

This presentation, originally developed for the 2014 convention of the 4Cs, (aka Columbia Council of Camera Clubs), evolved from earlier presentations that I made at the 2009 and 2011 conventions. In 2022 the 4Cs changed its scoring scale, and those changes are now incorporated.

This presentation has two purposes. First, it provides examples and guidelines for scoring images in 4Cs competitions. Second, it provides a conceptual framework for evaluating images.

All images are copyrighted by their respective makers. Please respect those copyrights.

This presentation is adapted from the convention version to include many of the talking points, so expect a lot of text.

Gordon Battaile  
September 2022

# The Question

It's natural for photographers who want to learn the craft to seek feedback on how they're doing.

Judging is the best tool we have for assessing and comparing large numbers of images, but it has to be done well to be useful.

So... How do you do it well?

# Judging Images In 4Cs Competitions

- I. Evaluating Images
- II. The 4Cs Scale
- III. The Principles of Judging
- IV. An Alternative Approach to Scoring

# Part I: Evaluating Images

Caveats apply:

1. There does not appear to be an authoritative source on the subject so this material is based on my own analysis.
2. This subject resists definitive treatment so the presentation continues to be a work in progress.



# The Five Attributes

Let's start with a question:

What are the levers of control for an image that can be changed *independently* to affect image quality?

These five are more or less in the order that you typically deal with them in making a photograph:



Attribute #1:  
Content

This means everything in the image: subject(s) and background. It does not mean how those elements are arranged (that's composition) or how they are lit (that's lighting).



While content changes often modify other attributes, it is possible to change *only* the content of an image (as this example does) and still transform its meaning or effectiveness.





Attribute #2:  
Lighting



For images that use ambient light, this attribute equates mostly to timing. For images with controlled lighting, it is more a function of technique. This example illustrates both aspects.



### Attribute #3: Composition

Once you've decided what elements you are photographing, composition is the process of arranging those elements in (or excluding some elements from) the frame.

Note that this is a bit different from a typical use of “composition” in a painting or graphic arts context, where it might include color or other aspects of the subject material.







Attribute #4:  
Camera Operation



This is one aspect of technique.

The example here only shows a difference in exposure, but focus, camera movement, and depth of field are also factors.





Attribute #5:  
Post-processing

This is another aspect of technique.



Post-processing does not have to change other attributes. For example, modifying the contrast would not affect anything else.

However, some post-processing does change other attributes: for example, cloning modifies the content.

# Image DNA

The five attributes could be collapsed into a set of three:

- Content (subject, background, and lighting)
- Composition
- Technique (camera operation and post-processing)

Either way, these are the materials out of which each photograph is built: an image's DNA.

Every other image attribute depends on the handling of these attributes.

# But Wait...

If those are the only attributes that control an image, where do these other ideas fit in??

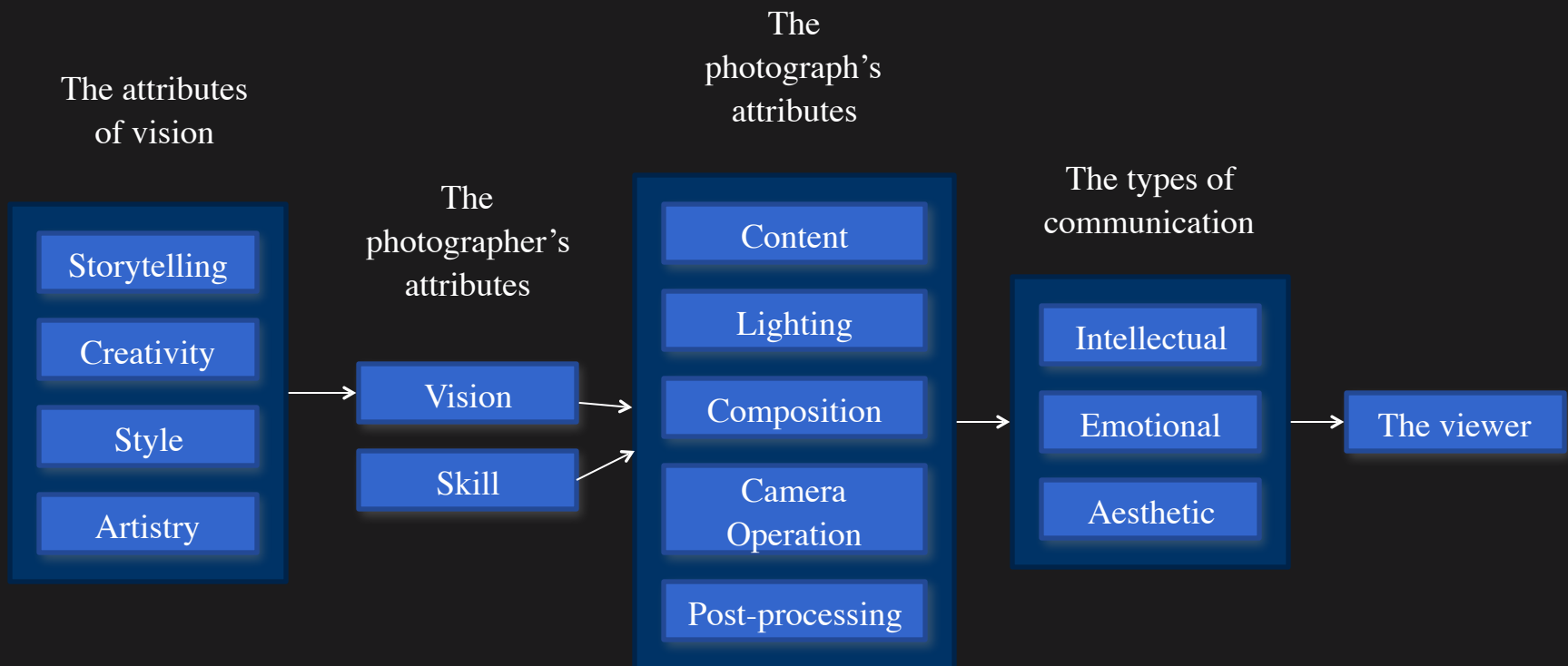
Artistry  
Creativity  
Skills  
Style  
Storytelling  
Vision

Unfortunately– but not surprisingly– the picture gets more complicated when you try to account for these...



# The Big Picture

Here is one of many possible ways to describe the attributes and their relationships:



# The Problem With The Big Picture

It's too complicated to be a useful framework for analyzing images in 3 seconds!

Furthermore, it implies (inaccurately) that every attribute always plays a role in a successful image. Not true: there are many routes from vision to viewer.

So is there a simpler way to look at things? Fortunately, Yes!

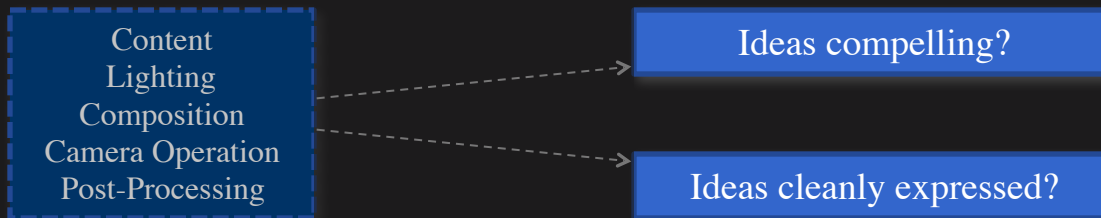
# A Simpler Framework

It is simpler but still accurate to approach image analysis by asking just two questions:

1. How compelling are the ideas in the image?
2. How cleanly expressed are the ideas in the image?

# A Simpler Framework (cont.)

Both answers depend on the handling of the five attributes:



However, looking at it this way helps clarify some key points:

- What matters in an image is what and how well it communicates.
- For any given image you only need to think about the attributes that are important to the ideas in that image.

Two examples follow.



## What are the primary ideas?

- Large plantation
- Fall color
- Backlit canopy
- Trunks picking up blue sky (color contrast)
- Small sapling (size contrast)

## Net message:

Long row of tall, skinny trees with nice color and light, and small sapling.

## How compelling?

So-so: the compositional emphasis on the height of the trees is not interesting, and neither is the sapling. The size of the plantation and contrast in lighting would be more interesting but are not well emphasized.

## How cleanly expressed?

- Sapling not well defined (not that it would help much)
- Detritus in foreground
- In other respects, reasonably clean





## What are the primary ideas?

- Large plantation
- Fall color
- Backlit canopy
- Trunks picking up blue sky (color contrast)
- Person as counterpoint for scale and contrast
- Posture suggests relaxation and enjoyment
- Looking down the row emphasizes length of the row

## Net message:

Long row of trees with nice color and light, with someone enjoying the scene





How compelling?

Much better than the vertical. The emphasis is now on the most interesting/appealing aspects: the length of the row and the color contrast. The figure adds visual and conceptual contrast and helps emphasize the appeal of the scene.

How cleanly expressed?

Also much better: all elements of the image relate directly to the main ideas being conveyed.

## Part II: The 4Cs Scale

As of September 2022, the 4Cs uses the scale 5-9 with half points in that range.



## Important:

The 4Cs does not require member clubs to use this scale for *internal* competitions, just 4Cs-sponsored inter-club competitions.

However, any and all clubs are welcome to use it for internal competitions as well. The 2022 scale update makes that more practical than it was with the previous scale.

# Scale Interpretation

1 - 4	5	5.5	6	6.5	7	7.5	8	8.5	9
Not used in 4Cs scoring	Noticeable problems offset strengths		Okay but with minor problems		Solid; Average 4Cs quality		Strong; rewards viewing		Outstanding; an exciting image

## Anticipated typical score distribution in 4Cs competitions

0%	1-4%	15-20%	50-65%	15-20%	1-4%
----	------	--------	--------	--------	------

Important point: The scale is intended to cover only the range of image quality typically seen in 4Cs competitions. It is not intended to cover quality levels above or below the typical 4Cs range.

With this scale, 9 is intended to be achievable in a typical competition: multiple award winners will likely have a score above 24, and there may be an occasional 27.

# The 4Cs Reference Image Project

In 2022 the 4Cs voted to move to a new scale. To illustrate the scale we repeated the process that the council used in 2014:

- Many different clubs and individuals within the 4Cs contributed candidate images, and I assembled a set of 180: 30 in each of six categories.
- 35 judges from around the council scored the images.
- In general, scores pulled significantly towards the middle of the scale. The average scores of the reference images reflect that.
- The reference set is composed of the images that were the best fit for the scores.
- Judges should familiarize themselves with the new scale before judging.

# The Judges

The 35 judges came from 12 different clubs:

- BCC: Robert R., Jim S., & Dave C.
- EPS: Anne C.
- FGCC: Carol B., Clem, P., Gene D., Mary Lane A., Bruce H., & Gordon B.
- FPCC: Doug F., John C., Jon F., Eloise C., Sharp T., Jan E., & Bob W.
- GPC: Irene F.
- KCCC: Kent M., & Bjorn F.
- MCC: Lisa D. and Bill C.
- WILD: Larry P., Susan S., & Rose C.
- YAAP: David M., Mark S., Ted C., & Kim T.
- SOPA: Howard H., Barb T., & Neal T.
- WVCC: Eileen H.
- VVCC: Art B. & Allen B.

Thank you judges!

# Reference Image Project – Overall Results

The average score for all images and all judges was 7.18. (The target was 7.0)

The distribution of scoring ranges (i.e., high score minus low score) across the 180 images:

0.5	1	1.5	2	2.5	3	3.5	4
0.5%	1.5%	3.6%	17%	15%	37%	9%	17%

Much like the 2014 judging process, the distribution is the most striking result:

- One sixth of the images had a 4 point scoring range (i.e., 5-9)
- Almost 2/3 of the images had a scoring range of at least 3 points

However, given the large number of judges, this result is not unexpected: every image has a bell curve of responses, and the more judges you add, the more complete the bell curve becomes.

# What about 9s?

The goal for the new scale is to make scores of 9 achievable. Three images in the candidate set had at least half the judges score them as 9, but if the scale used all of those as the standard, it would set the bar too high.

As a result, the examples of 9 with one exception have median scores of 8.5 and average scores in the 8.4-8.5 range. They nonetheless are of sufficient quality to represent scores of 9 on the new scale.

Judges should not take these as the *minimum* quality for scoring 9: they represent the mid-range for 9.

# The Reference Images

There are 30 reference images: six sets of five. There is a set for each of the following six general categories:

- Landscape
- Birds
- Wildlife
- Flowers
- Monochrome
- Altered Reality

The first four categories on the list are the main components of competition in the Traditional category.

Landscape  
Score: 5





Landscape  
Score: 6





Landscape  
Score: 7





Landscape  
Score: 8





Landscape  
Score: 9



Kaboom, by Mary Lane Anderson



Birds  
Score: 5



Birds  
Score: 6





Birds  
Score: 7





Birds  
Score: 8





Birds  
Score: 9



Room Mates, by Gene Davis

Wildlife  
Score: 5





Wildlife  
Score: 6





Wildlife  
Score: 7



Wildlife  
Score: 8





Wildlife  
Score: 9



Night Visitor, by Doug Fischer

Flowers  
Score: 5



Flowers  
Score: 6





Flowers  
Score: 7



Flowers  
Score: 8





Flowers  
Score: 9



Delicate Elegance, by Mary Lane Anderson

Monochrome  
Score: 5





Monochrome  
Score: 6





Monochrome  
Score: 7



Monochrome  
Score: 8





Monochrome  
Score: 9



Symbol of Death, by Katie Rupp

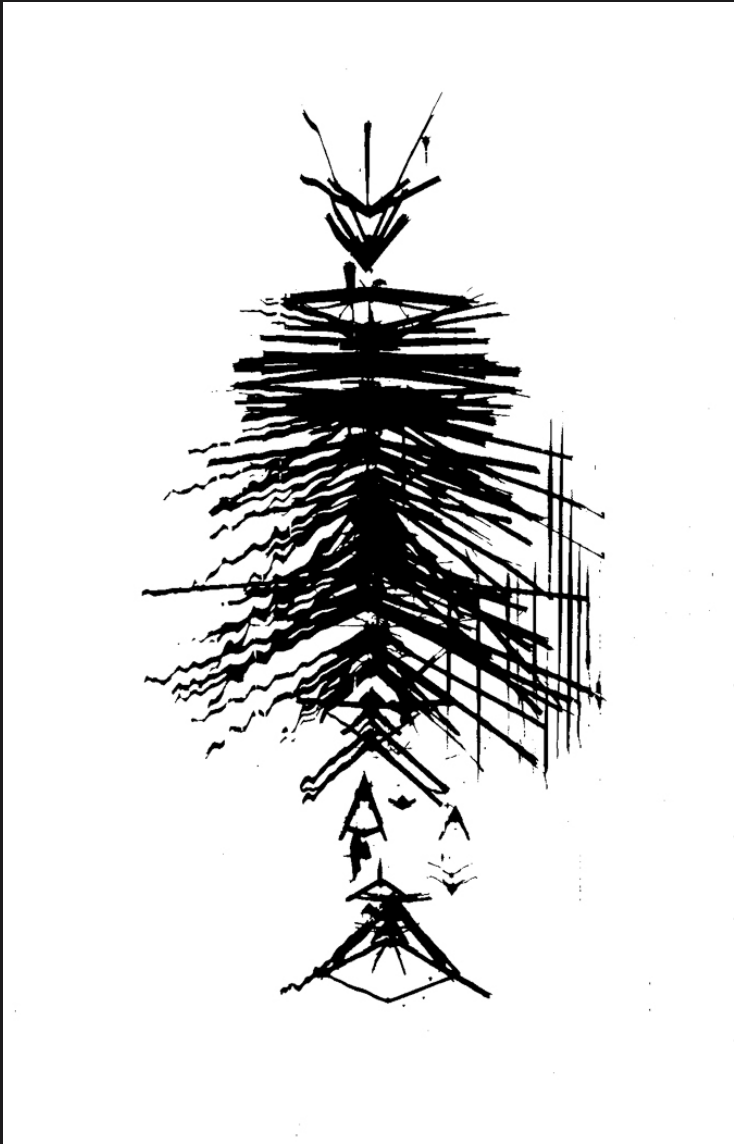


Altered Reality  
Score: 5



Altered Reality

Score: 6





Altered Reality  
Score: 7





Altered Reality

Score: 8



Altered Reality

Score: 9



Rainbow Columbine, by Carol Berget

# Highest Scoring Image



Average: 8.9

31 of 35 judges scored it  
as a 9 in the wildlife set.

Barred Leaf Frog  
Doug Fischer (FPCC)



## Part III: The Principles of Judging

I count 5 of them. All are variations of *Be Fair*.

# Principle #1

## Set aside your personal biases

Approach judging the work of others as you would want others to approach judging your work: conscientiously and-- to the extent possible-- without personal bias.



Gopher Snake

Pat Starr (PPS)

In the mid 1990's I participated in a judging of PSA images that included one of a rattlesnake. We were using a 5 point scale and the image got a 1-3-5. We all turned to look at the judge who gave it a 1, and she said "I don't like snakes!"

Think of it this way: judge based on what you know about photography, not based on your opinions about the merits of the subject or approach. Whether you like or don't like snakes is irrelevant.



## Principle #2

# Judge each image as a whole

Take an image's best and worst attributes into account, but do not use one or the other as the sole basis for judging.

For example, an image that has great composition but bad lighting is not as good as an image that handles both well, and is better than an image that handles neither well.

When there are widely varying scores on an image, it is a good bet that one or more of the judges is not observing this principle.



Newport Bridge

David Roy (PPS)

This image is one of a very small number to have been scored with a 5 point range: one judge gave it a 4, another gave it a 9. It seems likely that one judge scored it high based on the creative processing, and the other judge scored it low based on busy composition or processing artifacts. Both judges seem to have ignored the other balancing attributes of the image; considered as a whole it should probably be a 7 with the current 4Cs scale.

# Principle #3

## Analyze instead of just responding

The two most important sentences in image evaluation:

This image is effective *because...*

This image is not effective *because...*



Consider the image by Photojournalist Kevin Carter that won the Pulitzer Prize in 1994. (Please do an image search to see the image.) It shows a Somali child apparently collapsed on the ground from hunger with a vulture watching in the background.

By camera club standards, the lighting is harsh and the ground is littered in an unappealing way. But these actually reinforce the story of desperation told by the image: if the image featured the same subjects on a beautiful dune at sunset, the message would be confused and much less powerful.

## Principle #4 Be consistent

Do not change the way you are evaluating images in the middle of a judging session-- even if your scores are noticeably higher or lower than the scores of other judges. Doing so would unfairly distort the results.

# Principle #5

## Use the scale as described

If a scale (like the 4Cs scale) says that average images should be scored as 7, expect to give a lot of scores of 7.





Albino

Lori Montgomery  
(BMPC)

Some years ago this image scored 2-2-3 in a print competition, making it the lowest scoring image in memory. While the processing to emphasize noise may not be to every judge's taste, the 2-2-3 scores indicate a total failure to understand the scale.

# Applying the Principles

1. Do not give scores outside the range 5-9.
2. Scoring splits of 2 points are significant and merit discussion and possible rescoring.
3. Scoring splits of 2.5 points or more are unacceptable: images should always be discussed and rescored until the range is 2 points or less.

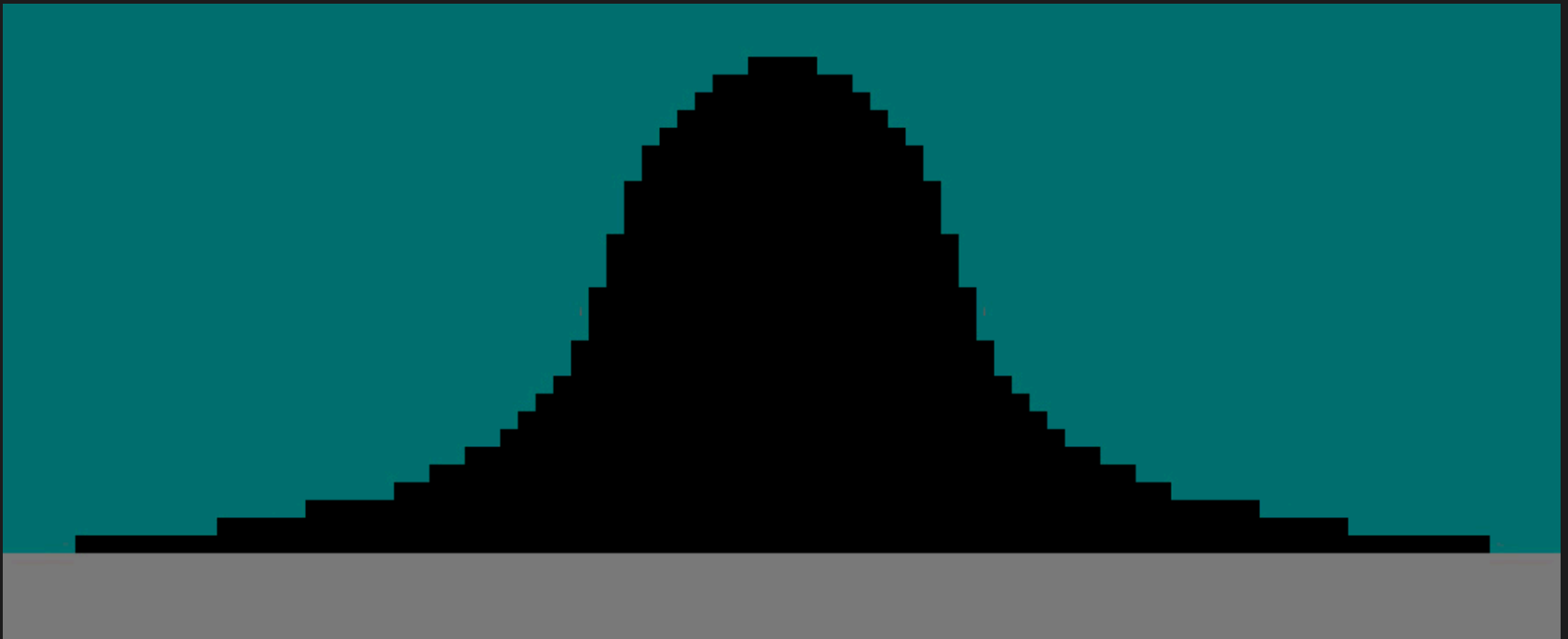
## Part IV: An Alternative Approach to Scoring

There is another way to look at the judging process based on the idea of the bell curve. It clarifies why it's useful to observe the principles of judging.



# Intro to the Bell Curve

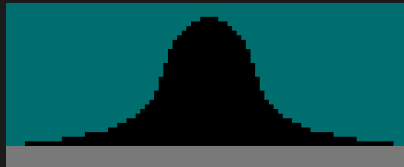
A bell curve is a typical distribution of data points around a center value, and gets its name from its bell-like shape. If for example you ask a large number of people to guess how many jelly beans are in a jar, the bell curve is the shape you get from graphing the guesses.



# Characteristics of a Bell Curve

Two key points:

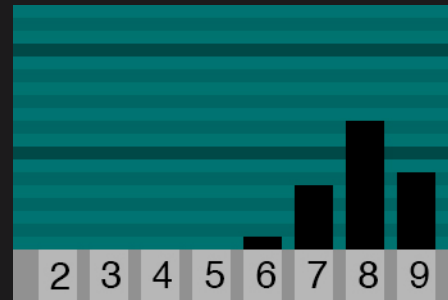
- The shape of the curve depends on the number of possible values and the number of graphed values.
- Once the number of data points (like jelly bean guesses) is large enough, the location of the center of the curve settles into a stable spot: after 1000 people have guessed, adding another 1000 won't significantly change the center of the curve unless you change how you are picking people to make guesses.



# Applying the Bell Curve to Judging

Judging produces bell curves-- if you have enough judges. They would be rough curves for 4Cs images because there are only 9 possible values (5-9 in half-point increments), but still bell curves.

The center of the curve would be a lower score for weaker images and a higher score for stronger images:

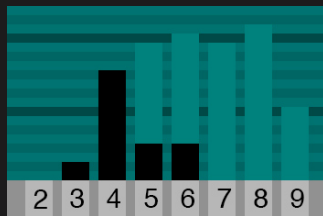


For very strong or very weak images you may end up with half a bell, or even just a spike at one end.

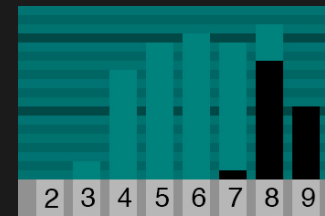
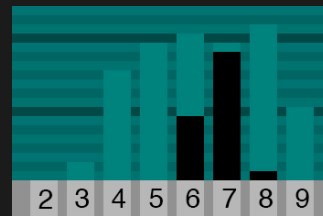
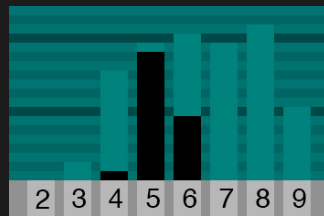


# Bell Curves in the 2014 Reference Judging

More judges will make a curve more bell-like in shape, but even with only 22 judges (a relatively small number) the results from the 2014 reference judging showed the bell shape. Here are four examples from it including the highest and lowest scoring images:



Lowest scoring image



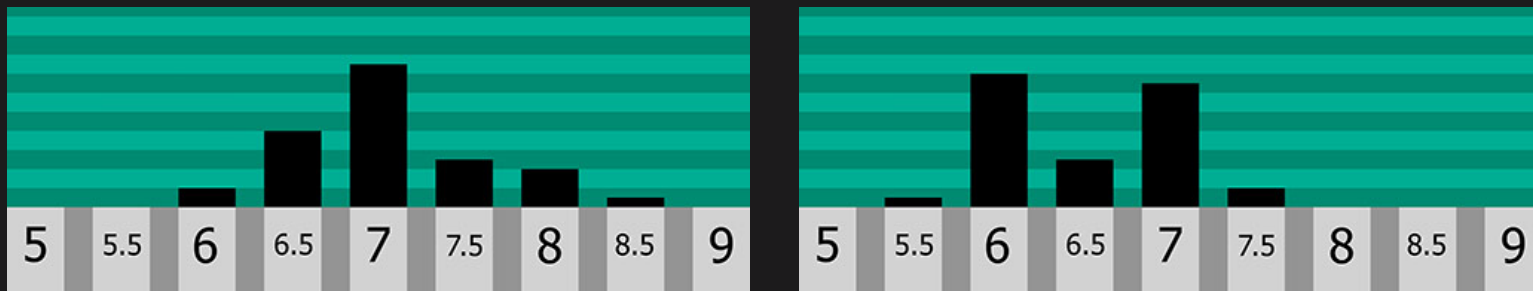
Highest scoring image

Though rough, these show the fundamental characteristics: center weighted with a taper off to the sides.

# Bell Curves in the 2022 Reference Judging

There were 35 judges in the 2022 process, but there was a wrinkle: some judges didn't use the half points. As a result, images that had an average score on a half point often had a bell curve with a dip in the middle.

For example, here are the results for two images with very good consensus between judges, one with an average of 7 and the other with an average of 6.5.

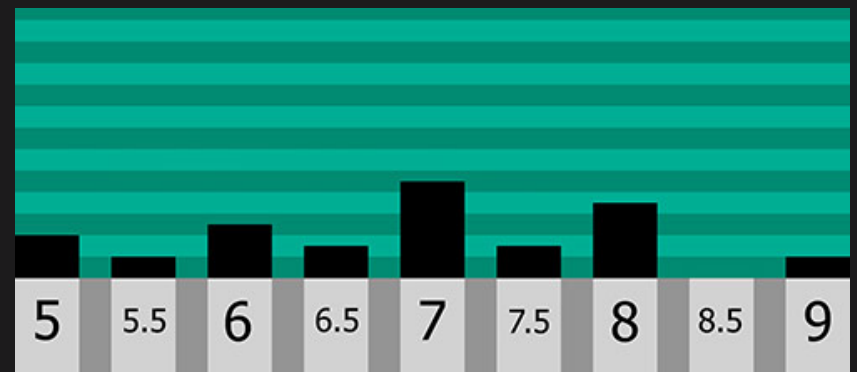


Even with this complication, the results show the same basic characteristic: scores clustered in the middle three scores and tapering off at the ends.



# Example From Reference Image Judging

This image, which was in the Altered Reality category, had the least consensus among the 180 images in the reference judging. Even so, it still shows the bell shape, albeit a flattened bell.



Average: 6.86



# Applying the Bell Curve to Judging: The Takeaway

Judging is subjective: for any given image some judges will score on the high end and others will score on the low end. But if enough people score an image, all the subjective opinions effectively average out to produce a “typical” score: the middle of the bell curve.

As a result, once you have enough judges the center of an image’s curve becomes stable: if you double the number of judges, it won’t change the center of the curve unless you change how you are picking judges.

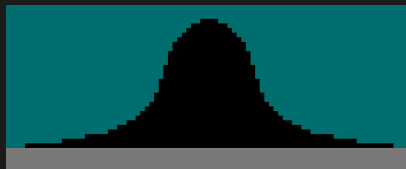
This leads to an important and very useful point...

# How Good is an Image?

Photography competitions exist to try to answer the question, “How good is this image?”

For each image, the middle of its bell curve is **THE DEFINITIVE ANSWER** ...at least for the group that generated the curve. Different groups (i.e., groups that are selected in a different way than a group like the 4Cs) will generate different curves and therefore different answers.

There is no objective answer; the best we can do is the normalized answer for a given group that a bell curve provides.



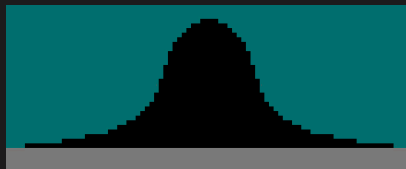
# The Problem Is...

There aren't enough judges in the 4Cs to have  
1000 people score every image!

However, we do two key things to approximate the result:

- We use three judges instead of one; this offsets the risk of using a single judge who deviates from the middle of the curve.
- We ask judges to observe the judging principles. If conscientiously observed, the judging principles will steer scores toward the middle of the bell curve.

But we could do better....





# The Solution (or Perhaps Part of It)

The most useful thing you can do as a judge is this:

- For each image, predict the middle of the its bell curve
- If your prediction is different from the score you would otherwise give, score based on your prediction, not your own opinion.



# Why??

It boils down to this:

- 4Cs judging is effectively anonymous. Competitors generally don't know the judges and usually don't even get the per-judge scores. Even if they get the individual judge scores, they don't know which judge was which. So they can't tie a given score to a particular judge.
- Competitors DON'T CARE about the opinion of a judge they don't know. They often care about the opinions of people they know, but that's not what 4Cs judging provides.
- Competitors DO want to know how good their images are. That's why people participate, and that's what predicting each image's bell curve provides. So the scores of an anonymous judge can still be really valuable— as long as they help predict the middle of each bell curve.

# Look at it this way...

If you submitted images to a competition judged by people you don't know, which of these would you prefer?

- A) judges who give you their personal opinions, or
- B) judges who make an effort to predict the typical response

Both sound reasonable...until you encounter a judge who's personal opinion is on the fringes of the bell curve. (9-9-5? Where did that 5 come from??)

Now turn that around. Since competition exists to benefit the competitors rather than the judges, the whole idea here is maximize benefit to the competitors by ensuring that YOU are not the judge who is on the fringes.

The best way to do that is to observe the guidelines and predict the middle of the bell curve.

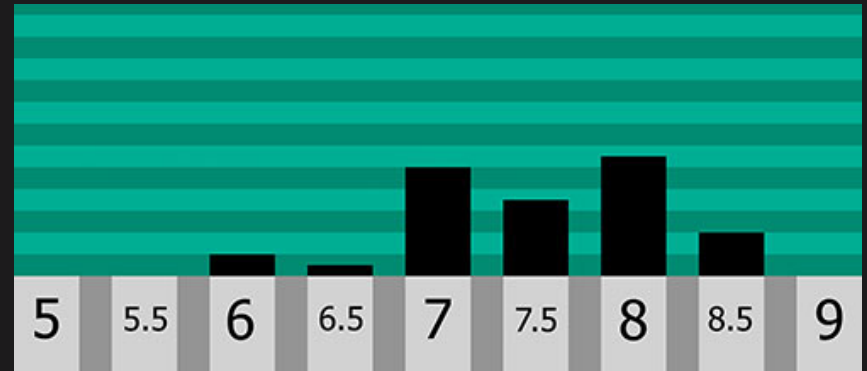


# Where is the Middle of this Bell Curve?



What is your prediction of the average score for this image?

# Where is the Middle of this Bell Curve?



Average: 7.51

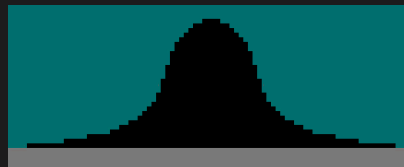
The consensus on this image was a bit better than average.

# Bell Curve Summary

In a 4Cs competition, what the competitors want to know is how good their images are, and the middle of the bell curve is the answer.

If you conscientiously observe the judging principles, your own score should already match your prediction of the middle of the bell curve, i.e., the average 4Cs score.

If your own score does not match your prediction, you would be doing the competitor a favor by scoring your prediction. It may stick in your craw but you'd be doing the right thing: *your prediction is more valuable than your opinion.*





# Judging Matters

Most competitors— even relatively new ones-- have a reasonable intuitive sense of image quality, i.e., where the bell curve is for a given image.

Even if the results within a competition are fair, if the scores overall are much lower than expected per the scale it sends a powerful and negative message to the competitors.

When that occurs, or even when an individual judge's scores deviate from the bell curve, it can create a degree of frustration that can turn people off of competition, off of camera club, and even off of photography.

# Practice Helps!

Judging is like playing a piano: you get better with practice. Meterites CC devised a really useful approach to this:

- Everyone scores all images on a sheet of paper.
- Three sheets are drawn at random and used for the official scores. Alternatively, you can draw five and discard the high and low scores.
- Everyone gets the scores back to compare to the official scores. If there is interest, you can discuss cases where there are disagreements.

You can leave names off the papers (and not return them) if you want to do anonymous judging, but that eliminates some of the benefit for feedback.

# Final Thoughts

It's natural for photographers who want to learn the craft to seek feedback on how they're doing.

Judging is the best tool we have for assessing and comparing large numbers of images, but it has to be done well to be useful.

If you are judging, please make every effort to be fair and observe the guidelines.



*End*

Summary reference slides follow

# Summary 1: Evaluating Images

For both making and evaluating salon photographs there are two key questions:

1. Are the ideas in the image compelling?
2. Are the ideas in the image cleanly expressed?

The answers to these two questions are determined by these attributes, which determine all image characteristics:

1. Content (i.e., everything in the frame)
2. Lighting
3. Composition
4. Camera Operation
5. Post-Processing

## Summary 2: 4Cs Scale

1 - 4	5	5.5	6	6.5	7	7.5	8	8.5	9
Not used in 4Cs scoring	Noticeable problems offset strengths		Okay but with minor problems		Solid; Average 4Cs quality		Strong; rewards viewing		Outstanding; an exciting image

### Anticipated typical score distribution in 4Cs competitions

0%	1-4%	15-20%	50-65%	15-20%	1-4%
----	------	--------	--------	--------	------

Important point: The scale is intended to cover only the range of image quality typically seen in 4Cs competitions. It is not intended to cover quality levels above or below the typical 4Cs range.

With this scale, 9 is intended to be achievable in a typical competition: multiple award winners will likely have a score above 24, and there may be an occasional 27.

# Summary 3: The Principles of Judging

These all are variations on *Be Fair*:

1. Set aside your personal biases
2. Judge each image as a whole
3. Analyze instead of just responding
4. Be consistent
5. Use the scale as described



# Summary 4: The Bell Curve

Applying the idea of the bell curve in judging will help make the result more useful:

- Judging any image with many judges would produce a bell curve for that image, and the curve would be stable.
- The middle of the bell curve for an image is the best answer we have to the question “How good is this image?”
- Judging is useful only if it answers that question, and consequently only if it approximates the middle of an image’s bell curve.
- The most useful thing a judge can do is predict the middle of the bell curve for that image.
- Conscientious application of the judging principles should produce that middle score; when there is a difference between your score and your prediction, score your prediction.