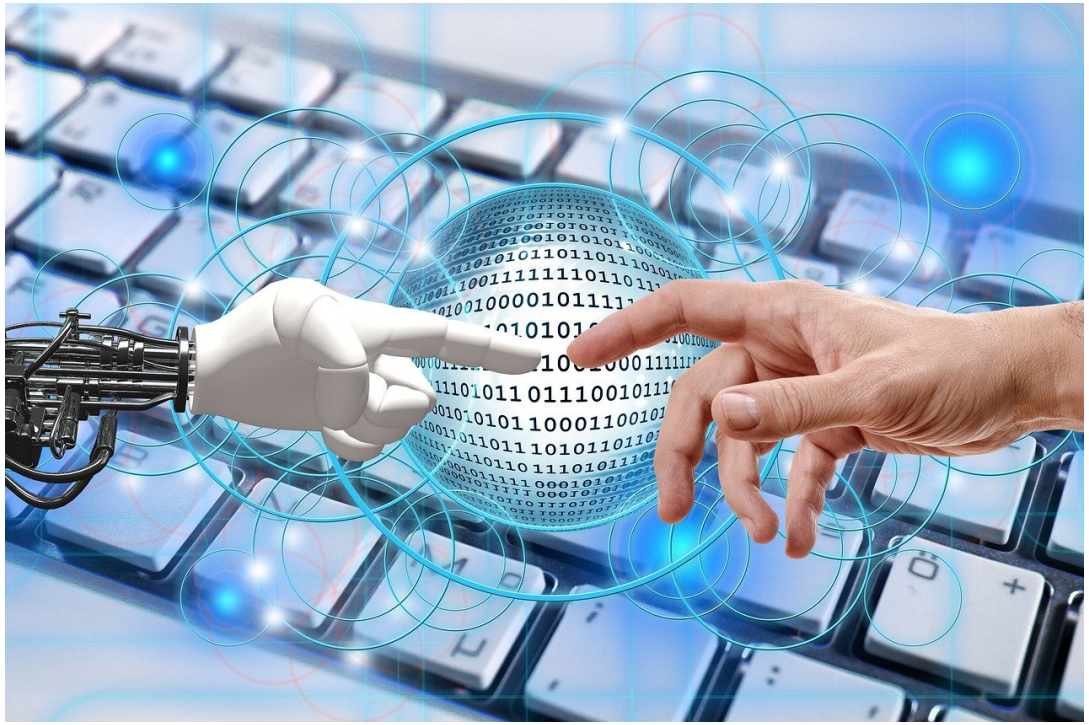
# Power Analysis



Prof. Dr. Teena Hassan
teena.hassan@h-brs.de

Department of Computer Science
Hochschule Bonn-Rhein-Sieg
Sankt Augustin

27 June 2024

A²S

# Learning Goals

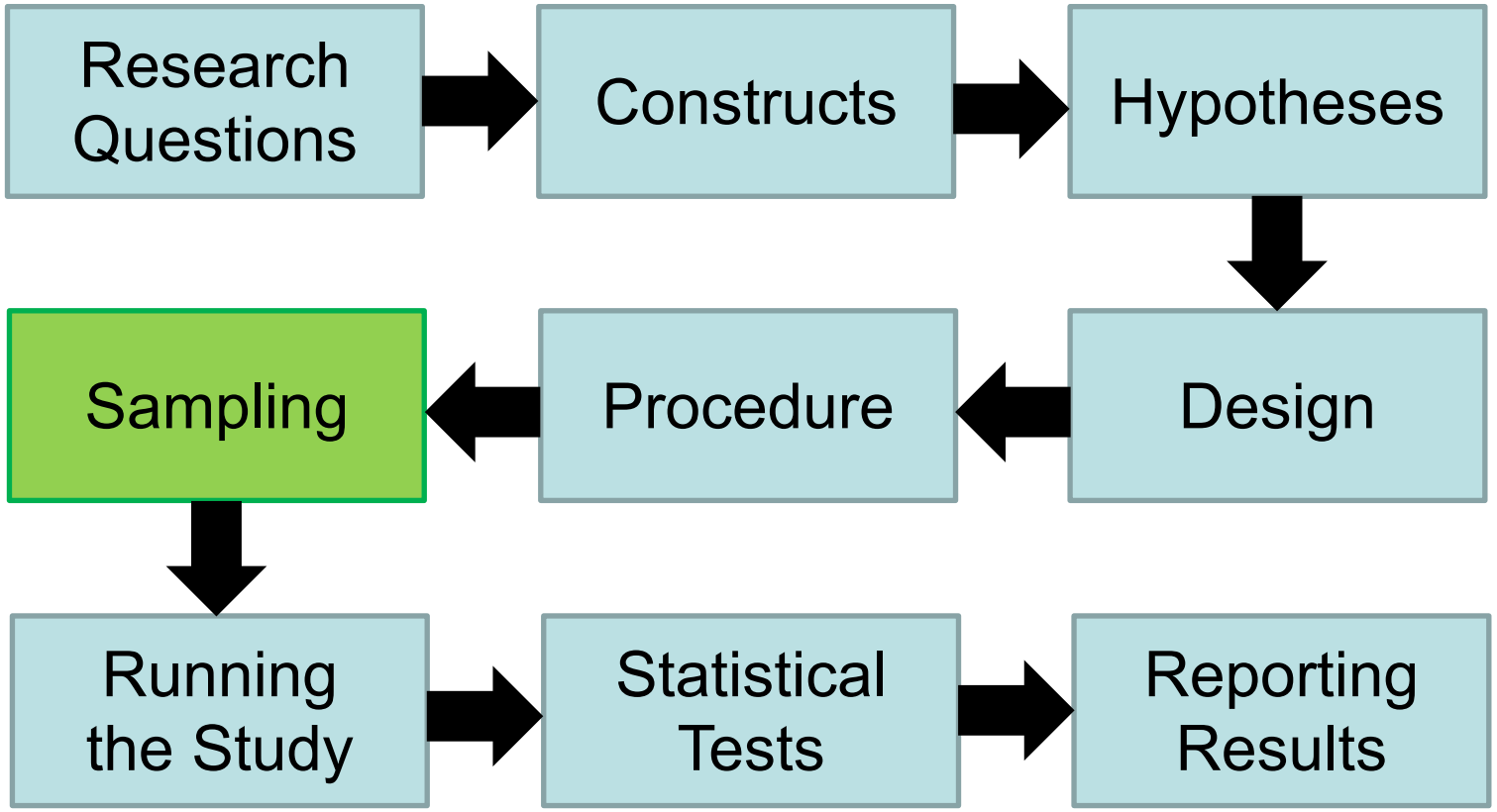
- At the end of today's lecture, you will be able to:
  1. Identify the factors that influence the **sample size** of participants required for an experiment.
  2. Apply **power analysis** to determine sample size for validating hypotheses in HRI experiments.

# Primary Reference

- Guy Hoffman and Xuan Zhao. 2020. A Primer for Conducting Experiments in Human Robot Interaction ACM Trans. Hum. Robot Interact . 10, 1, Article 6 (October 2020), 31 pages https://doi.org/10.1145/3412374

- Your reading assignment for last week!

# Hypothesis-Driven Experimental Research

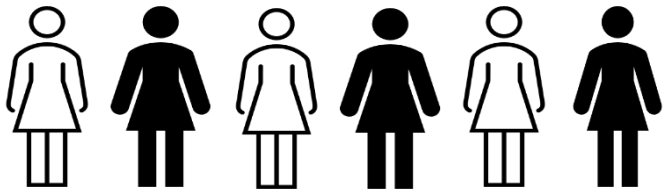



(Hoffman & Zhao, 2020)

# Make a Guess...

➢Which design would require more samples to achieve the same statistical power?

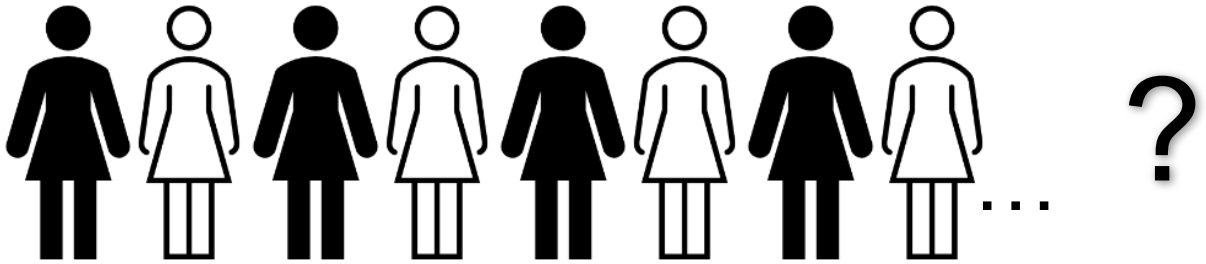
A. Within-subjects design



B. Between-subjects design

# How Many Samples are Enough?

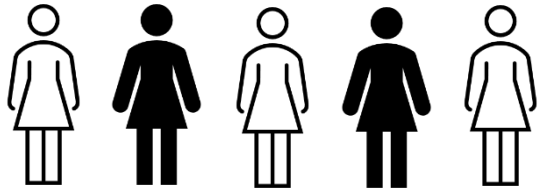
- Very crucial question in empirical research.
- How many samples are needed to draw scientifically sound conclusions?



# We Need…

## Power Analysis

Simply put, a **statistically sound** method to determine **the sample size** and **power** of an experiment.



Sample size

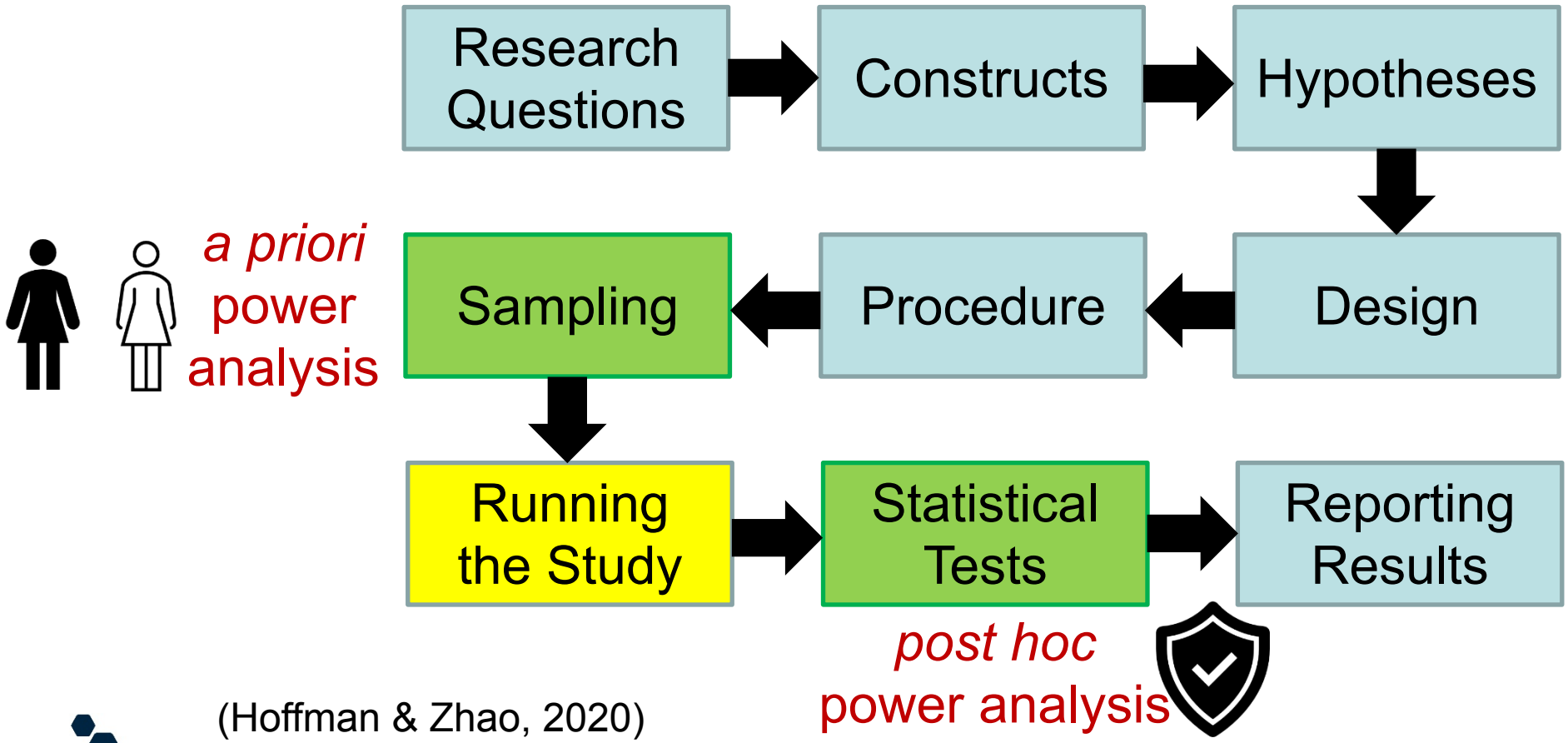
Power

# Why is Power Analysis Important in HRI?

- Humans are **unpredictable** and **exhibit variability** in their behavior.
  - Inter- and intrapersonal variability
- When we conduct an experiment:
  - How can we be sure that the **effects that we observed are meaningful** and not noise?
- **We should perform a priori and post hoc power analyses.**

(Bartlett et al., 2022)

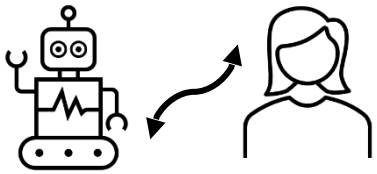
# Hypothesis-driven Experimental Research




(Hoffman & Zhao, 2020)

# Hypothesis H1 Revisited

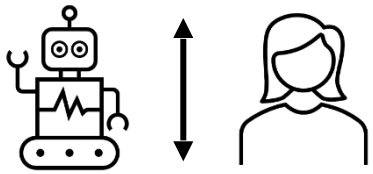
**Level of users‘ trust in:**



**Cond-A**

>

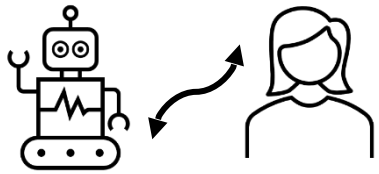
**Level of users‘ trust in:**



**Cond-B**

# The Null Hypothesis H0

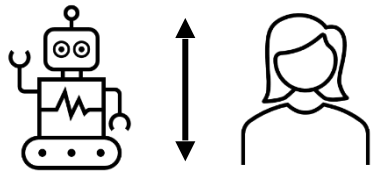
**Level of users‘ trust in:**



**Cond-A**

=

**Level of users‘ trust in:**



**Cond-B**

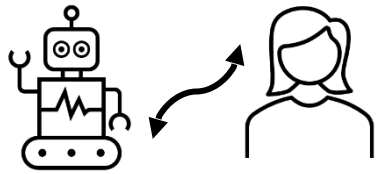
➢ **H0** posits there is no difference between users‘ level of trust in cond-A and cond-B.

Null Hypothesis

# Objective of the Experiment

- **Can we reject H0:**

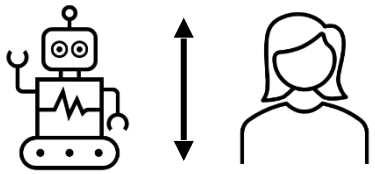
**Level of users‘ trust in:**



**Cond-A**

=

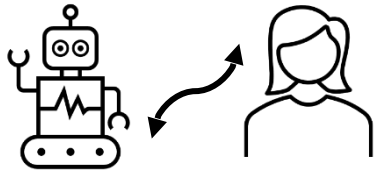
**Level of users‘ trust in:**



**Cond-B**

- **And accept H1:**

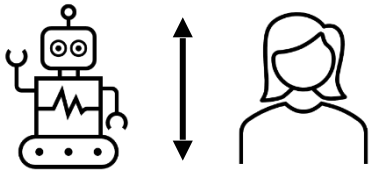
**Level of users‘ trust in:**



**Cond-A**

>

**Level of users‘ trust in:**

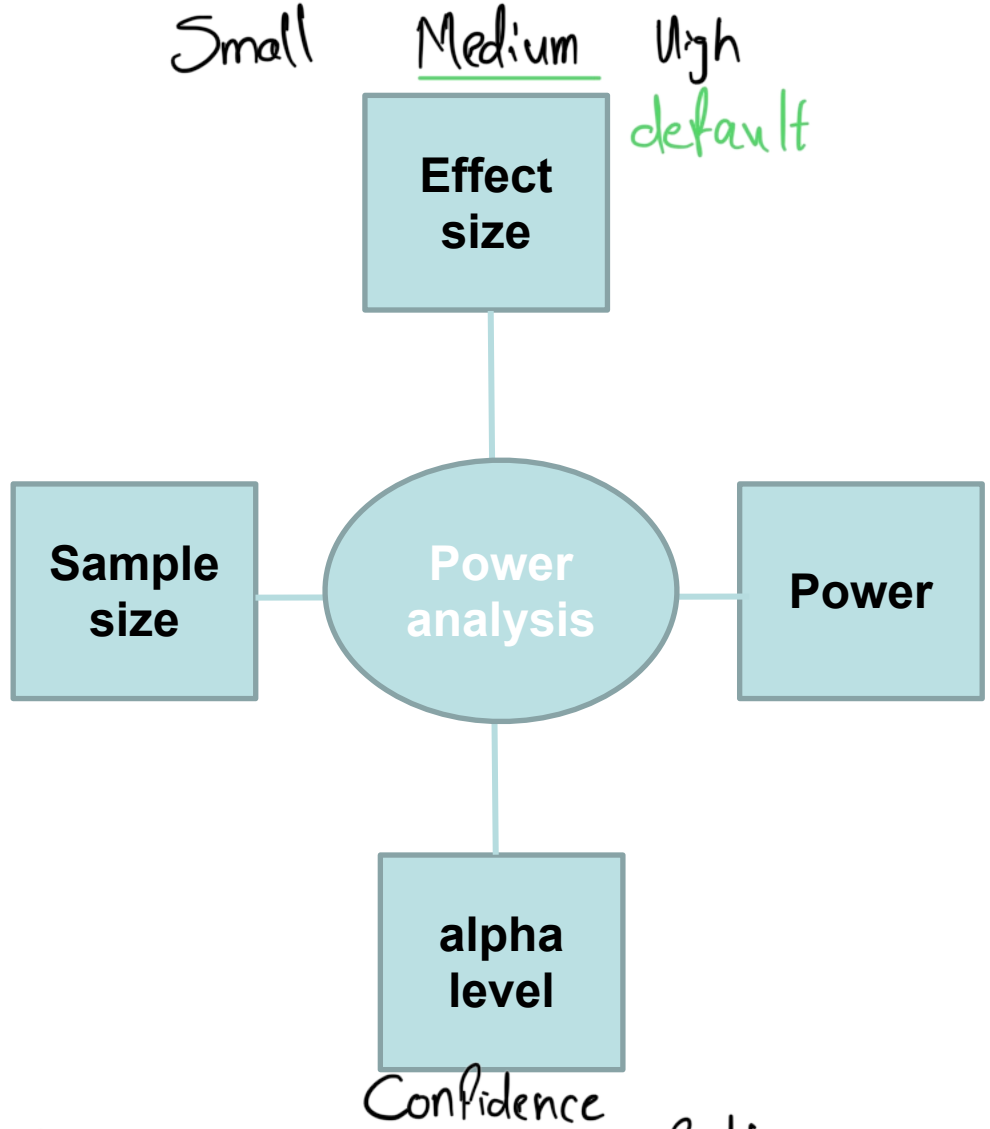


**Cond-B**

# Power Analysis

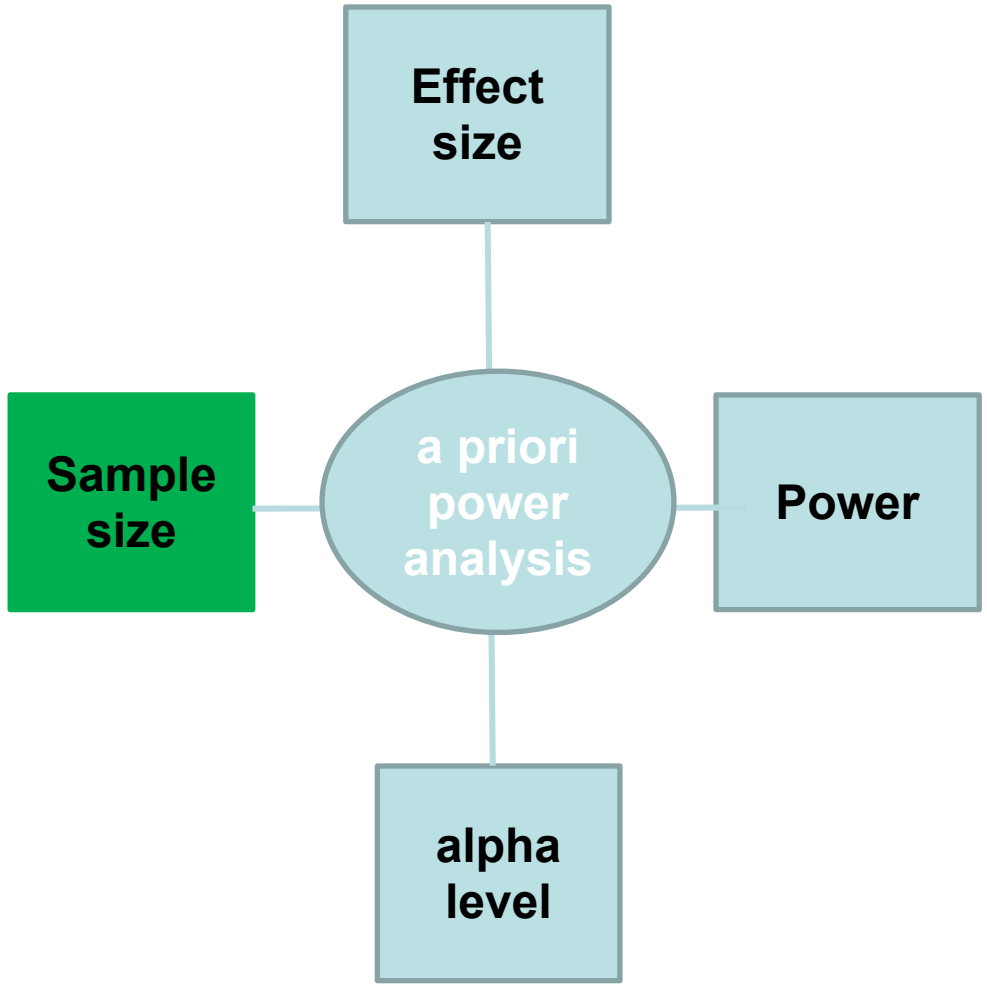
- Examines the relationship between four parameters:
  1. alpha level
  2. Power
  3. Effect size
  4. Sample size

- Given any three of these parameters, we can predict the fourth.



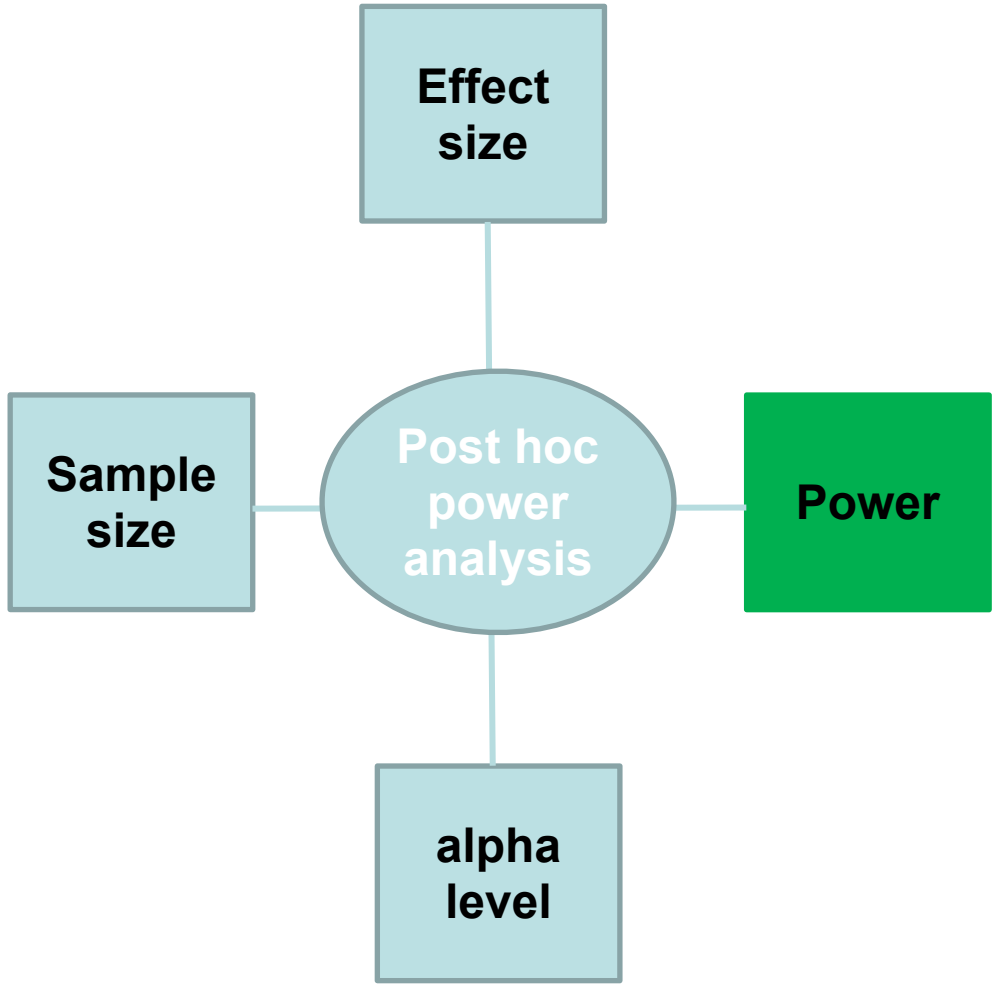
# *a priori* Power Analysis

- Given:
  - Expected or desired values for:
    1. alpha level
    2. Power
    3. Effect size
- Predict, sample size needed for the experiment.

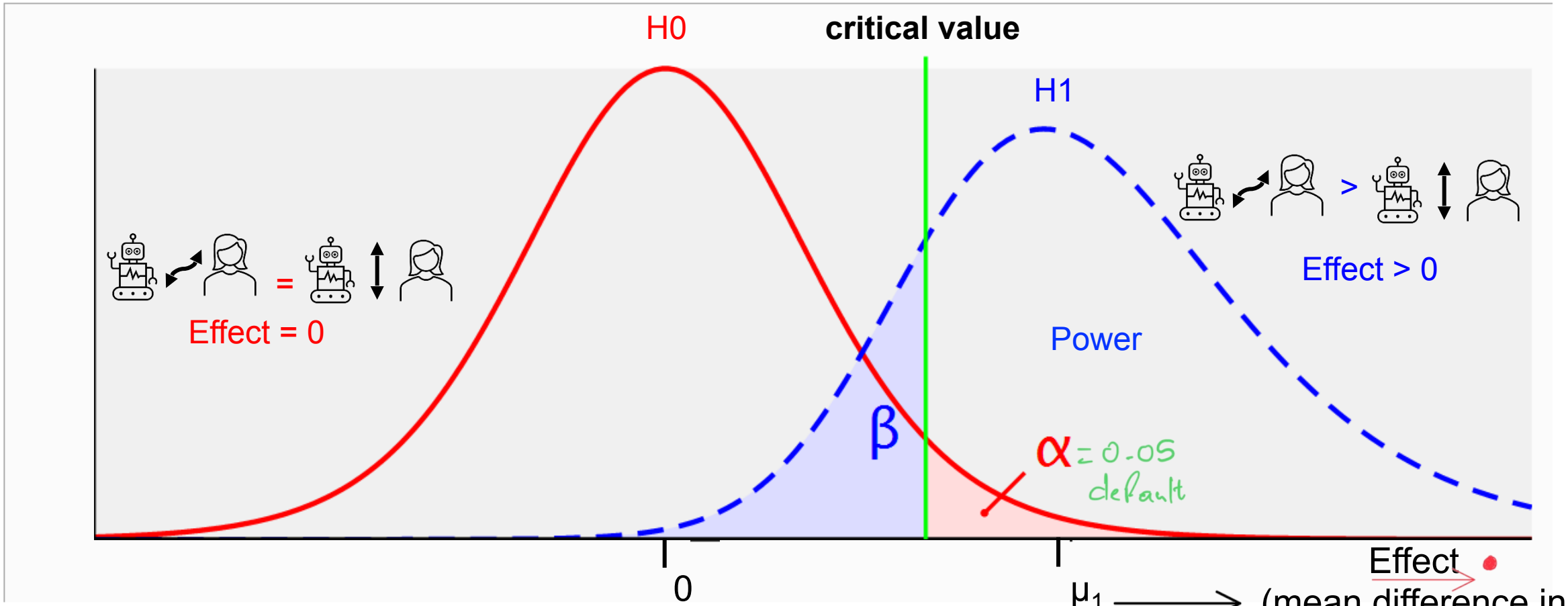


# *post hoc* Power Analysis

- Done after the experiment.
- Given:
  - Values computed from experimental data:
    1. Effect size
    2. Sample size
    3. alpha level
- Predict statistical power of the experiment.



# Alpha Level, Power, Effect Size




Screenshot from G*Power Software

To get the same power after reducing alpha, we should increase the sample size.

# Errors Induced by Sampling – Type I, Type II

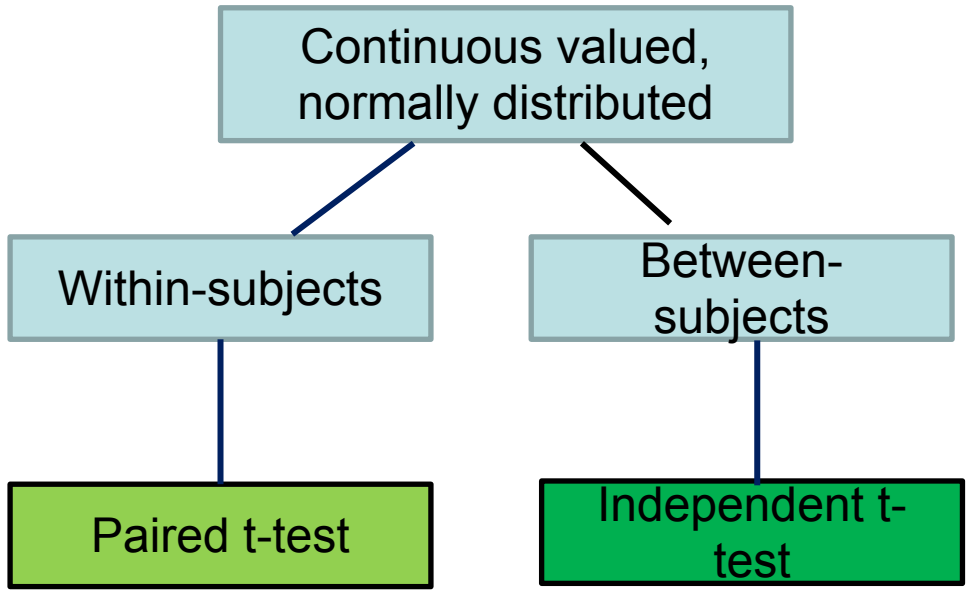
| | **Reject H0<br>Accept H1** | **Accept H0<br>Reject H1** |
|---|---|---|
| **H0 is True<br>H1 is False** | α<br>False positive rate<br>(Type I error) | 1- α<br>True negative rate |
| **H0 is False<br>H1 is True** | 1- β<br>True positive rate<br>**(POWER)** | β<br>False negative rate<br>(Type II error) |

Holds in the general population

Holds in the sampled population

# Estimating Effect Size

- **Effect size is the difference between outcomes measured in different conditions.**
- Effect size computation depends on the method chosen for the statistical test.



# Computing Effects Using t-tests

- Independent t-test
  - Between subject
    - Two separate groups of subjects A and B
    - Each group exposed to only one condition
    - Two sampling distributions
  - Computes means of dependent variable in either distribution: $\mu_A$, $\mu_B$
  - Difference between means: $\Delta = \mu_A - \mu_B$
  - Effect: Compares $\Delta$ to zero.
  - Effect size (Cohen's d): $\Delta$ / $\sigma$pooled

$$\sigma_{pooled} = \sqrt{\left(\sum(a - \mu_A)^2 + \sum(b - \mu_B)^2\right) / (n_A + n_B - 2)}$$

- Paired t-test
  - Within subject
    - One group of subjects
    - Each subject exposed to both conditions A and B
  - One sampling distribution S based on difference $\Delta_s$ in response of each subject s in condition A and condition B.
    - $\Delta_s = a_s - b_s$
    - $\mu_S$= Mean $\Delta_s$ over all s
  - Effect: Compares $\mu_S$ to zero.
  - Effect size (Cohen's d): $\mu_S / \sigma_S$

# *a priori* Power Analysis

- Commonly used values:
  - alpha: 0.05 or 0.01
  - power: 0.8 (HRI), 0.95 (clinical)
  - effect-size (for t-tests): 0.5

| | Small effect size | Medium effect size | Large effect size |
|---|---|---|---|
| **Cohen's d** | 0.2 | 0.5 | 0.8 |

Jacob Cohen. 1992. A power primer. Psychol. Bull. 112, 1 (1992).

# Conclusion

- In this lecture, you learned to:
  1. Identify the factors that influence the **sample size** of participants required for an experiment.
  2. Apply **power analysis** to determine sample size for validating hypotheses in HRI experiments.

# References


- Guy Hoffman and Xuan Zhao. 2020. **A Primer for Conducting Experiments in Human–Robot Interaction**. ACM Trans. Hum.-Robot Interact. 10, 1, Article 6 (October 2020), 31 pages. https://doi.org/10.1145/3412374

- Madeleine E. Bartlett, C. E. R. Edmunds, Tony Belpaeme, and Serge Thill. 2022. **Have I Got the Power? Analysing and Reporting Statistical Power in HRI.** *J. Hum.-Robot Interact.* 11, 2, Article 16 (June 2022), 16 pages. https://doi.org/10.1145/3495246

# Next Week

- Lecture on “Statistical Tests"
  - Thursday, 4.07.2024 at 9 am in B060, Sankt Augustin