EuroRV^3 2017

Perception in Visualization
12-13 June, Barcelona, Spain

TU Delft Delft University of Technology

From a user study to a valid claim

How to test your hypothesis and avoid common pitfalls

Niels de Hoon, Elmar Eisemann, Anna Vilanova

Introduction

Find support by means of a user evaluation for a claim made on a visualization

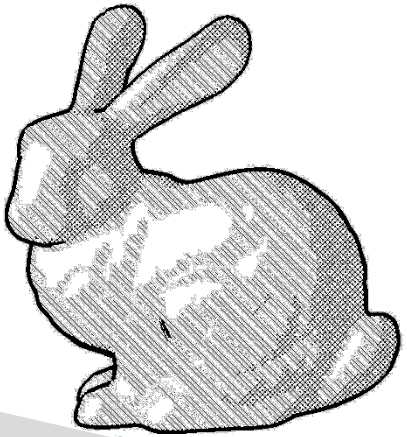
An accessible summary of the statistical tools that can be used

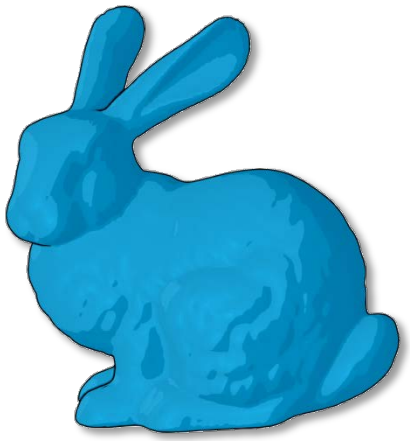
Common pitfalls and how to avoid them

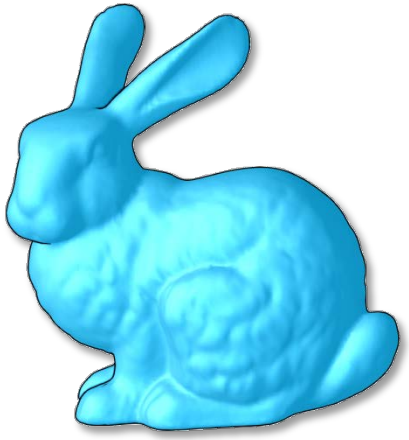
Introduction

User-based quality measures:

- Perception
- Effectiveness
- Task performance







Introduction

The number of user-based evaluations of visualizations has been increasing[1,2]

Previous work indicates when[3,4] to perform a user study and how it should be conducted[5,6]

1: Tory M., Möller T.: Human factors in visualization research.
2: Isenberg T., Isenberg P., Chen J., Sedlmair M., Möller T.: A systematic review on the practice of evaluating visualization.
3: Munzer T.: A nested model for visualization design and validation.
4: Smit N. N., Lawonn K.: An introduction to evaluation in medical visualization.
5: Glaβer S., Saalfeld P., Berg P., Merten N., Preim B.: How to evaluate medical visualizations on the example of 3d aneurysm surfaces.
6: Carpendale S.: Evaluating Information Visualizations

Evaluation steps

- Formulate a hypothesis
- Define the user study
- Find the right (amount of) participants
- Conduct the user study
- Statistical analysis

Evaluation steps

- Formulate a hypothesis

We would like to reject the hypothesis (strongest conclusion)

E.g.: in the justice system
Null hypothesis: suspect = innocent
Alternative hypothesis: suspect ≠ innocent

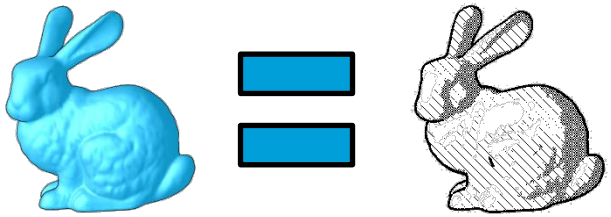
We need enough evidence to reject the null hypothesis

Evaluation steps

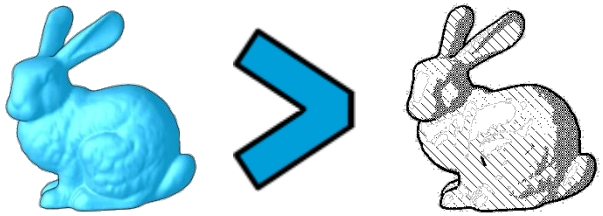
- Formulate hypothesis

By conducting the user study we want to find support for a claim that holds for our visualization

Null hypothesis:



Alternative hypothesis:



Our technique **State of the art**

Shape perception techniques

Evaluation steps

- Formulate hypothesis
- Define the user study

Questionaire?

Task performance?

Quantitative proof?

Evaluation steps

- Formulate hypothesis
- Define the user study
- Find the right (amount of) participants

Domain experts/laymen?

How many do we need?

How many can we find?

Evaluation steps

- Formulate a hypothesis
- Define the user study
- Find the right (amount of) participants
- Conduct the user study

Question/Task	User 1	User 2	…
Question 1	4.2	4.5	
Question 2	3.9	3.6	
…			
Task 1	30.6	32.1	
Task 2	15.9	14.3	
…			

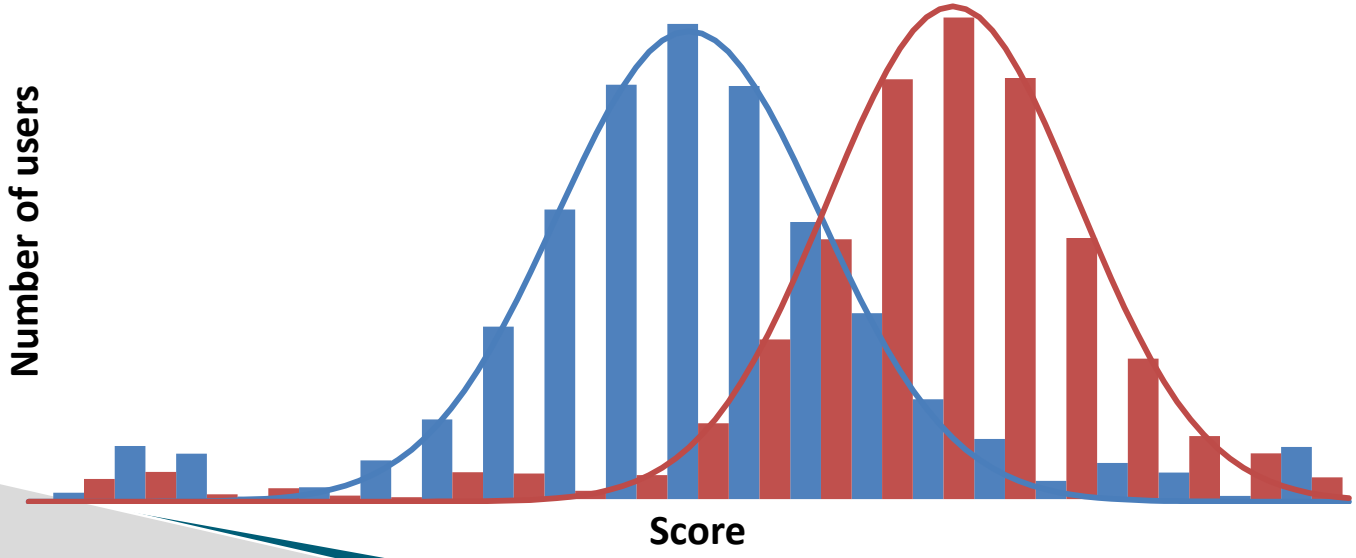
Evaluation steps

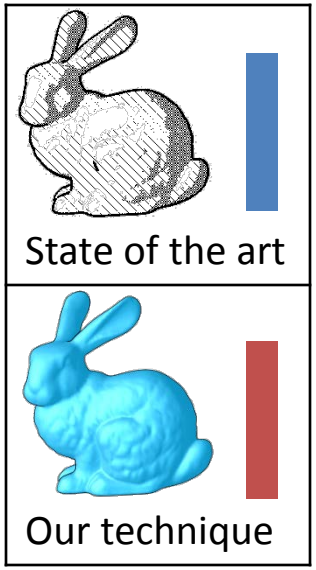
- Formulate a hypothesis
- Define the user study
- Find the right (amount of) participants
- Conduct the user study
- Statistical analysis

How do we show our experiment supports our claim?

Data

Question/Task	User 1	User 2	…
Question 1	4.2	4.5	
Question 2	3.9	3.6	
…			
Task 1	30.6	32.1	
Task 2	15.9	14.3	
…			





Data assumptions

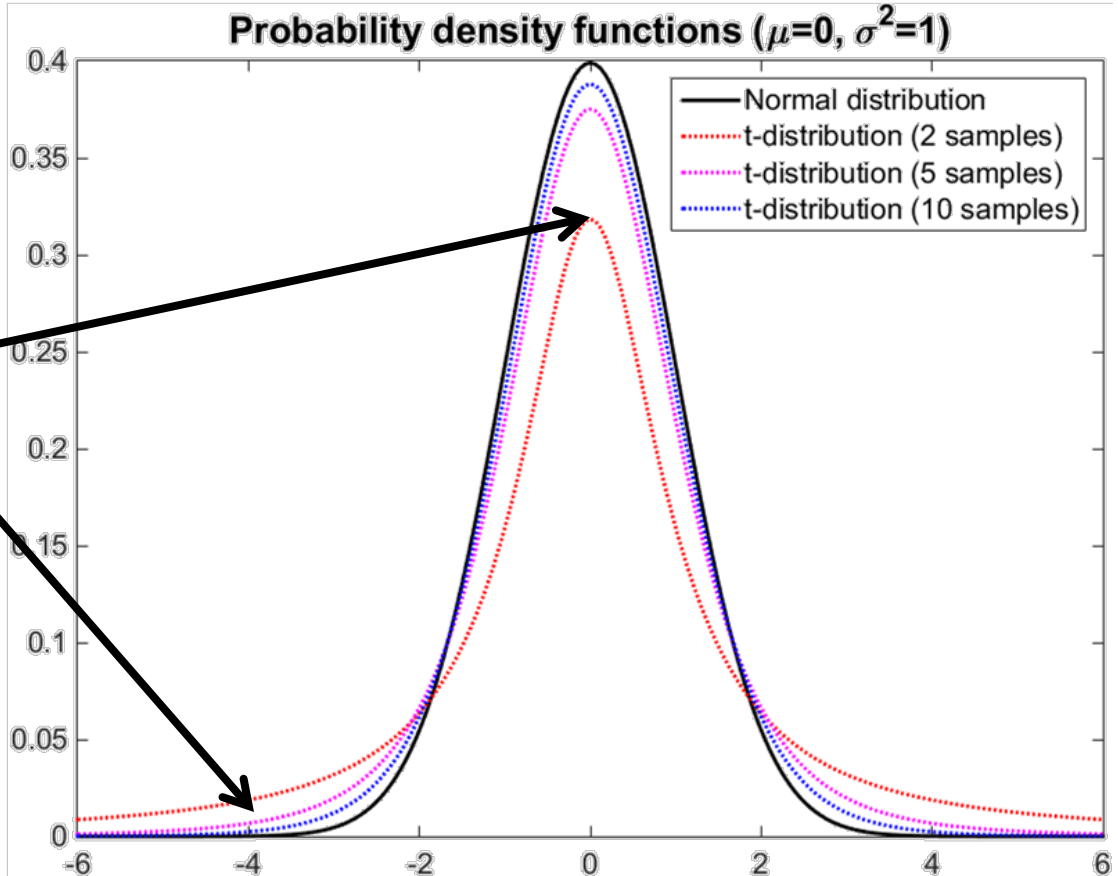
- Assume we have a user study with a small number of participants
- The mean and variance are unknown
- The distribution of the data is assumed to be a normal distribution

Use the t-distribution

t-Distribution

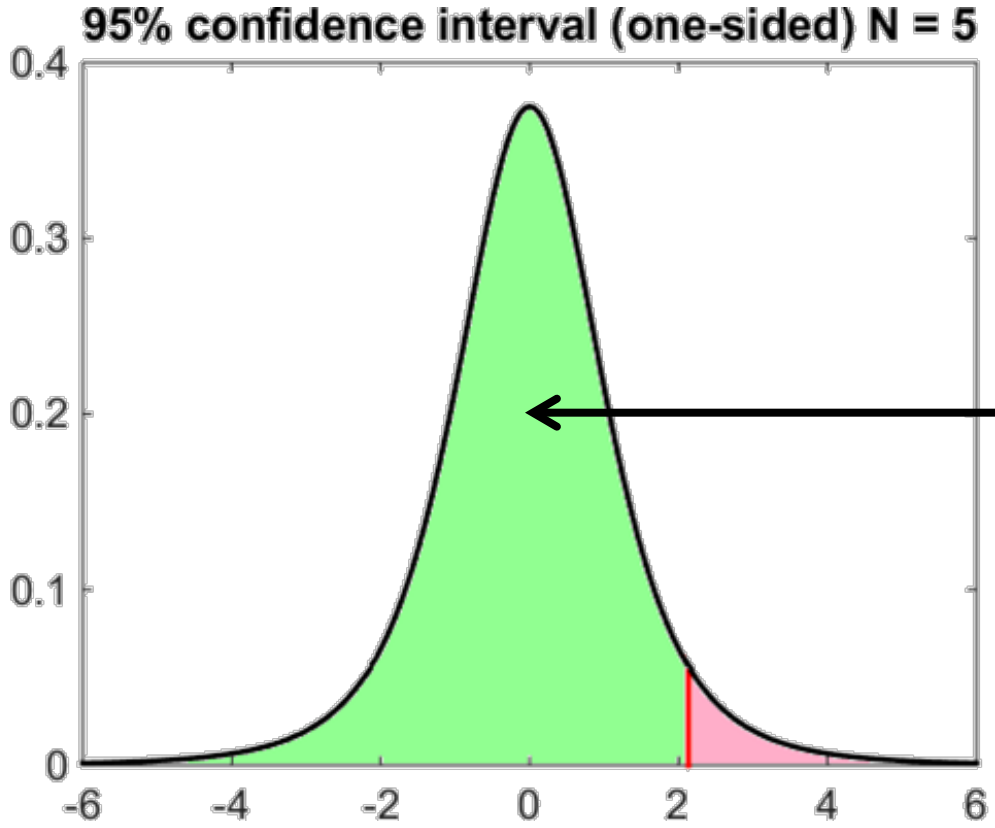
Describes the samples drawn from a normal distribution without knowledge on both the mean and variance

Lower number of samples result in lower probabilities and a wider spread



Confidence interval

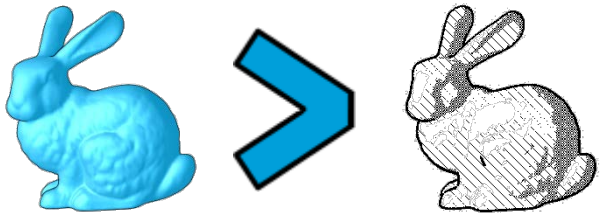
From the distribution we can estimate for which we have 95% confidence the mean lies within this interval

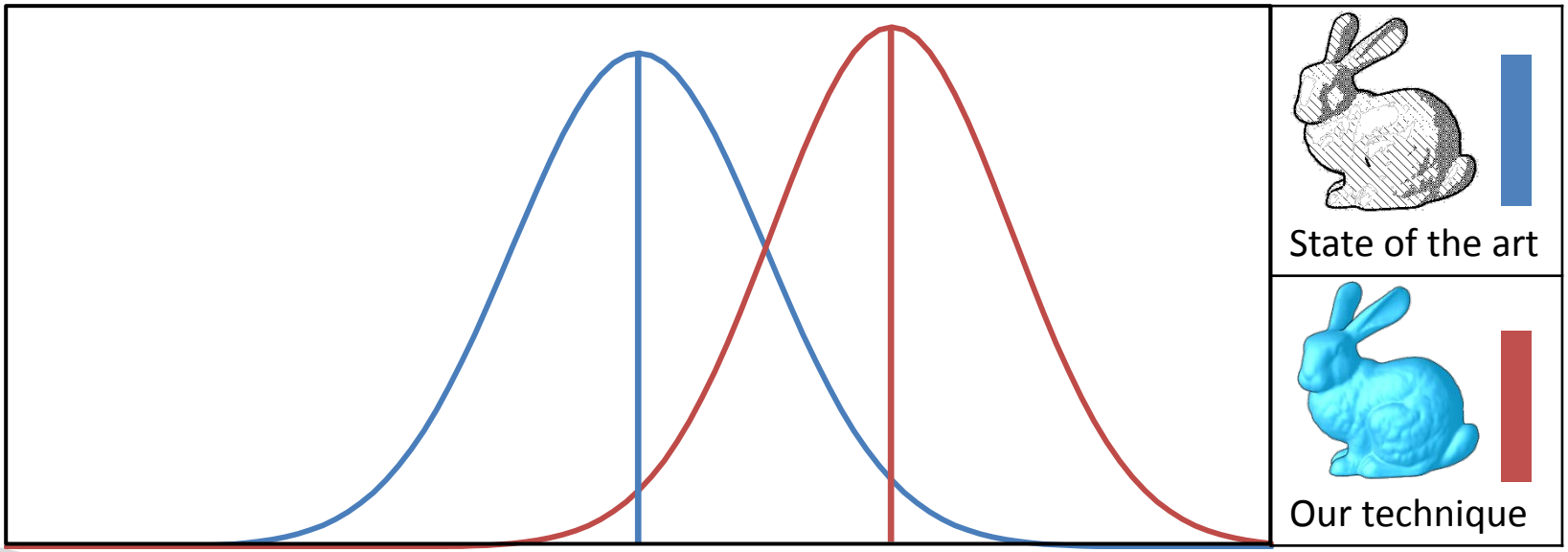


$$\int (distribution) = 0.95$$

Note: for the *t*-distribution the confidence interval will be bigger when less samples are available

Hypothesis





p-value

Assume H_0 is true

Minimize the probability when redoing the experiment we find a value that is at least as extreme as the one we found

This probability is the *p*-value

Reduce the probability of a *false positive*

False positve

- The probability of a false positive should be small,
  e.g. we do not want to convict an innocent person
- Stronger conclusion (more significant)

Hypothesis testing

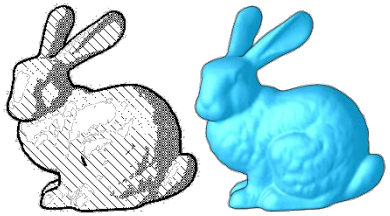
- When we cannot reject the null hypothesis, the null hypothesis is not necessarily true
- In this case we lack evidence to reject the hypothesis
- Therefore we *fail to reject* the hypothesis
- This conclusion is weak, it is not the same as saying that it was proven, since it was only not disproved.

Hypothesis testing pitfalls

The hypothesis should be clear before the user study is conducted

- Helps design the user study
- Clear impact of questions on outcome
- Helps to avoid fine tuning the hypothesis

E.g.: Which shading technique provides a better shape perception

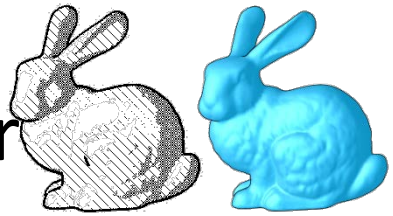


Hypothesis testing pitfalls

Be aware of the limitations of the data

- A user study is a high level evaluation
- Conclusions on underlying details can be difficult to derive

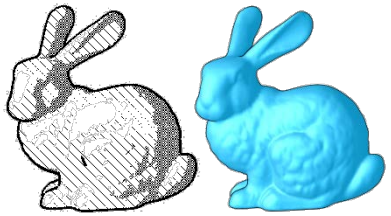
E.g.: We cannot determine from a single user study why a technique works better



Hypothesis testing pitfalls

The hypothesis should be testable

- The hypothesis should be based on something that can be measured
- “Our tool increases productivity” instead of “Our tool encourages exploration”

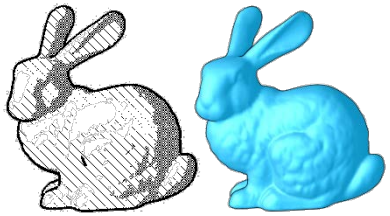


Hypothesis testing pitfalls

The hypothesis be should supported by reason

- Why a certain result is expected to be found
- Reduces the probability of a false positive

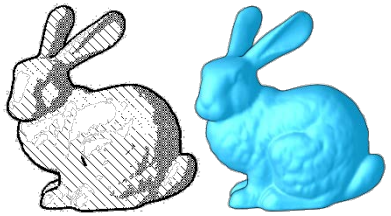
E.g.: Both techniques are intended to visualize shape



Hypothesis testing pitfalls

The number of hypotheses should be small

- The probability of a false positive increases with the number of hypotheses

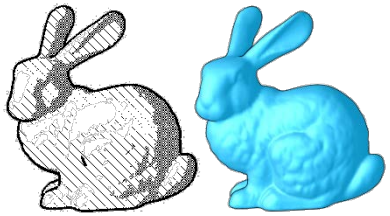


Hypothesis testing pitfalls

Find the right participants

- Laymen opinions are less usable for domain specific tools
- Attempt to sample the full user population

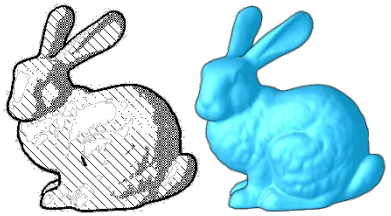
E.g.: Laymen may be less familiar with NPR rendering techniques



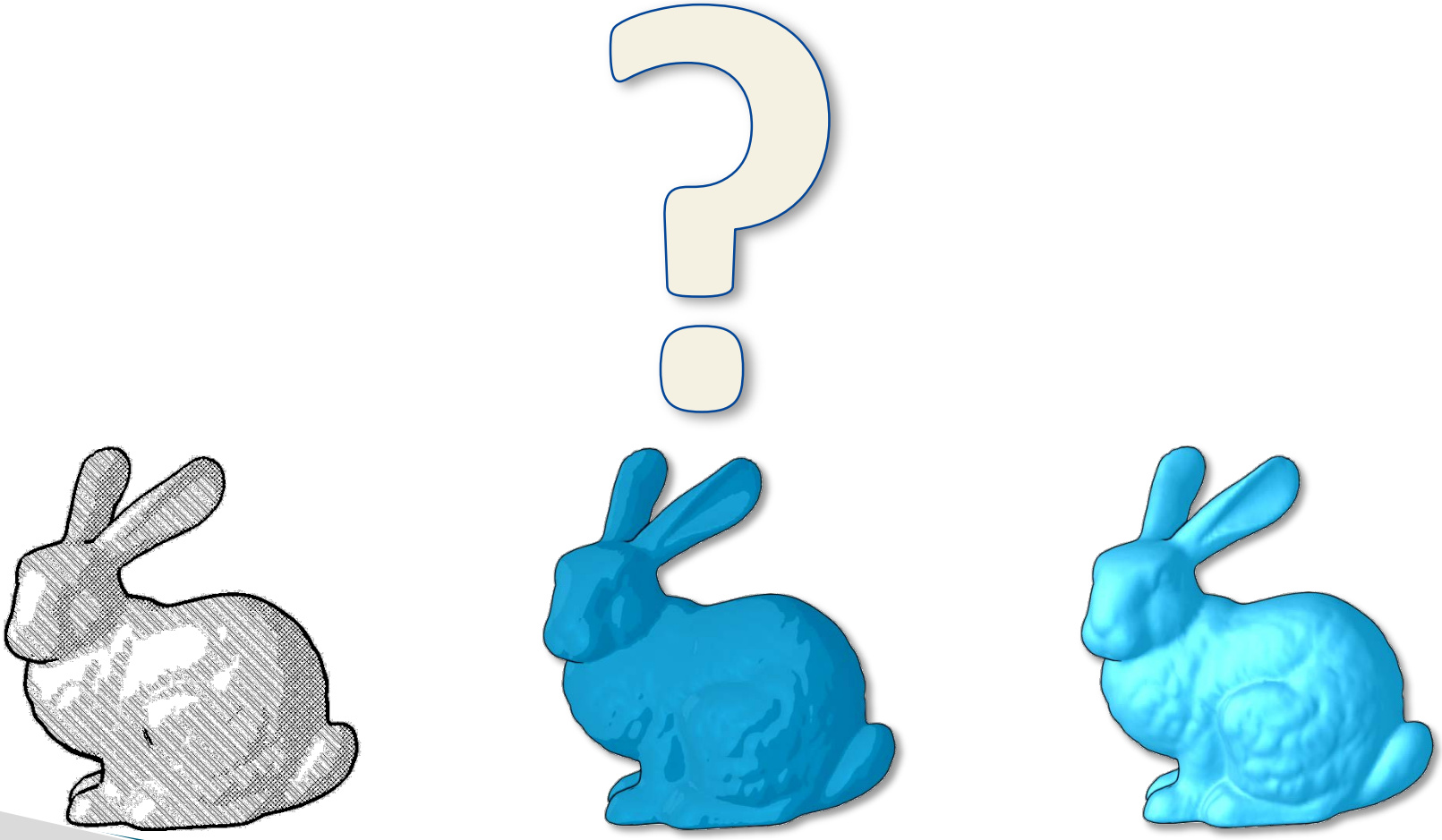
Hypothesis testing pitfalls

Use the right number participants

- Adding users to make results significant increases the probability of a false positive



Questions?



N.H.L.C.deHoon@tudelft.nl