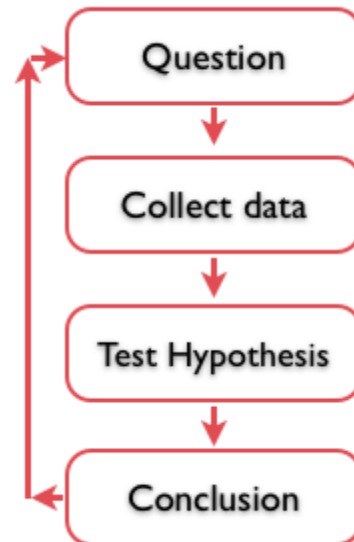# Human Subject Studies: Logistics, Design, Evaluation

The Research Process

1. The Research Problem

- Identify the research problem/question
  - Literature review
  - Open questions left unanswered by previous researchers

1. The Research Problem

- Defining the research problem helps you to formulate a **research hypothesis**, which is tested against the **null hypothesis**.

  - **Research hypothesis:** is the statement created by researchers when they speculate upon the outcome of a research or experiment.

  - **Null hypothesis:** a statement which the researcher tries to disprove, reject or nullify.

2. Data Collection

- Preparation: Make Hypothesis Testable (Operationalization)
- Preparation: Design the Study
- Conduct the Experiment or Observation

3. Test Hypothesis

- Organize the Data

- Analyze the Results

- Check if the Results Support your Hypothesis

4. Conclusion

- Look for Other Possible Explanations
- Generalize to the Real World
- Suggestions to Further Research

Evaluation Metrics

- Many possible metrics, dependent on what is being studied:
  - Percentage of tasks successfully completed
  - Accuracy
  - Time required for completion
  - Number of errors that occurred
  - Deviation from planned route/behavior/trajectory
  - Number of successfully completed subtasks
  - Fan out: number of robots controlled by a single human
  - Operator response time
  - Level of robot autonomy
  - Operator workload or effort
  - Accuracy of user mental model of the task/device
  - Level of operator trust in the system
  - …..

Subjective measures common in surveys and observations

- **Likert Scale:**

Strongly Disagree	Disagree	Undecided	Agree	Strongly Agree
(1)	(2)	(3)	(4)	(5)

- **Task workload:** NASA-TLX is a subjective workload assessment tool that allows developers to perform subjective workload assessments on operator(s) working with various human-machine systems.

- **Trust:** The Working Alliance Inventory is a questionnaire commonly used in therapy and other helping relationships that tracks trust and belief in a common goal that the therapist and patient have for one another. Has been adapted to robotics as well.

- **Anxiety:** The Negative Attitudes toward Robots Scale (NARS) has been put forward as a tool for measuring people's anxiety toward robots.

Study Variables

- The goal of good experimental design is to isolate the variables of interest such that we can draw concrete conclusions about their relationship.
  - **Independent variables** are aspects of the experiment that the experimenter will purposely manipulate in order to test these hypotheses.
  - **Dependent Variables** are aspects of the experiment that can be measured/controlled in order to see the effects of the manipulation of the independent variable(s).

- The first step in designing a study is to formulate the hypotheses in order to decide what independent and dependent variables are necessary to address them.
  - "I expect that changing [insert independent variable] will have an impact on [insert dependent variable]"

Performing Experimental Research

- Determine the **sampling groups** to study
  - Typically there is one **control group**, and the other groups are used to test experimental conditions

- Determine experiment design

- Conduct **pilot study**
  - small scale test of the study to ensure you're getting the kind of data you need

- Conduct the experiment
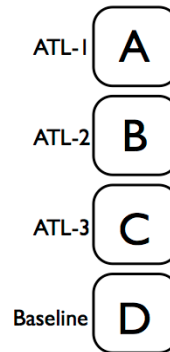
- Perform analysis
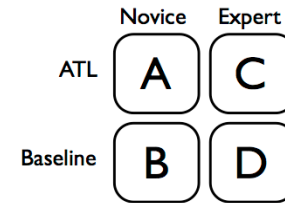
Experiment Type

Two Group

IV: Algorithm Version

ATL   A

Baseline   B

Multi-Group

IV: Algorithm Version

ATL-1   A

ATL-2   B

ATL-3   C

Baseline   D
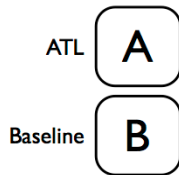
Factorial

IV: Algorithm Version
     Expertise

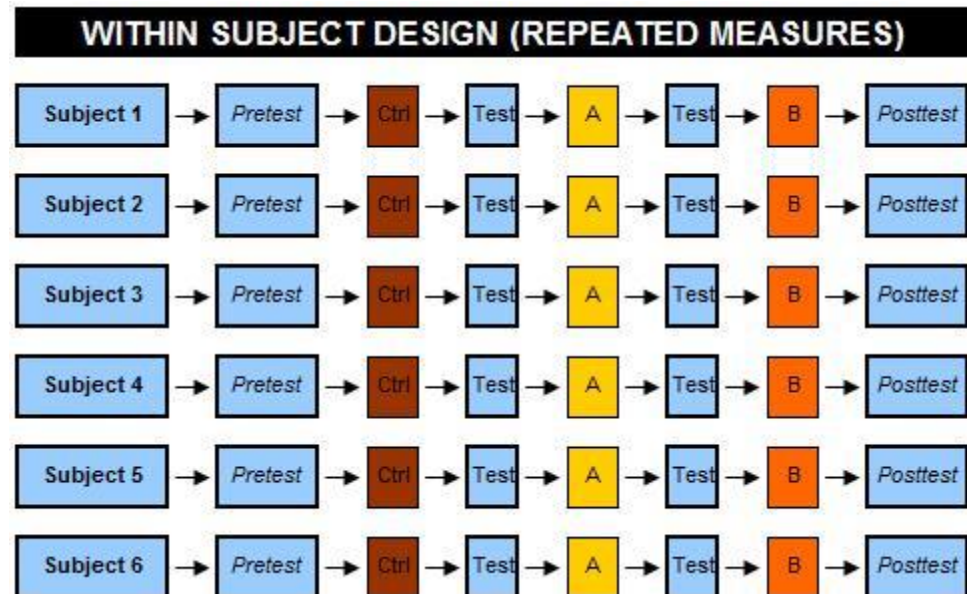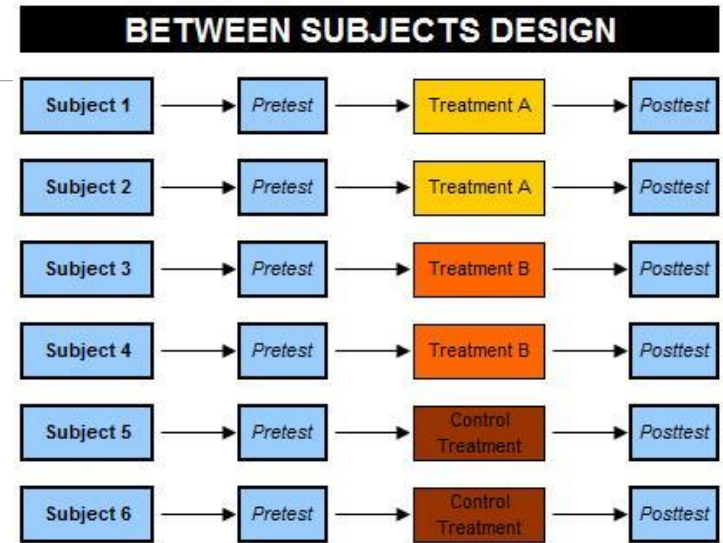|	Novice	Expert
ATL	A	C
Baseline	B	D

Often referred to as a 2x2 design (two independent variables, each with two possible values)

Experiment Design

- Between subjects design
  - Participants are part of a single group/treatment.

- Within subjects design
  - Every single participant is subjected to every treatment, including the control

Examples

- Between subjects design
  - Students in a class are separated into three groups, each using a different textbook. Grades at the end determine which book is better.
  - What variables need to be randomized?

- Within subjects design
  - Each student in a class tries all three textbooks and reports which one he/she prefers.

Novelty Effect

- Novelty effects can be particularly strong in HRI
  - Many participants may have never seen or interacted with a robot before.
  - We don't want the effects of this first interaction (so-called novelty effects) to completely overpower the manipulation of the independent variable in our study.

- One way to reduce impact:
  - Include a practice session with the robot as part of your study
  - Data from this practice session would not be recorded or used in our analysis.

Advantages

- Between subjects design
  - Different conditions are evaluated independently, reducing chance of one condition influencing another
  - Shorter studies, better compliance from participants
  - Easier to design

- Within subjects design
  - Fewer subjects required
  - Less chance of individuals skewing results

The Good Samaritan Experiment

- Biblical story: *A traveler is beaten, robbed, and left half dead along the road. First a rabbi and then a Levite come by, but both avoid the man. Finally, a Samaritan comes by. Samaritans and Jews generally despised each other, but the Samaritan helps the injured man.*

- What possessed the priest and the Levite to pass by the injured man by the side of the road? Possibly they were in a hurry and were filled with busy, important thoughts. Maybe the Samaritan was in less of a hurry. Or maybe the virtues that the religious leaders espoused were not something they followed themselves (unlike the Samaritan).

The Good Samaritan Experiment

- 1978 study on altruistic behavior

- Variables to be tested
    - the relative haste of the participant
    - how occupied their minds were with other matters (it has been argued that, because the thoughts of the rabbi and the Levite were on religious and spiritual matters, they might have been too distracted to stop and help)

The Good Samaritan Experiment

- Three hypotheses that they wanted to test:

  - People thinking about religion and higher principles would be no more inclined to show helping behavior than laymen.

  - People in a rush would be much less likely to show helping behavior.

  - People who are religious for personal gain would be less likely to help than people who are religious because they want to gain some spiritual and personal insights into the meaning of life.

How would you design this experiment?

The Good Samaritan Experiment

- Participants:
  - Religious studies students on a study course
  - Asked to fill in a questionnaire about religious affiliations and beliefs, to help evaluate and judge the findings of hypothesis 3.

The Good Samaritan Experiment

- Participants were given some religious teaching and instruction and then were told to travel from one building to the next. Between the two buildings was a man lying injured and appearing to be in desperate need of assistance.

- Variable 1: the amount of urgency impressed upon the subjects, with some being told not to rush and others being informed that speed was of the essence.

- Variable 2: the relative mindset of the subject, with one group being told that they would be giving lectures on procedures in the seminary, the others that they would be giving a talk about the 'Good Samaritan'.

The Good Samaritan Experiment

• Analysis based on a six point scale of assessing helping behavior

0=failed to notice victim as in need
1=perceived need but did not offer aid
2=did not stop but helped indirectly (told the aide on their arrival)
3=stopped and asked if victim needed help
4=after stopping, insisted on taking victim inside and then left him.
5=refused to leave victim, or insisted on taking him somewhere

Results

- When the subject was in no hurry, nearly 66% of people stopped to lend assistance. When the subject was in a rush, this dropped to 10%.

- People who were on the way to deliver a speech about helping others were nearly twice as likely to help as those delivering other sermons, showing that the thoughts of the individual were a factor in dictating helping behavior.

- Religious beliefs did not appear to make much difference on the results; being religious for personal gain, or as part of a spiritual quest, did not appear to make much of a noticeable impact on the amount of helping behavior shown.