CS-200 - INTRODUCTION TO HUMAN-COMPUTER INTERACTION

Lecture 15
Evaluation

16/3/2019

DATA ANALYSIS AND STATISTICS

DATA ANALYSIS AND STATISTICS

- How do you know the results you get are meaningful (and not just down to chance)?
- Lots of statistical tests: e.g.:
    - t-test
    - Anova
    - Kruskall-Walis
    - Wilcoxon
    - Friedman
- Easily a whole course on this

DATA ANALYSIS AND STATISTICS

- You'll usually end up with a "p" value
- This value represents the probability that the effect did not occur by chance
- Between 0 and 1

DATA ANALYSIS AND STATISTICS

- You also need to set a risk level - known as the alpha (**α**) level

- "Rule of thumb": set **α as 0.05**

    - So, five times out of a hundred you would find a statistically significant difference between the data sets even if there was none (i.e., by "chance")

DATA ANALYSIS AND STATISTICS

- Remember the null hypothesis?
- Significant result:
    - Means the null hypothesis is highly unlikely

$$p<0.05$$

- Non-significant results
    - Does not mean the null hypothesis is true
    - But, the results achieved could be a chance finding

DATA ANALYSIS AND STATISTICS

There are a lot of things to take into account when choosing a statistical test…

- Parametric vs un-parametric tests
- Matched / unmatched, paired / unpaired samples
- One group, two groups, 3+ groups
- Post-hoc analysis …

DATA ANALYSIS AND STATISTICS

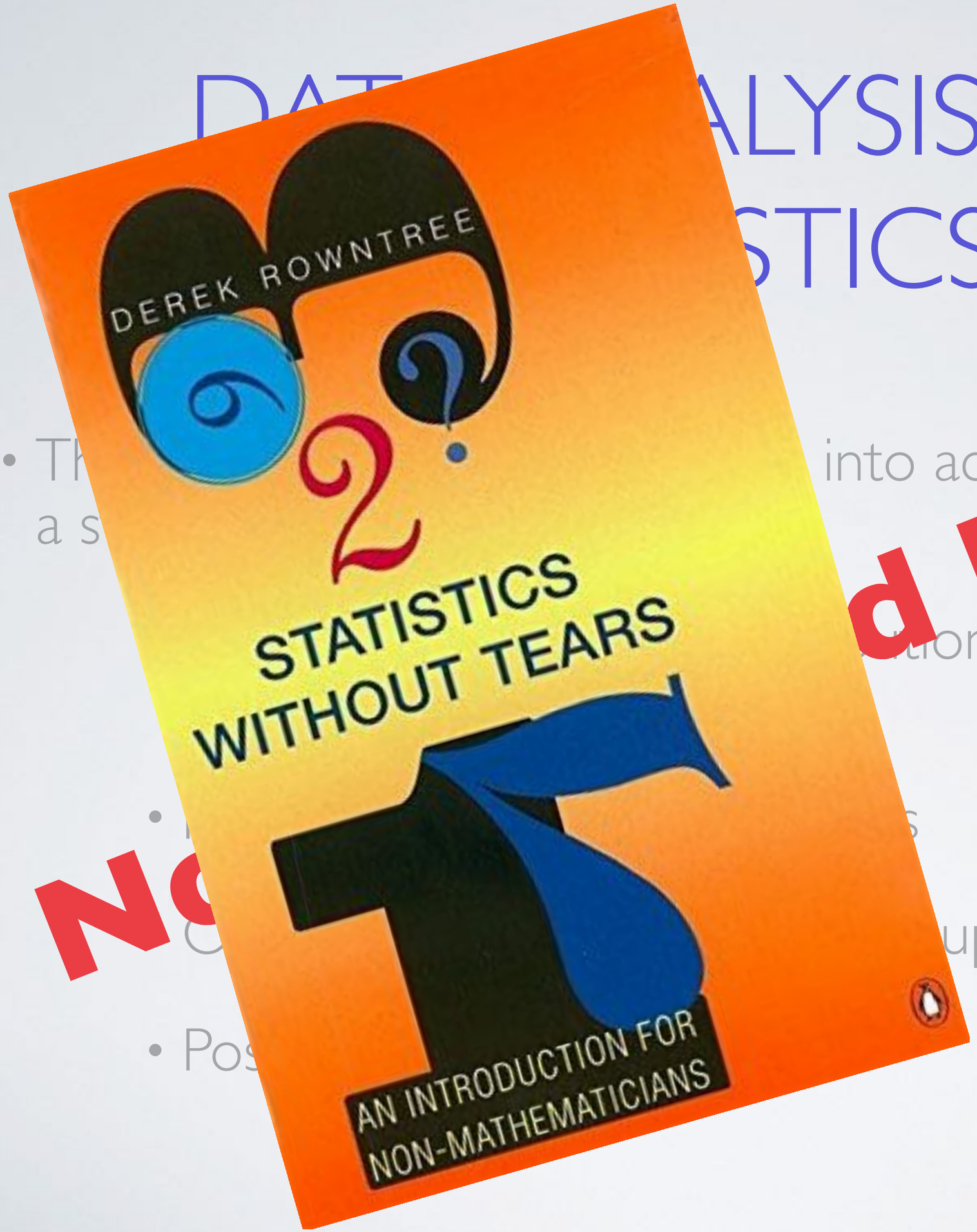
There are a lot of things to take into account when choosing a statistical test…

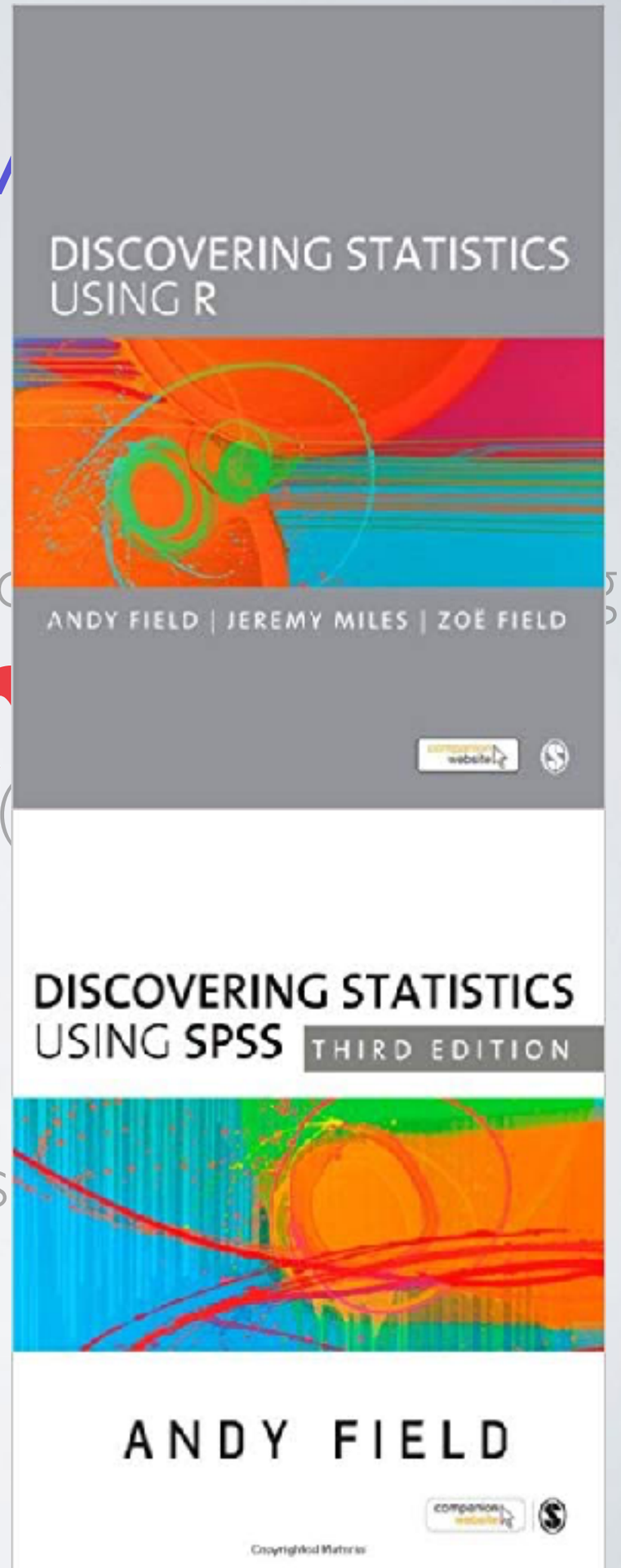
- Parametric vs un-parametric tests)
- Matched / unmatched, paired / unpaired samples
- One group, two groups, 3+ groups
- Post-hoc analysis

Not Covered Here…

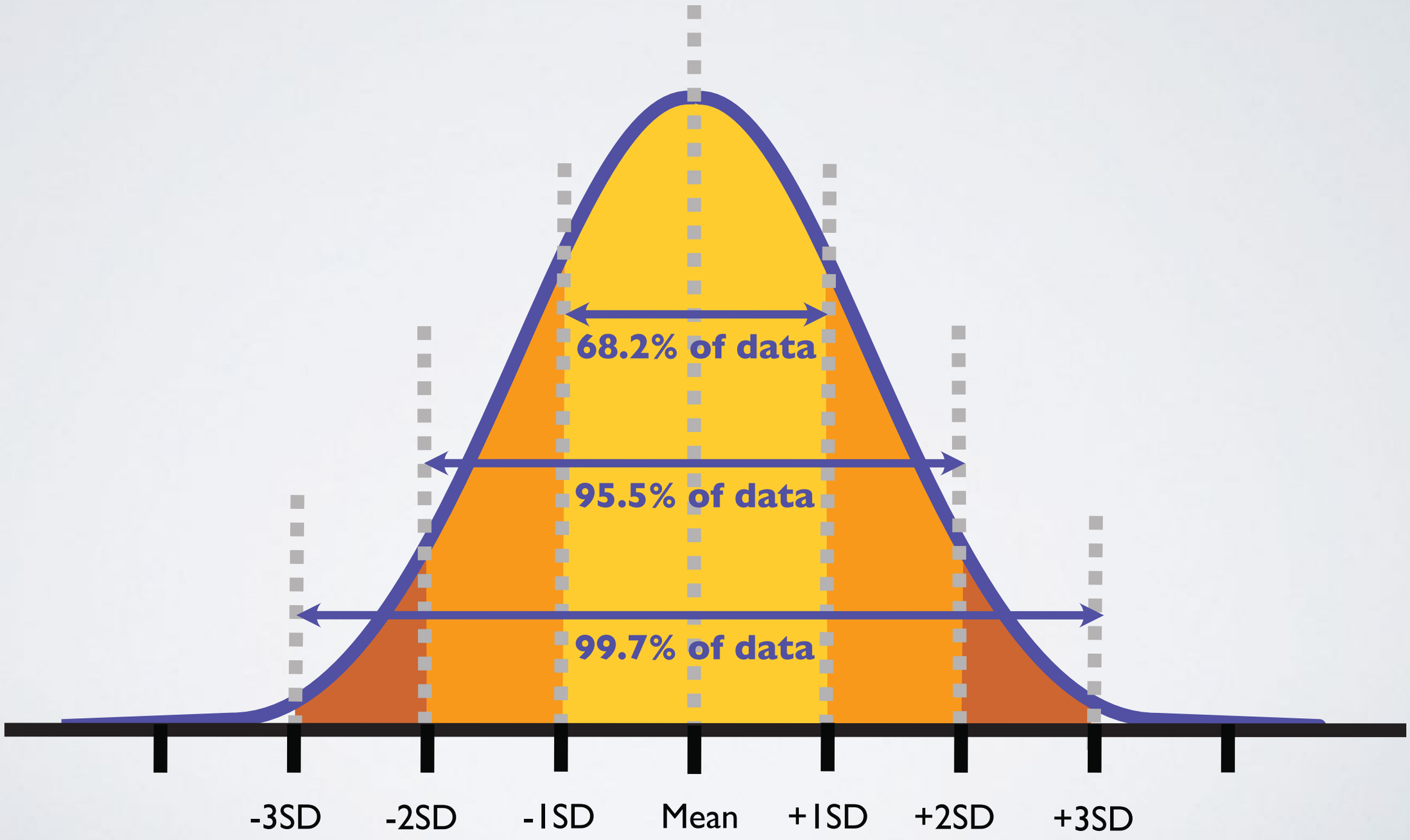
DAT ALYSIS A
STICS

- Th into acc g
  a s
- I
- C ups
- Pos





NORMAL DISTRIBUTION



T-TEST

EXAMPLE: T-TEST

- You want to find out whether Welsh or English people drink a different amount of beer per week.
- Research Question:

Who drinks the most beer in a week - the English or the Welsh?

- It is unrealistic to ask every English and Welsh person how much beer they drink, so we just take a sample of each (e.g., 300 Welsh and 300 English)
- Ask all 600 people how many pints of beer do they drink in an average week

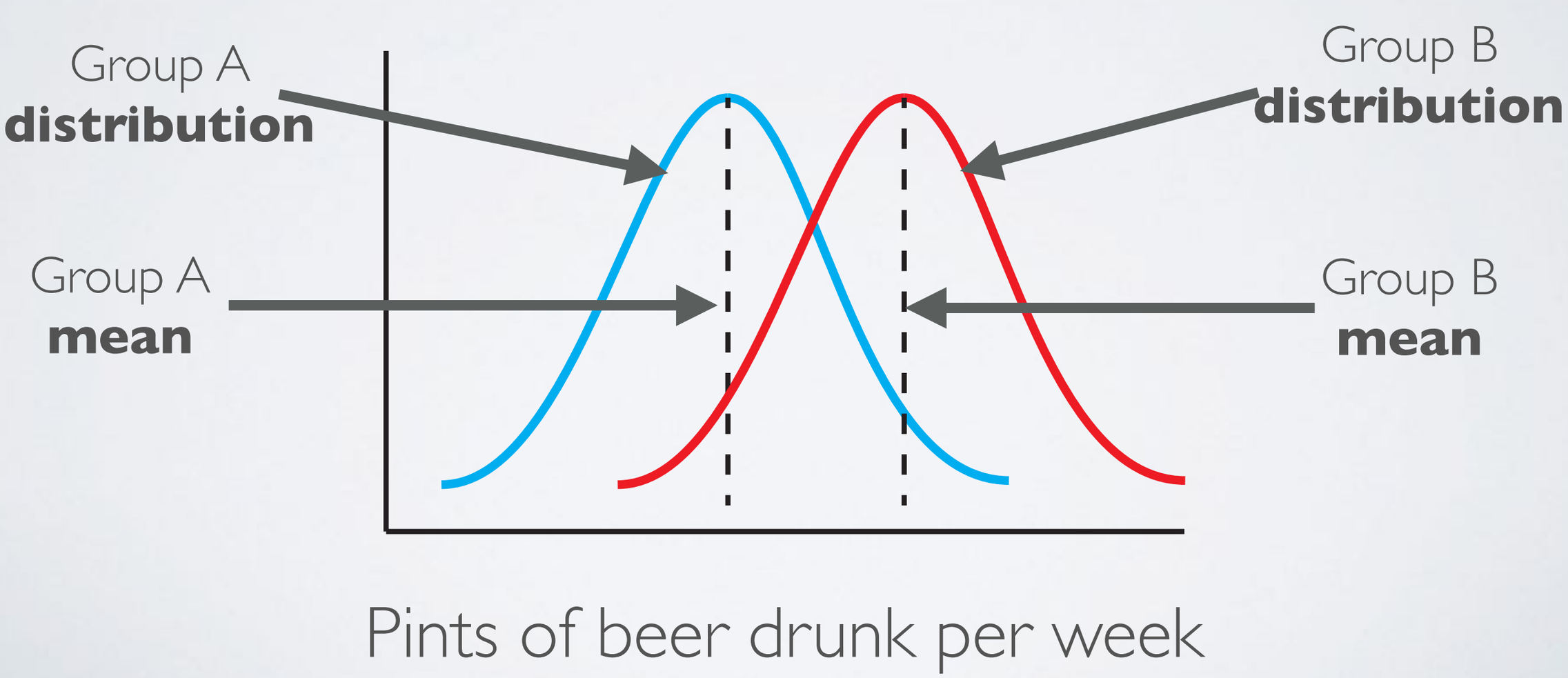
EXAMPLE: T-TEST

- So…. the t-test asks whether the difference is probably representative of a real the real difference between ALL Welsh and English people or
- If it is likely to be a meaningless statistical fluke
- A difference is more likely to be significant if:
    - the difference between the means is large
    - the sample size is large
    - the standard deviation is low

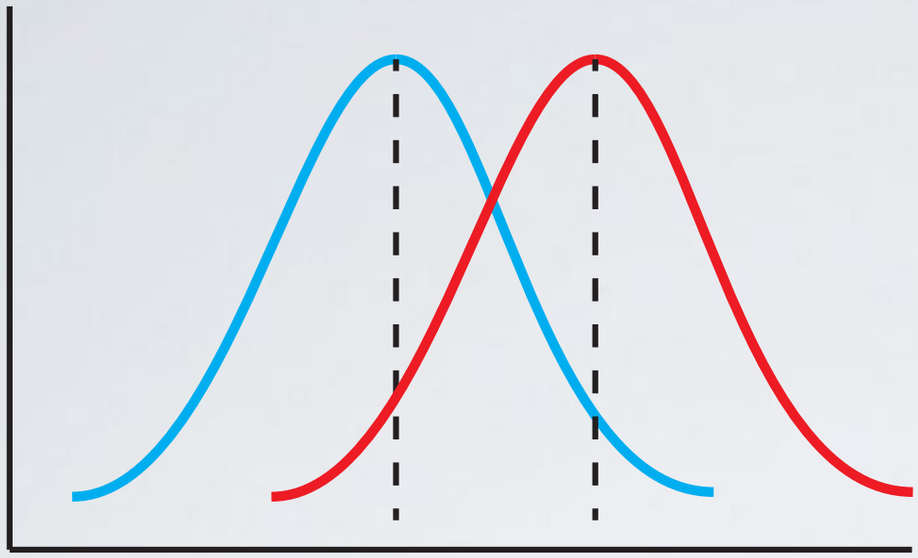
EXAMPLE: T-TEST

GOAL:

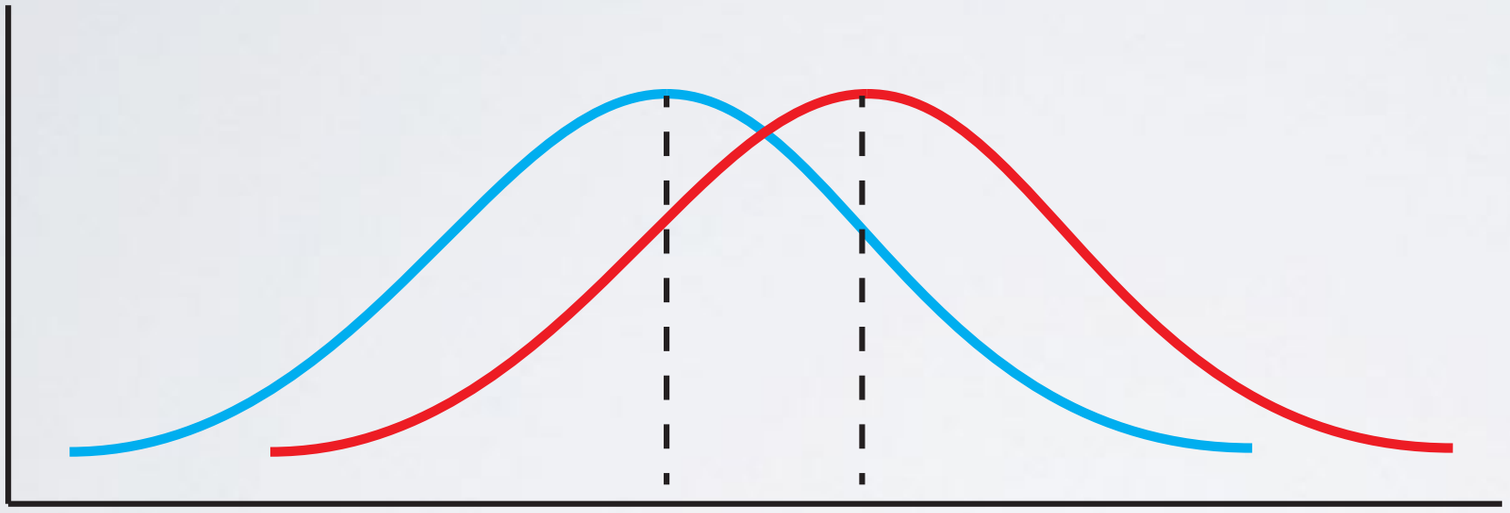
To assess whether the means of two groups are statistically different from each other.



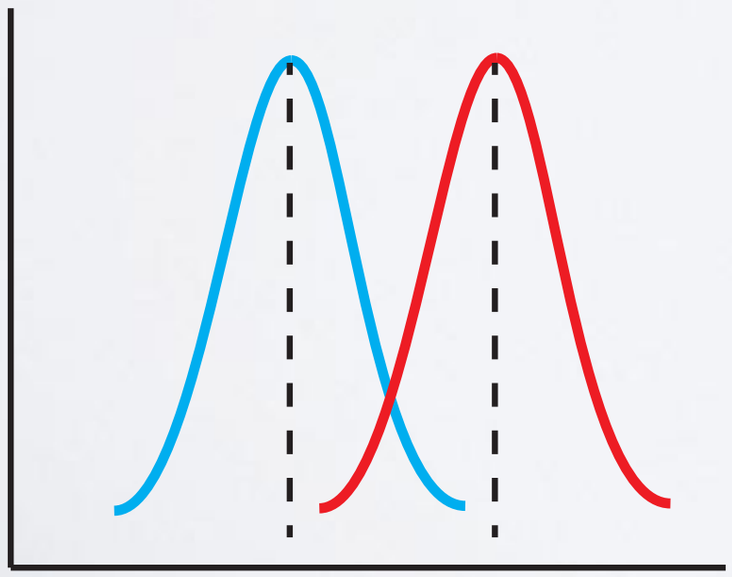
VARIABILITY



medium variability



high variability



low variability

VARIABILITY

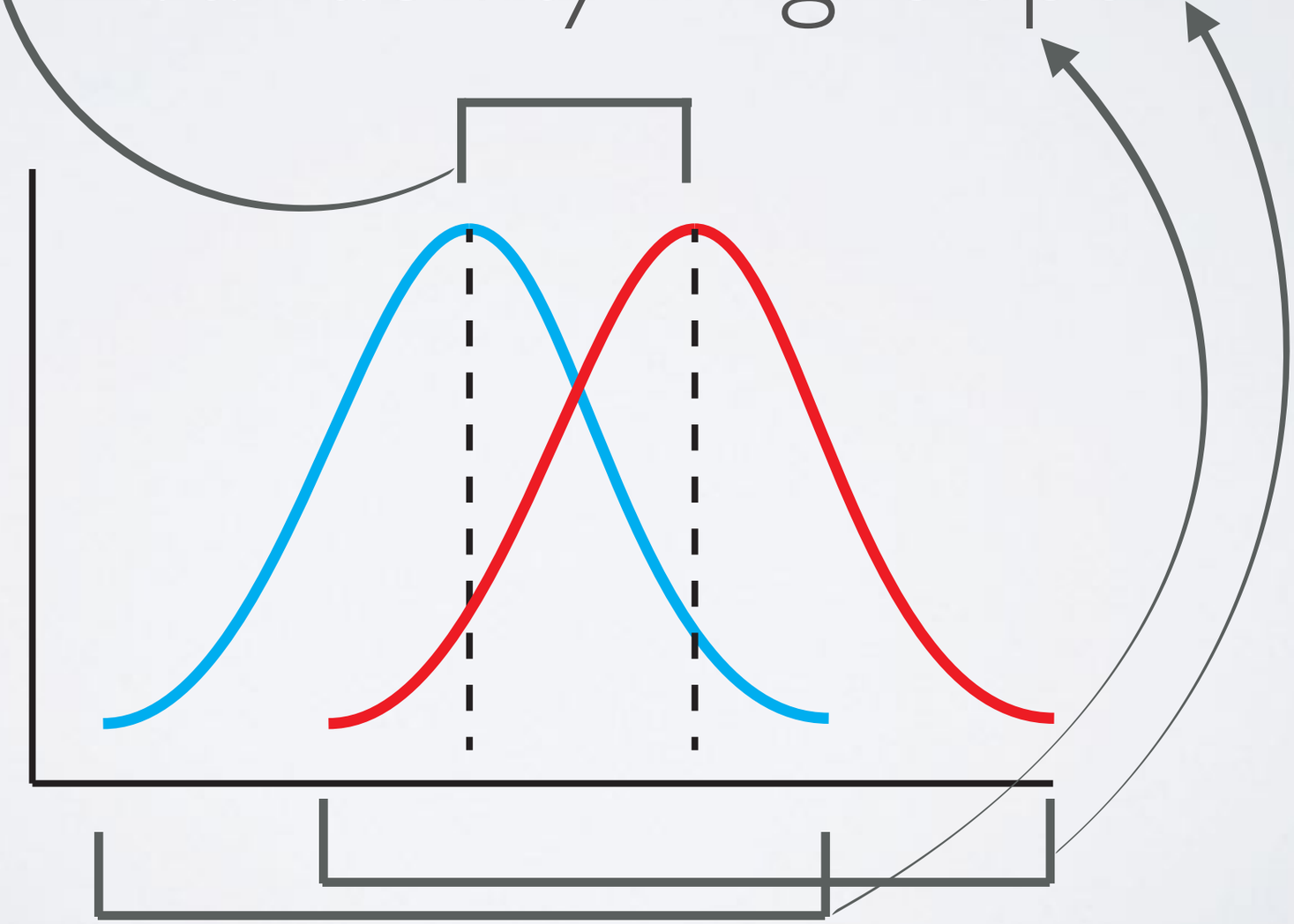
medium variability

- So…
  - When we are looking at the differences between scores between two groups, we must also judge the difference in the means **relative to the spread or variability of their scores**
  - t-test does this

low variability

T-TEST FORMULA

$$t = \frac{\text{difference between group means}}{\text{variability in groups}}$$



T-TEST FORMULA

$$t = \frac{GrA_m - GrB_m}{\sqrt{\frac{GrA_{sd}}{GrA_n} + \frac{GrB_{sd}}{GrB_n}}}$$

GrA_m = Mean of Group A

GrB_m = Mean of Group B

GrA_{sd} = Standard deviation of Group A

GrB_{sd} = Standard deviation of Group B

GrA_n = Total values in Group A

GrB_n = Total values in Group B

ONE- OR TWO-TAILED?

Remember our one and two tailed hypotheses?

One-tailed hypothesis:

System A is faster to complete a task than System B

The Welsh drink more beer in a week than the English

Two-tailed hypothesis:

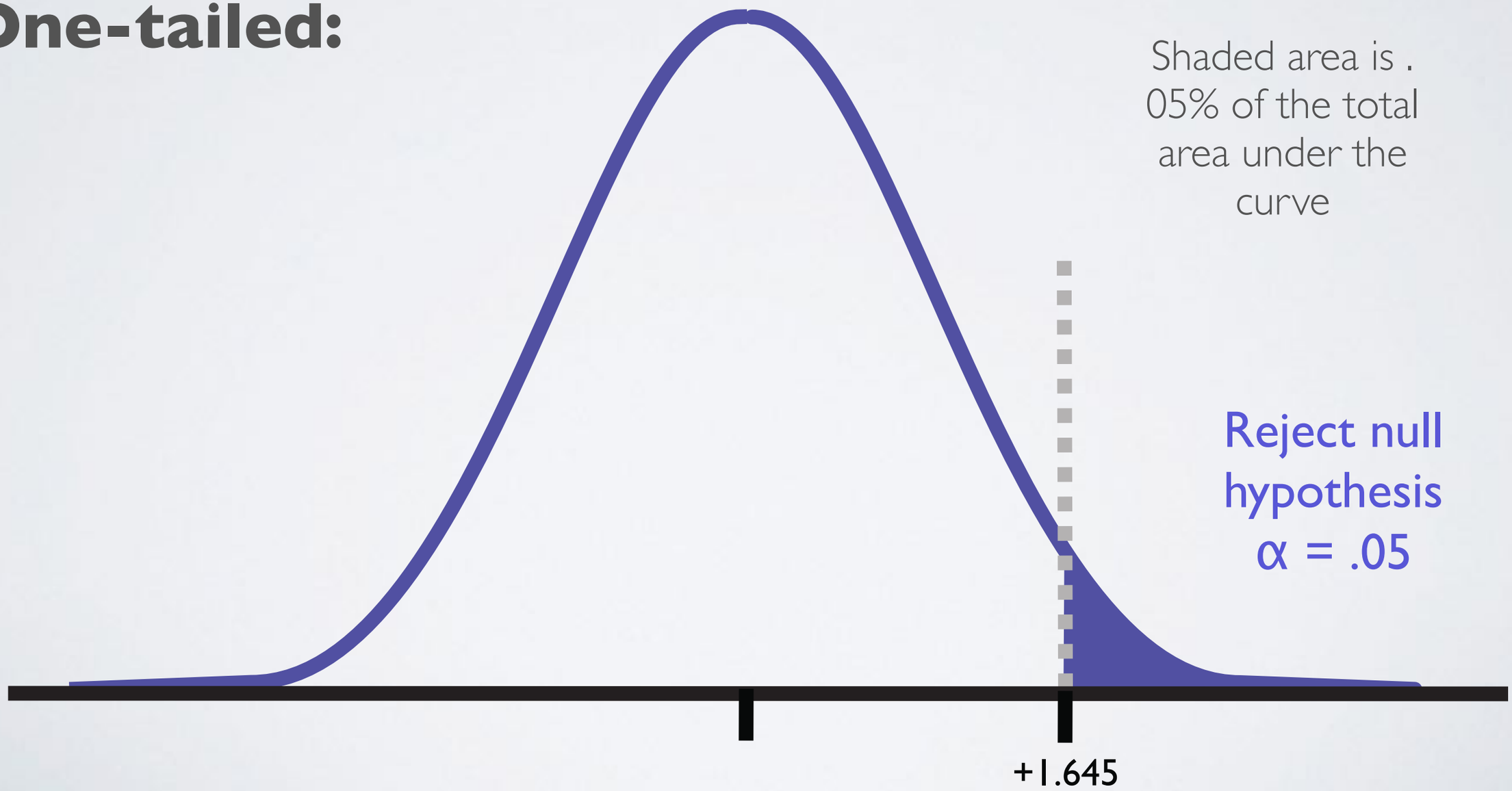
One of the systems will be faster than the other

Either the English or the Welsh will drink more beer in a week than the other

ONE- OR TWO-TAILED?

Significance level (α) = .05

One-tailed:



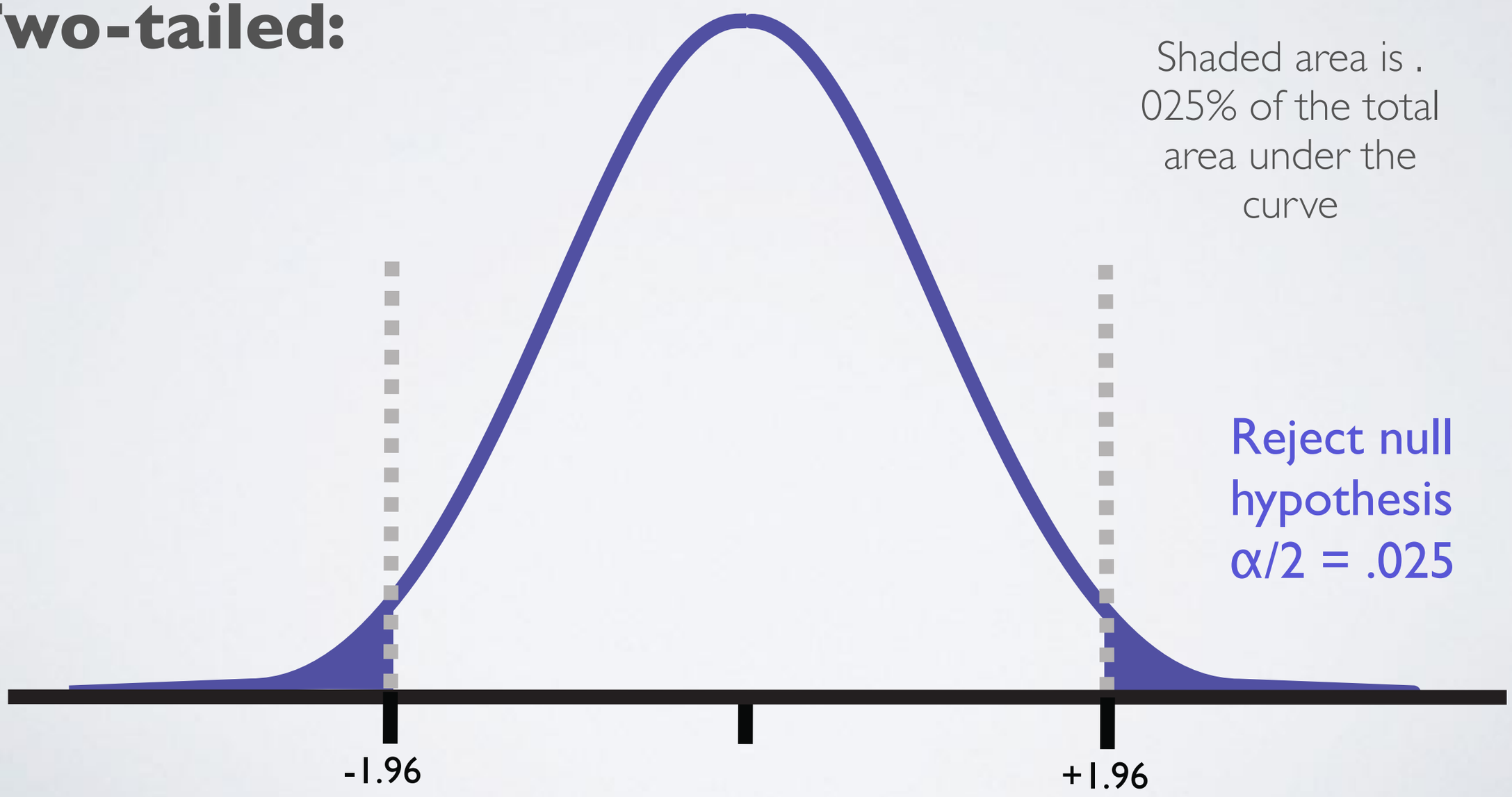
Shaded area is .05% of the total area under the curve

Reject null hypothesis
α = .05

ONE- OR TWO-TAILED?

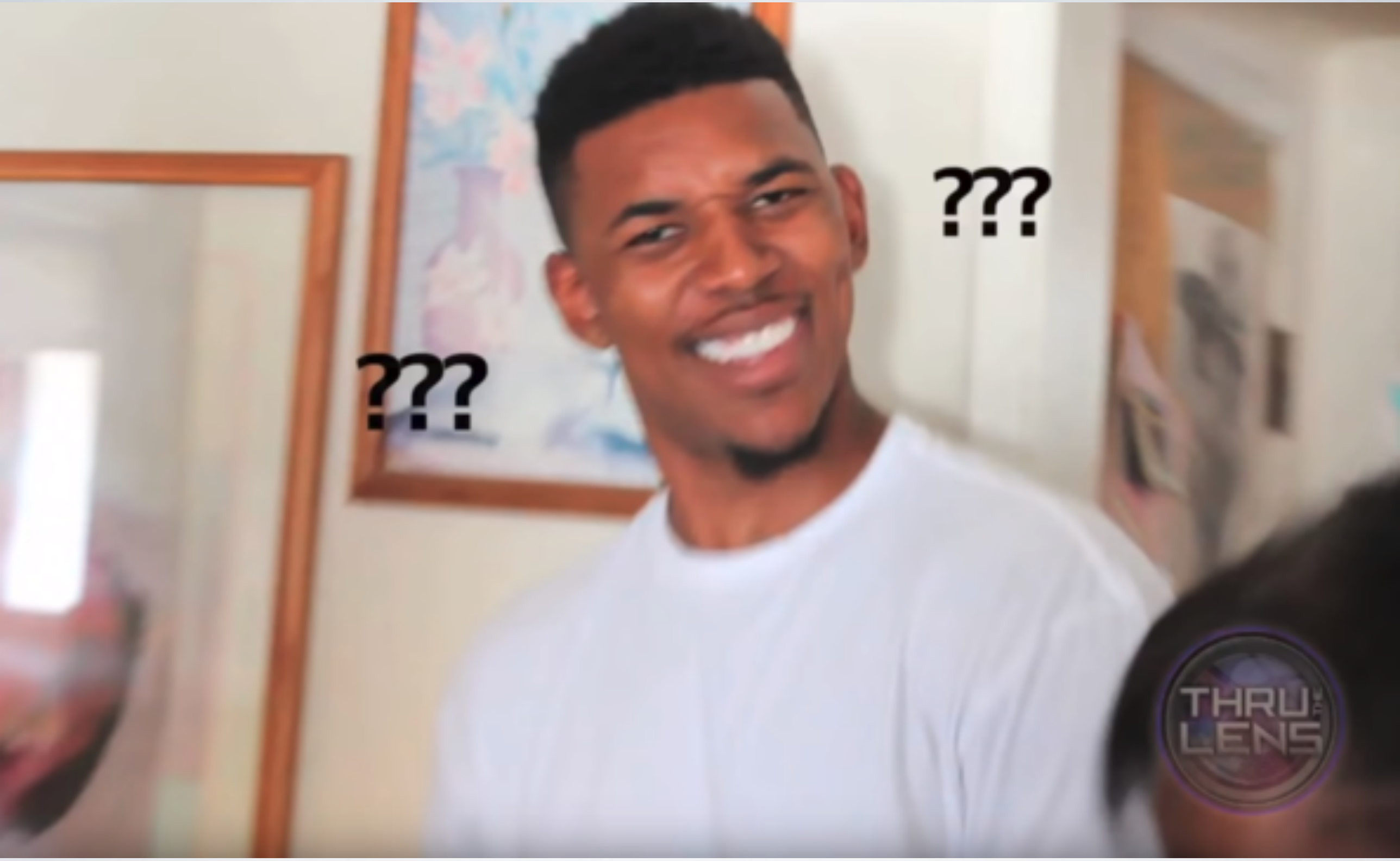
Significance level (**α**) = .05

Two-tailed:



Shaded area is .025% of the total area under the curve

Reject null hypothesis
α/2 = .025

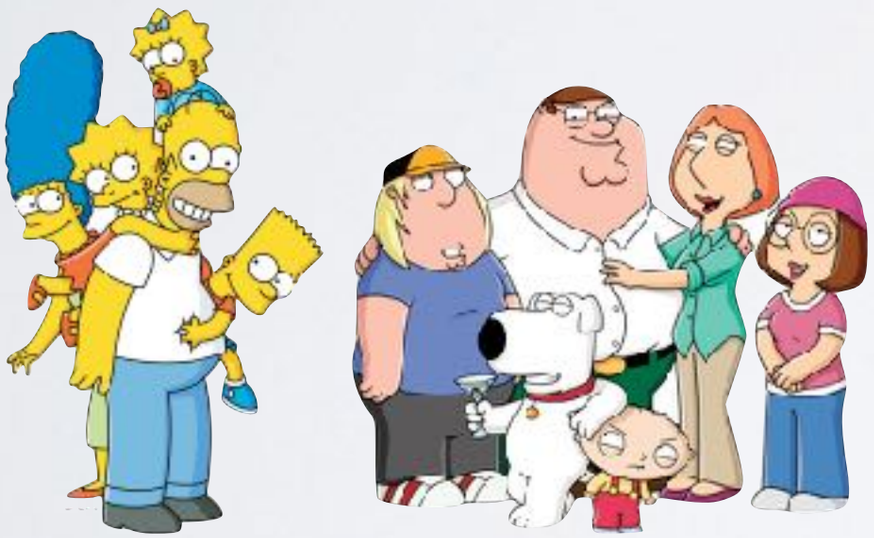


www.youtube.com/watch?v=AGh66ZPpOSQ

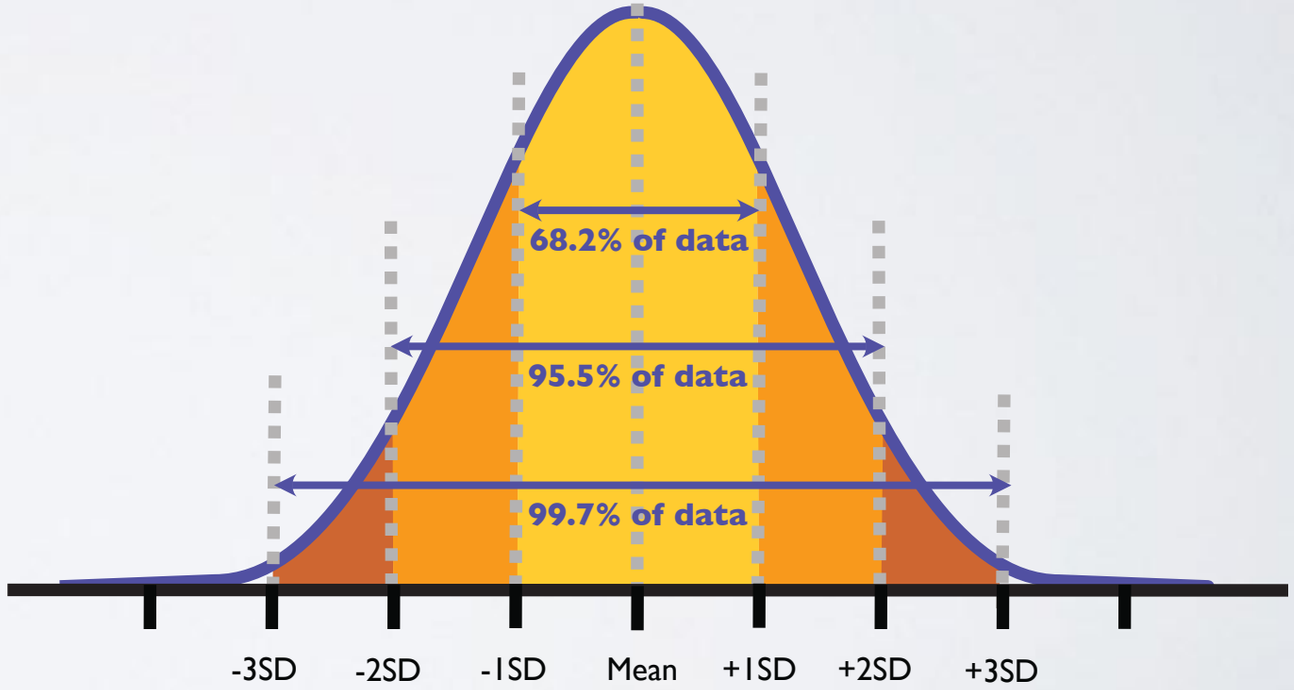
RECAP: UN-PAIRED T-TEST

WHEN TO USE:

- When comparing two un-paired (between-subject) groups from a gaussian (normal) population



between-groups



normal distribution

RECAP: UN-PAIRED T-TEST

ASSUMPTIONS:

- The data is normally distributed
- The data comes from different groups of people (if it came from the same group of people - i.e., within-groups - we would use a *paired* t-test)
- The dependent variable is measured on an incremental level, such as ratios or intervals.
- The independent variables must consist of two related groups or matched pairs.

EXPERIMENT

A SCIENTIFIC EXPERIMENT

- End users with an interactive prototype in a controlled environment
- Goal is to scientifically prove some performance aspect of a design
  - Define research question
    - "Which system is faster to use?" (Quantitative)
    - "Which system is more accurate?" (Quantitative)
    - "Which system is more user friendly?" (Qualitative)
  - Hypotheses
    - "System A is faster than System B"
    - "System A is more accurate than System B"
    - "System A is more user friendly than System B"
  - Identify dependent and independent variables
    - Dependent: time taken to complete task, errors made
    - Independent: system type

A SCIENTIFIC EXPERIMENT

- Select number and type of users
    - Between groups – one group uses System A the other group System B
    - Within groups – all participants do all conditions (requires counterbalancing)
- Define tasks
    - Access representative set of functions (repeat
        - e.g., perform input task on System A and on System B (repeat n times)
- Gather data
    - Pre-study data gathering (e.g., demographics, experience, current opinions)
    - Logging, timing, observation etc.
    - Post-study data gathering (e.g., qualitative questioning, quantitative scoring)
- Analyse results
    - Cluster qualitative data - find patterns
    - Carry out statistical tests on quantitative data to see if differences in performance are due to chance or something actually happening

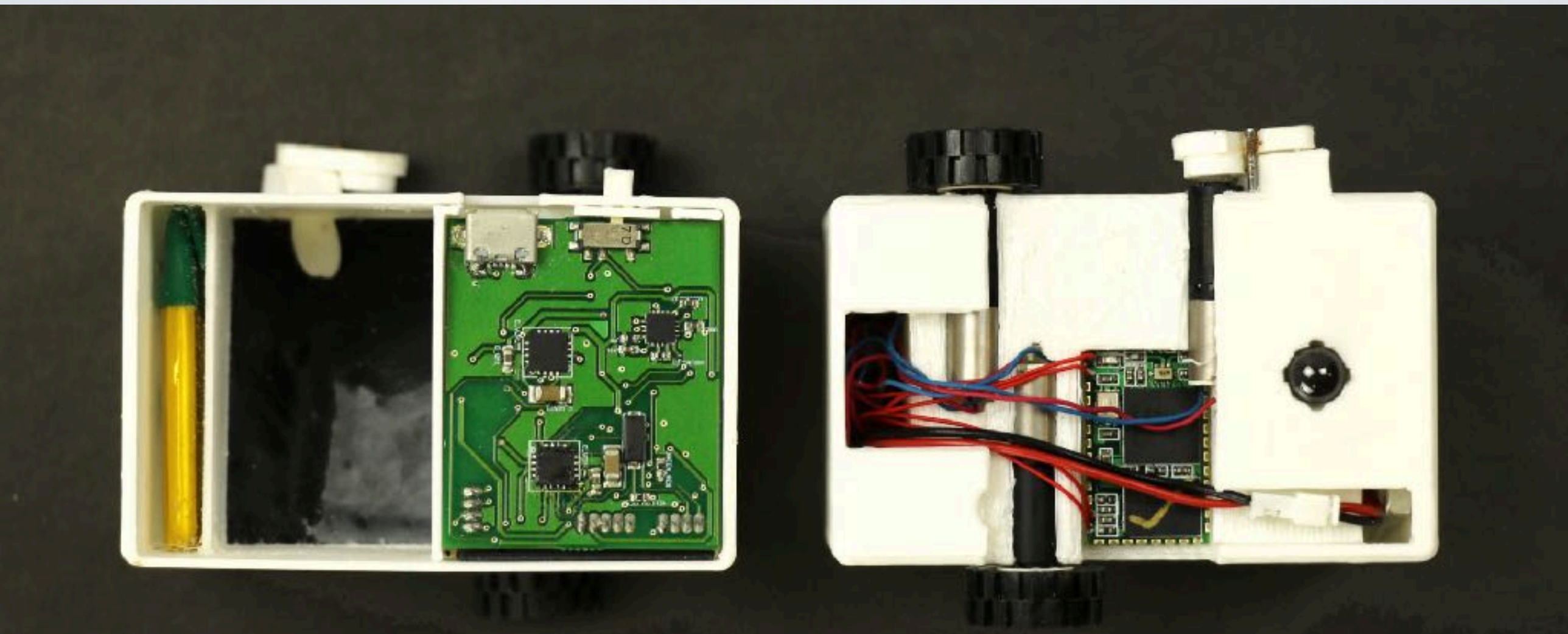
LABORATORY STUDY

Steps

1. Welcome and information to participants
2. Ethics and consent
3. Pre-study data capture
4. Tasks and measures
5. Post-study data capture
6. Incentive payment

EXAMPLE: SUSTAINABOT

EXAMPLE: SUSTAINABOT



EXAMPLE: SUSTAINABOT

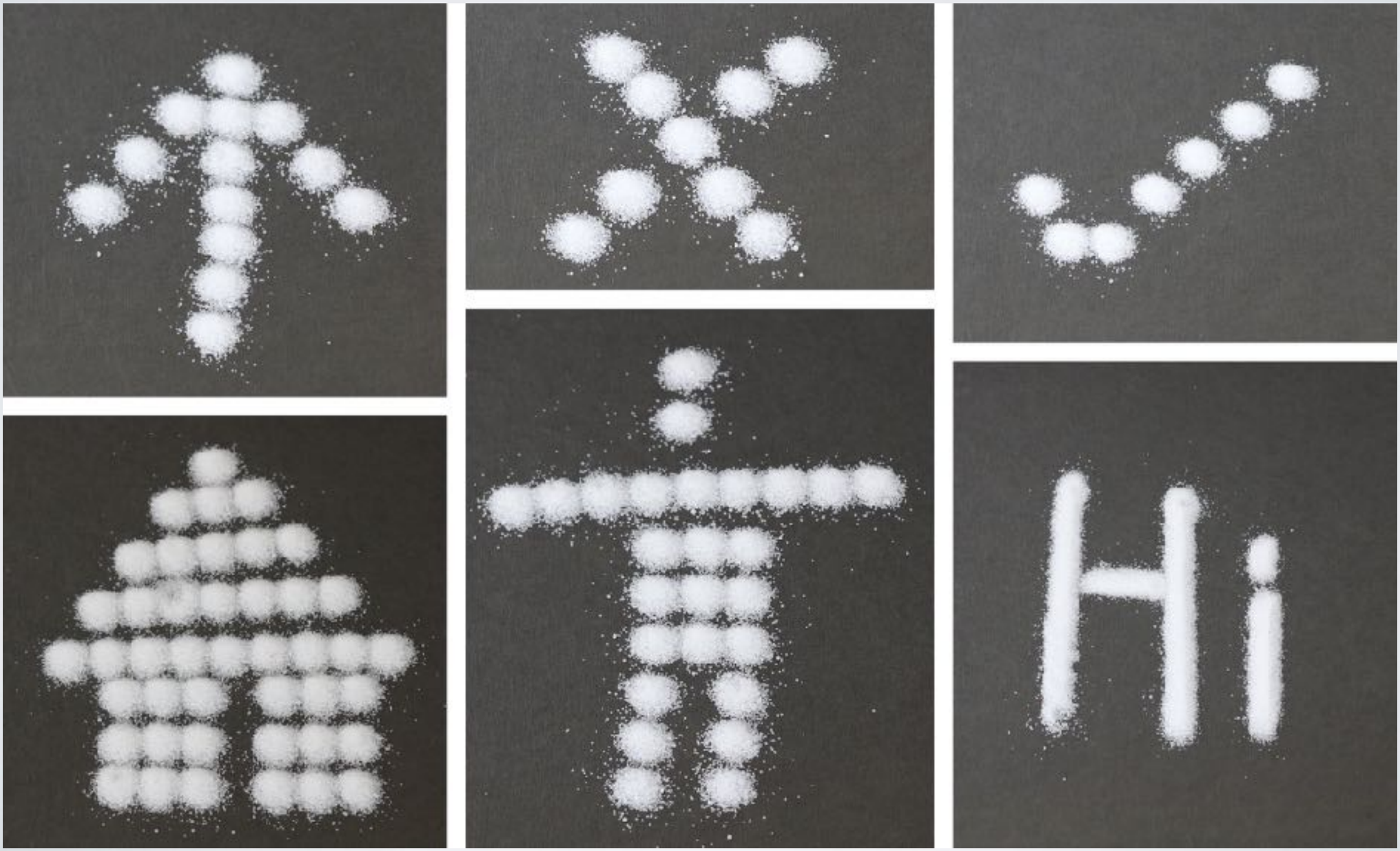
Sustainabot – Exploring the Use of Everyday Foodstuffs as Output and Input for and with Emergent Users

Simon Robinson, Jennifer Pearson,
Mark D. Holton, Shashank Ahire, Matt Jones

FIT Lab, Computational Foundry, Swansea University
Industrial Design Centre, IIT Bombay

CHI 2019

EXAMPLE: SUSTAINABOT



EXAMPLE: SUSTAINABOT

- Setting: Lab environment
- Participants: 23 emergent users (India) / 144 UK users
- Incentive: ₹500 / £5
- Time: ~30 mins
- Dependent variable: target users
- Study Type: between-groups

EXAMPLE: SUSTAINABOT

- Research questions:
    - "How well can users recognise the images drawn by Sustainabot?"
    - "What situations in which Sustainabot could be useful to users?"
    - "What types of content might users want to print?"
    - "Where might Sustainabot be used?"
    - "What materials might be used"
    - "What potential problems could be identified?"

EXAMPLE: SUSTAINABOT

Procedure:

- Welcome / ethics
- Pre study demographics
- Concept video / explanation of purpose of device
- Standard image demonstration
- Tasks (x5)
    - Show a video of Sustainabot "printing" an image and ask participants to:
        - State what they think each one is, and
        - Rate each one out of 10 for how recognisable it is
- Post study interview

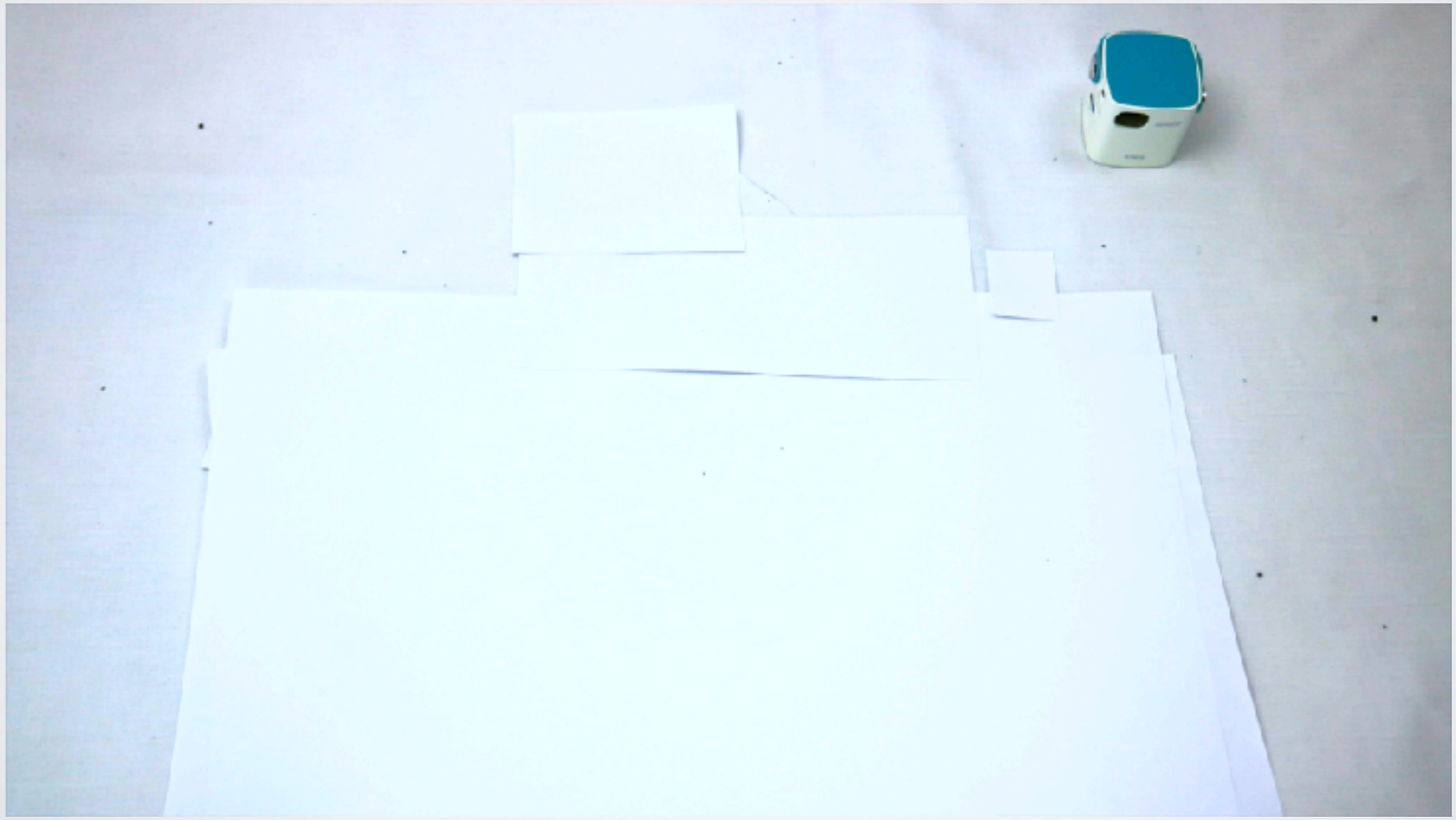
EXAMPLE: SUSTAINABOT

Demographics / prior experience questions:

- Age
- Gender
- Occupation
- Phone ownership
- TV / projection ownership (and size if appropriate)
- Computer ownership
- Rangoli experience
- Have they ever wanted to show something large but not been able to?

EXAMPLE: SUSTAINABOT

Concept video (a mock-up that doesn't actually work, but demonstrates how end product might work)



EXAMPLE: SUSTAINABOT

Post study analysis:

- Cluster main qualitative points into groups to answer research questions
- Determine % of image recognition according to image
- Get average rating for each image + perform statistical analysis.

EXAMPLE: SUSTAINABOT

Quantitative results:

	India		UK	
	Correct (%)	Rating	Correct (%)	Rating
Tick	65	5.4	78	5.5
Arrow	91	6.5	100	6.7
Person	83	5.0	96	5.3
Face	100	5.6	97	5.1
House	100	6.3	97	5.5

Plus lots of comments that help us answer research questions…

IN CASE YOU'RE INTERESTED…

Sustainabot – Exploring the Use of Everyday Foodstuffs as Output and Input for and with Emergent Users

Simon Robinson
FIT Lab, Computational Foundry
Swansea University, Swansea, UK
s.n.w.robinson@swansea.ac.uk

Jennifer Pearson
FIT Lab, Computational Foundry
Swansea University, Swansea, UK
j.pearson@swansea.ac.uk

Mark D. Holton
Computational Foundry
Swansea University, Swansea, UK
m.d.holton@swansea.ac.uk

Shashank Ahire
Industrial Design Centre
IIT Bombay, Mumbai, India
ahire.shashank@iitb.ac.in

Matt Jones
FIT Lab, Computational Foundry
Swansea University, Swansea, UK
matt.jones@swansea.ac.uk

ABSTRACT

Mainstream digital interactions are spread over a plethora of devices and form-factors, from mobiles to laptops; printouts to large screens. For emergent users, however, such abundance of choice is rarely accessible or affordable. In particular, viewing mobile content on a larger screen, or printing out copies, is often not available. In this paper we present Sustainabot – a small robot printer that uses everyday materials to print shapes and patterns from mobile phones. Sustainabot was proposed and developed by and with emergent users through a series of co-creation workshops. We begin by discussing this process, then detail the open-source mobile printer prototype. We carried out two evaluations of Sustainabot, the first focused on printing with materials in situ, and the second on understandability of its output. We present these results, and discuss opportunities and challenges for similar developments. We conclude by highlighting where and how similar devices could be used in future.

CCS CONCEPTS

• **Human-centered computing** → Interaction techniques; • **Computer systems organization** → Robotics; Embedded systems.

KEYWORDS

Robots, everyday materials, interaction, emergent users.

ACM Reference Format:

Simon Robinson, Jennifer Pearson, Mark D. Holton, Shashank Ahire, and Matt Jones. 2019. Sustainabot – Exploring the Use of Everyday Foodstuffs as Output and Input for and with Emergent Users. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3290605.3300456

1 INTRODUCTION

PARTICIPANTS

- Pick a representative population
    - e.g., computer-literate adults, people with poor eye-sight, children, left-handed people, lower-literate etc.
- To be suitable, participants must be:
    - Members of the desired population
    - Selected at random from the population (not your friends!)
- How many?