This Image Intentionally Left Blank:
Mundane Images Increase Citizen Science Participation

Alex Bowyer, Veronica Maidel, Chris Lintott, Alexandra Swanson, Grant Miller
Zooniverse / University of Oxford, UK / Adler Planetarium, Chicago

Introduction & Background

Snapshot Serengeti asks volunteers to identify animals in camera trap photographs. We would like to encourage users to stay longer and classify more images. In a previous experiment, we inadvertently removed "blank" images, and found that contributions dropped. Could it be that blank images are important for motivation?

Research Question:

What impact do blank images have on engagement?

Do blank images (left) encourage or deter users when interspersed with non-blank images (right)?

Results & Conclusions So Far

Null hypothesis: "There are no differences in mean session length between the cohorts."

Kruskal-Wallis test result: $p < 0.05$ => Strong evidence against the null hypothesis =>

The alternative hypothesis is: "The cohorts have different mean session lengths."

Dunn's post-hoc test: 80% cohort confirmed significantly different from 0%, 20% cohorts.

=> Session lengths are longer with 80% blank images than with 0% or 20% blanks.

We also calculated the *effect size* of percentage blanks upon session length, along with confidence intervals, comparing each cohort to control using *Glass' Δ*. We found a **decreasing negative effect size from moderate to small as % blanks increases**:
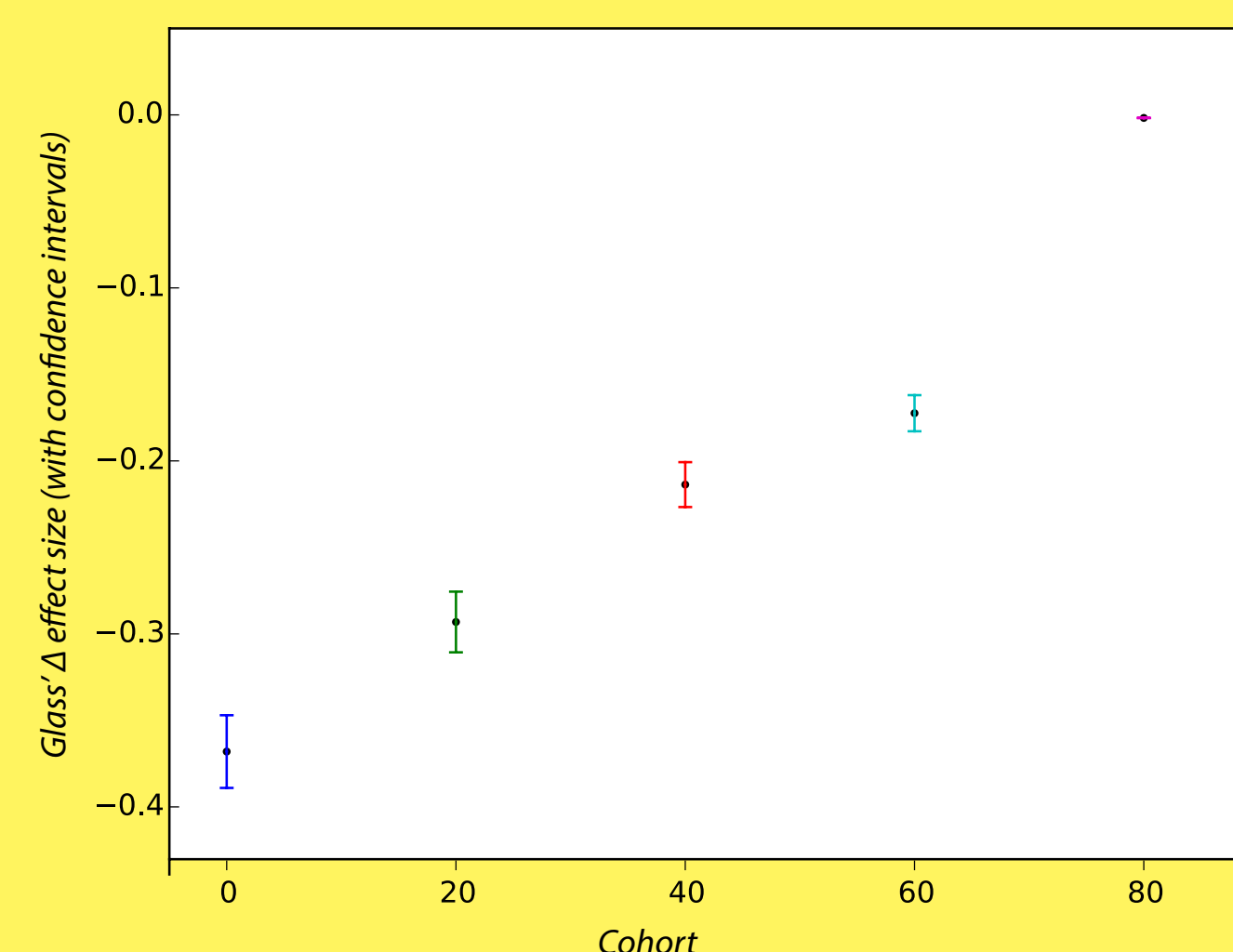
The effect size of percentage blanks upon session length, relative to control
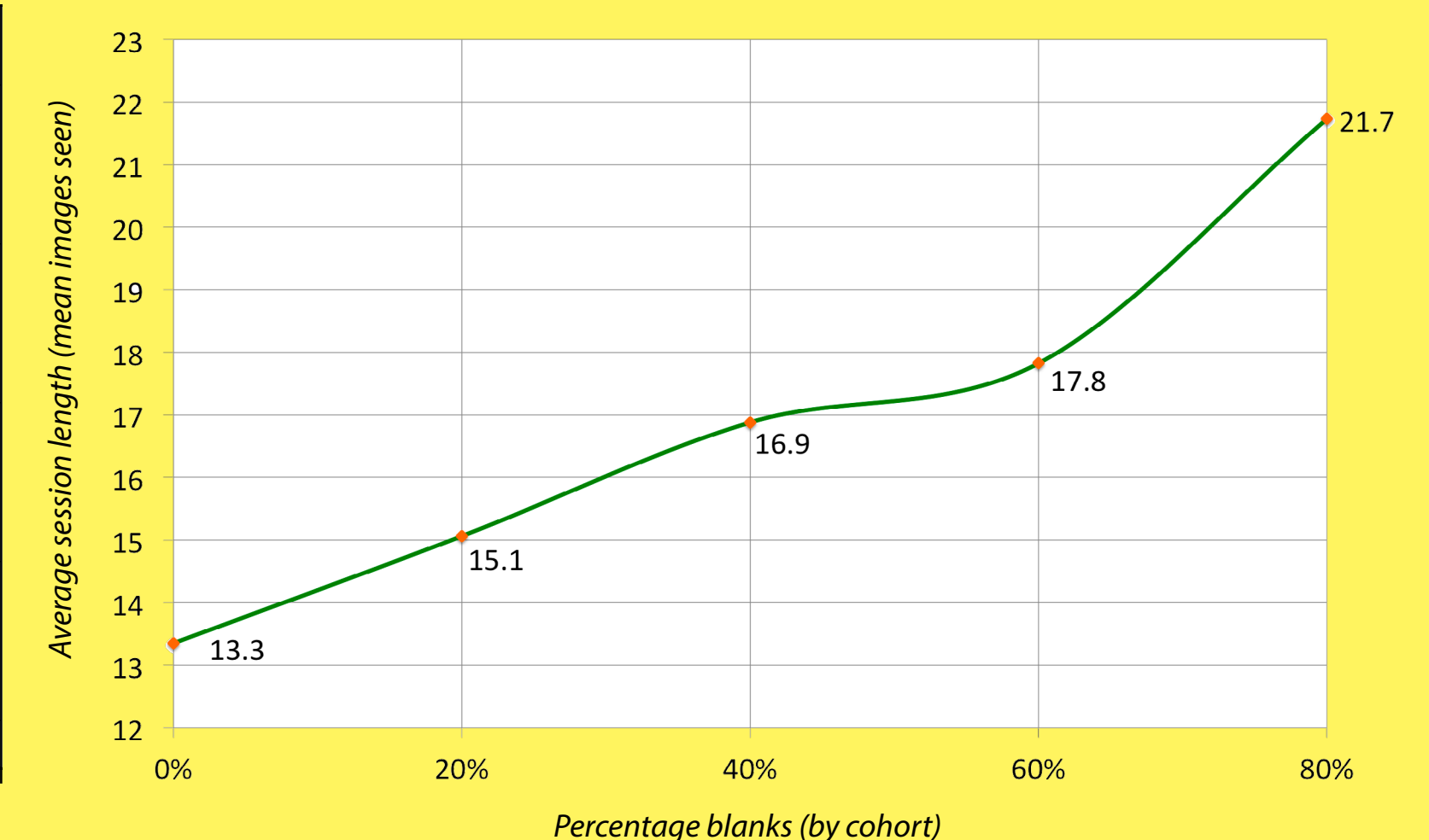
A visualisation of the increasing mean session length for each percentage of blanks (summarised per cohort)
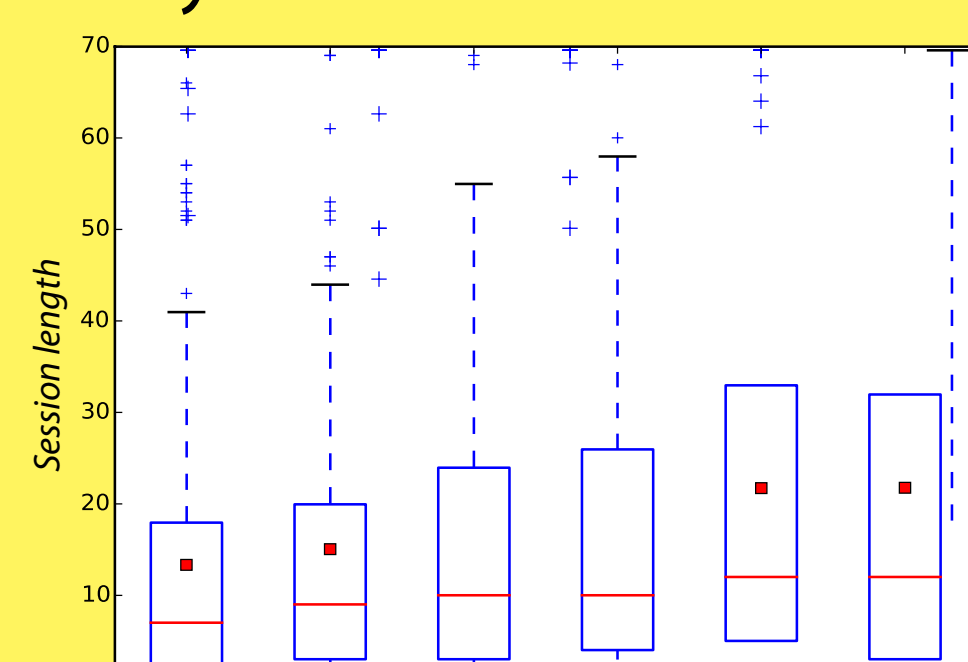
The Experiment

Normally, users are shown images at random, most of which are blank. We built a framework[1] to control which images were shown. Each user was assigned to cohort. Each cohort was allocated images with a defined percentage of blanks.

The experiment ran for 3 weeks in July 2015, with 1,882 participants. Cohorts were defined and filled as follows:

Cohort	User Sessions
0% blanks	293
20% blanks	240
40% blanks	225
60% blanks	228
80% blanks	223
Control (71% blank)	297

Analysis of the data shows a difference between cohorts:

Box-and-whisker plot: session lengths

Cohort	Significant?
0%	Yes: $p < 0.01$
20%	Yes: $p < 0.01$
40%	Yes: $p < 0.01$
60%	Yes: $p < 0.05$
80%	No: $p > 0.05$

T-Tests: Differences from control, by session length

[1] Our experimental framework consisted of two components: **Geordi** (http://github.com/zooniverse/geordi) - a REST-based user analytics collector, and the **Zooniverse Experiment Server** (http://github.com/zooniverse/experiment-server) - a PlanOut-based experiment management system.

Can psychology explain it?

Intermittent rewards[1] can increase contribution, especially when delivered randomly. This effect is seen both in gambling and computer games.

The theory is that more "misses" make each "hit" seem more valuable.

Rare, unpredictable "hits" motivate you to seek them out and keep going.

The task becomes fun, more of a *hunt*.

In our case, an animal seen after a series of blanks gives an *immediate payoff*. [2]

[1] Kendall, 1974 - "Prefence for intermittent reinforcement"
[2] Kaufmann, 2001 - "Worker motivation in crowdsourcing and human computation"

What do volunteers say?

Some quotes from a recent survey:

*"It's like panning for gold... Just one more and you will **find something interesting**"*

*"I like looking for **something out of the ordinary**."*

*"There is a certain thrill in **coming across a good image** on Snapshot Serengeti"*

*"You click Next and think 'Just one more...' but then **something interesting comes up**."*

*"I'll carry on **until I see an interesting one**."*

*"I never know **what animal might show up next**."*

Questions being investigated - Please share your thoughts!

1. Is this correlation or causation?
Can we be certain that having more blank images *causes* the increase in engagement?

2. At what percentage of blanks does the effect upon session length reverse?
It seems that between 80% and 100% mean session length should peak and decline.

3. Is there historical support for the finding?
Do historical records of all past classifications on the site also show the same correlation?

4. Do other aspects of the images influence engagement?
Perhaps variety, complexity, difficulty, aesthetics, or novelty are important...

5. How can we better control percentage blanks seen, while retaining randomness?
Each cohort experienced a wide variation from the mean, especially with short sessions.

ZOONIVERSE