

# Camera Phone Image Quality (CPIQ) ~~~ Subjective Image Quality Evaluation

*James H. Clark*

*Vista Point Technologies*

# Image Quality

Image quality is the perceptually weighted combination of all visually significant attributes of an image when considered in its *marketplace* or *application*

- Current CPIQ focus is:
  - Consumer mobile imaging products / applications
    - Quality of the captured still image
- Current CPIQ focus is not:
  - Quality of the handset display
  - Video quality
  - Professional / commercial photography
  - Scientific / technical photography
  - Data capture / bar code / machine vision applications
  - Image utility, transfer efficiency, storage density, etc.



AP / Shizuo Kambayashi

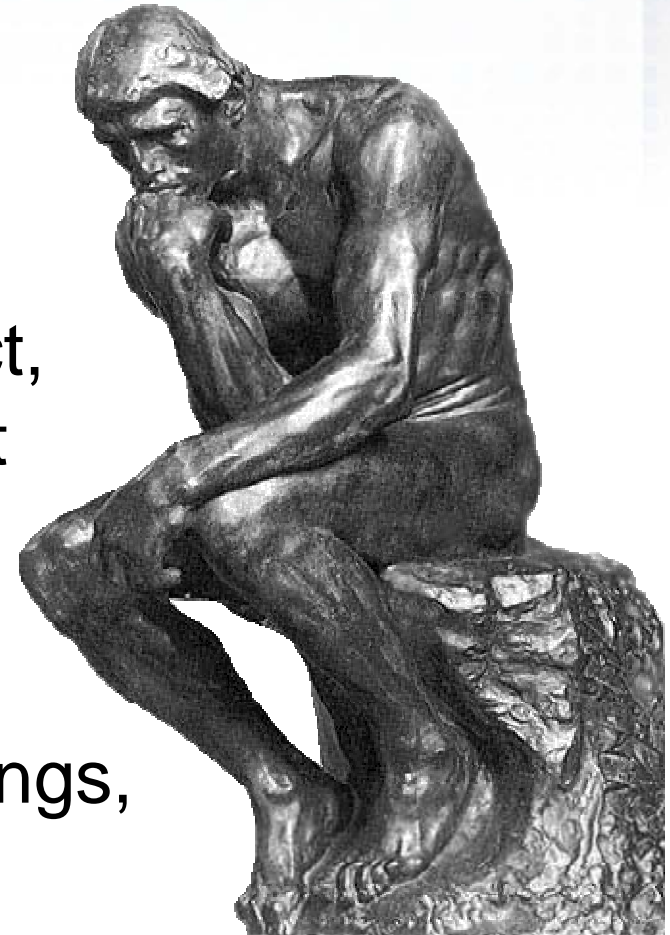
# Subjective vs. Objective measurement

## ***Subjective:***

- Existing in the mind
- Belonging to the thinking subject, rather than the object of thought

## ***Objective:***

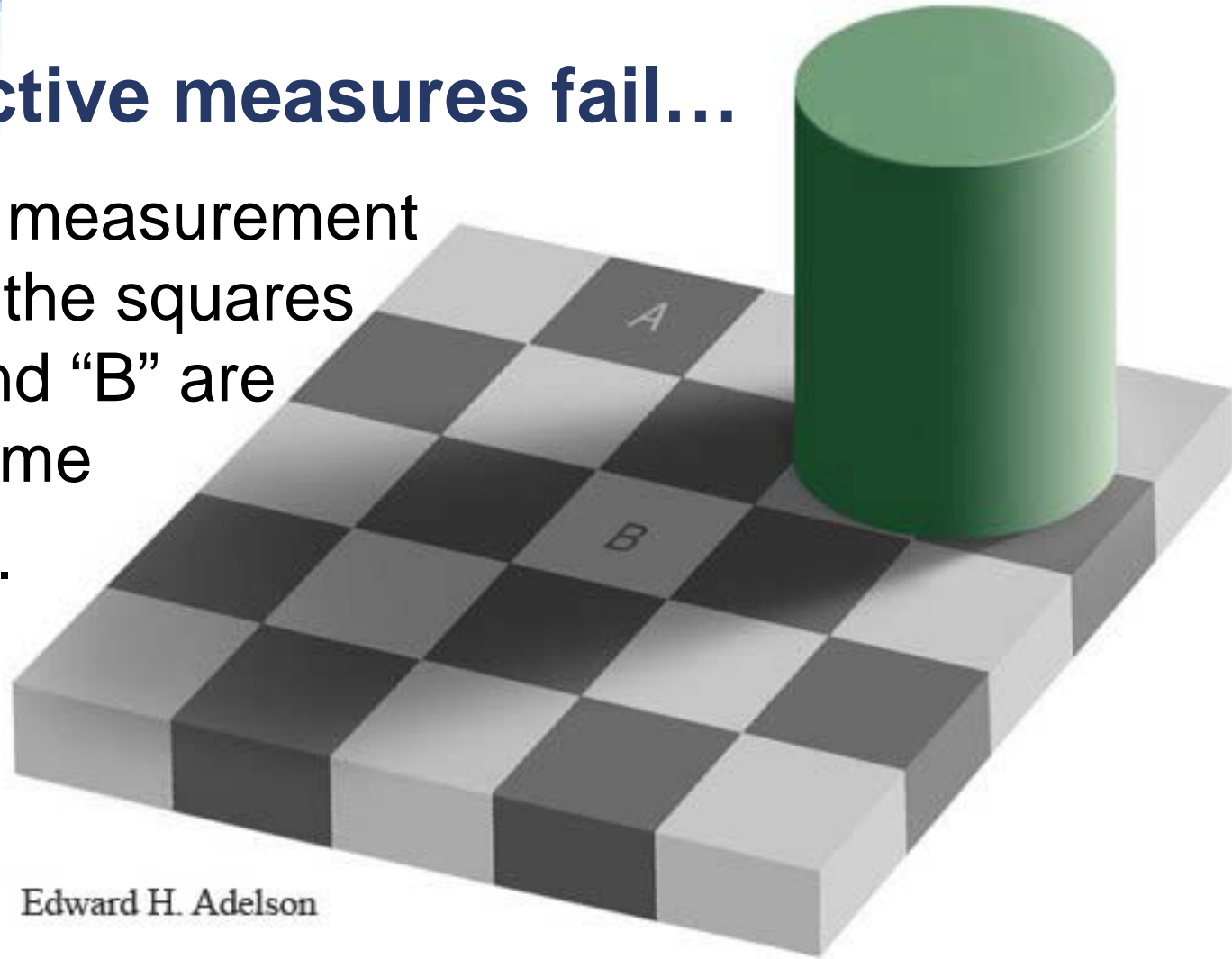
- Not influenced by personal feelings, interpretations or prejudice



*Le Penseur par Auguste Rodin*

## When objective measures fail...

The objective measurement indicates that the squares marked “A” and “B” are exactly the same shade of gray.



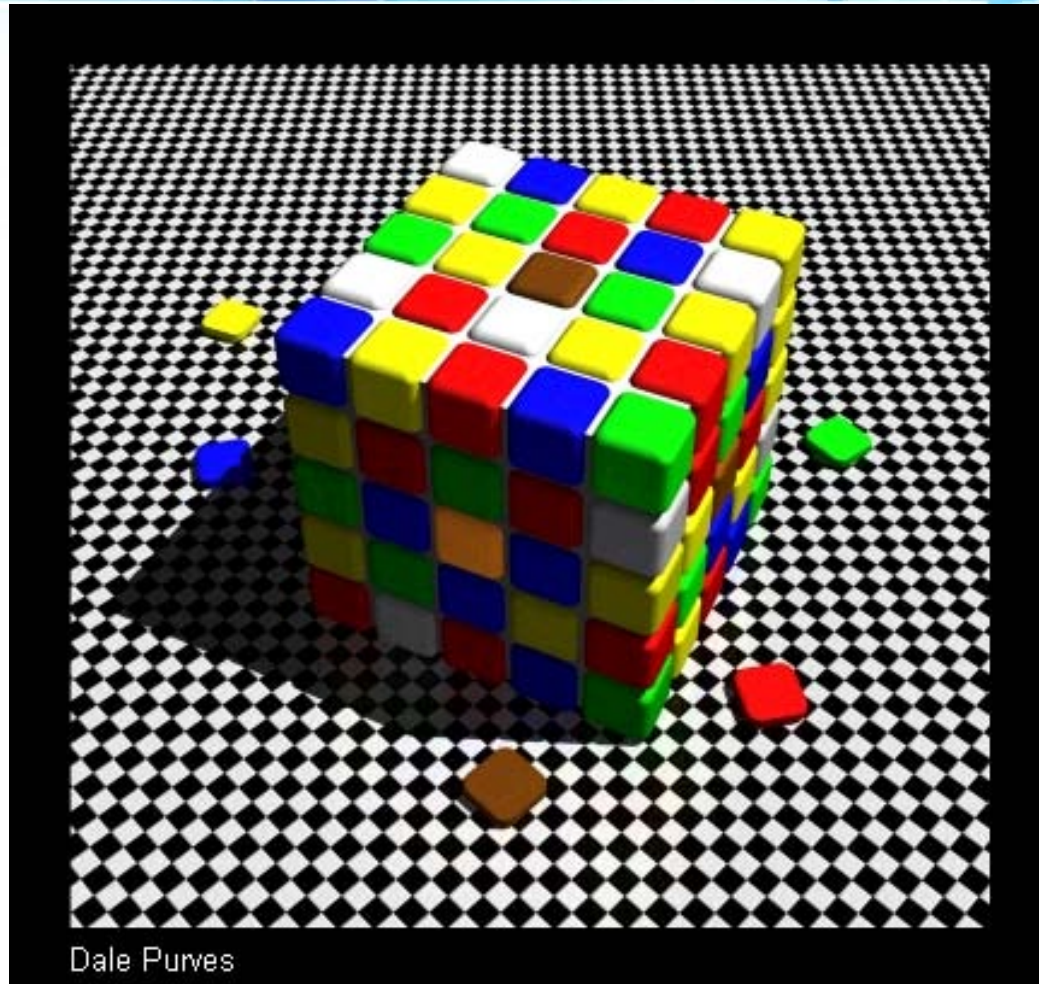
Edward H. Adelson

*Well, who you gonna believe, me or your own eyes? – Chico Marx*

## Another example...

The human visual system perceives color based not on the light that actually reaches the eye, but in large part on the memory of reflectances and illuminances that usually would have generated the stimulus in the past.

*Neurobiologists Beau Lotto and Dale Purves*



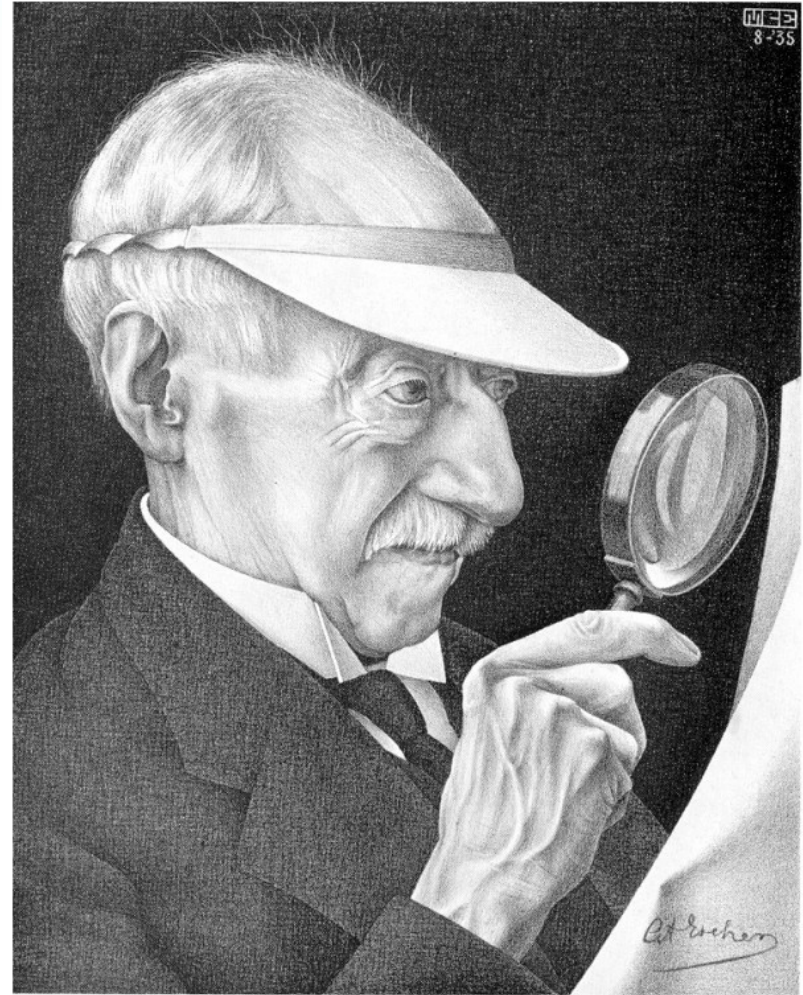
The center squares in the top and front cube faces are the same color.

# Psychophysics

- From Greek,  $\psi\upsilon\chi\omicron$   $\phi\upsilon\sigma\iota\kappa\omicron\varsigma$  (mental and physical)
- The study of the relationship between stimulation and response

# Psychometrics

- Using humans as meters



# Subjective Evaluation Methods Frequently used in Imaging

- Categorical Scaling
- Acceptability / Improvement Needed
- Anchored Scaling
- Forced-choice Paired Comparison
- Preference Scaling Paired Comparison

# Categorical Scaling

- Test images are binned according to category labels
- Most category scales are nominal or ordinal
- If labels are chosen carefully, an interval scale can result
- For example, a categorical scale found to behave like an interval scale, and used in the photo industry for years:
  - Excellent
  - Very Good
  - Good
  - Fair
  - Poor
  - Not Worth Keeping

# Acceptability / Improvement Needed

- A type of categorical scale
- Minimum 2 categories
  - Acceptable
  - Unacceptable
- More categories allow observers to express degrees of acceptability / improvement needed
- Requires scenario to establish context for making judgment

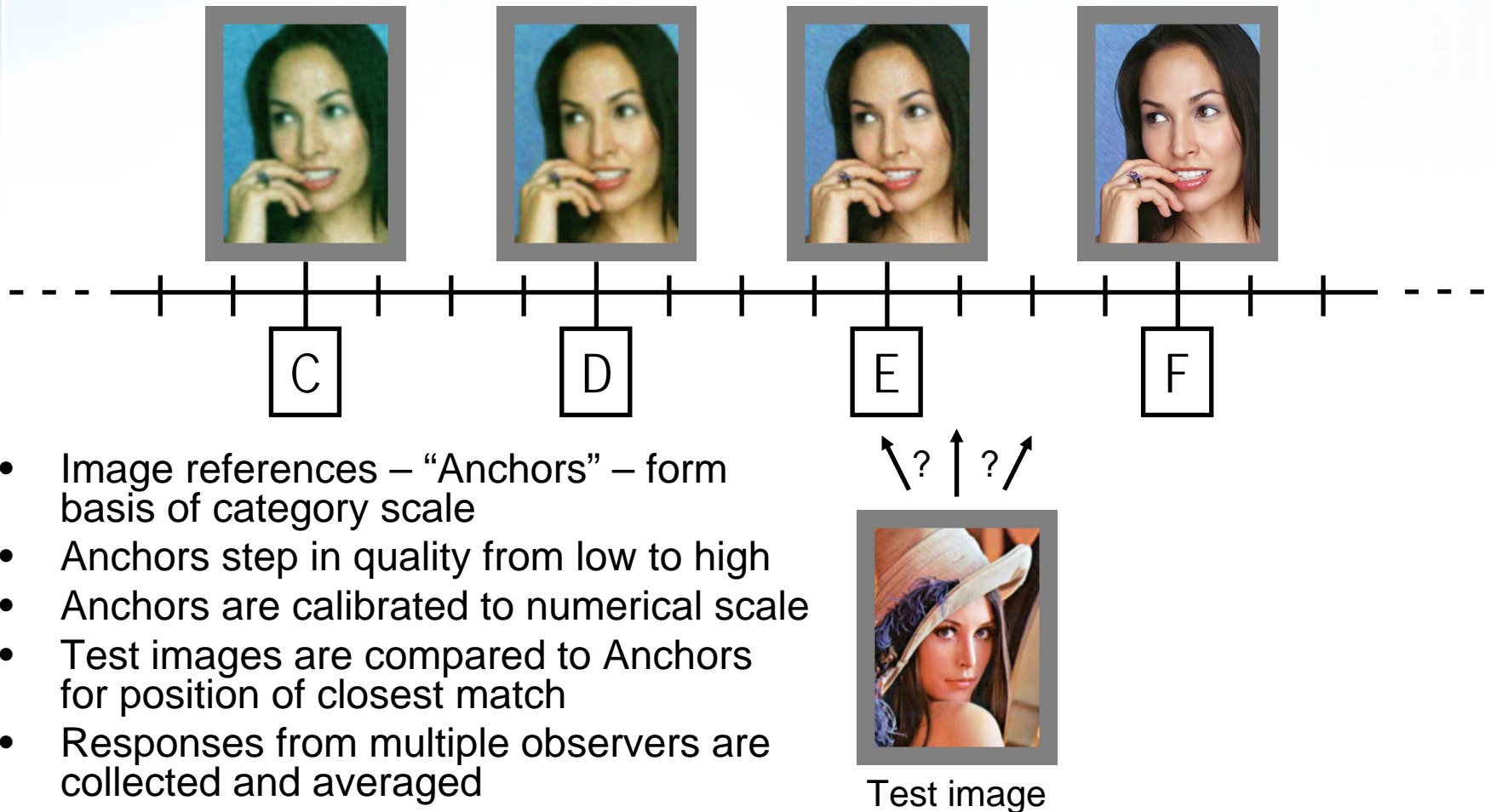
## Acceptability Rating

- - Completely Acceptable, No Improvement Needed
- - Acceptable, Slight Improvement Needed
- **Threshold Of Acceptability**
- - Unacceptable, Significant Improvement Needed
- - Completely Unacceptable, Of Little Or No Value

# Anchored Scaling

- A type of categorical scale with images used as category markers
- When the quality of the anchor images has been calibrated, the scale is considered to be quantitative
- An example is ISO 20462-3:2005, Photography - Psychophysical experimental methods for estimating image quality - Part 3: Quality ruler method

# Anchored Scaling Concept



# Anchored Scaling

## ADVANTAGES

- Provides a numerical score that can be transformed into other scales
- Test results may be compared across time, location – if same references are used each time
- Can efficiently and effectively evaluate single or numerous products
- Less subjective than other methods
- Scale covers a broad IQ range

## DISADVANTAGES

- Low sensitivity to very small differences in test samples
- Test samples not compared directly to each other – differences must be inferred
- ‘End of scale’ effects
- Best if test samples have content similar to anchor images
- Development of anchor sets (calibrated graphical references) is time-consuming and tedious

# Forced-choice Paired Comparison

- Most sensitive technique regularly used for visual psychometric studies
- Each test sample is paired with every other test sample and the observer must select one
- Called forced-choice because “no difference” or “no preference” responses are not allowed
- Analysis breaks down if sample differences are large enough to give unanimous judgments
- Many comparisons and many observers required for validity.

# Preference Scaling Paired Comparison

- A paired comparison technique in which the observer can express degrees of preference, including “no preference”
- Expands the useful quality range, while simultaneously decreasing the number of observers needed, compared to forced-choice paired comparison



# Subjective Method Pros & Cons

Method	Advantages	Disadvantages
Categorical Scaling	<ul style="list-style-type: none"> <li>• Simple to conduct and analyze</li> <li>• Meaningful to non-technical audience</li> </ul>	<ul style="list-style-type: none"> <li>• Requires large numbers of observers for reliability</li> <li>• May not transfer to other languages</li> </ul>
Acceptability	<ul style="list-style-type: none"> <li>• Simple to conduct and analyze</li> <li>• Meaningful to non-technical audience</li> </ul>	<ul style="list-style-type: none"> <li>• Requires large numbers of observers for reliability</li> <li>• Acceptability' drifts over time</li> </ul>
Anchored Scaling	<ul style="list-style-type: none"> <li>• Can provide absolute IQ scores</li> <li>• Good for large numbers of test samples</li> </ul>	<ul style="list-style-type: none"> <li>• Test samples not compared directly</li> <li>• Anchors difficult to produce</li> </ul>
Forced-choice Paired Comparison	<ul style="list-style-type: none"> <li>• Good for detecting very small differences</li> </ul>	<ul style="list-style-type: none"> <li>• Limited to relatively small numbers of samples</li> <li>• Analysis is complex</li> </ul>
Preference Scaling Paired Comparison	<ul style="list-style-type: none"> <li>• Simple to conduct and analyze</li> <li>• Good for large numbers of samples</li> <li>• No need for calibrated anchors</li> <li>• Sensitive -- test samples are compared directly</li> </ul>	<ul style="list-style-type: none"> <li>• Scores are relative, not absolute</li> </ul>

# Resource Requirements for Subjective Evaluation

- Observer Pool
  - Color and visual acuity tested
  - Naive to test objectives
  - Not working in any image quality field
  - Committed to spending 1 hour per week
- Controlled Viewing Environment
  - ISO standard lighting
    - D50 (5000K)
    - 500 - 2000 lux
  - Calibrated monitors
  - Neutral gray surround



Photos courtesy of Hewlett-Packard

