. *Systematic Data Collection*, vol. 10 of *Qualitative Research Methods*. SAGE Publications Inc., Thousand Oaks, CA.

WILLATS, J., AND DURAND, F. 2005. Defining Pictorial Style: Lessons from Linguistics and Computer Graphics. *Axiomathes 15*, 3 (Sept.), 319–351.

WINKENBACH, G. A., AND SALESIN, D. H. 1994. Computer-Generated Pen-and-Ink Illustration. In *Proc. of SIGGRAPH 94*, ACM Press, New York, 91–100.

ZANDER, J., ISENBERG, T., SCHLECHTWEG, S., AND STROTHOTTE, T. 2004. High Quality Hatching. *Computer Graphics Forum 23*, 3 (Sept.), 421–430.