# Counting on AI: Rater Instructions

## Overview

The raters are the most important parts of studies that use ratings for data. In the jargon of Psychology research methods, raters count as "equipment," just like MRI machines or DNA sequencers.

A tried and true way to make sure that ratings are consistent across different raters is to "calibrate" ourselves. Fortunately, we don't need to have our liquid helium levels filled up to do this (like MRI machines), we just need to receive the same training. Also fortunately, the only training we need is to read this document!

## Why the study uses these object types

The images contain three types of objects to count:

- bocce balls
- medicine balls
- hot air balloons

These are a little different from the types that we all talked about in our meetings, so I want to explain why. But first, like a good scientist (= uptight person), I want to define the trerms:

- **bocce ball**: Bocce is a lawn bowling-type game where players throw/roll 4-inch wide, hard balls at a smaller ball. The AI-generated images sometimes, but rarely, include smaller balls mixed in with the 4-inch ones. Both sizes can be called "bocce balls." In a typical game each player uses different colored balls.
- **medicine ball**: Wikipedia tells me that medicine balls were invented over 150 years ago to torture gym students. They are heavy, soft balls about 14 inches across. They can come in different colors and textures.
- **hot air balloon**: OK, Wikipedia told me everything about all three of these things. There are several varieties of hot air balloons, but the most commonly used type is about 60 feet tall (just the balloon, not including the basket underneath). They come in all sorts of colors, but usually stick to the same classic hot-air-ballon shape.

In the end I decided not to use tennis balls or basketballs, even though the images of tennis balls were

nice.

The images of basketballs were weird, because sometimes the basketballs would be morphed together to look like a dividing embryo. I tried different types of similarly-sized balls, including "playground balls," but the best images seemed to come from "medicine balls." I couldn't help but wonder if it was easier for the AI model to draw medicine balls because they can come in any color – it didn't have to try to "find" many regions of the random image that it starts with that were all the same color.

That made me realize that asking the AI model for tennis balls was imposing a similar constraint – tennis balls are all the same color. If hot air balloons and medicine balls are allowed to be any color, then I wanted the smallest object type to follow the same rule. That lead me to bocce balls, which are different colors so that players can tell theirs apart from others'.

I would rather have used objects a little more familiar than bocce balls and medicine balls, which don't really come up in conversation that often (at least not for me), but I was happy with their visual properties.

# Count up to 20

The AI model was asked to draw images with 1 through 16 of the different objects. However, the AI model often gets the number wrong, and we'd like to know exactly how it goes wrong. Therefore, our job is to count all of the objects in the image, even if there are more than 16.

An image to rate and the "more than 20" question

However, sometimes the AI model's image is so obviously wrong that it just doesn't seem worth it to count all its mistakes. So, **if and only if we're sure** that there are more than 20 objects in an image, we can click an option for "Definitely more than 20" which will make the response slider-thingies disappear for that trial, so that we don't have to give a precise answer.

Recommendation: If it's not immediately clear whether there are more than 20 objects, start to count them, and then stop if you get close to 20 and there are still lots of objects left.

## DO count objects even if only a little bit is visible

Even if only a tiny sliver of an object is visible (because most of it is outside of the frame, or most of it is

hiding behind another object), **please do count it**.

# Problematic counting situations that need judgment

Sometimes there will be things in an image that might be objects, or might not be. We'll just have to make our best judgment in those cases.

Here are some cases that I noticed while making the images:

## Ghostly objects

Especially with hot air balloons, there can be semi-transparent "ghost balloons" sometimes.

Recommendation: if it looks more like a balloon than anything else, even if it's transparent, then count it.

## Merged objects

Sometimes two objects will look morphed together like a dividing embryo.

Recommendation: if it looks like the two things would separate if you picked up one end and shook it, then count them as two objects.

## Far, tiny objects

Sometimes there will be very small and/or blurry things in the far background that might or might not be objects.

Recommendation: if it could possibly be something other than the type of object you're counting, don't count it. Otherwise, do count it.

# How to count boundary clipped objects

The second response slider-thingy asks for the number of boundary clipped items. "Boundary" just means "the edge of the picture" in this case. If less than **50% of an object** is shown in the frame of the picture, it counts as boundary clipped.
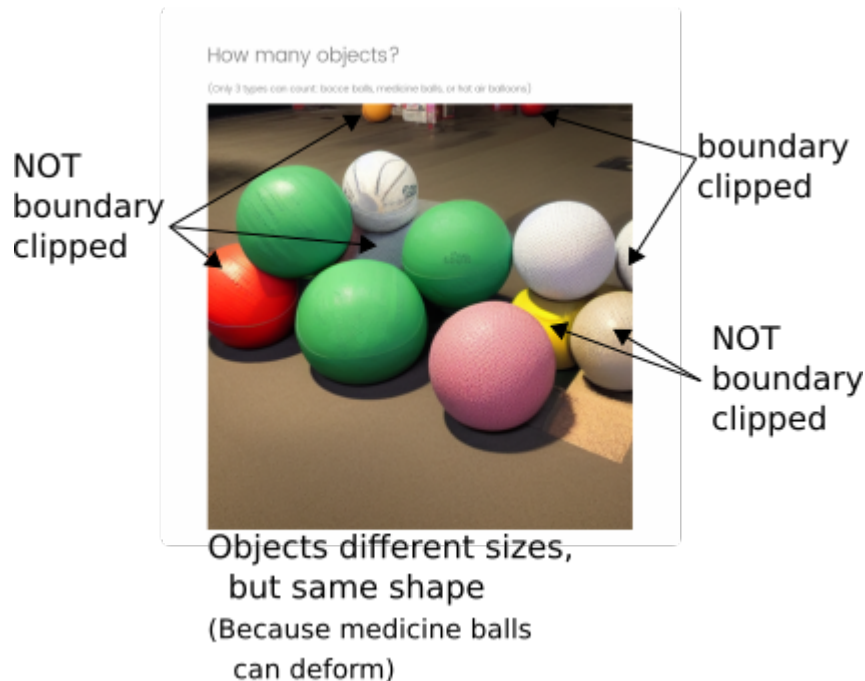
Boundary clipping examples

Note: Objects that are partly occluded by (hidden by) *other objects* **don't count** as boundary clipped.

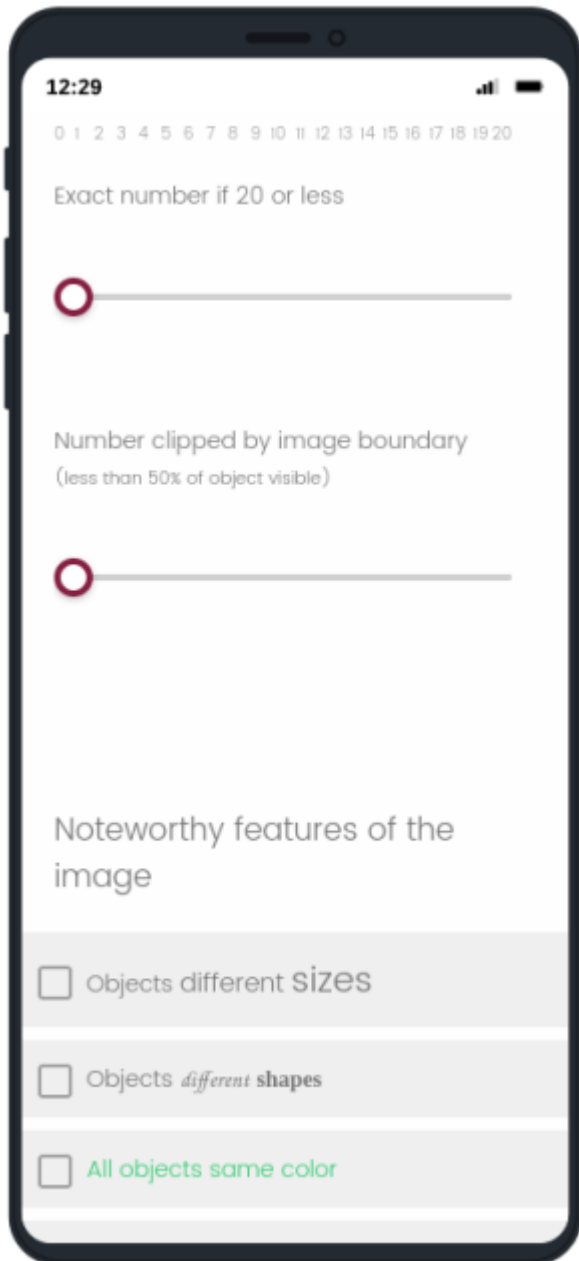In the example below, I counted 13 medicine balls, with 2 boundary clipped.

# Consistency of objects

The following are questions to help us assess how consistently the AI model has drawn the objects in a picture.

## Different sizes

This means **real world size**, **NOT retinal size**. Retinal size is basically how many inches the objects takes up on the screen, and this varies drastically depending on the camera angle.

Imagine that there is a picture of two hot air balloons that are the same real world size (say, 60 feet tall). If one is far away, it might only take up 1/4" on the screen, while one that is near the front of the picture might take up 2". **These two balloons would be rated to be *the same* size.**

The slider-thingies and some of the check boxes

Recommendation: Allow a little tolerance so that objects that are very nearly, but not exactly the same size can count as "the same size." When objects are noticeably different size, then rate them as "different sizes."

# Different shapes

Recommendation: Again, allow a little tolerance for objects that are very nearly the same shape to count as "the same shape."

## All same color

Most of the time, the objects will be different colors. When all of the objects of one of our key types (bocce ball, medicine ball, or hot air balloon) are the same color, then select this response.

## Other object type present

For example, if an image shows a person along with a bunch of bocce balls, then select this response. The same goes for an image of 10 medicine balls plus one hammer.

However, **don't count the background** as a different type of object. For example, the grass that bocce balls are sitting on doesn't count as a different type of object. Any trees in the background of an image of hot air balloons don't count, either.

Things from the background that *would* count include:

- A house on the ground below a bunch of hot air balloons. (A house is a man-made thing, which makes it different from natural elements belonging to the background.)
- A bowl lying on the grass that holds several bocce balls. (A bowl is not a typical part of the background for images of bocce balls.)

Just use your judgment and it will be fine.

# When to leave a comment

## When an other object type is present

If you check the box for "Other object type present," please add a comment reading "other: X" where X is the name of the other type of object.

## Noteworthy image features

The final response field allows us to check a box and then enter a comment about the image.

Recommendation: Don't do this too often.

Good reasons for leaving comments include:

- You're particularly unsure about your response
- The image was ambiguous or confusing ("I can't tell what that thing in the corner is supposed to

be!")
- The image contained something truly awful (not just weird, but offensive or disgusting)
- The objects formed an unusual pattern. (Not just any pattern, but a truly noteworthy one. Examples: if all the objects form a nice square grid, you **don't** need to comment on that, but if the objects form the numeral "9" please do comment on that.)

From:
https://wiki.anthonycate.org/ - **Visual Cognitive Neuroscience**

Permanent link:
**https://wiki.anthonycate.org/doku.php?id=research:rater_instructions**

Last update: **2024/02/11 20:08**