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), there has been considerable progress on 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 in terms of scientific illustration.

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]. It tries to break free from the constraint of (photo-)realism that many other rendering techniques strive for and generates images and animations that at least in 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 a lot of 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 traditionally created hand-drawn illustrations compared to computer-generated non-photorealistic illustrations. The purpose of this study was to improve our understanding of the differences between both types of images at the current stage of NPR research and open up or validate 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 for 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 first discuss some related work in Section 2. Then we briefly explain the design of the observational study in Section 3. Afterwards, Section 4 presents some initial results and general observations. In the following Section 5 we then discuss these results and interpret them. Finally, we conclude with a summary and suggest some future work in Section 6.

2 Related Work

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

Communication in architecture and design: The first evaluation of NPR was published by [Schumann et al. 1996]. They performed a study with 54 architects to compare the usability of computer-generated images with respect to communicative goals during design concept development. Subjects were shown three different images portraying the same architectural object: a CAD plot image (wire-frame with hidden-surfaces removed), a constant shading image and a NPR image generated by a sketch-renderer developed by the authors [Strothotte et al. 1994]. Their results showed that those three kinds of images have very different effects on viewers with very positive aspects reported about the use of NPR images.

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., iso-surfaces and direct volume rendering).

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 of these images 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 learned two times faster than photographs and caricatures were learned one and a half times faster than photographs.

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

*e-mail: {isenberg|pneumann|sheelagh|mario}@cpsc.ucalgary.ca
†e-mail: jorgej@acm.org or jaj@rtr.inesc-id.pt

influence judgment in non-trivial ways.

In a similar vein, [Halper et al. 2003a; Halper et al. 2003b] suggested applying a theory of psychology to NPR in terms of biological, social, and environmental paradigms emerging from user 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, which would encourage use of NPR systems by non-experts towards intent-driven illustration.

Space perception in immersive environments: [Gooch and Willemsen 2002] presented the first work to examine and evaluate space perception in a functional, non-photorealistically rendered immersive environment. Their experiments involved (1) direct walking tasks in a physical hallway and (2) NPR renditions (silhouettes, boundaries, and creases) of a scaled model of the same hallway, visualized through a head-mounted display. Their study successfully indicated the degree to which NPR images are capable of conveying a veridical sense of spatial organization.

Shape and data depiction from textures: [Kim et al. 2004] describe two comprehensive experiments to 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 the "principal direction grid" pattern more effective at shape perception and that oblique viewpoints seem to favor classification.

In a study on a similar subject, [Jackson et al. 2003] examined how well different visualization techniques allow viewers to comprehend a flow field. However, in order to back up a previous study based on numeric measurements, [Jackson et al. 2003] asked a professional graphic designer to comment on the visualization. They found that the subjective results of the designer's critique corresponded quite 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 previous approaches, our paper is the first to compare ink-based non-photorealistic renditions with hand-made pen-and-ink illustrations in an observational user study using three different groups of participants.

3 Description of Observational Study

The goal of our observational study was to gain understanding in the way viewers evaluate traditional hand-drawn illustrations compared to computer-generated techniques. We chose a combination of user observation 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 second part consisted of a semi-structured interview in which we asked about the participants' assessment of the shown images in more detail.

3.1 Subjects

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 work with 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* who are 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 Materials

For this study we concentrated on 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). It consisted of one separate Win32 application for each of the tree models. Each model could be rotated using the mouse and a virtual trackball in order 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.

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. Table 1 gives an overview of the hand-drawn illustrations and their authors as well as the NPR techniques used to create images for this study.

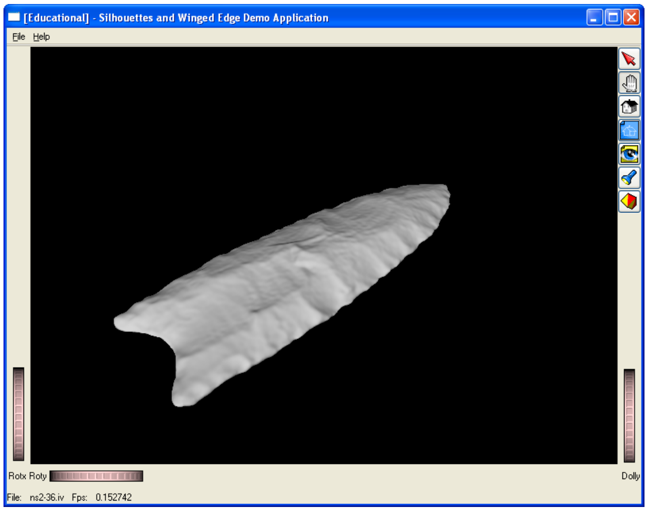
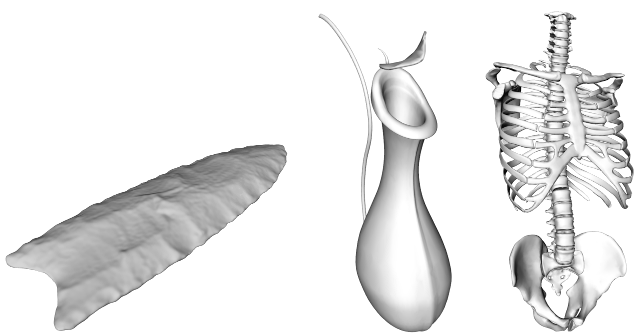


Figure 1: Screenshot of simple viewer provided to the illustrators.



(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.

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 out again in their original size (using a HP LaserJet 4100 PS with a maximum resolution of 1200 dpi). In addition, we also printed all computer-generated images on the same type of paper, using slightly different sizes for the images 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 printing was 216 mm by 279 mm (Letter sized) color laser paper with 105 $\frac{g}{m^2}$ weight.

3.3 Methodology

Each study session involved a single participant working through three stages. Data was recorded using video and audio and also field notes were taken. The three stages were as follows:

Stage 1: *Pile-sorting task.* Participants were given the 30 illustrations with a different random order of images 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 the size of the illustrations 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. Participants were asked to think-aloud during the sorting in order to be able to capture their thoughts on video.

Stage 2: *Semi-structured interview.* After the sorting was finished, the piles were spread on the table so that each image was visible and could be discussed by itself or in connection to its category (see Figure 4). In a first question, the participants were 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 and why?
- Which images do you like least and why?
- In what context would you use any of these images? Where would you like to see them and why?
- Which images would you use in a university level textbook and why?
- Which images would you use in a textbook for children (late kindergarten to early elementary school) and why?
- Which images have the most computer-generated feel and why?
- Which images look most hand-drawn and 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.

Figure 4: Discussion session after the pile-sorting.

4 Results and Discussion

We present the results of the study according to the three stages of the experiment.

4.1 Pile sort task

The pile sorting stage was on average 33:38 minutes long; 29:54 minutes for general students, 34:02 minutes for NPR students,

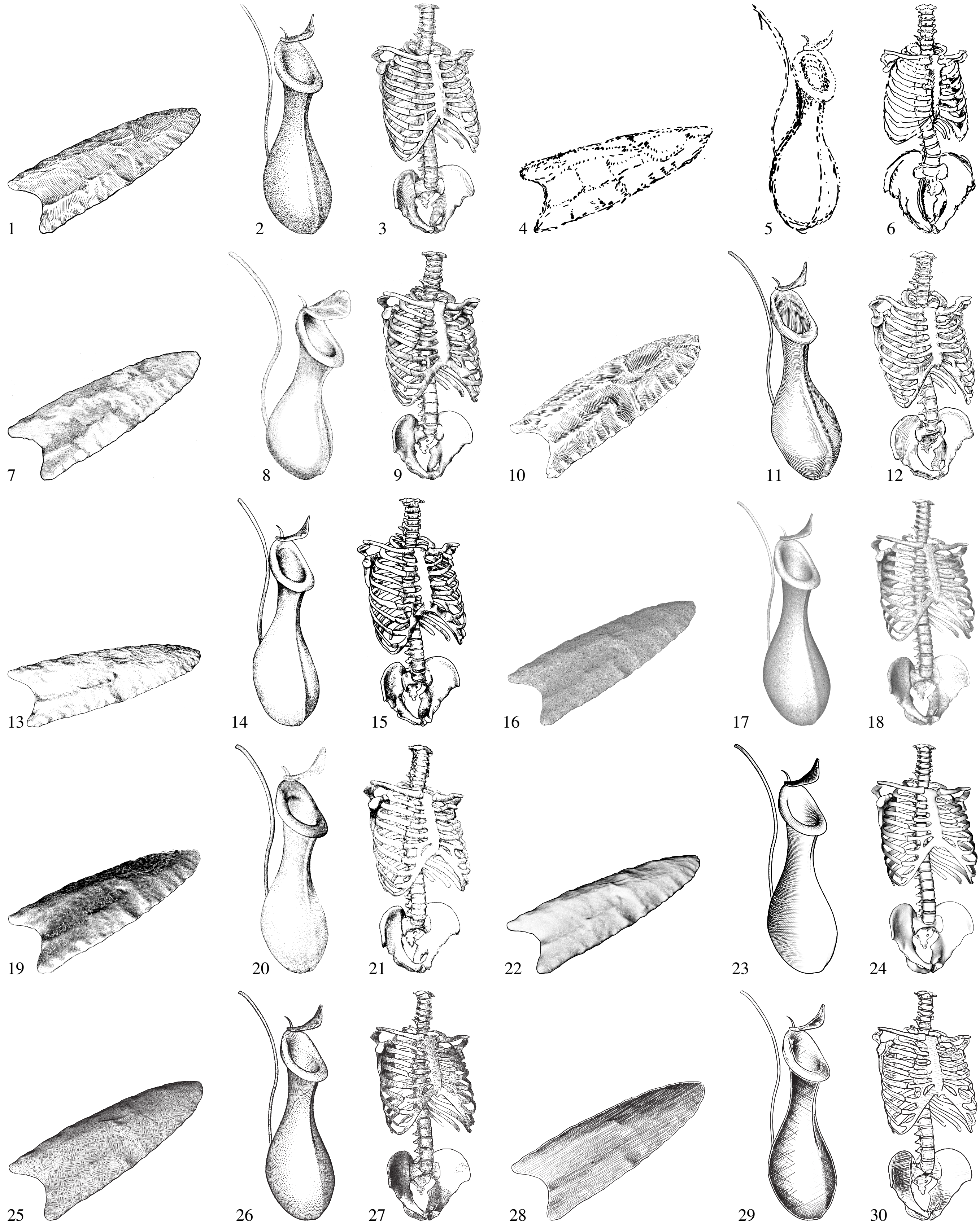


Figure 3: All study images with their numbers used throughout the paper. All images are copyright of their respective authors (refer to Table 1), used with permission.

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 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 Figure 3.

and 36:57 minutes for artists. In general, most people used a categorization related to drawing/rendering style based on the type of mark being used (lines styles vs. dots). Other categorizations included the amount of realism, detail, information content, object orientation, or the overall look and feel of an image using aesthetic criteria. We did not find significantly different piling methods between our three user groups. None of the participants set out to make piles that clearly distinguished computer-generated from hand-drawn images. However, several categorizations included piles that participants later 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 usually not used to group them.

An image-by-image similarity matrix was created from each individual's categorization by tabulating the co-occurrence of items in piles so that items that were together were counted as being similar. These similarity matrices were then combined per group as well as for all participants and analyzed with hierarchical clustering using average linkage (between-groups) with chi-square as the dissimilarity measure.

Figure 5 gives an overview of the hierarchical cluster analysis. This dendrogram displays the different cluster steps on the *y*-axis and the data items on the *x*-axis. The earlier data items appear in the same subtree the higher is the calculated similarity between them. We highlighted a few clusters that we will discuss in more detail.

- Cluster 1 includes three hand-drawn images by the same artist which almost all participants clearly distinguished from the other images. Piles that included these images were titled: *"less clear, simple drafts, sketchy, incomplete drawing, artistic, not very detailed, general shape, little tone."*
- Cluster 2 shows that participants often distinguished according to the mark used in the image. Except for Images 22 and 24, all images in this cluster use stippling or a mixture of lines and dots as the drawing style. Image 22 and 24 fall out since they are grouped with Images 16, 17, and 18 (Cluster 5). These five images were among seven rendered images most often described as having the *"most computer-generated feel"* to them. They are also very detailed, using a high number of lines or dots which was likely the reason for participants to often put them together in piles.
- Cluster 3 consists of illustrations using different hatching-styles, both computer-generated and hand-drawn. However, the computer-generated images in this cluster were among those most often described as looking hand-drawn.
- Cluster 4 has no clear distinction between hand-drawn and computer-generated images. They are likely united by their sole use of stippling or a mixture of lines and dots.

In this graph, several smaller clusters can be found and grouped according to similar criteria. At the early clustering stages it can be seen that often illustrations by the same artists or algorithms are grouped together and that also the type of object often played a role.

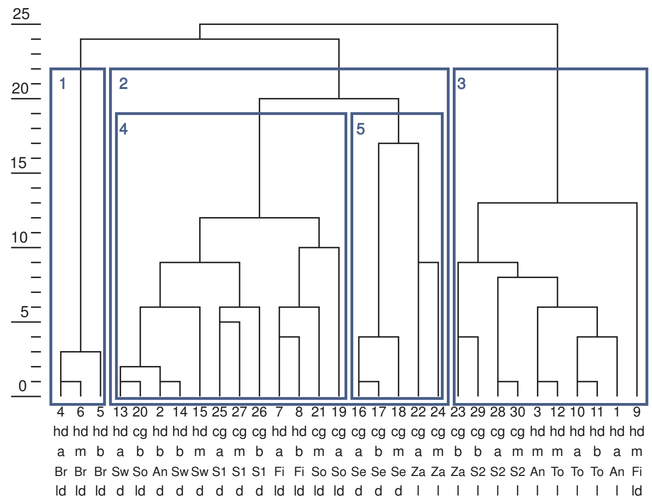


Figure 5: Cluster graph for all participants. The first data line contains the image number (refer to Table 1 and Figure 3), the second line describes whether the image was hand-drawn (hd) or computer-generated (cg), the third line the type of model (archaeological, biological, or medical), the forth line the first two letters of the illustrator or first NPR algorithm author's last name, and the last line whether it contains mainly dots (d), lines (l), or a mixture (ld). Five clusters are emphasized using blue boxes and are explained in the text.

4.2 Semi-structured interview

During the interview part we established the participants' criteria for building piles and then asked more specific questions about the images that were spread out on the table. 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 their 'liking' of an image. 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 2). 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. This shows that liking an image was not necessarily related to whether it was hand-drawn or computer-generated.

Images with significantly outstanding dislike scores were Image 4, 5, 6, and 28. This is explained by the fact that participants were mostly answering this question in the context of scientific illustration. These images were described as too sketchy and defining shape less well in the context of learning and teaching. The first three of these images were also clearly distinguished in a cluster (Cluster 1). However, these images were identified as being valuable in an arts context and as having an interesting drawing style to look at.

In what context would you like to see these images? In what context would you use them? This questions was mostly answered in context to specific images or categories. All participants mentioned that most of the images were useful for instructive purposes, as in scientific or artistic textbooks, as diagrams in classes, for museum displays, magazines, periodicals, journals, papers, or for encyclopedias and dictionaries. Other suggestions included usage in comics or graphic novels, 3D computer programs or computer games, bathroom tiles or restaurant signs, Web pages, advertisement, or as art displays.

Which images would you use in a university level textbook?

Table 3 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 such. They were described as *"highly detailed, more realistic, better shading, good texture display, traditional illustration style, clear, good sense for 3D, ... "*. However, not all computer-generated images that stood out as such were also named to be good for university textbooks (Images 24, 25, and 26). The images users named in this question often corresponded to the ones they liked the most which also shows the context in which participants answered the "like" question.

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

Table 3: Images named most suitable for including in a university level textbook.

Which images would you use for children's textbooks? Images named most often to be suitable for children's textbooks (kindergarten to early elementary school level) were Images 11, 12, 17, 1, and 18 which is significantly different from what was recommended for the university level (see Table 4). In general, we received very 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 than adults.

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

Table 4: Images named most suitable for including in a children's textbook.

Which images look most computer-generated and which look most hand-drawn? 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 5).

The results show that many hand-drawn images were often correctly recognized as such (meaning that they stood out as hand-drawn), most significantly 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 significantly 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 be computer-generated (about 0.7 images per participant on average with Image 7 being named most often). General students named 1.25 hand-drawn images per participant on average to stand out as computer-generated as opposed to 0.375 and 0.5 for NPR students and artists, respectively. Computer-generated images were a bit more often thought to stand out as hand-drawn (about 2.9 images per participant on average), the images that were mostly responsible for this were Images 28, 29, 30, and 23. These images used longer lines with a lower density than images using lines that were seen rather as computer-generated. Figure 5 shows that these are also those computer-generated ones grouped in Cluster 3 with only hand-made images. Again, general students named 4.125 computer-generated images per participant on average to stand out hand-drawn as opposed to 2.125 and 2.5 for NPR students and artists, respectively. Also, these numbers are much bigger than the inverse case (hand-drawn images that stand out as computer-generated).

Participants in the general and NPR category had overall 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. 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

The insight gained into both traditional hand-drawn and computer-generated scientific illustration was very significant. Our study design lead to participants discussing their thoughts and interpretations of the images freely with us. The three user groups mostly 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 much more broadly 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. The results of the study showed that participants could mostly distinguish between

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 2: Answers of all participants on whether a certain image is particularly liked or particularly disliked in increasing order.

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

hand-drawn and computer-generated images but that those found to look very computer-generated were still often thought to be very successful in the context of scientific illustration. In the following section we will discuss our results in more detail.

5.1 Interpretation of Results

In general, the non-photorealistic pen-and-ink illustrations that 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, even 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; Sousa et al. 2004] were very often recognized as computer-generated illustrations. However, we cannot deduct that the rendering technique 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. As an example, the two highly detailed images created with [Zander et al. 2004] (Images 22 and 24) were judged differently than Image 23 that was created with the same technique but used fewer lines. They were also often piled differently from each other as can be seen in Figure 5.

However, it was quite significant that the three hatched RENDER-BOTS images [Schlechtweg et al. 2005] frequently stood out as hand-drawings, Image 30 by more than half and Image 28 even by two thirds 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 due to the variation in 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 can 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 very three-dimensional, *"close to 3D objects"*, use a lot of shading and lighting and seem to apply a lot of attention to these. A very significant 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 5. 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.

However, the fact the 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 which were characterized as very 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. In general, we noticed a tendency of participants to recommend images that they liked for usage in a university textbook. This was due to the fact that participants answered this question many times in the context of scientific illustration so that the images they thought worked best were also the ones they liked best. Participants who answered using aesthetic criteria often chose different, mostly hand-drawn, images.

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 they convey a lot of information. In this case the individual mark that creates the image is not as important and the overall appearance is important. On the other hand, people enjoy illustrations with character because for these the artistic appearance is appreciated. Here, the individual lines are very important because they convey information such as shape or material and add the mentioned *"character"* and *"life"* to the images.

5.2 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, we have to be clear about the goals or application areas for our renditions. In terms of good scientific illustration, the tested NPR images were partially quite successful in shape and surface depiction 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 the shown images. On the other hand, the goal could be to create images that look hand-drawn. For this case we will discuss several suggestions for improvement in the following. However, in terms of scientific illustration, looking hand-drawn might not necessarily be the main goal to strive for. 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. Different questions that we should ask ourselves are:

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 more used to seeing certain illustration styles. 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 significantly 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 parts of the illustration technique will become apparent to the viewer and might lead them to different conclusions about the illustrations and the information they can gain from them. A good rendering technique could, for example, suggest different parameters for changing viewing contexts.

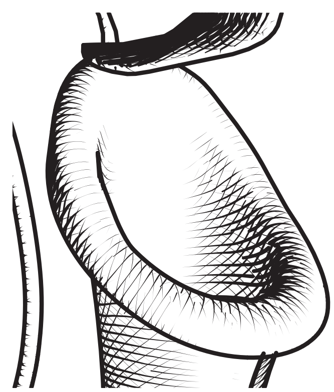
Strive for high-quality output. In some cases computer-generated images were recognized as such because participants noticed larger pixels when the NPR images were created based on pixels and used a lower resolution. However, for print reproduction it is essential that NPR techniques do not rely on pixel primitives as marks because at the required high resolutions individual pixels will be very small (i. e., one pixel 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 or as vector graphics.

Develop NPR techniques to portray material properties. One of the still unsolved problems in NPR pen-and-ink rendering is how to portray object materials using rendering techniques. So far, we have concentrated on conveying tone and shading through our mark properties and placement but we are generally lacking a way to show the 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 should find a way to incorporate this into NPR pen-and-ink techniques so that people using the created tools would be able to choose from a material library much like in photorealistic rendering.

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 (see Figure 6). 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. So we think that one possible way for future work in NPR is to not only work on non-photorealistic rendering but also on non-photorealistic models. If done right, this would not only give non-photorealistic renditions more illustrative power but also give them the *"character"* that many participants were missing in them.

(a) Detail from Image 11.



(b) Detail from Image 23.

Figure 6: Details from both a hand-drawn and a computer-generated image that 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 7). Participants called those *"snowflakes"*, *"worm holes"*, or *"bands"*. The 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 rotations and using a more random placement for the dots. Hatched images were often recognized as being computer-generated due to parallel line placement, mathematically looking curves and line intersections, and less variation in line weight. Interestingly, the RENDERBOTS algorithm [Schlechtweg et al. 2005] that uses a more random approach to line placement created images with the most hand-drawn appearance.

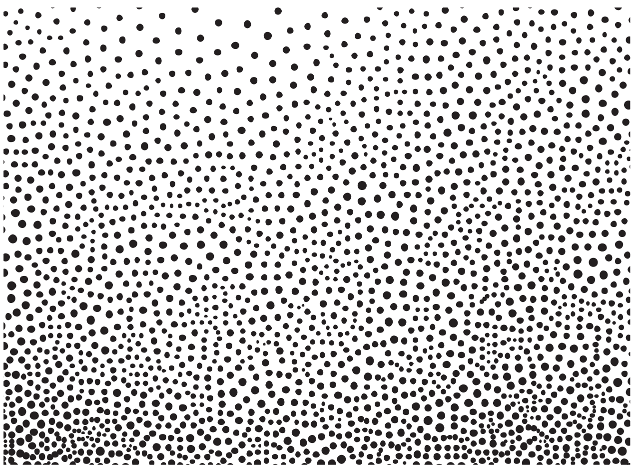
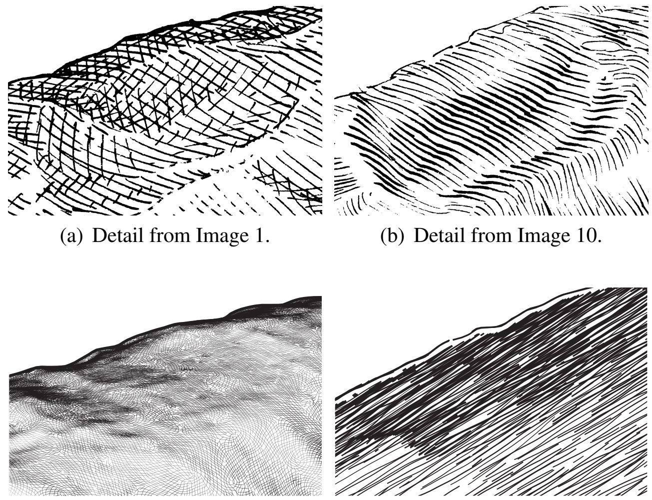


Figure 7: Detail from Image 16 showing the linear and circular patterns generated by a Voronoi relaxation based stippling as well as the perfectly round dots.

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. In order to make illustrations appear more expressive and interesting we should, therefore, pay a lot more attention to how lines or dots are placed, something that also contributes to the "character" of an image as mentioned above. We have to, for example, work 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 8). This could help to reduce the amount of marks necessary to reach a certain level of detail. In another example, one artist mentioned that Image 26 (botanic image stippled with RENDERBOTS) looked particularly computer-generated because the dots seemed to get bigger as they were more spaced out in brighter regions as opposed to being closer together in darker areas. On close examination one can find that this is, in fact, not really the case but that the dot size remains constant except in very bright areas when it rapidly decreases. However, when looked at from farther away the dots indeed seem to get bigger in brighter regions. We are not quite clear on what really causes this effect (we suspect a reason in perception) but we could not see it in hand-stippled examples in our image set although the dot size there also seems 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 parame-



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

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

Figure 8: Details from hand-drawn images compared with computer-generated details of the same region.

ters 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 for the end-user. 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.

5.3 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 which has already been discussed in Section 5.1. 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. The right geometric models that our tools can handle are sometimes hard to obtain or to create and the usability and intuitivity 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 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 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 tool to generate stipple images, 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, 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 Compendium of IEEE Visualization and IEEE Information Visualization 2005*, T. A. Keahey, Ed., 111–112.

DUKE, D. J., BARNARD, P. J., HALPER, N., AND MELLIN, M. 2003. Rendering and Affect. *Computer Graphics Forum (Proceedings of Eurographics 2003, Granada, Spain, September 1–6, 2003) 22*, 3, 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 *Proceedings of Second International Symposium on Non Photorealistic Animation and Rendering (NPAR 2002, Annecy, France, June 3–5, 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: Interaktion in Bewegung (September 8–10, 2003, Stuttgart)*, Teubner Verlag, Stuttgart, Leipzig, Wiesbaden, J. Ziegler and G. Szwillus, Eds., 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 (April 4–5, 2003)*, Deutscher Universitäts-Verlag, Wiesbaden, J. Schneider, T. Strothotte, and W. Marotzki, Eds., 67–78.

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.

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 *Proceedings of CHI'96 Conference on Human Factors in Computing Systems (Vancouver, Canda, Apr. 1996)*, ACM Press, New York, 35–42.

SECORD, A. 2002. Weighted Voronoi Stippling. In *Proceedings of Second International Symposium on Non Photorealistic Animation and Rendering (NPAR 2002, Annecy, France, June 3–5, 2002)*, ACM Press, New York, 37–44.

SOUSA, M. C., FOSTER, K., WYVILL, B., AND SAMAVATI, F. 2003. Precise Ink Drawing of 3D Models. *Computer Graphics Forum (Proceedings of Eurographics 2003, Granada, Spain, September 1–6, 2003) 22*, 3 (Sept.), 369–379.

SOUSA, M. C., SAMAVATI, F. F., AND BRUNN, M. 2004. Depicting Shape Features with Directional Strokes and Spotlighting. In *Proceedings of Computer Graphics International 2004*, IEEE, 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 (Proceedings of Eurographics 1994, Oslo, Norway, Sept. 1994) 13*, 3, 455–466.

TIETJEN, C., ISENBERG, T., AND PREIM, B. 2005. Combining Silhouettes, Shading, and Volume Rendering for Surgery Education and Planning. In *Data Visualization 2005: Proceedings of the Eurographics/IEEE VGTC Symposium on Visualization (EuroVis 2005, June 1–3, 2005, Leeds, England, UK)*, Eurographics Association, Aire-la-Ville, Switzerland, K. W. Brodlie, D. J. Duke, and K. I. Joy, Eds., Eurographics Workshop Series, 303–310, 335.

WELLER, S. C., AND ROMNEY, A. K. 1988. *Systematic Data Collection*, vol. 10 of *Qualitative Research Methods*. SAGE Publications Inc., Newbury Park, Beverly Hills, London, New Delhi.

ZANDER, J., ISENBERG, T., SCHLECHTWEG, S., AND STROTHOTTE, T. 2004. High Quality Hatching. *Computer Graphics Forum (Proceedings of Eurographics 2004, Grenoble, France, August 30–September 3, 2004) 23*, 3 (Sept.), 421–430.