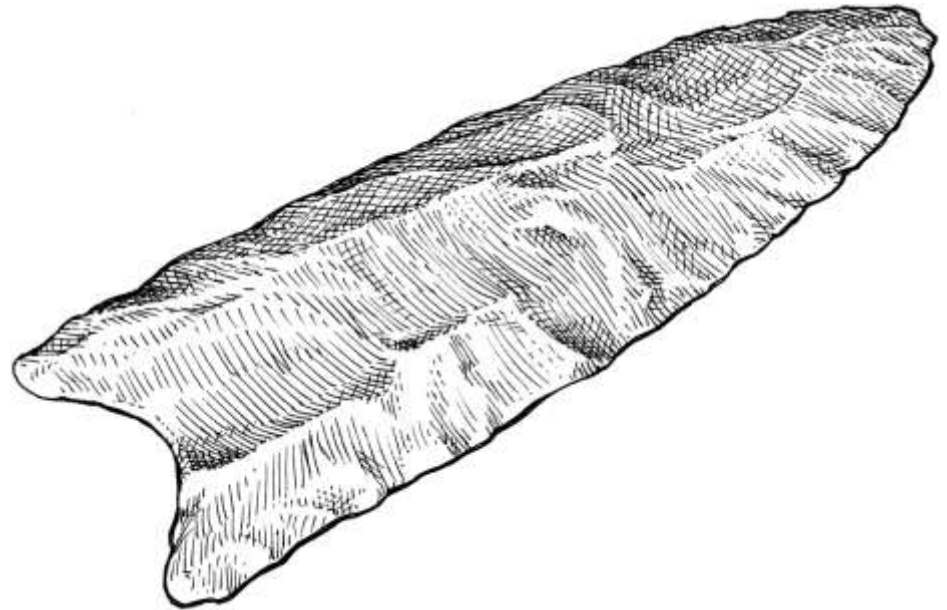


# Evaluating Non-Photorealistic and Illustrative Visualizations

Tobias Isenberg

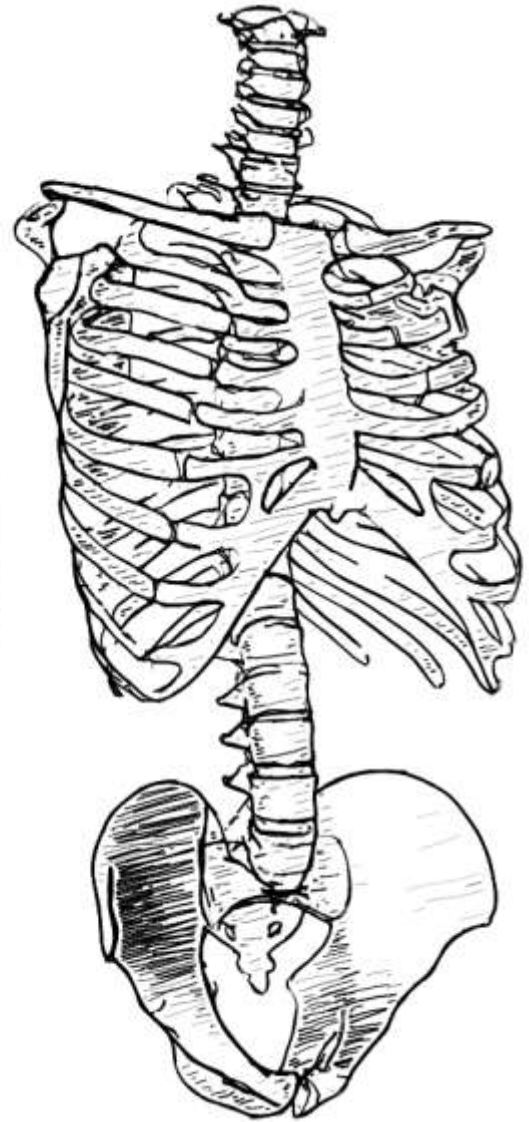
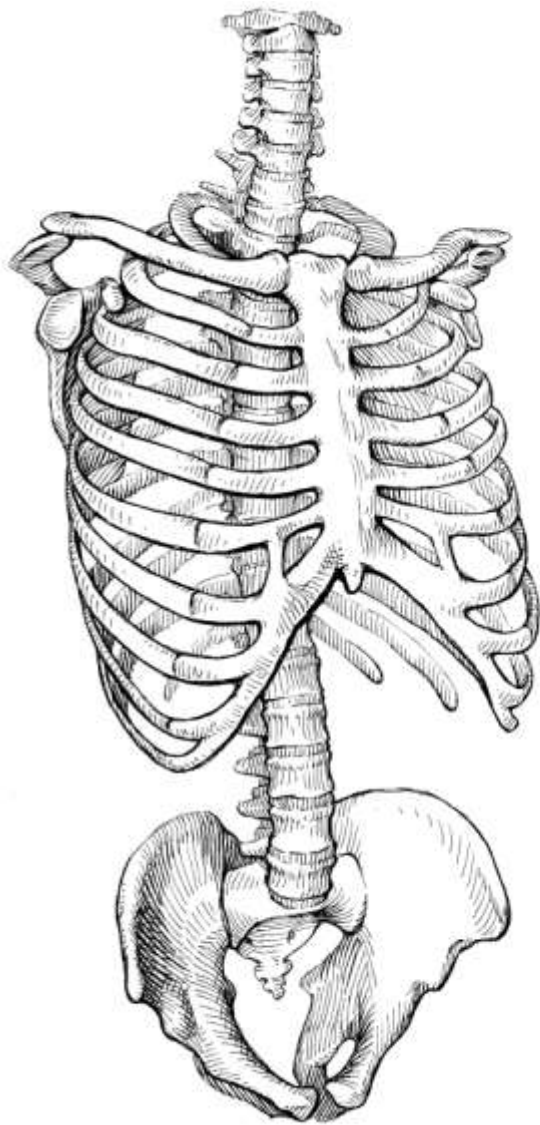


university of  
 groningen

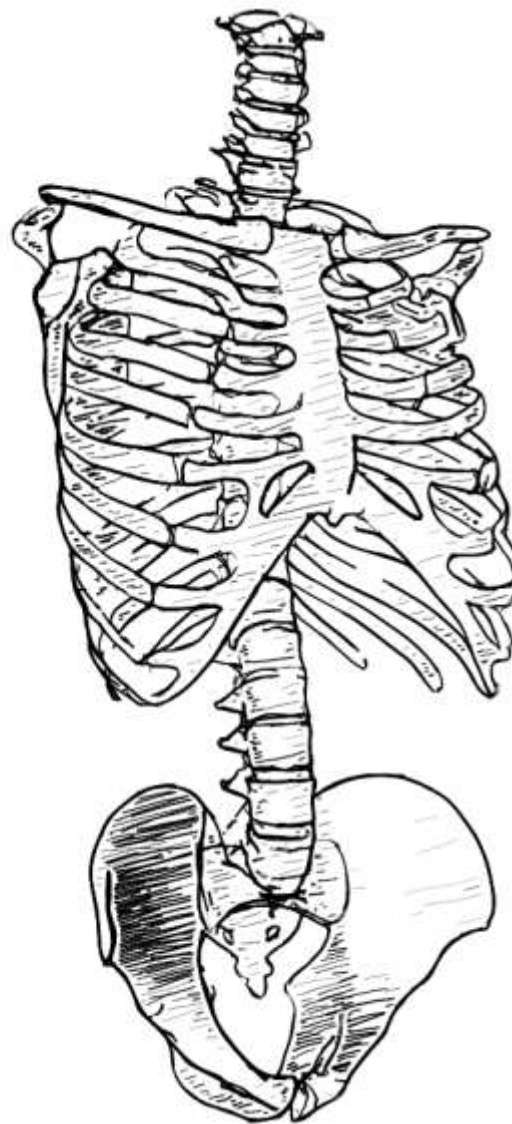
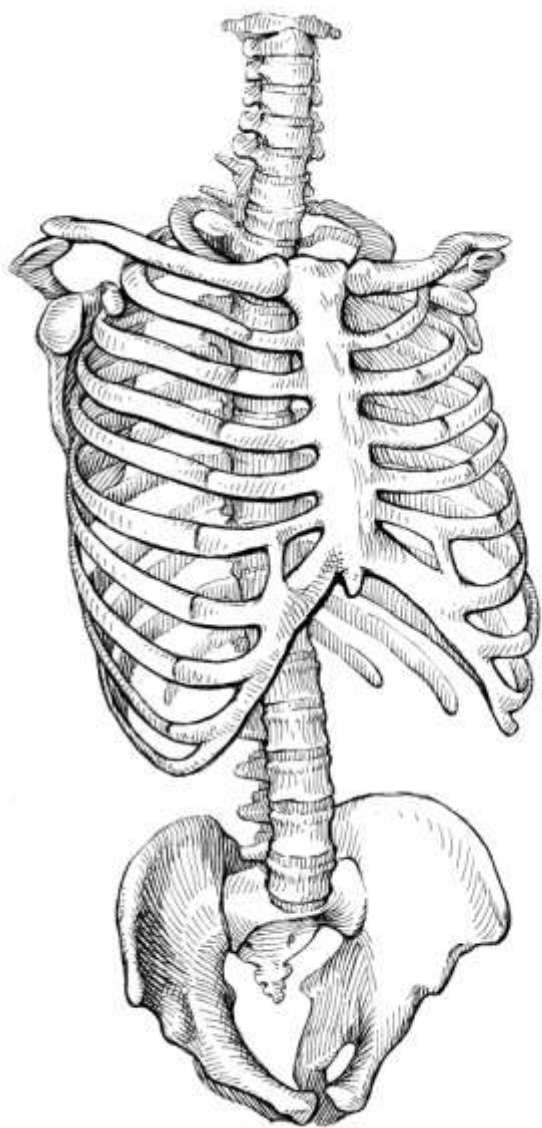
IllustraVis 2009

If you want to use content  
from these slides, you have  
to ask the respective author  
for permission!

# Which of these images do you like?



# Computer-Generated or Hand-Drawn?



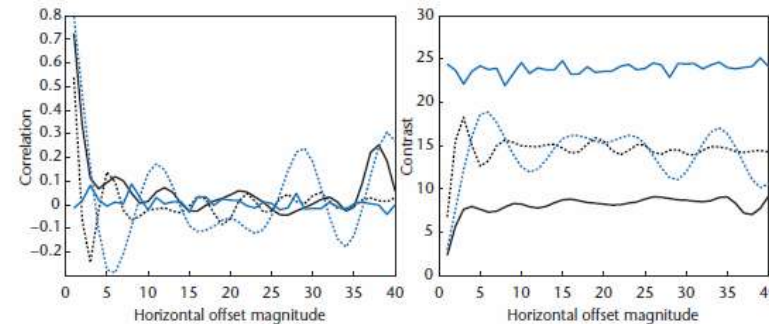
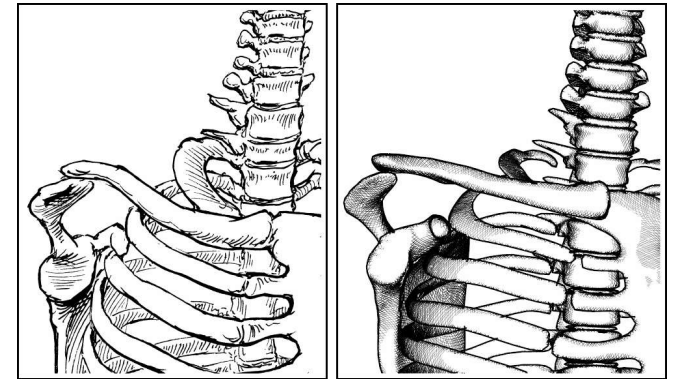
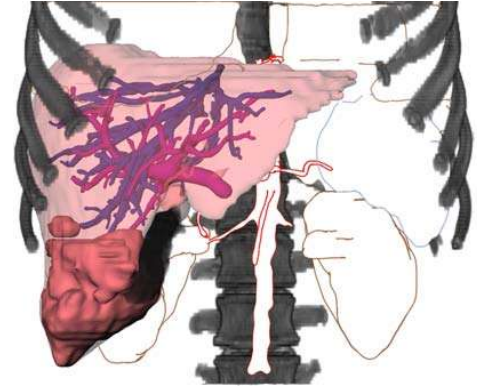
# How to evaluate illustrative rendering?

---

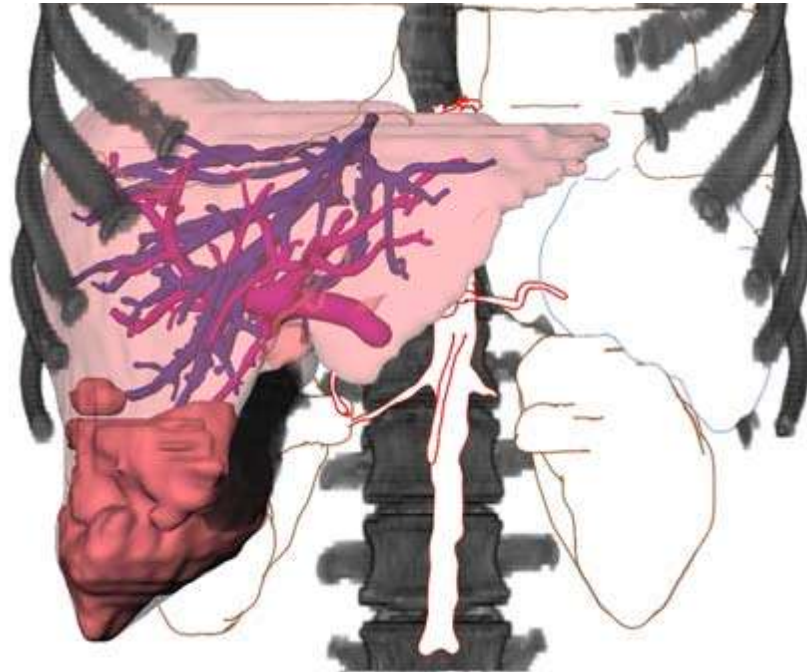
- measure some quality?
  - issue: what is the right quality to measure?
  - technique: statistical approaches
- ask people, ask an expert?
  - issue: what are the right questions?
    - avoid bias
  - techniques: qualitative, ethnographic
- mixed approaches?
  - issue: how to validate current algorithmic approaches?
  - technique: directly compare drawings by people with those generated by an algorithm

# Overview

- evaluation of hybrid illustrations with medical doctors [Tietjen et al., 2005]
- evaluation of pen-and-ink styles using an ethnographic technique [Isenberg et al., 2006]
- statistical evaluation of stipple rendering [Maciejewski et al., 2008]

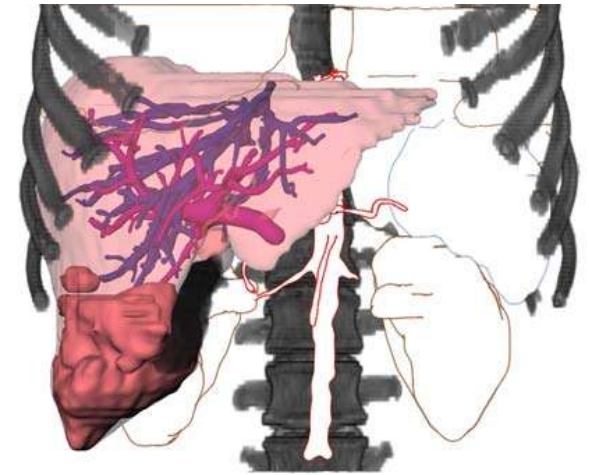
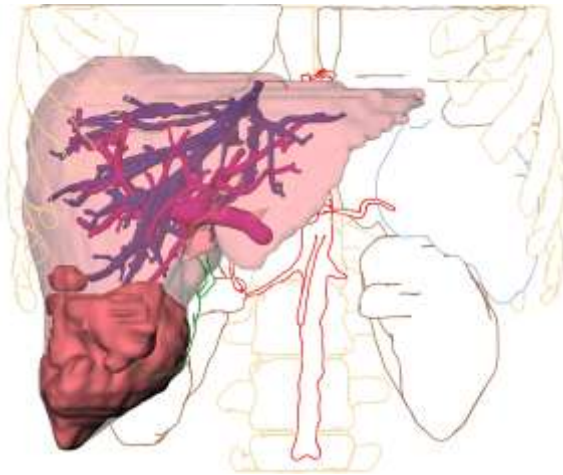
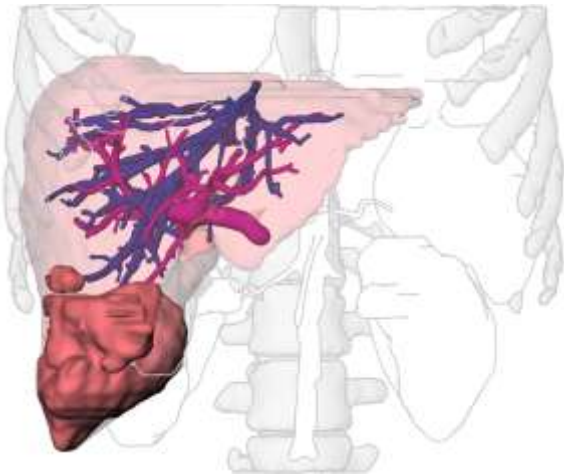


# Evaluation of Hybrid Illustrations with Medical Doctors

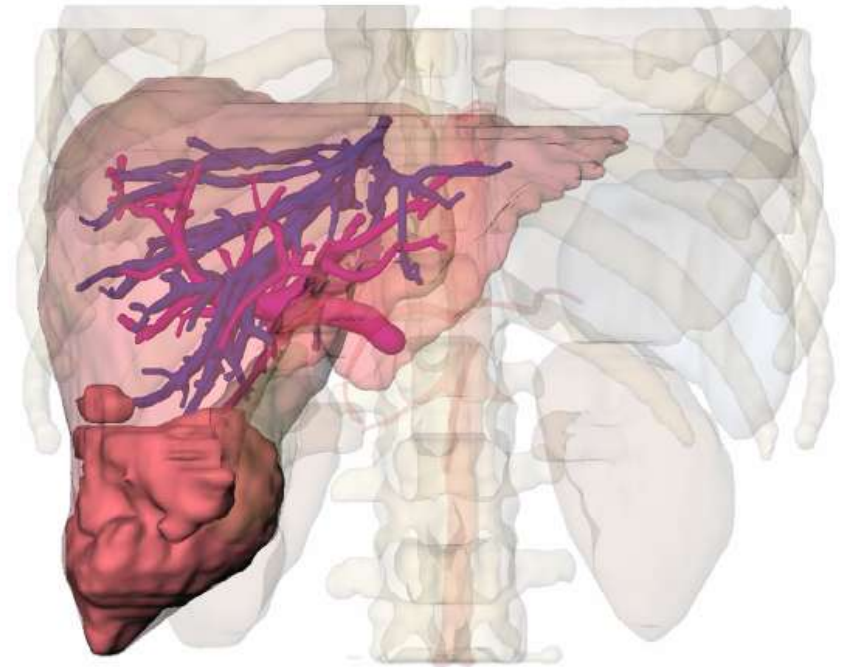
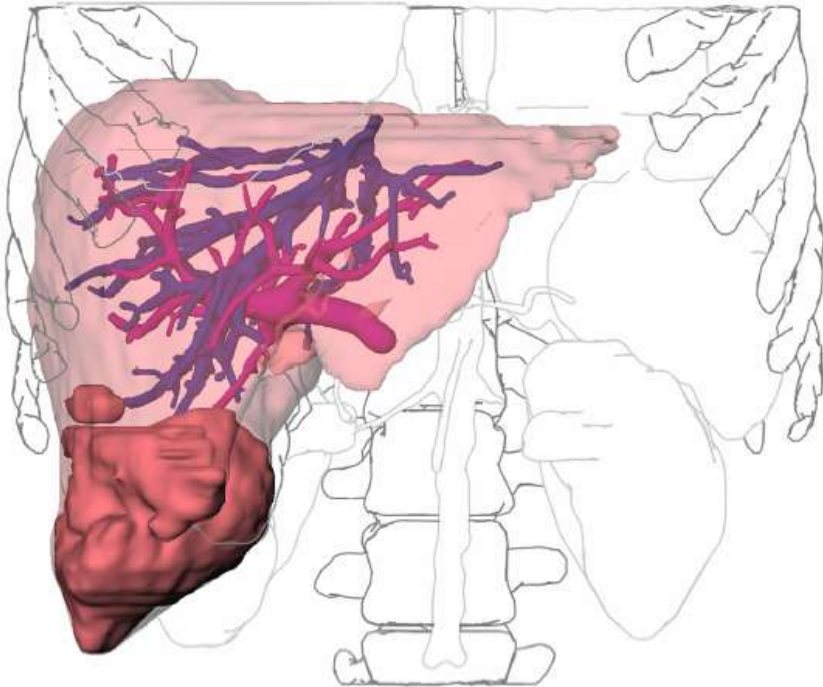


# Evaluation of Medical Illustration

- objects in focus, near the focus, and as context
- combinations of lines, shading, volume rendering
- comparison of different combinations of styles
  - specific application domain: experts in liver surgery
  - questionnaire-based evaluation, comparing 2 images per page: general impression, specific tasks/problems



# Evaluation of Medical Illustration



Welches Bild gefällt Ihnen auf den ersten Blick besser?

*Auf dieser Seite geht es um die direkte Gegenüberstellung der beiden Visualisierungstechniken.*

Wie gut ist die Leber von den umgebenden Strukturen zu unterscheiden?  
(gar nicht (--)) bis sehr gut (++)

Können Sie die Lage der Leber zum Brustkorb einschätzen?  
(nein, überhaupt nicht (--)) bis ja, sehr gut (++)

Wie gut sind die extrahepatischen Strukturen untereinander differenzierbar?  
(gar nicht (--)) bis sehr gut (++)

Mit welchem Bild würden sie sich auf eine Tumorsektion vorbereiten wollen?

links

--                      ++

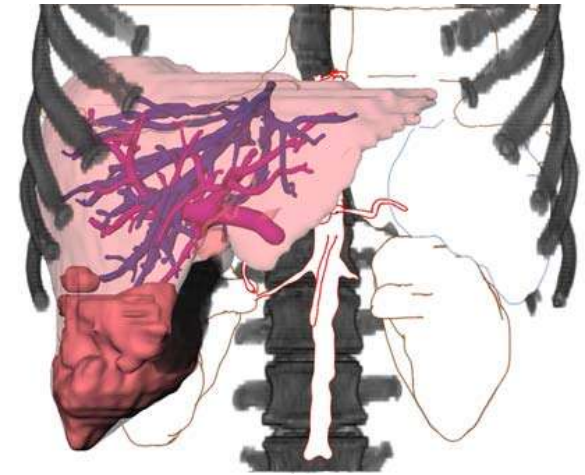
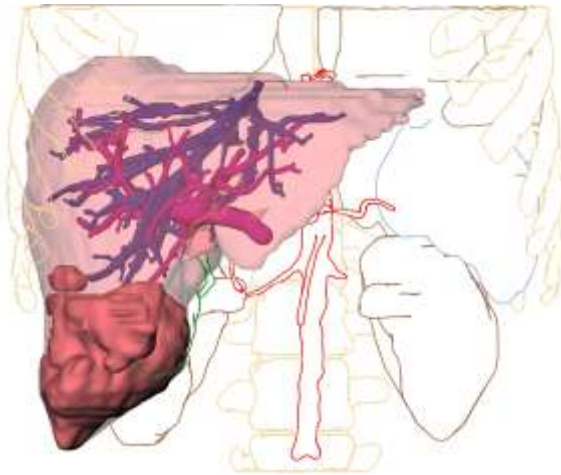
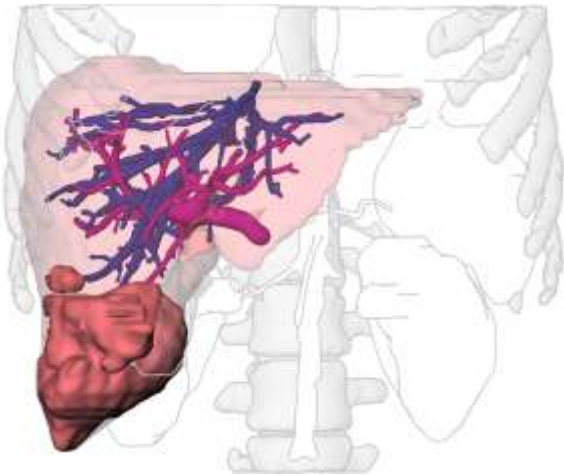
rechts

--                      ++



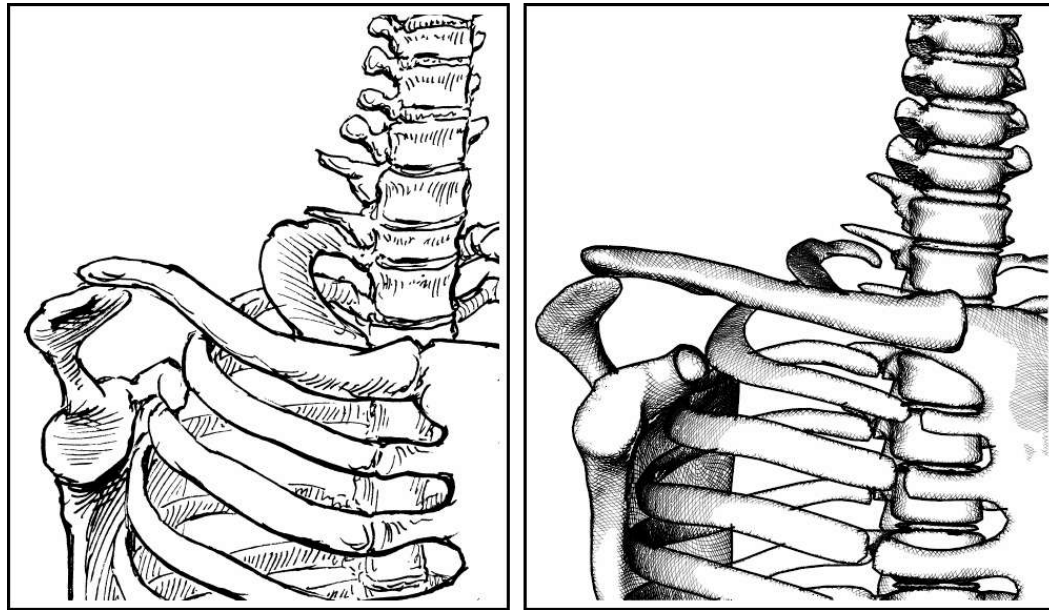
# Evaluation of Medical Illustration: Results

- less context information, but context is necessary
- silhouette representation of context appropriate
- lay people's opinions similar
  - only silhouettes not good, some shading important
  - colored silhouettes better, also slight shading good

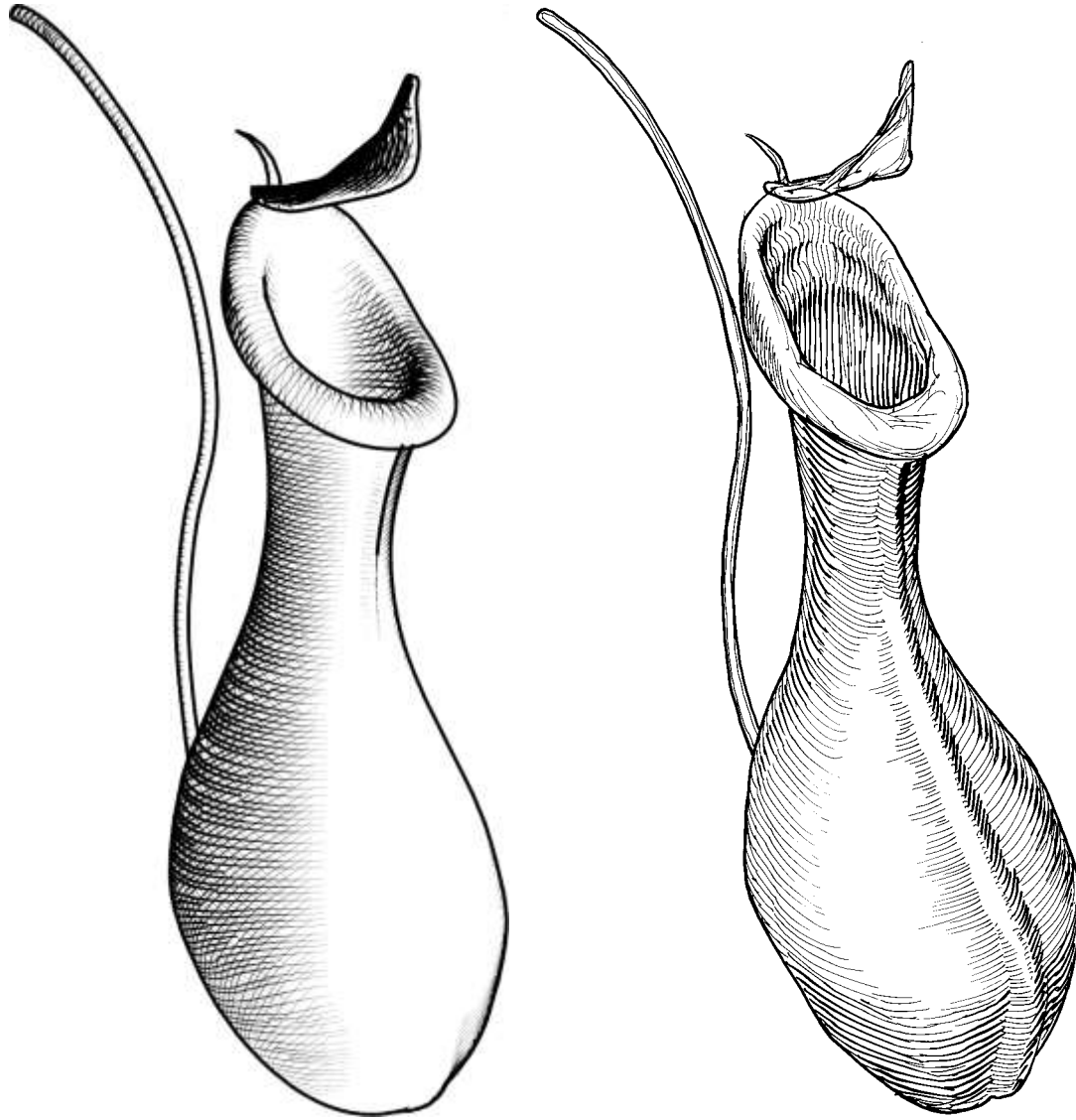


with G. Hojen & D. F. Heim

# Evaluation of Pen-and-Ink using an Ethnographic Technique



# Introduction



with P. Neumann, S. Carpendale, M. Sousa & J. Jorge

# Studying How People See Illustrations

---

- viewing/evaluating/understanding illustrations
  - complex process
  - difficult to analyze
  
- approach: exploratory observation
  - qualitative study, non-numeric results
  - no pre-determined hypothesis
  - criteria & terminology determined by participants
  - observing participants' actions & opinions

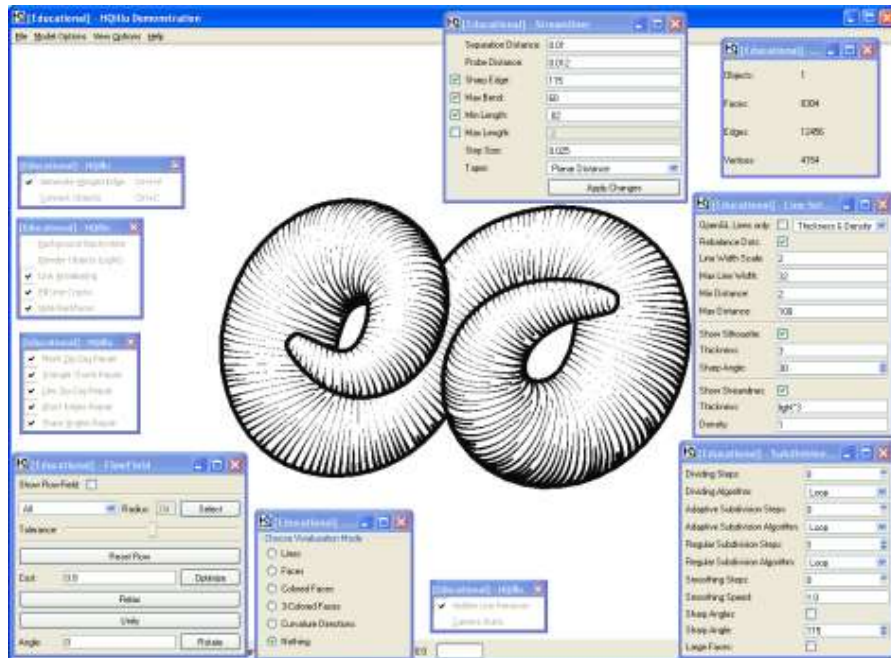
# Side Note: Ethnographic Studies

---

- qualitative evaluation technique
  - extraction of non-numeric criteria
  - e.g., opinions, feelings, concepts, common practices, etc.
  - major goal: not to bias/influence people
- bias already by wording of questions:
  - What do you like about this image?
  - Do you think this image lets you understand things well?
  - Why do you think this illustration is bad?
- question-based interviews will always bias people
- use techniques instead that do not ask questions but extract opinions/concepts otherwise

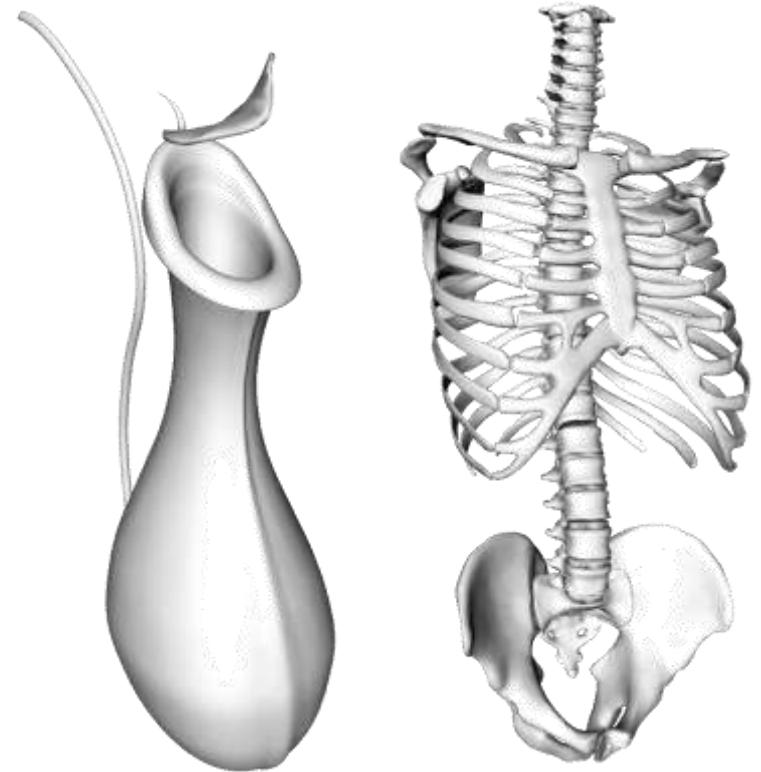
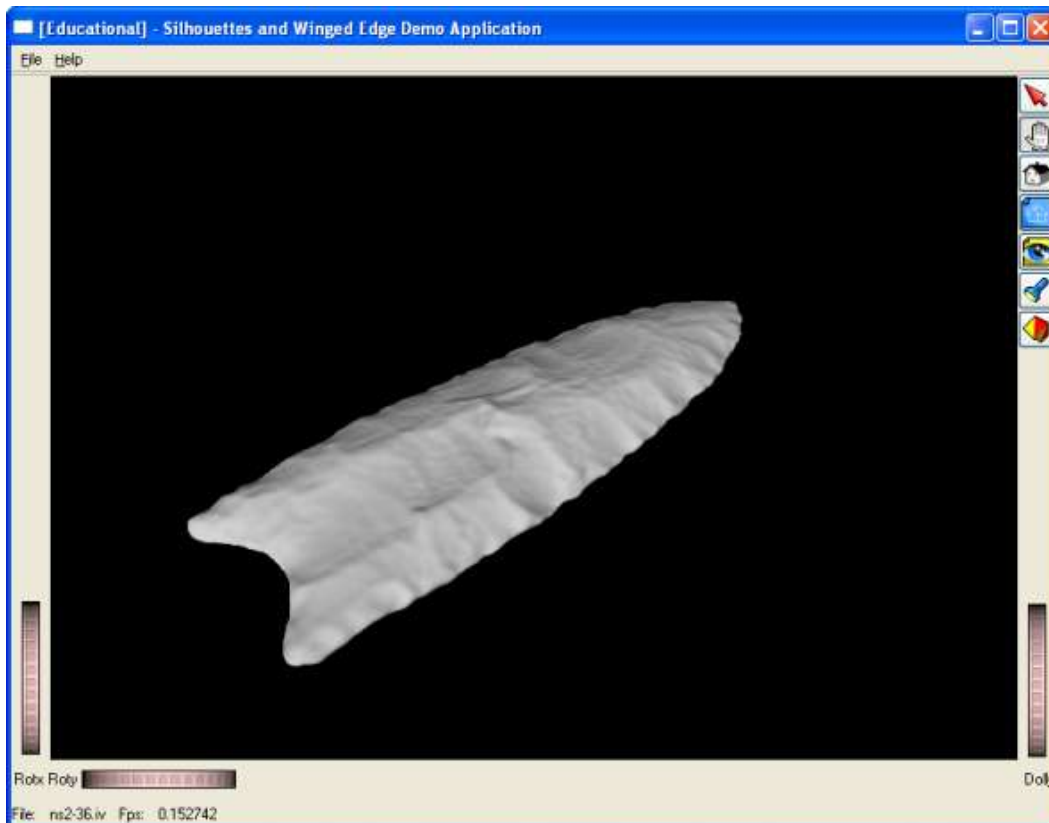
# Ethnographic Evaluation of NPR

- scope of the evaluation: compare NPR pen-and-ink techniques with ones created by illustrators by hand
- 5 NPR techniques compared to 5 illustrators: hatching and stippling techniques

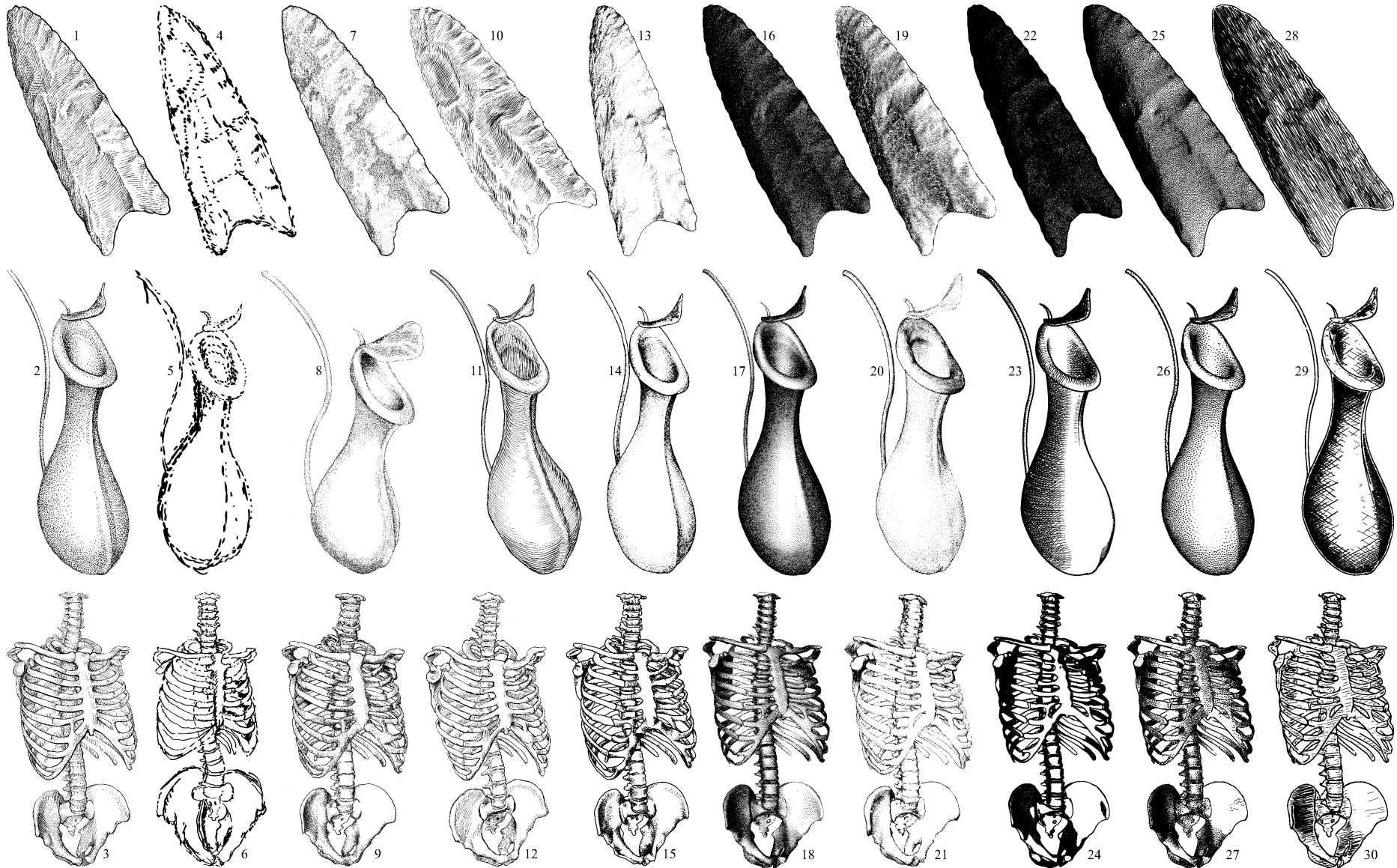


# Ethnographic Evaluation of NPR

- different domains, comparability between images
  - 3 different shapes: archaeology, anatomy, biology
  - same default few for each object



# Overview of Study Images





# Study Setup: Participants

---

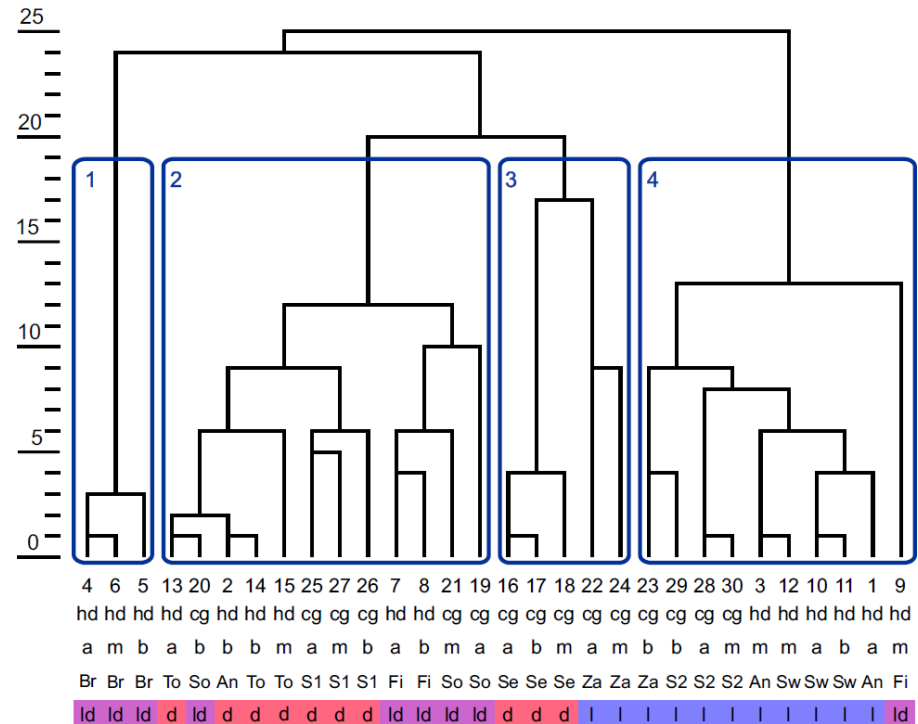
- four main groups identified:
  1. domain experts:  
scientists etc. who know their field
  2. professional illustrators:  
know how to create good illustrations
  3. illustration “end users:”  
learn with created illustrations
  4. NPR researchers:  
develop methods to generate illustrations with computers
- in our study: groups 2–4; mainly graduate students
- 8 participants per group;  $\Sigma$  24 participants

# Study Setup: Procedure

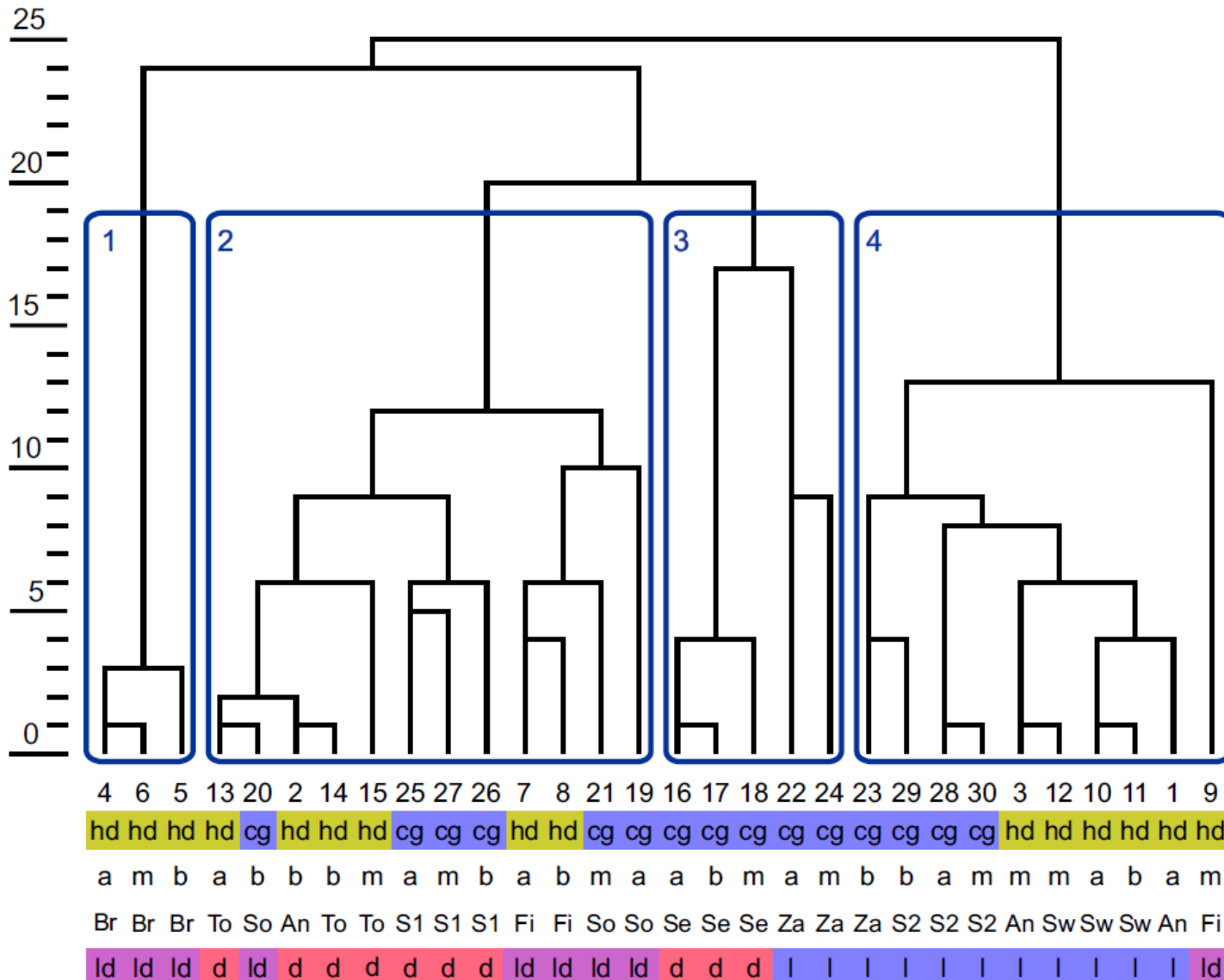


# Results: Cluster Analysis

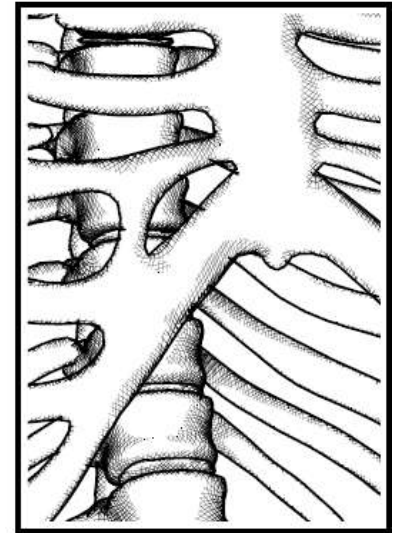
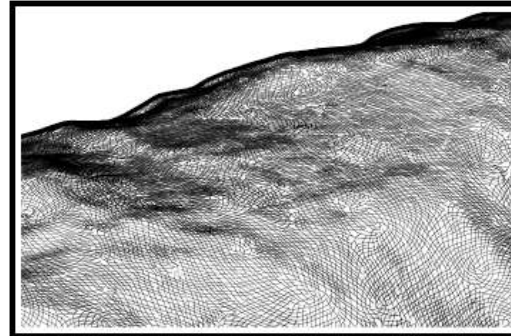
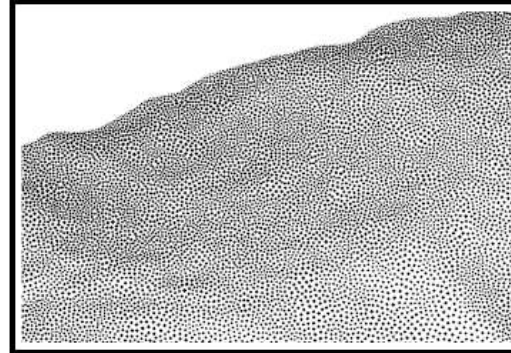
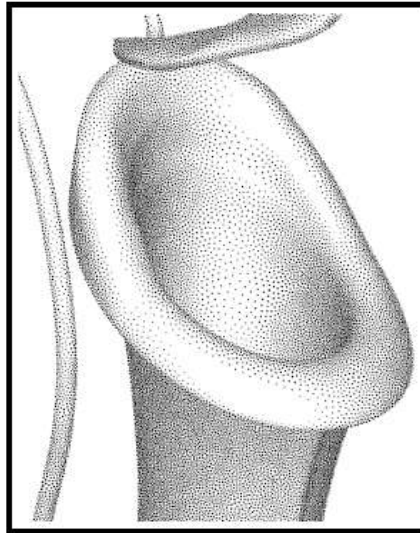
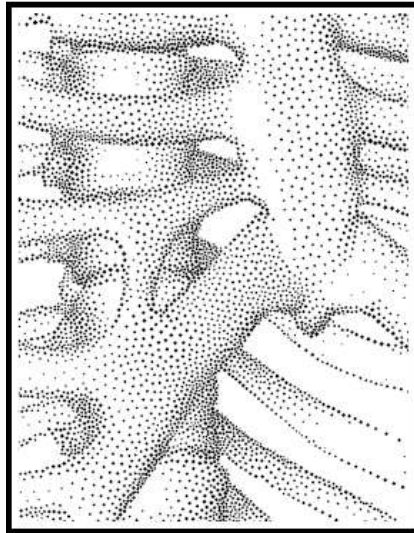
- categorization by drawing/rendering style by most ppl.
- criteria less often used: realism/detail, aesthetics, information contents, and orientation
- no significant differences between how the three groups categorized
- cluster graph from correlation table
- four main clusters



# Results: Style and Detail

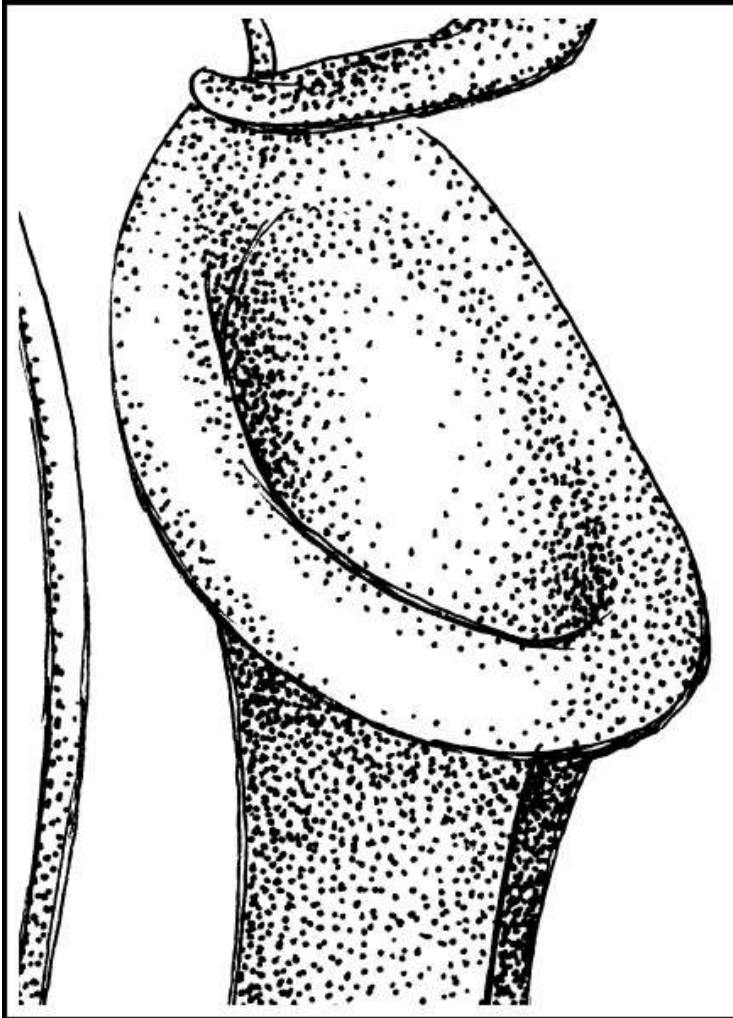


# Results: Style and Detail (Cluster 3)

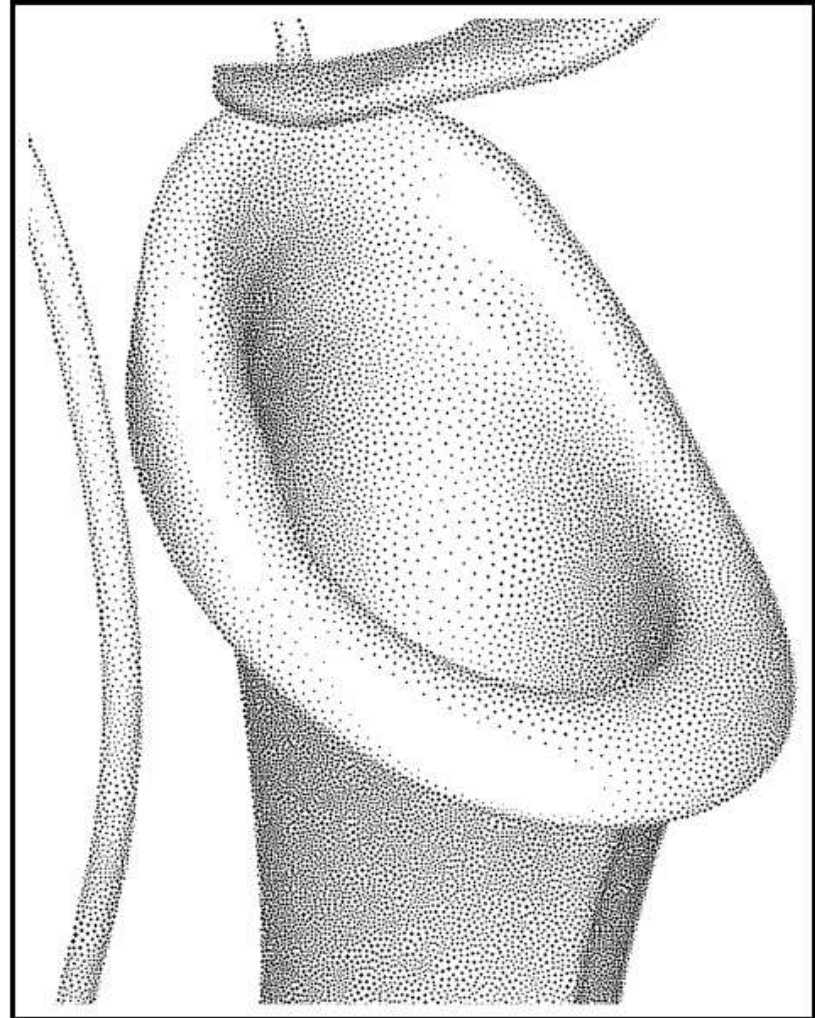


# Results: Style and Detail

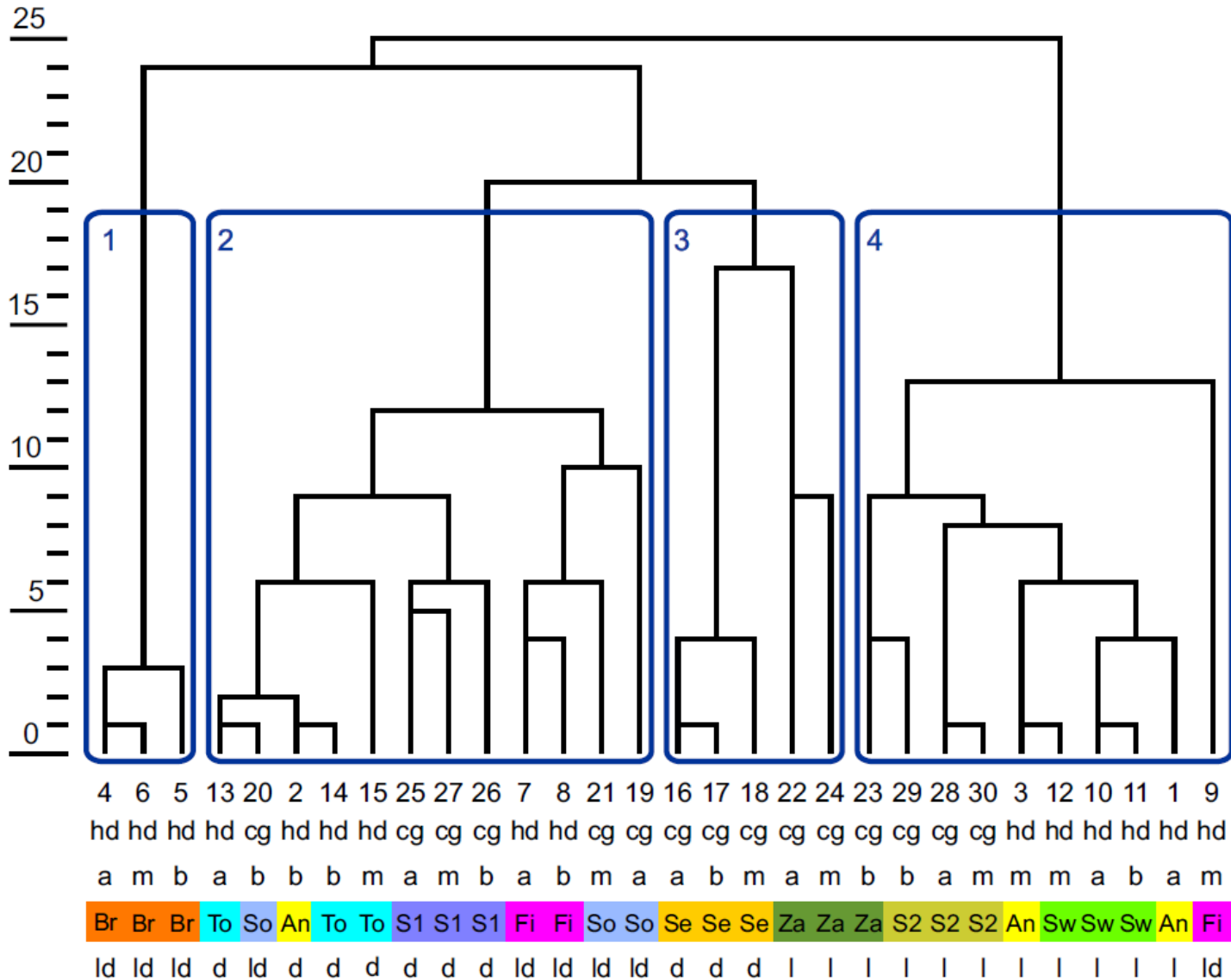
hand-drawn:



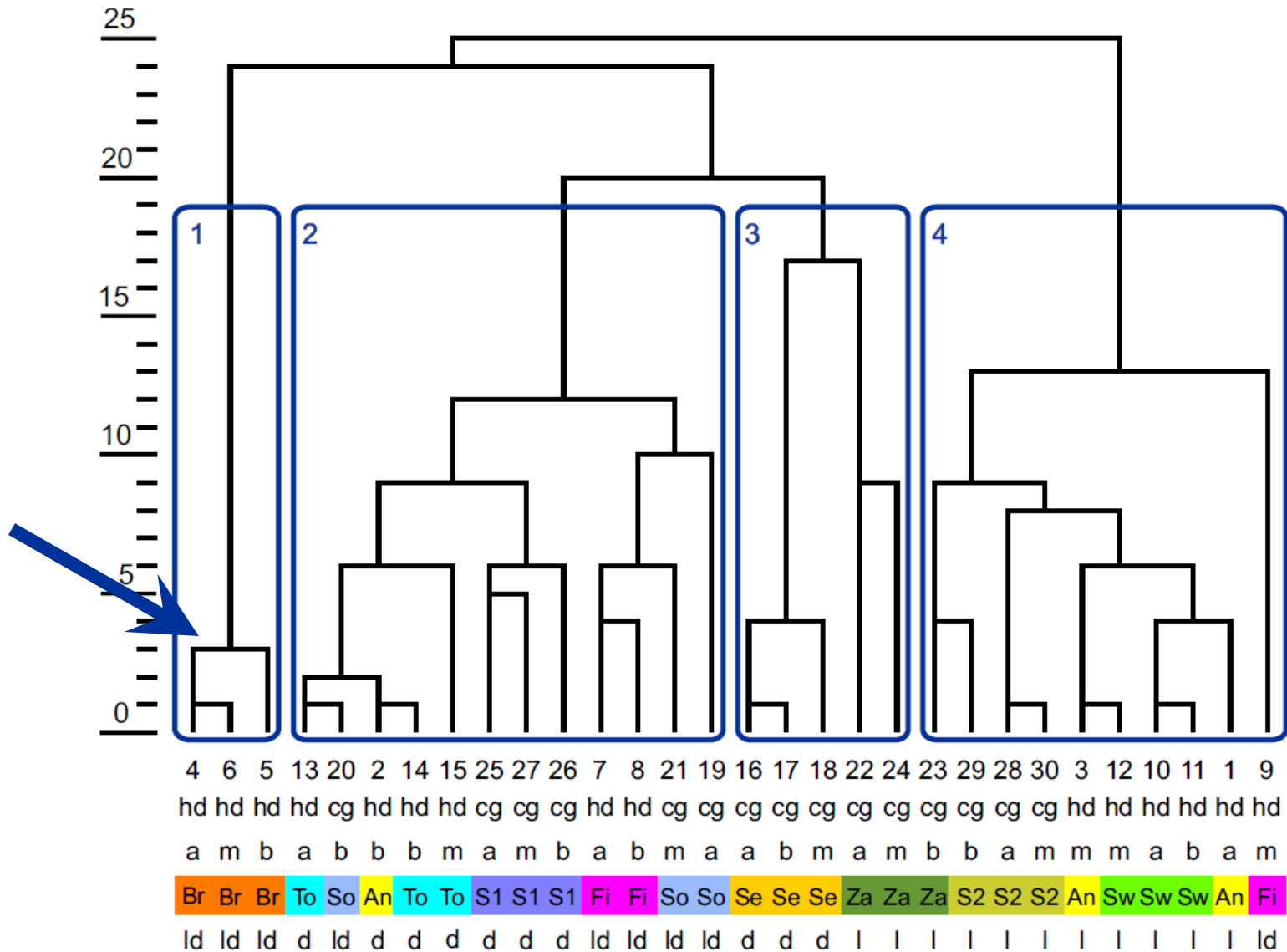
computer-generated:



# Results: Categorization by Artist/Algorithm

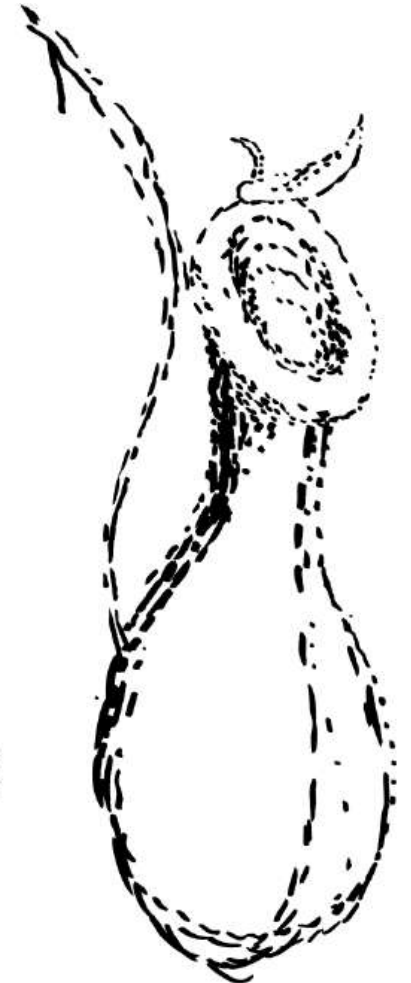
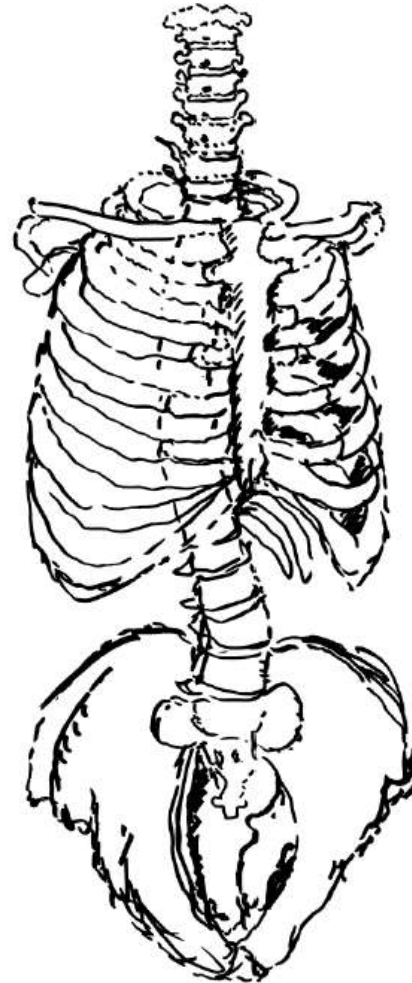


# Results: Cluster 1 – Loose and Sketchy

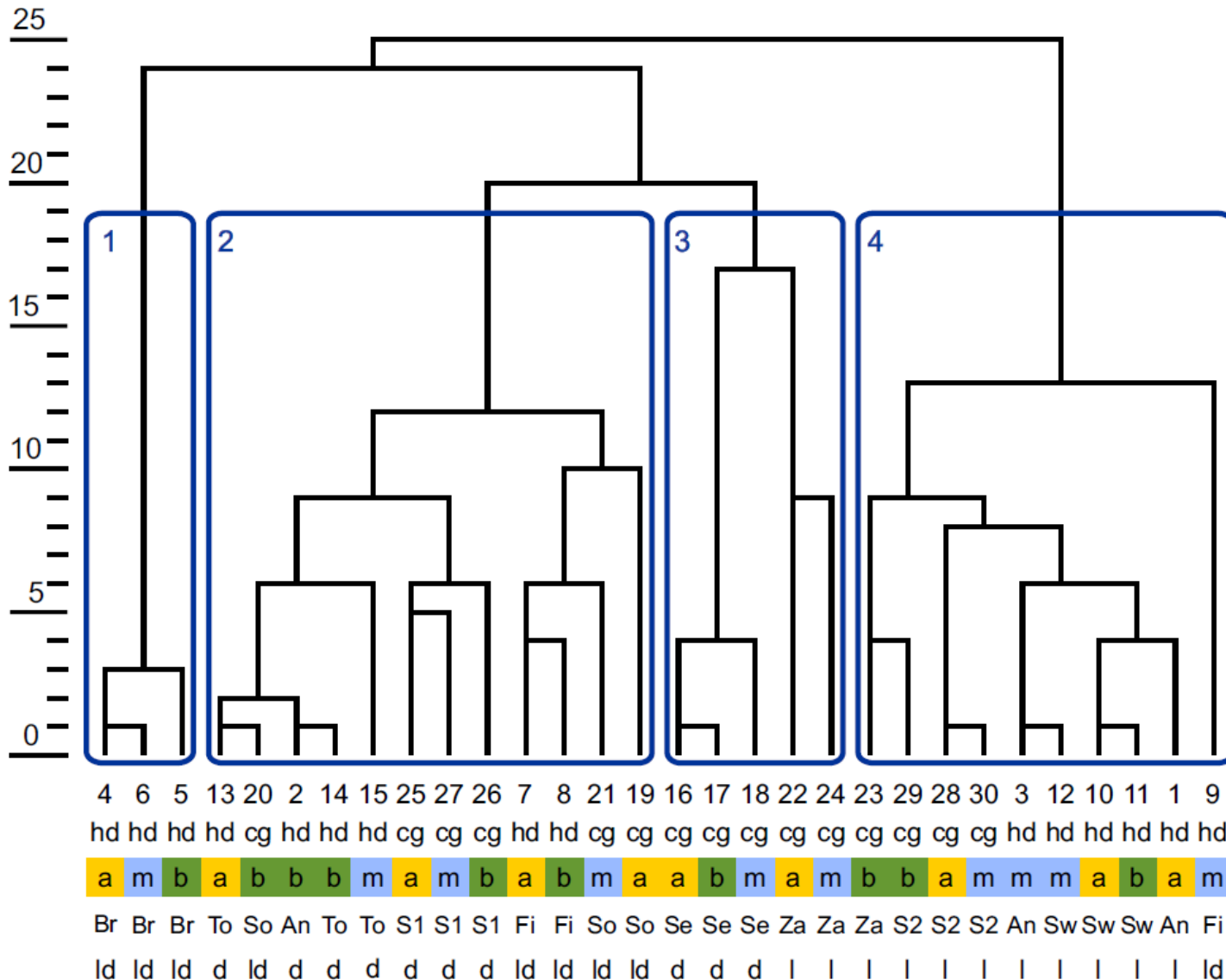




# Results: Cluster 1 – Loose and Sketchy

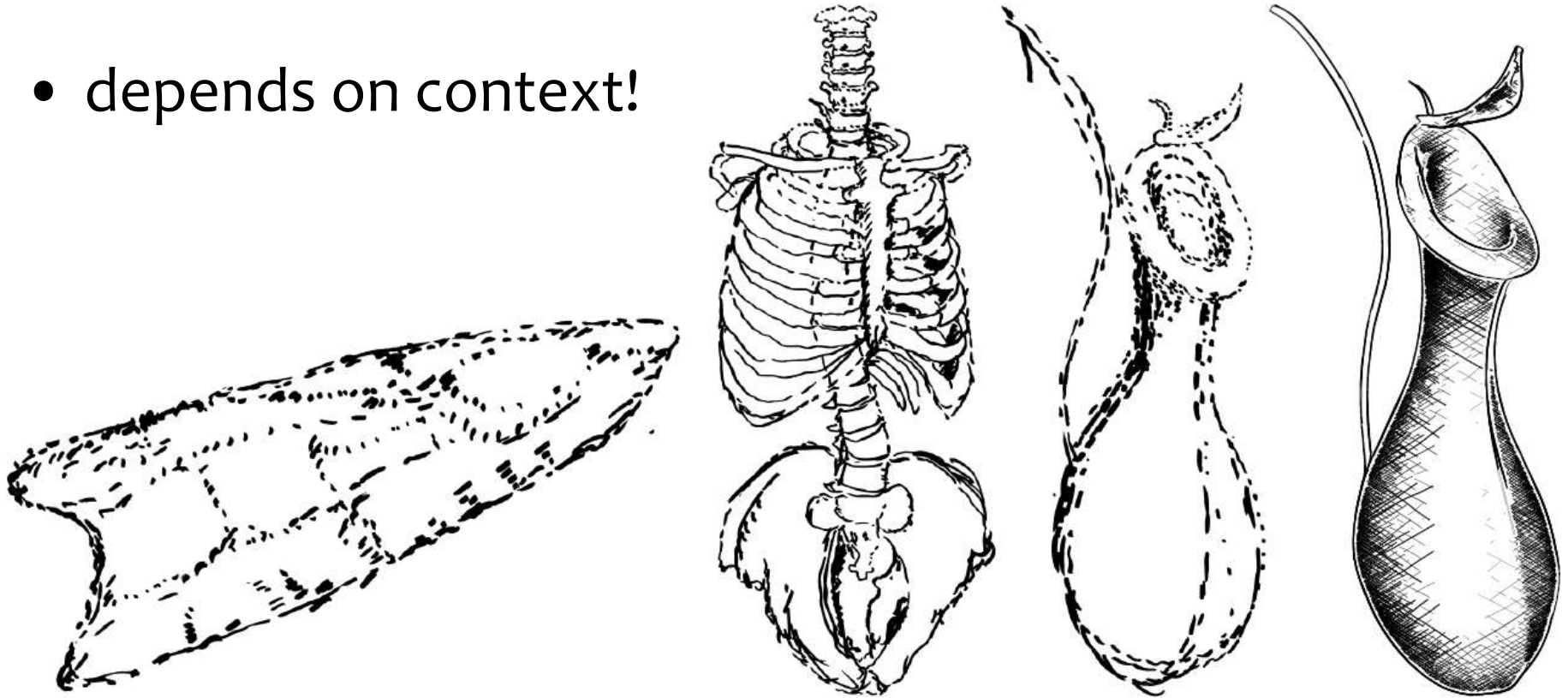


# Results: No Clustering w.r.t. Model



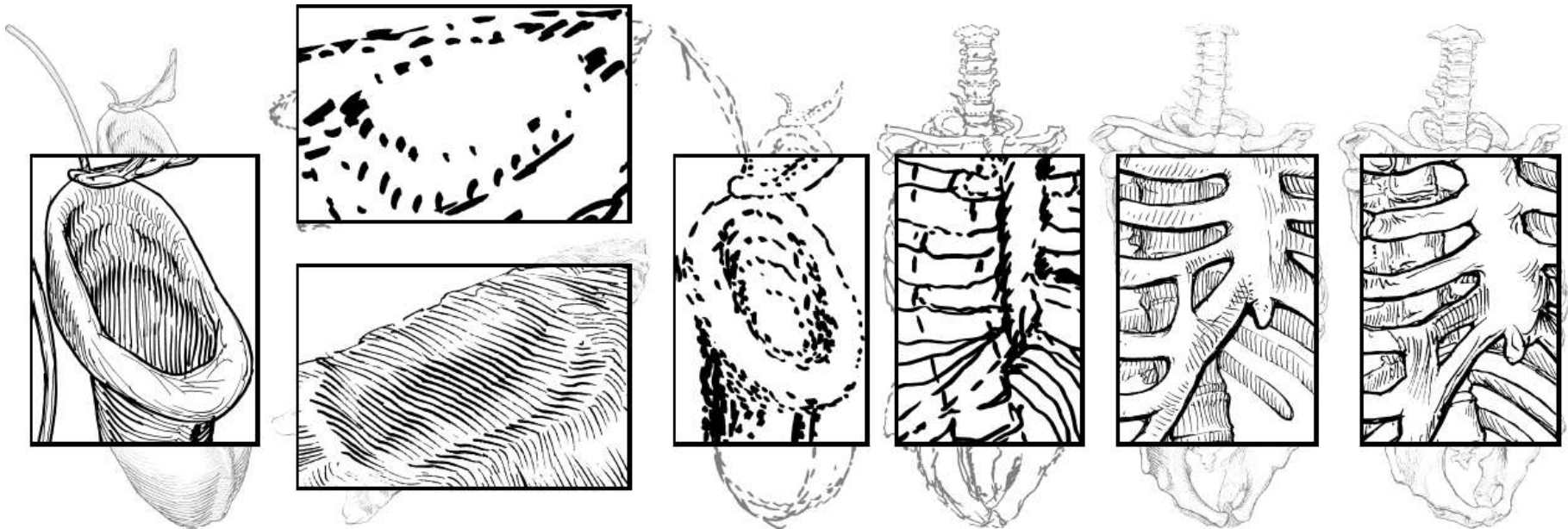
# Image Liking and Appeal

- no clear favorites, neither hand-drawn nor cg
- least favorites (named by  $\geq 58\%$ , all others  $\leq 25\%$ ):
- depends on context!



# Results: Images Looking CG or Hand-Drawn

- many hand-drawn images stood out as such – lines:



83%

83% / 71%

79%

79%

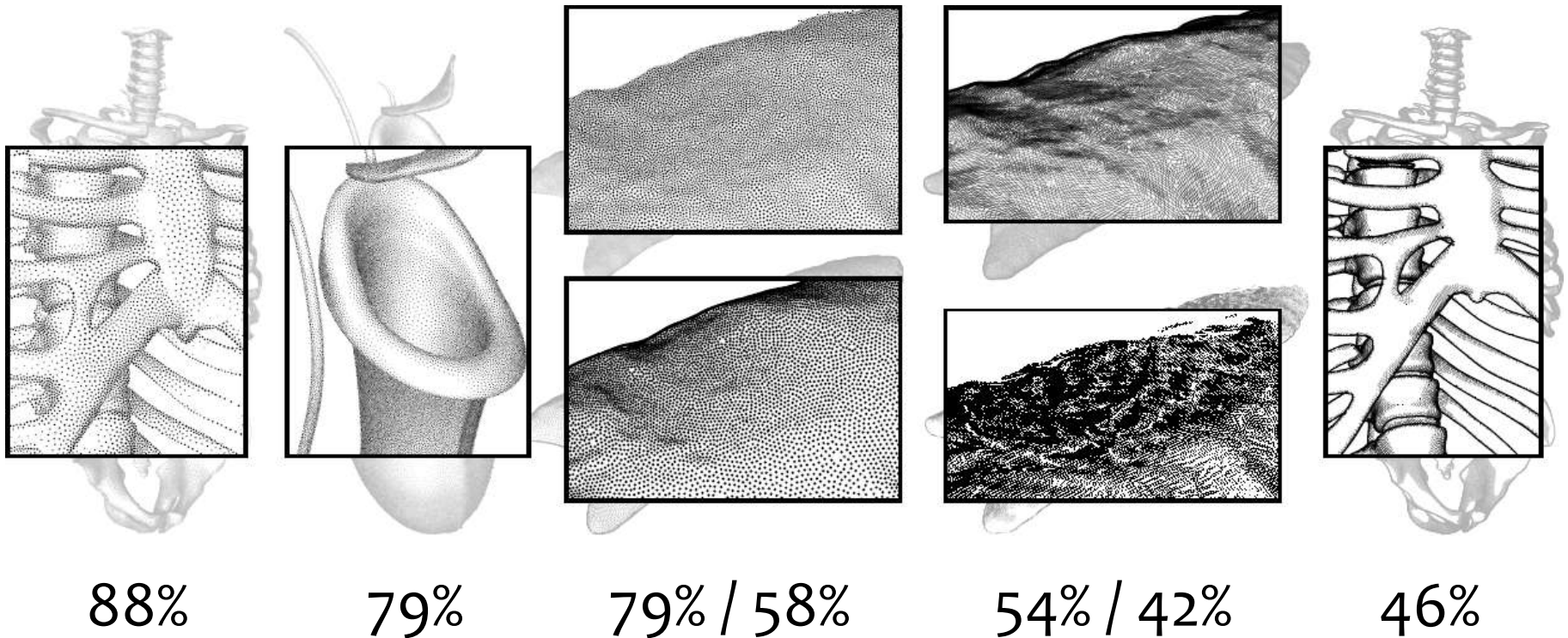
67%

58%

- hand-drawn images less often named to stand out as such: stippling or mix of stippling with lines

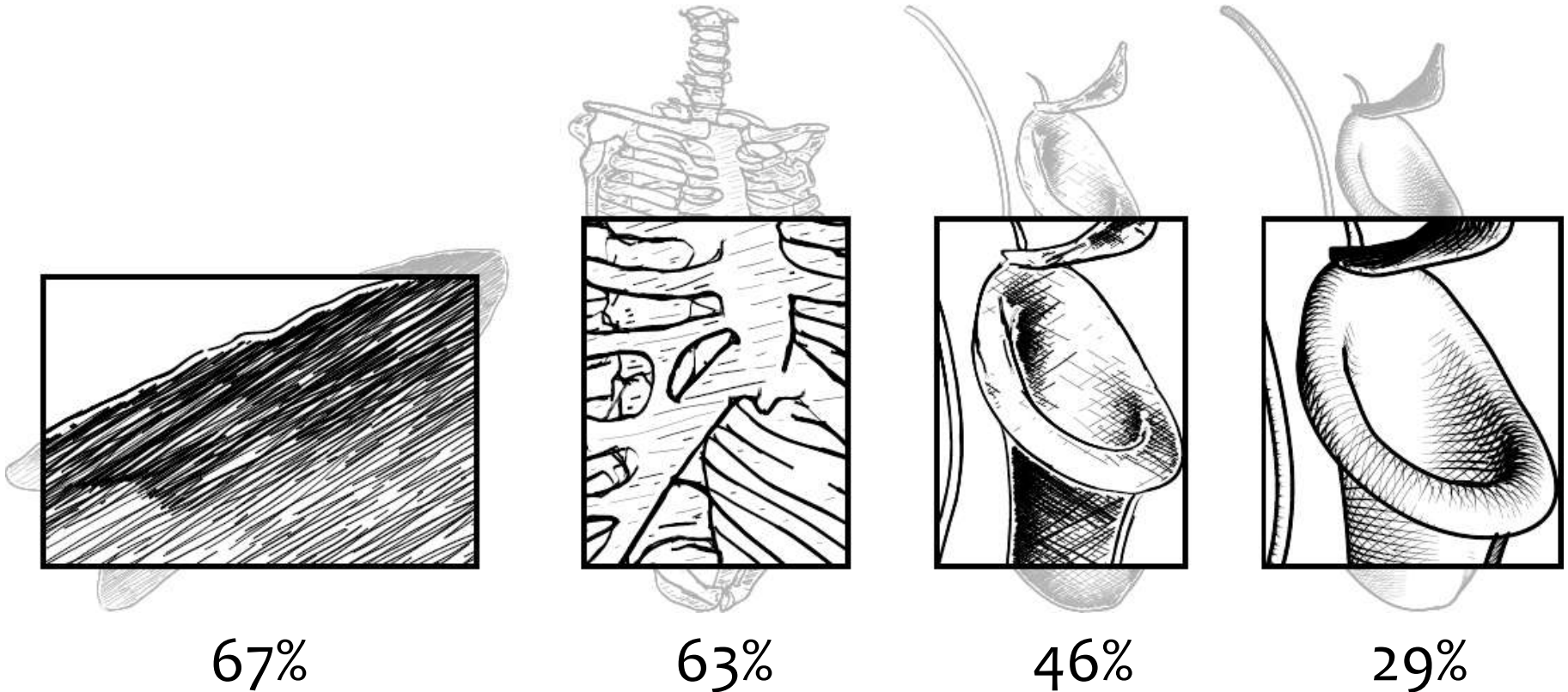
# Results: Images Looking CG or Hand-Drawn

- cg images often named to stand out as such: stippling or high-resolution lines



# Results: Images Looking CG or Hand-Drawn

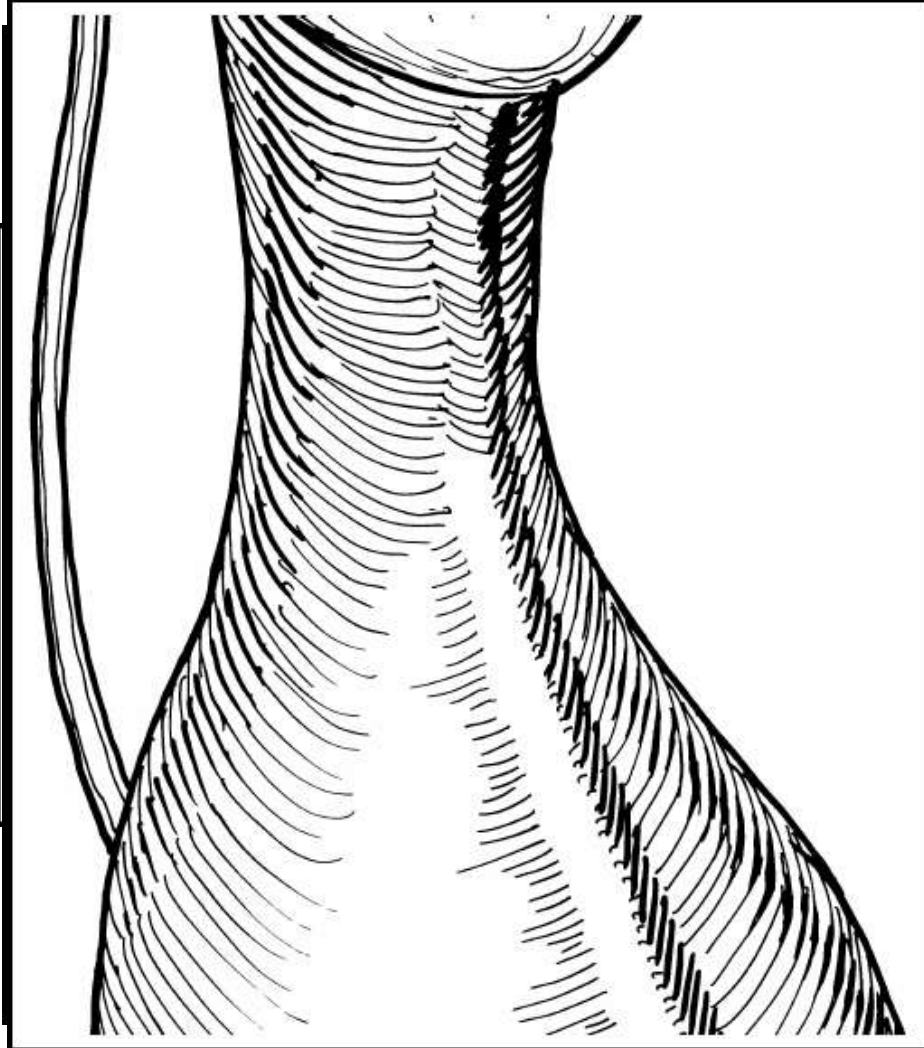
- hand-drawn images rarely thought to be cg ( $\leq 13\%$ )
- some cg images frequently thought to be hand-drawn:



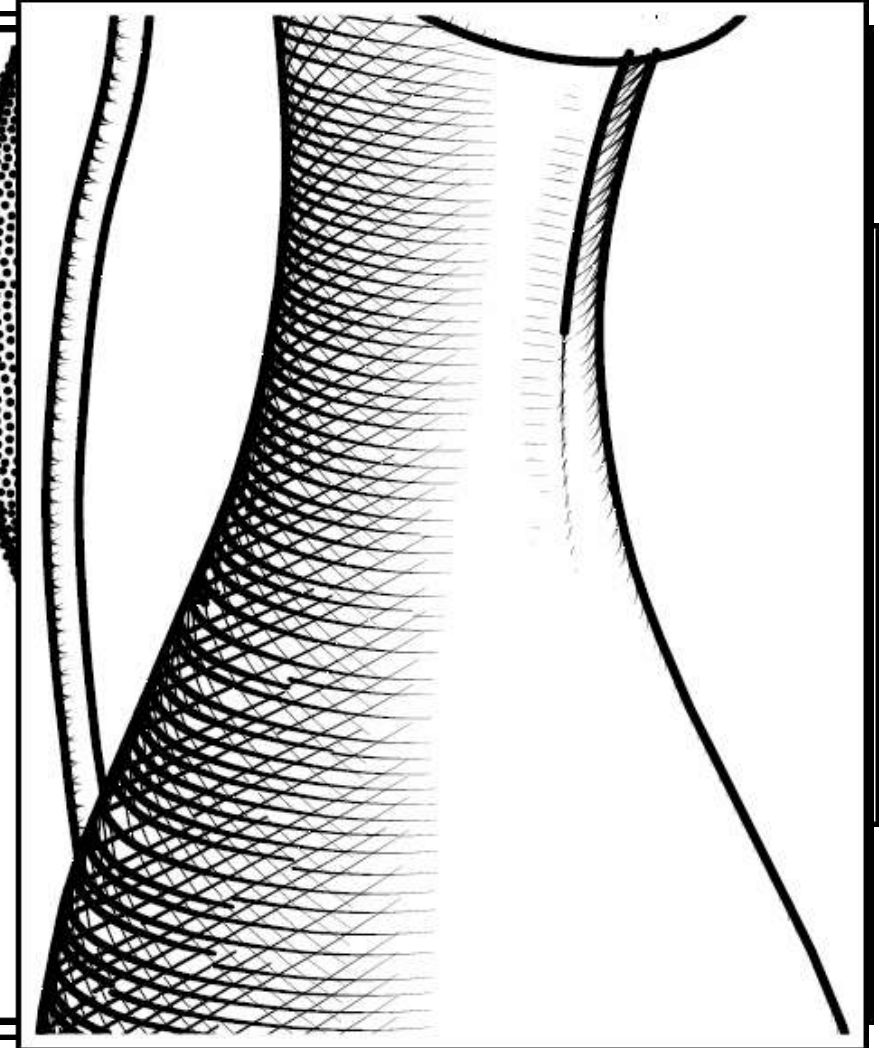
- randomness, longer and less dense lines, lower detail

# Other Observations (from Discussion)

hand-drawn:  
hand-drawn:

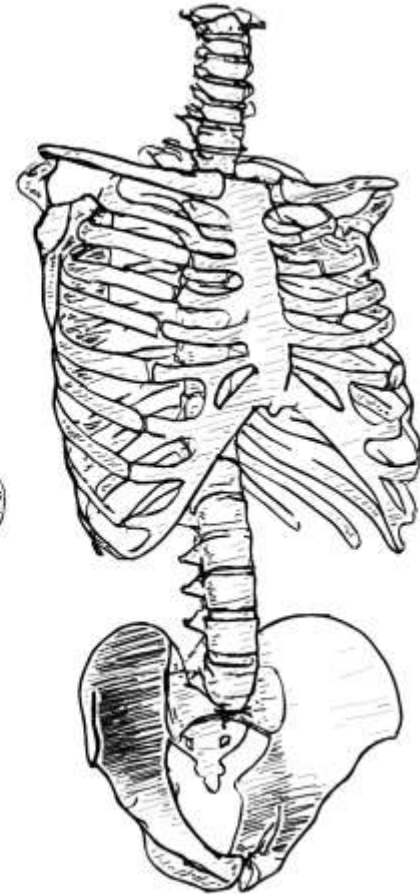
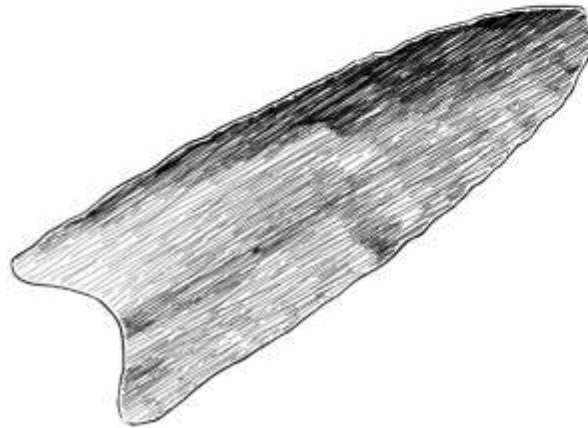


computer-generated:  
computer-generated:



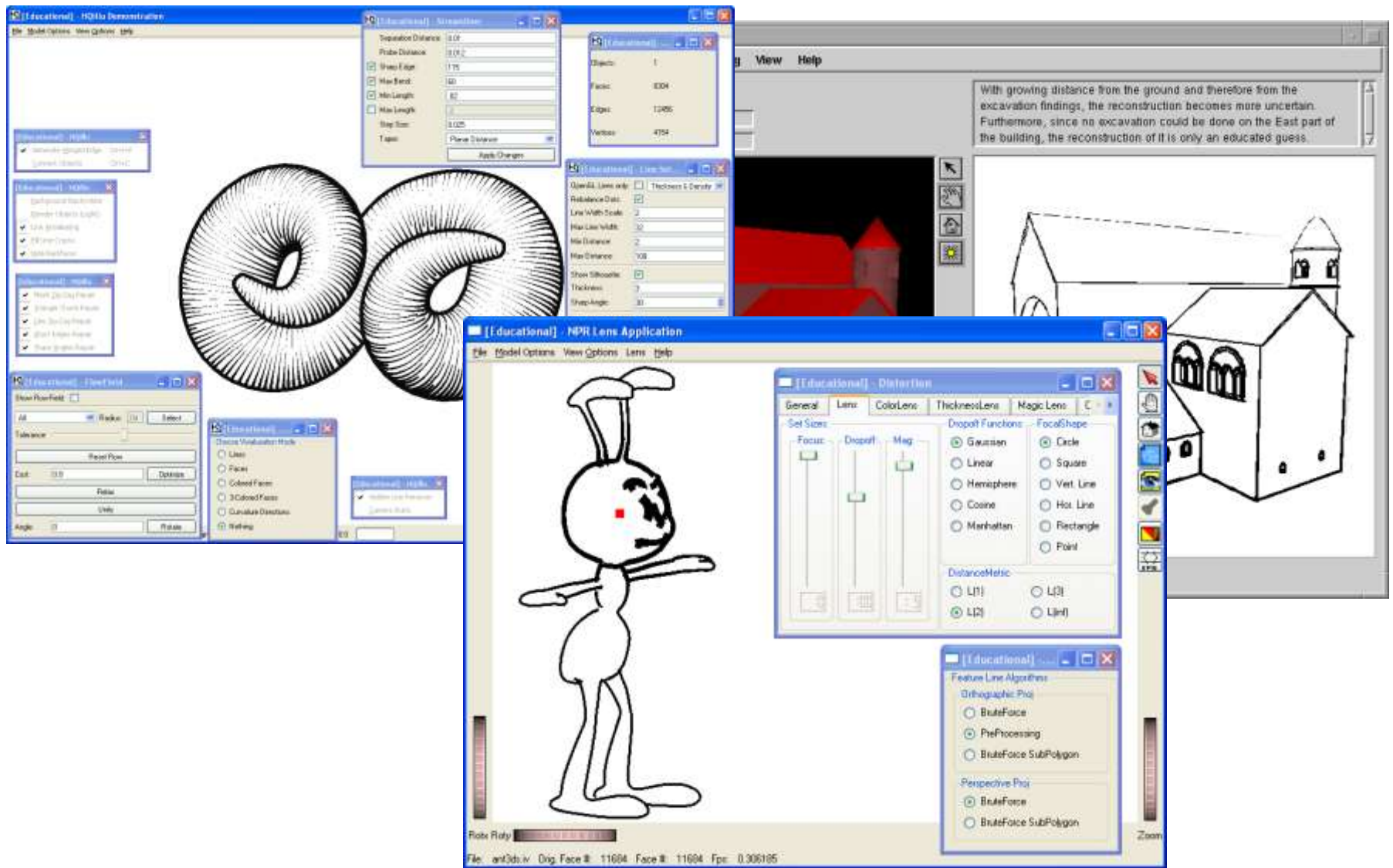
# Results

- NPR-Turing test not passed, CG images recognized as such (named standing out as hand-drawn by  $\leq 29\%$ )
- some almost always recognized as CG: detail, 3D shading, exactness
- depending on algorithm & parameters
- one major exception: RenderBots hatching “sketchy,” “simplified,” “not show shape well”

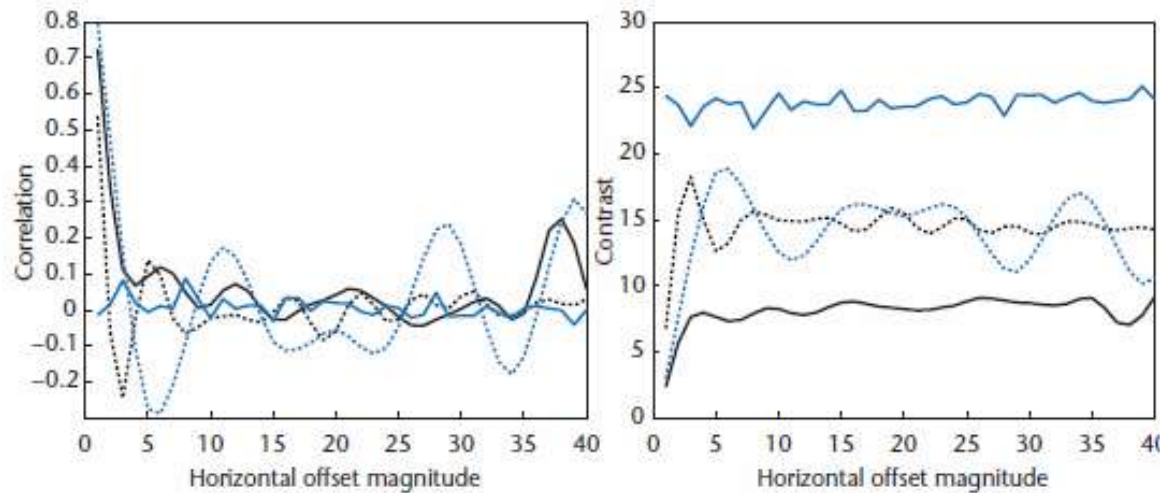




# Other Result: Interfaces and Interaction



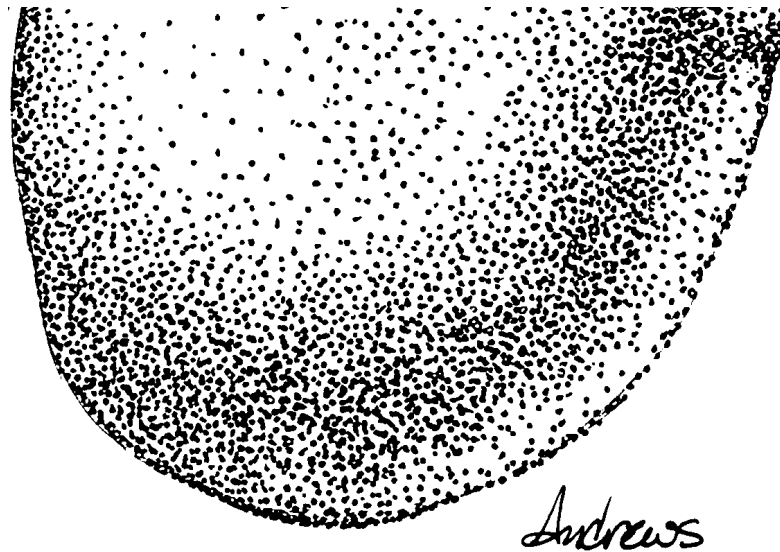
# Statistical Evaluation of Stippling



# Statistical Evaluation of Stippling

---

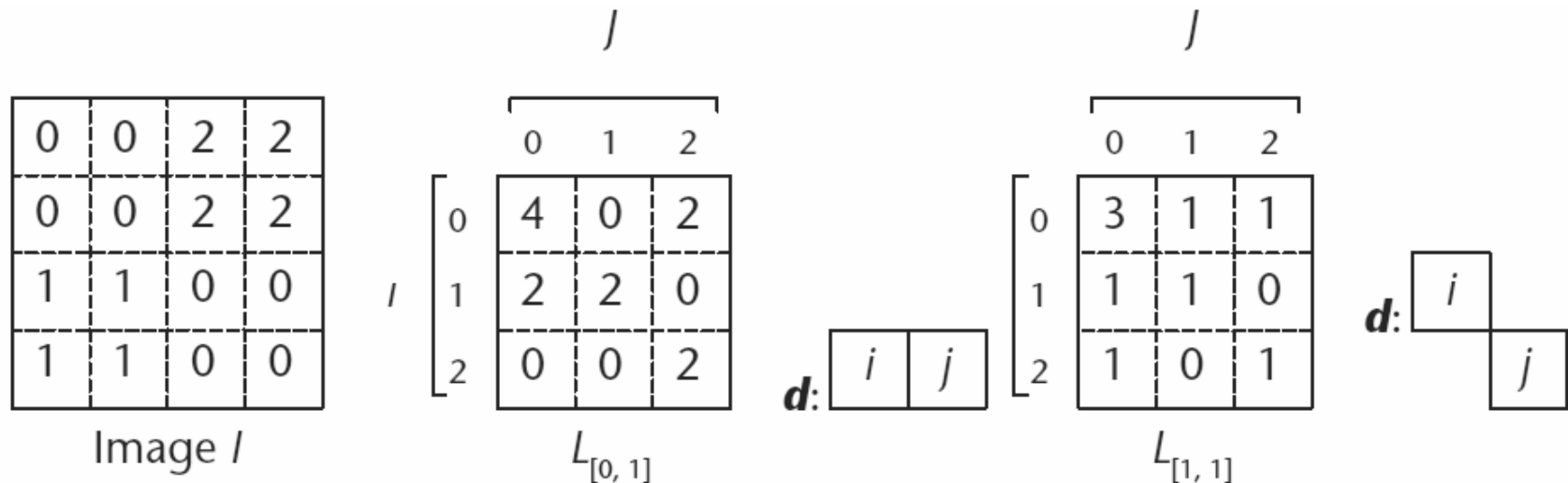
- what characterizes a hand-drawing style?
- stippling: distribution of stipple points
  - can be analyzed with respect to each other
  - statistical metrics to analyze properties of distribution



- goal: compare hand-drawn stippling to CG images

# Statistical Evaluation of Stippling

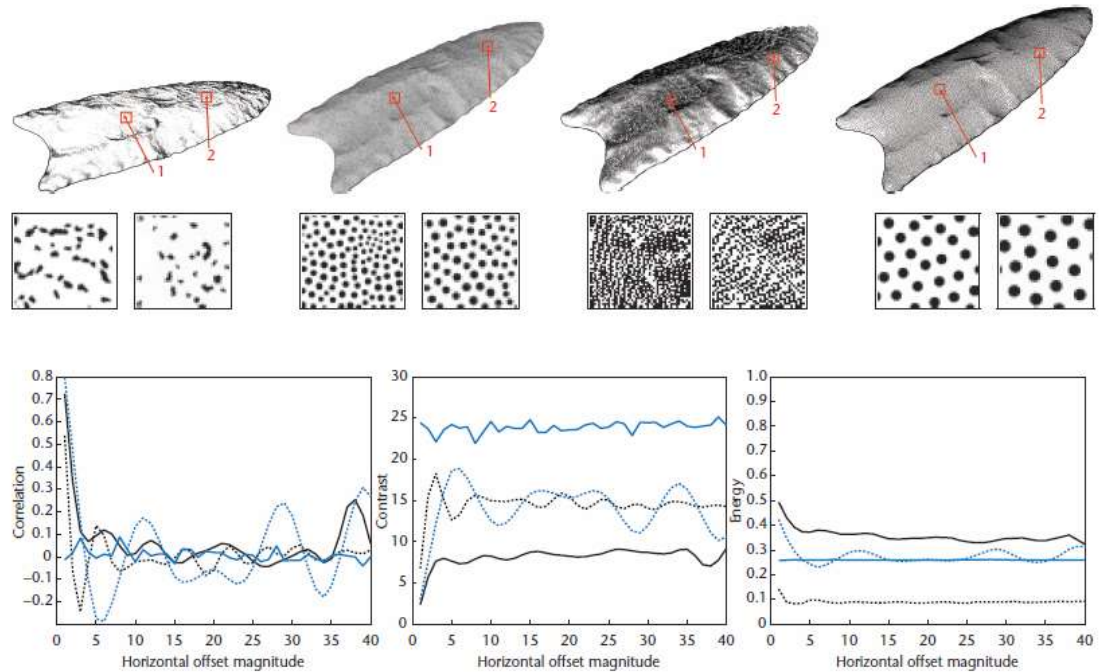
- based on gray-level co-occurrence matrix (GLCM)
  - 2D array recording the number of co-occurrences of gray level values in given spatial relationship
  - based on given offset vector, example:



- probability that a given gray value occurs in certain spatial relationship with respect to other gray value

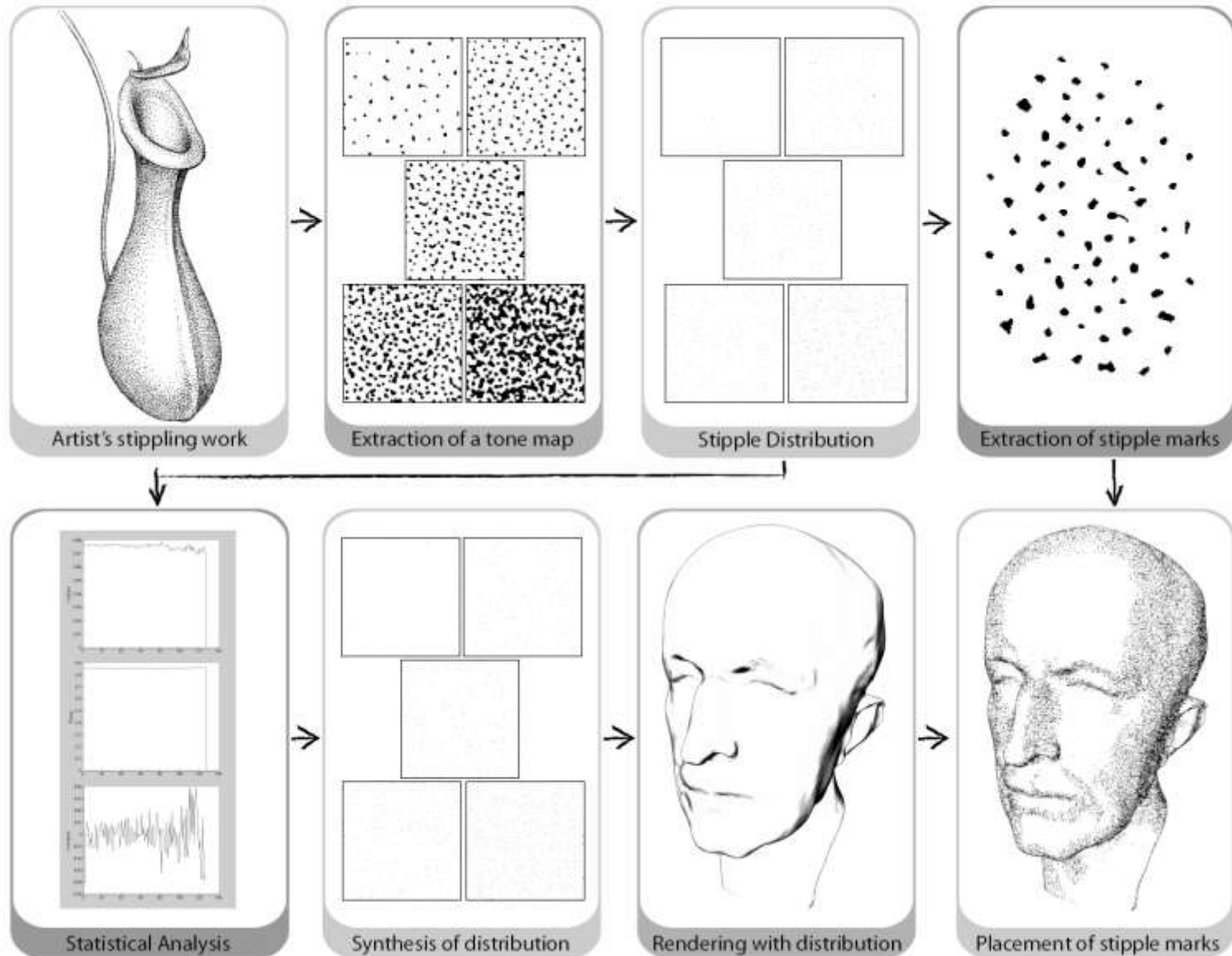
# Statistical Evaluation of Stippling

- CG images:
  - higher correlation of stipple placement to distance from other stipples in certain distances



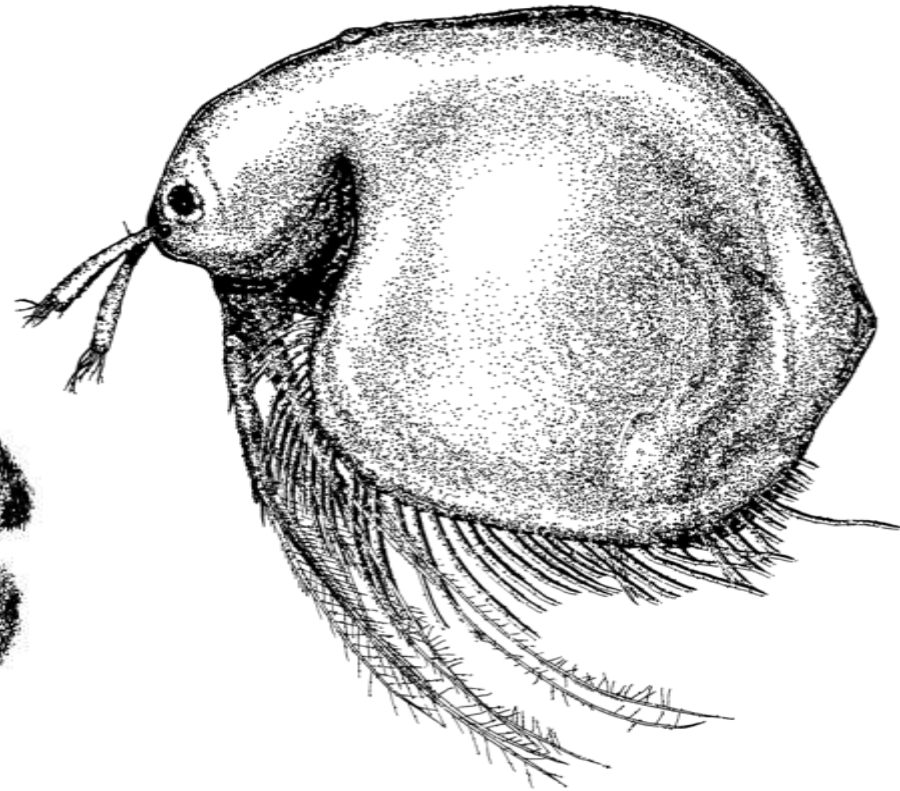
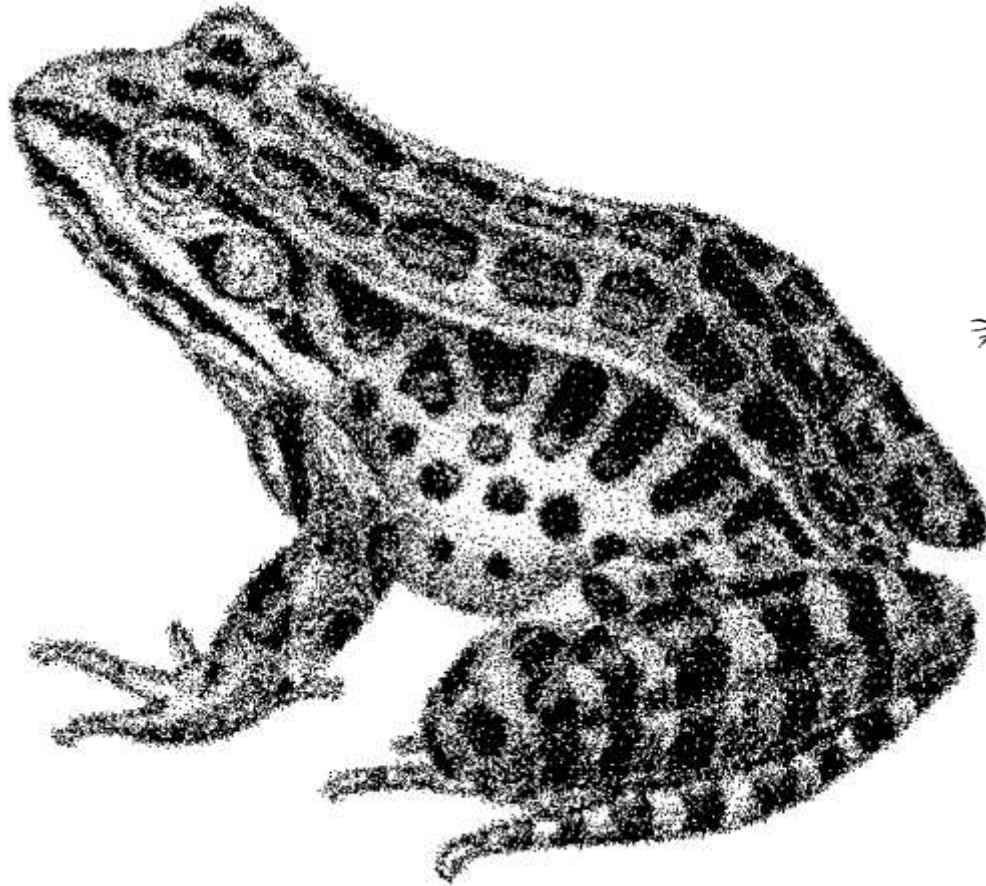
- other results:
  - artifacts of CG stippling can be found in statistics
  - hand-drawn stippling has similarities to natural textures

# Statistical Evaluation – Next Step: Synthesis

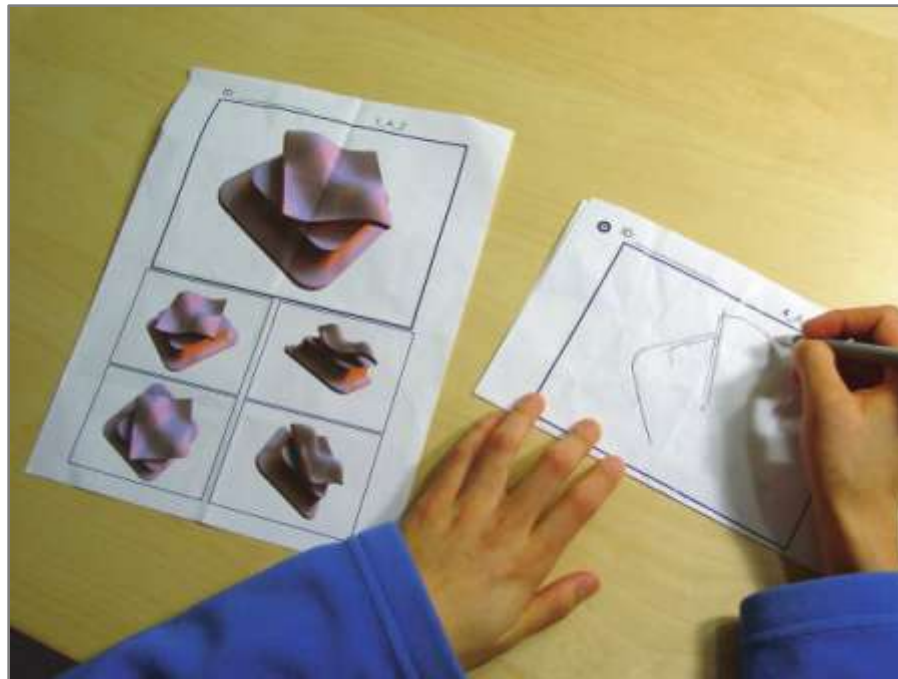


# Statistical Evaluation – Next Step: Synthesis

---



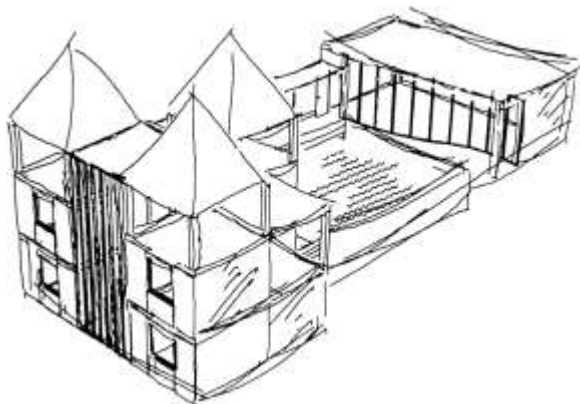
# Other Evaluation Approaches





# Other Evaluation Approaches (Selection)

**NPR in architecture**



[Schumann et al., 1996]

**space perception**



[Gooch & Willemsen, 2002]

**psychology of NPR**



[Halper et al., 2003]

**influence on gaze**



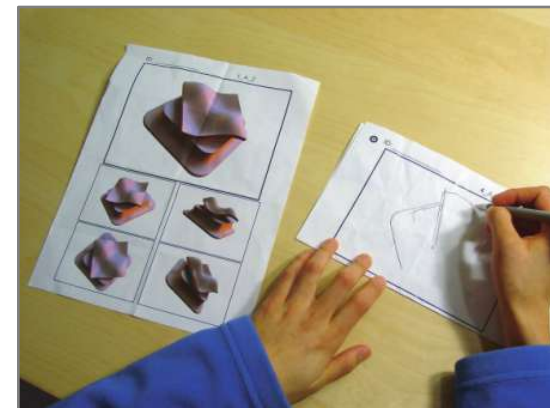
[Santella & DeCarlo, 2004]

**facial illustration**



[Gooch et al., 2004]

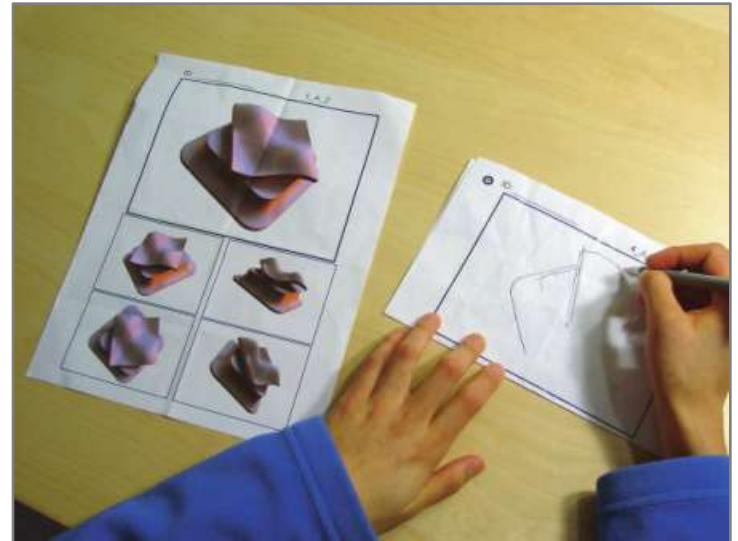
**where do people draw lines**



[Cole et al., 2008]

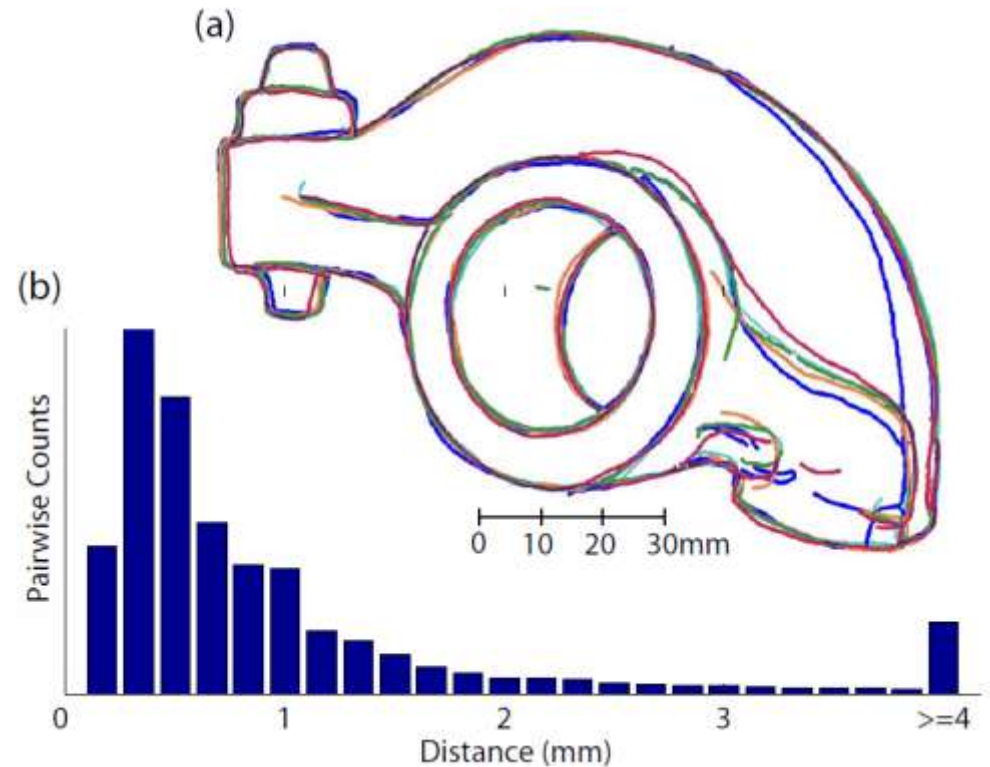
# Where do people draw lines? [Cole et al., 2008]

- goal: correlate people's line drawings with NPR
  - compare hand-drawings with current NPR line rendering concepts (silhouettes and feature lines)
  - derive algorithms that predict where artists would draw lines with certain probabilities
- approach: let people draw shapes from CG images
  - 2 steps: free drawing in frame and tracing a faint copy
  - traced images scanned and registered within frame
  - post-processing to obtain one pixel wide lines
  - 29 artists, each person drew 12 shapes



# Where do people draw lines? Results

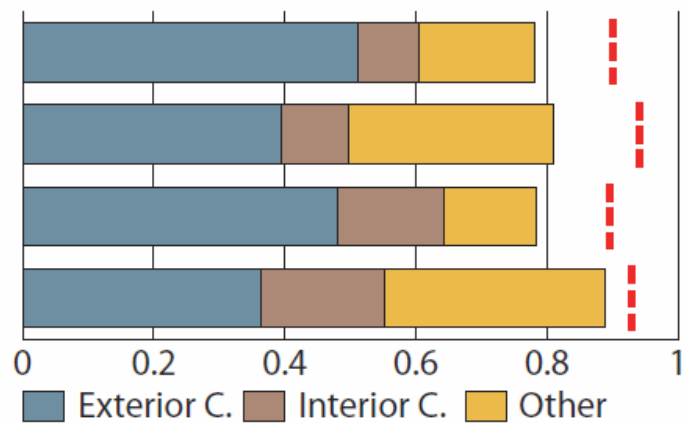
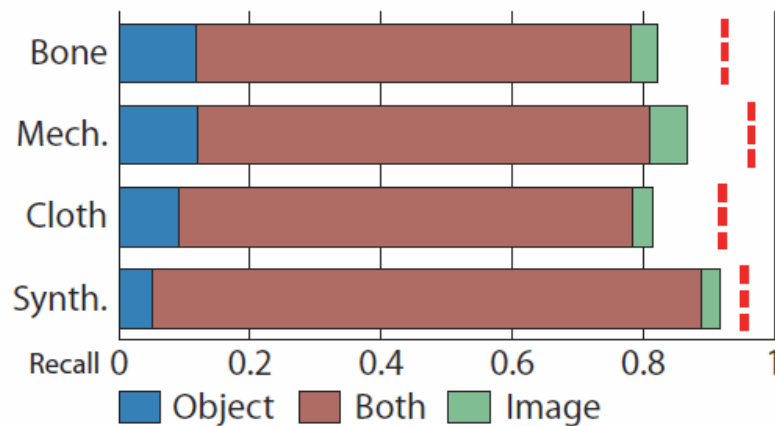
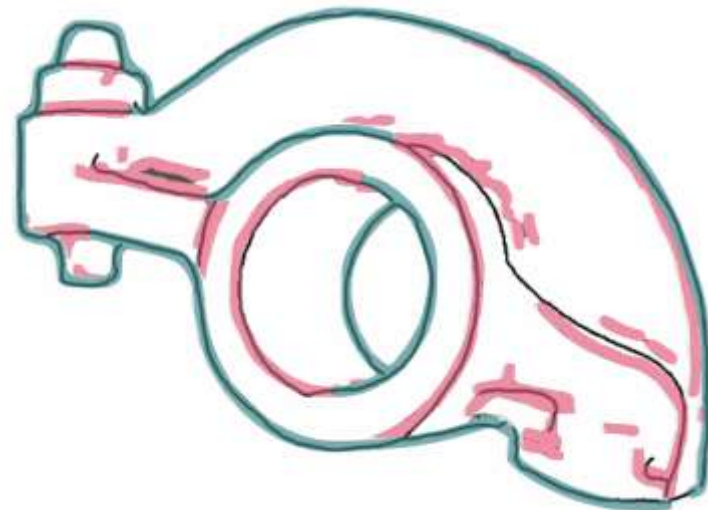
- images very similar to each other
  - 75% within 1mm for pair-wise comparisons
  - (a): five drawings overlaid in different colors
  - (b): pair-wise closest distance



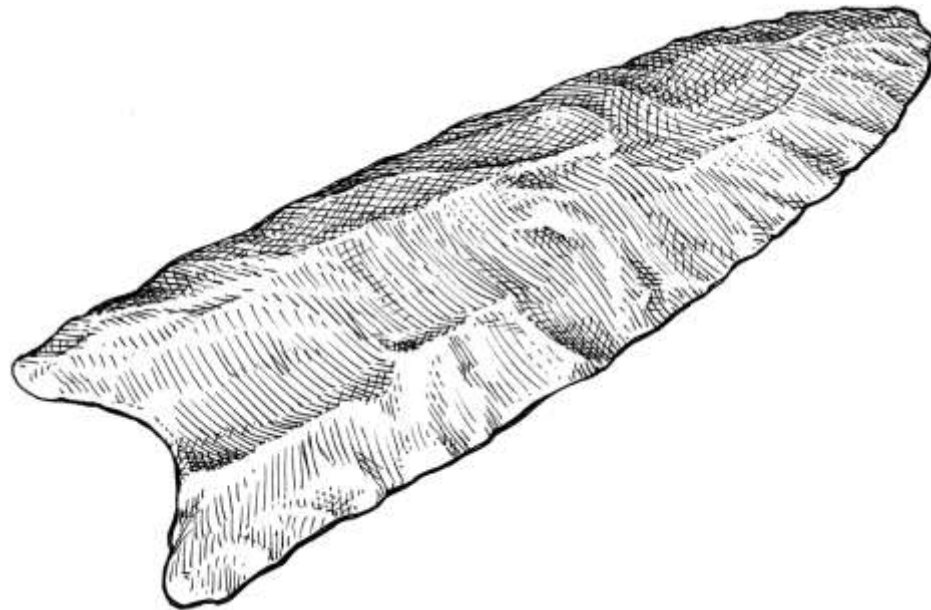
# Where do people draw lines? Results

- many hand-drawn lines are part of the CG line zoo; lines are near (1mm):

- silhouettes/  
occluding contours
- suggestive contours
- apparent ridges
- image edges



# Summary

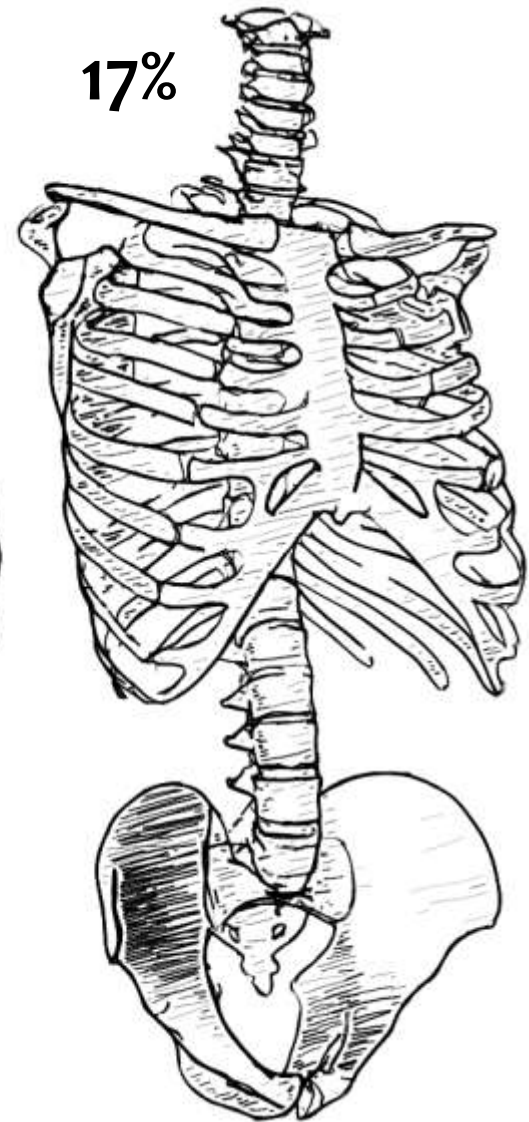
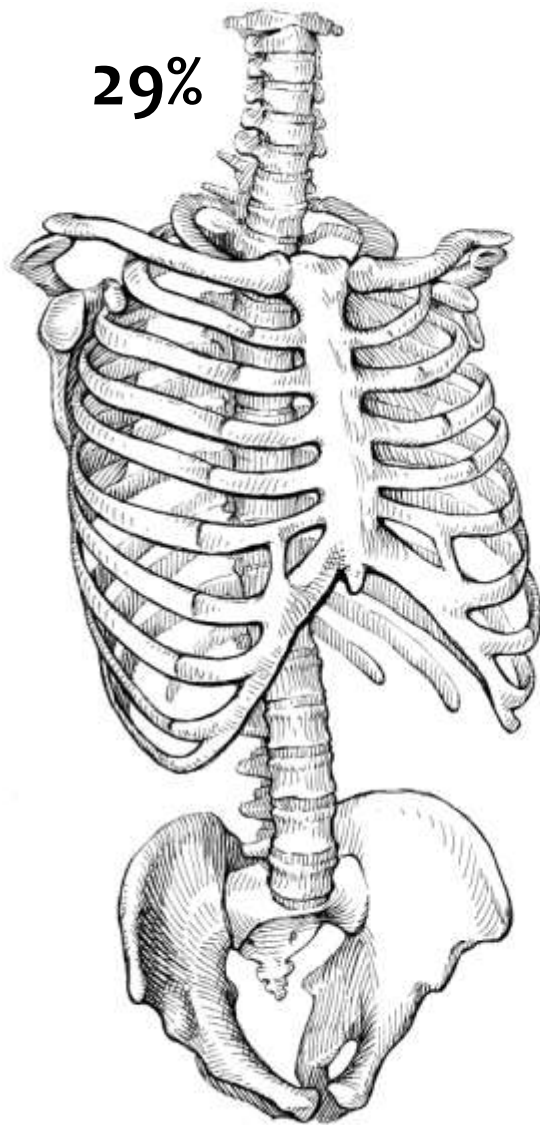


# Evaluating NPR – Summary

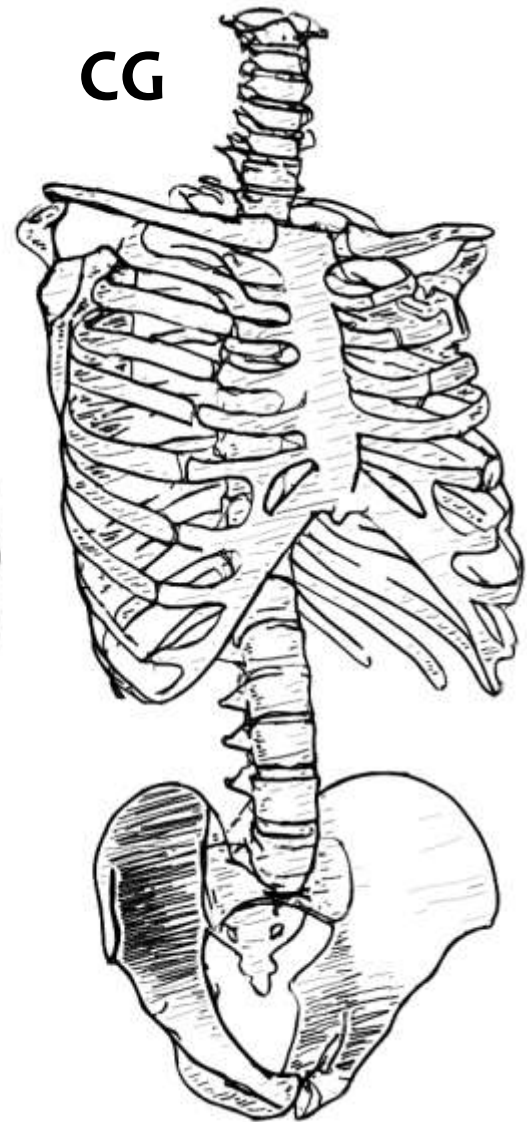
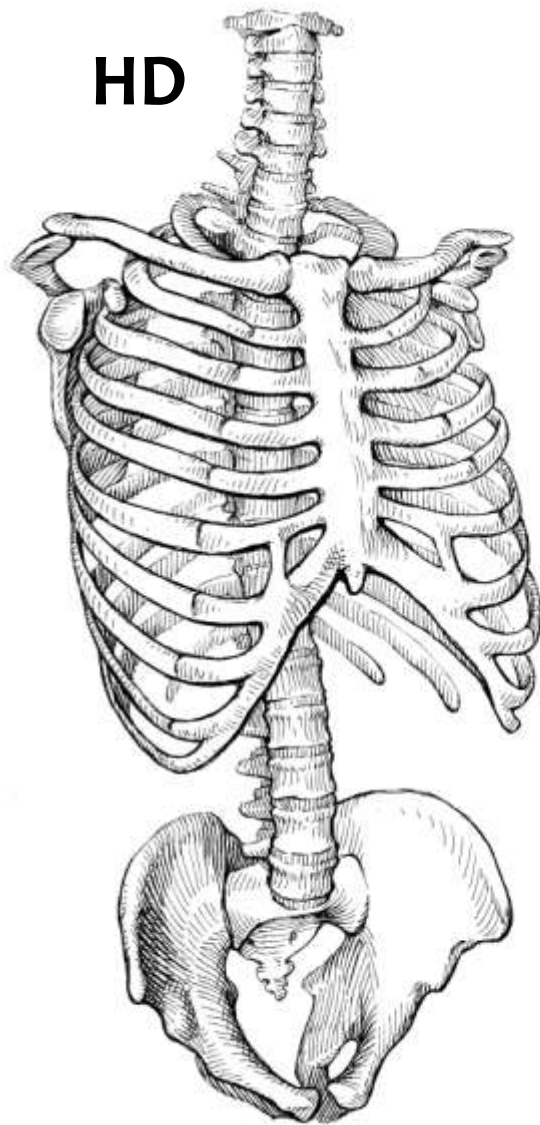
---

- new insight on where NPR techniques are applicable
- new insight on what people think about NPR images
- new insight on usefulness of NPR for specific domains and applications
- different techniques
  - qualitative and quantitative study techniques
  - experiments often w.r.t. given goal/purpose/domain
- (potential) ultimate goal:
  - algorithms to evaluate the produced images
  - algorithms to produce better images for a given purpose

# Results: Percentages Liked

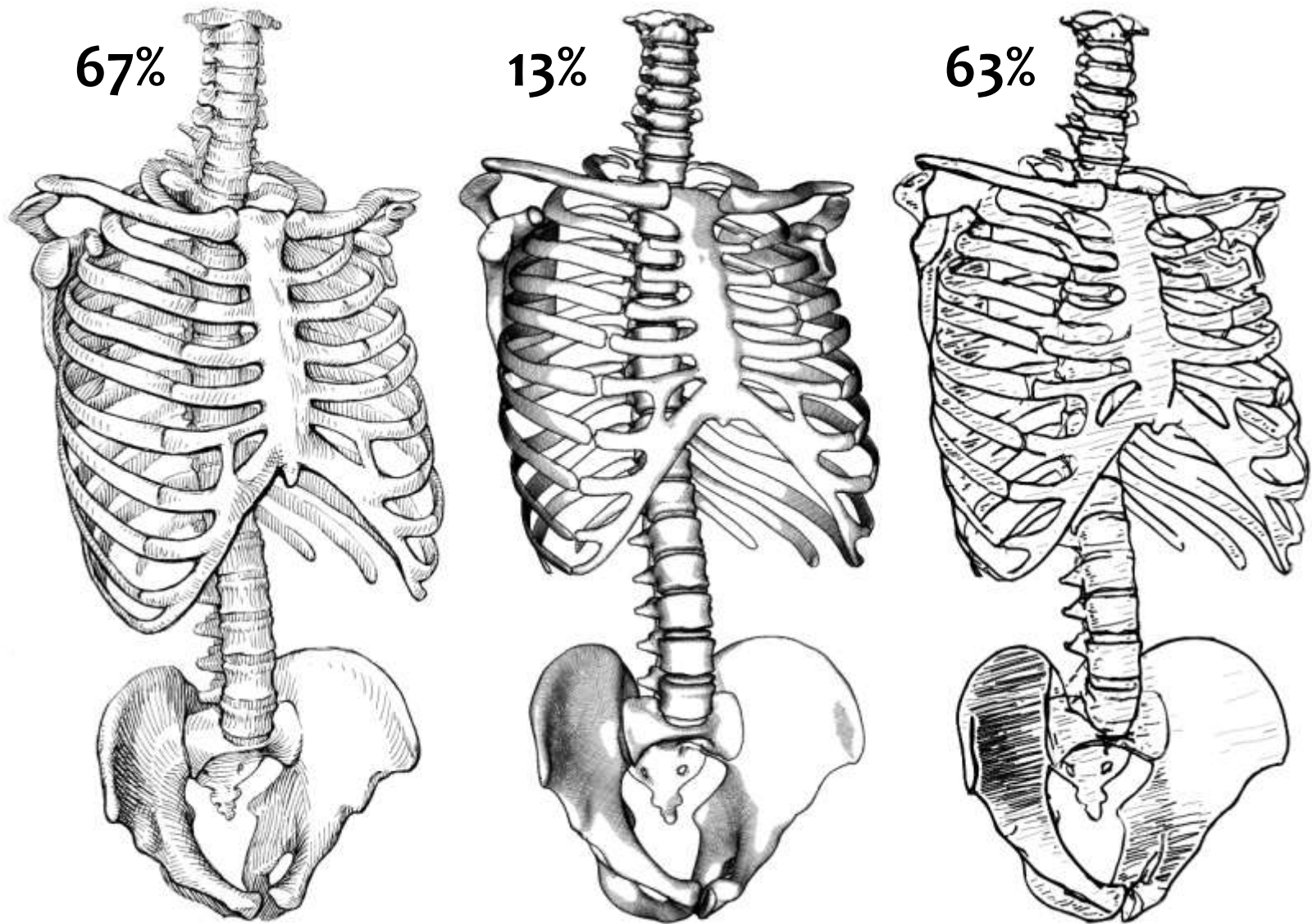


# Results: Hand-Drawn vs. CG

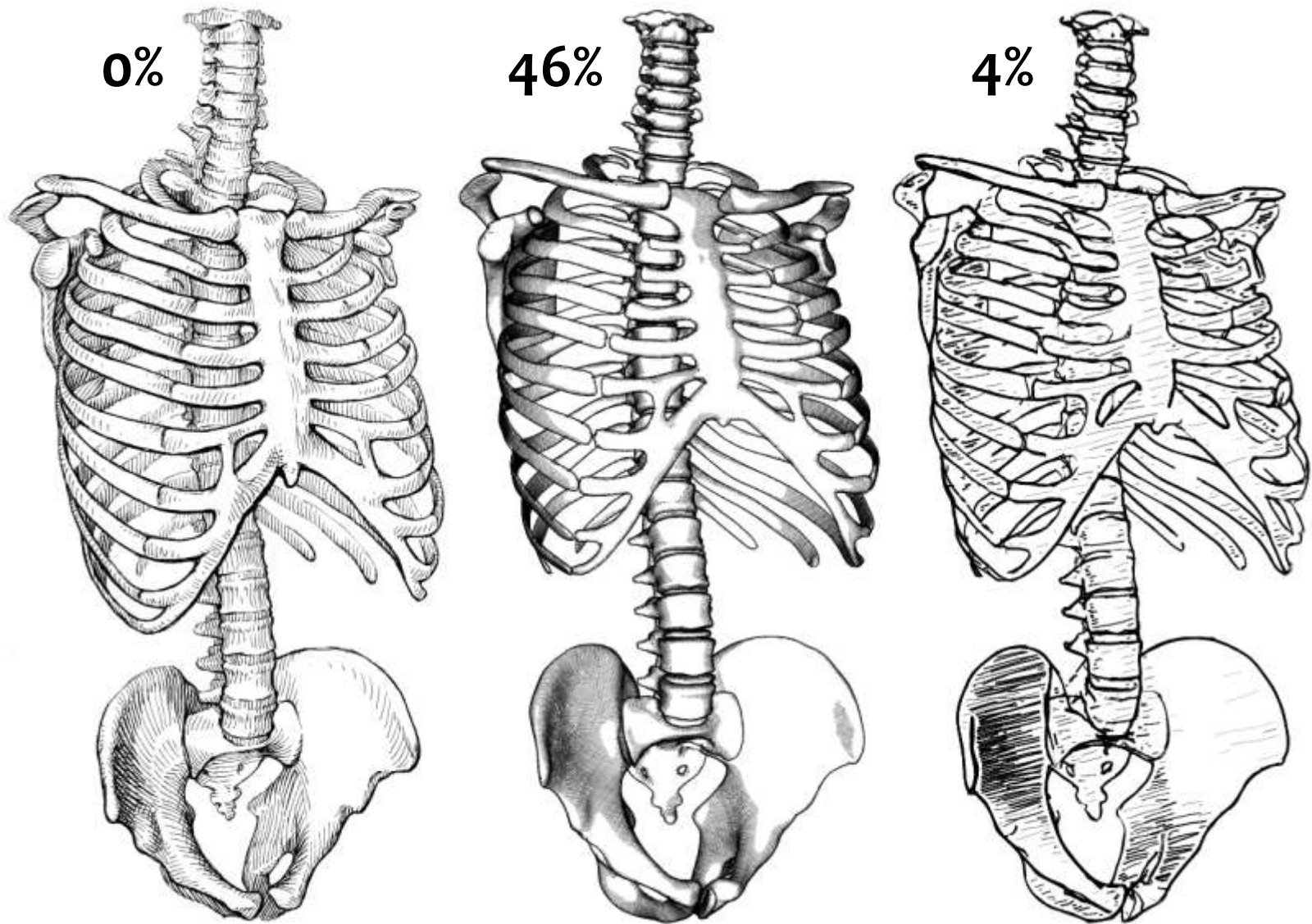




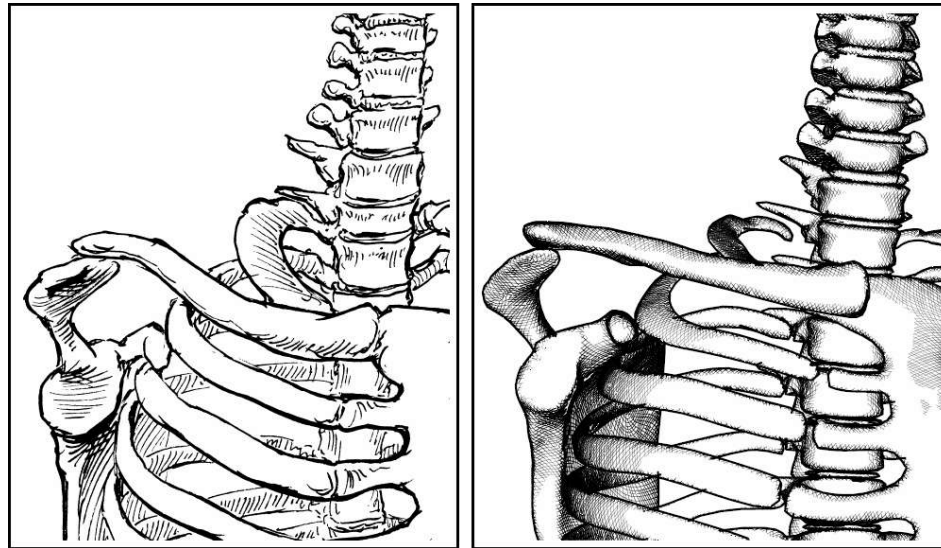
# Results: Stood out as Hand-Drawn



# Results: Stood out as CG



# Thanks for your attention!



- Thanks to all collaborators, in particular:  
William M. Andrews, Sheelagh Carpendale, Wei Chen,  
David Ebert, Petra Isenberg, Joaquim A. Jorge,  
SungYe Kim, Ross Maciejewski, Bernhard Preim,  
Mario Costa Sousa, Christian Tietjen
- Thanks to illustrators and study participants