

Other types of studies...

Case Studies

- an in depth study of a particular situation rather than a sweeping statistical survey.
- a study of how something works in a very specific situation
 - hard to generalize results generalized
 - provides in-depth analysis and useful insight for the case being examined
- Analysis: often more opinion based than statistical

Longitudinal Study

- A study performed over a long period of time
 - Psychology: years and decades
 - Robotics: weeks

Surveys

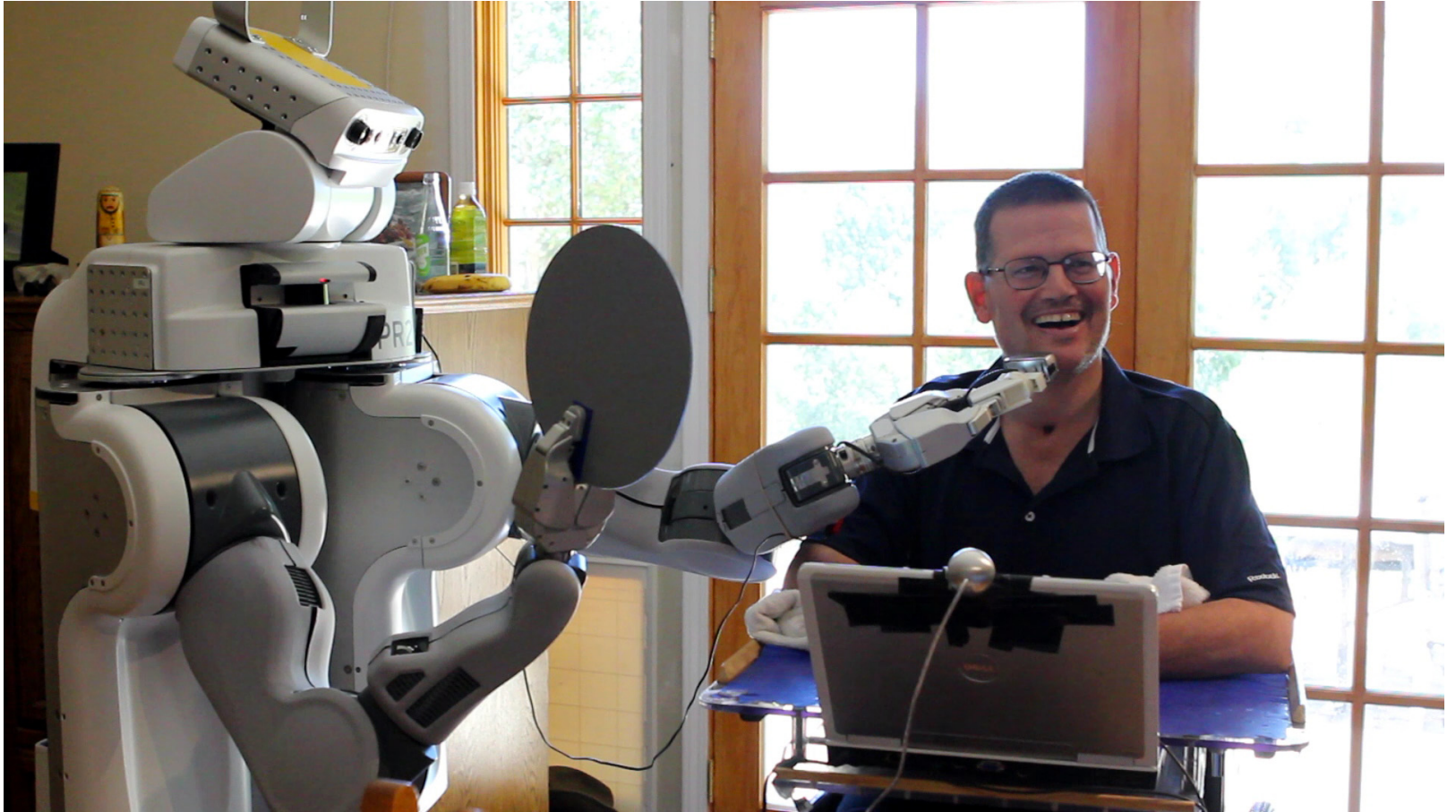
- Participants: recruited on campus, online, etc
- Types of questions: multiple choice, open answer, likert scale (“on a scale from 1 to 10...”)
- Groups of questions

Wizard of Oz

- *laboratory studies of **simulated** systems*
- The experimenter intercepts the subject's input to the robot and may provide responses as if they were coming from the robot itself.
- Main purpose is to simulate potential future systems that
 - are expensive to prototype
 - current technologies do not yet enable

Papers we've read...

Robots for Humanity



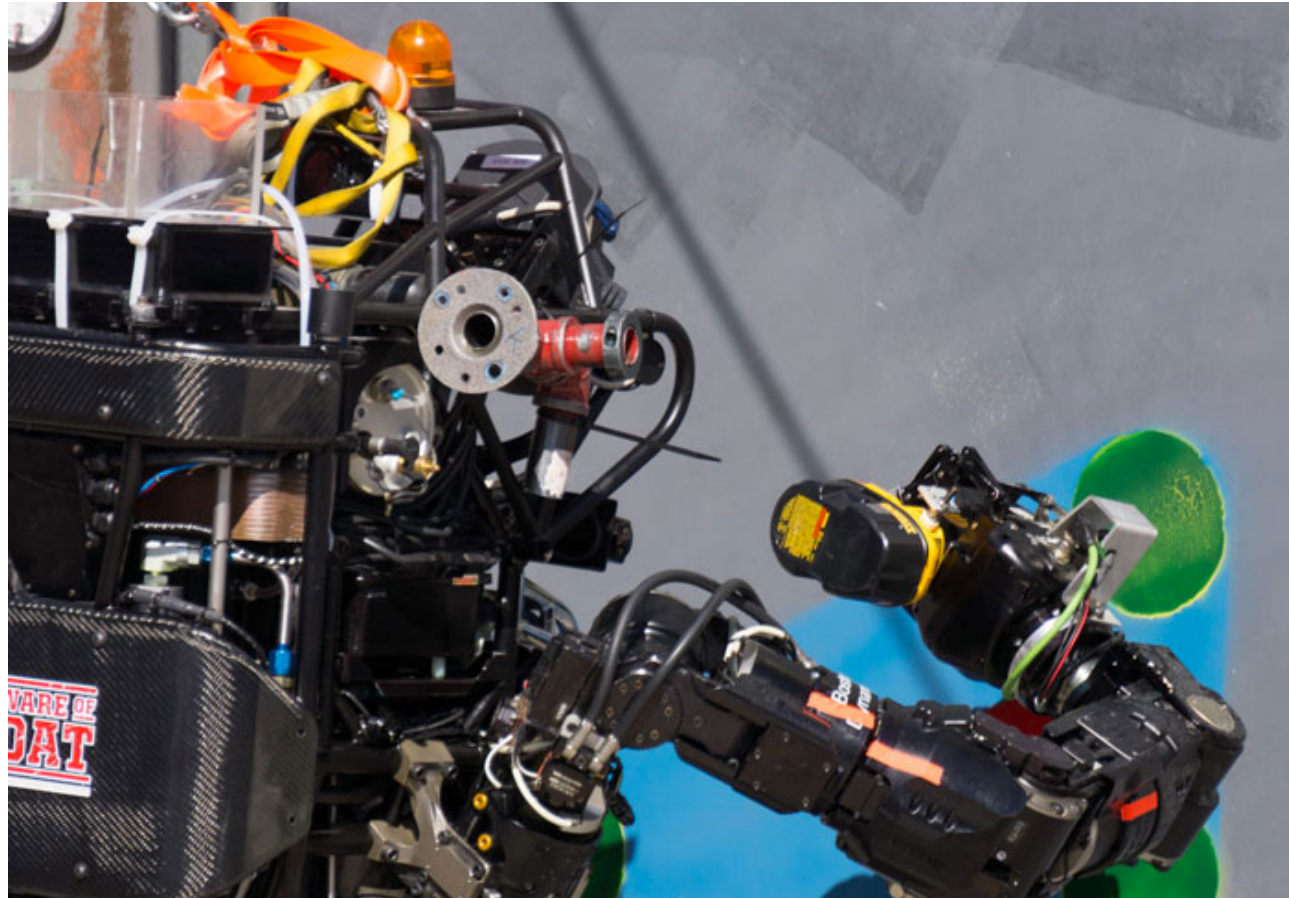
Enabling Building Service Robots to Guide Blind People

- [Paper](#)

Autonomous City Explorer

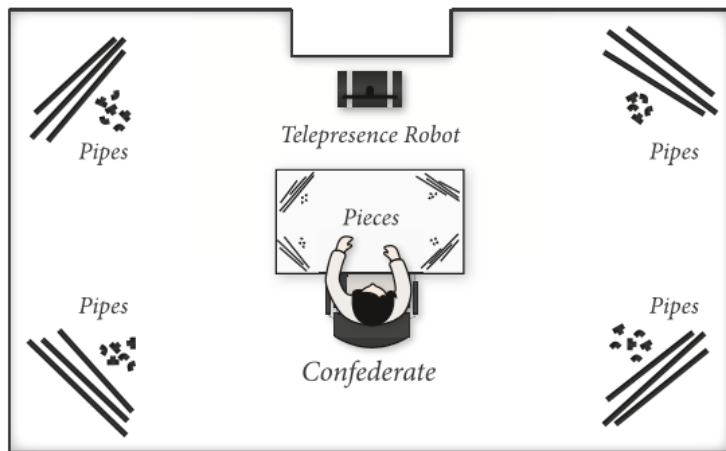


Analysis of HRI at the DRC Trials

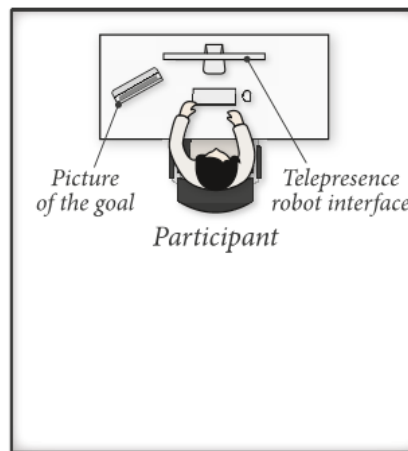


Bodies in Motion: Mobility, Presence, and Task Awareness in Telepresence

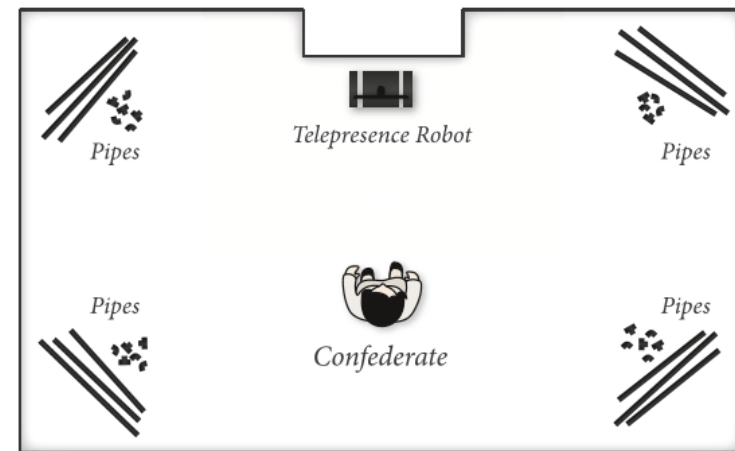
Local Environment, *Low-Mobility Task*



Remote Environment



Local Environment, *High-Mobility Task*



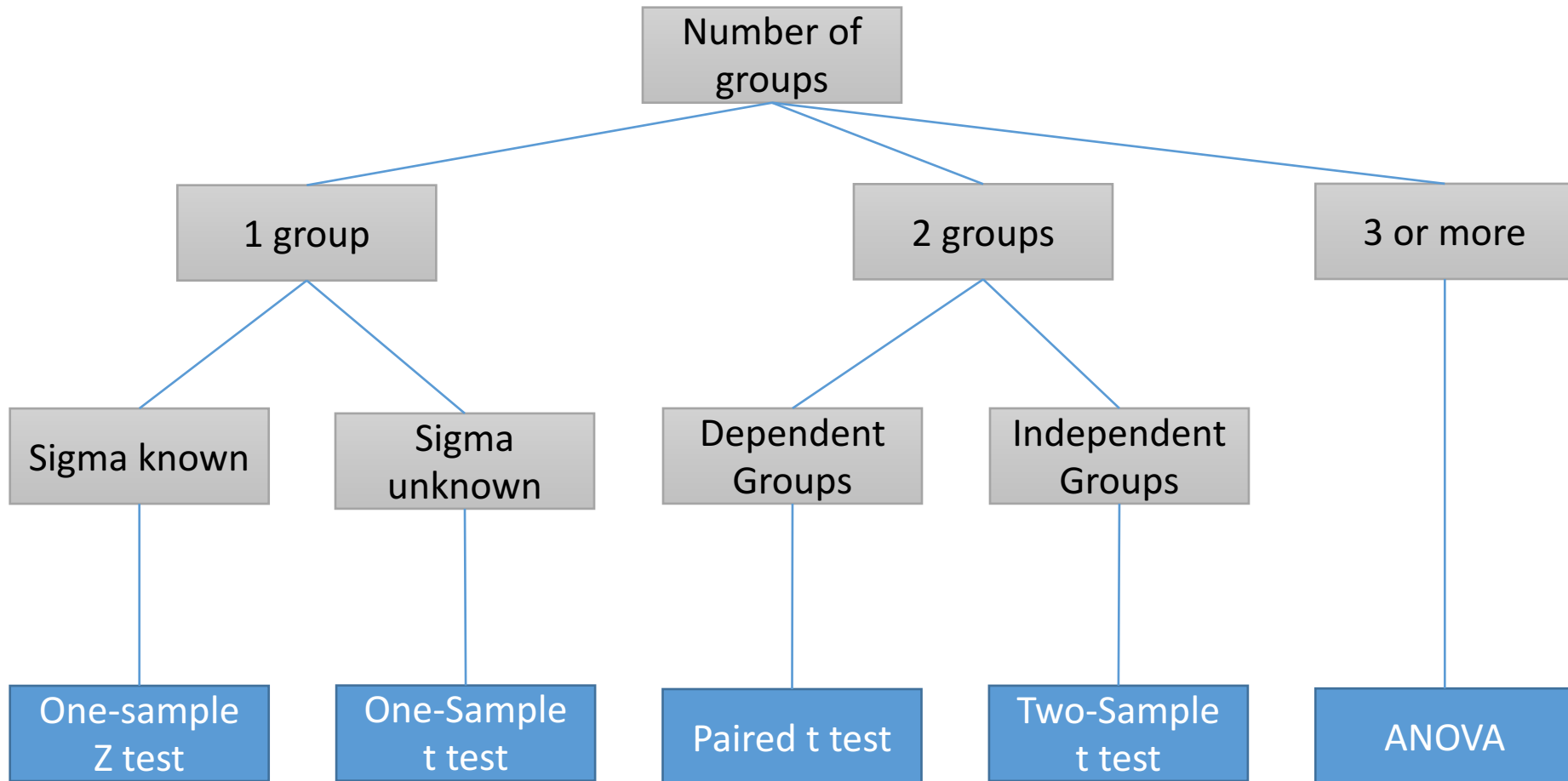
- [Paper](#)

Exploring Use Cases for Telepresence Robots



- [Paper](#)

Quantitative Data Analysis



Categorical Data Analysis

Yes/No, gender categories, survey questions, Likert

Chi-squared test

Mann-Whitney-
Wilcoxon test

t test

(Treats each category independently, thus completely ignoring the ordinal scale of Likert. Thus one of the others preferred for Likert.)

P values

- P values = the probability that the observed result was obtained by chance
 - i.e. when the null hypothesis is true
- **α level** is set *a priori* (Usually 0.05)
- If **$p < \alpha$** level then we reject the null hypothesis and accept the experimental hypothesis
 - 95% certain that our experimental effect is genuine
- If however, **$p > \alpha$** level then we reject the experimental hypothesis and accept the null hypothesis

P values

$P < 0.05$ *

$P < 0.01$ **

$P < 0.001$ ***

Values of p	Inference
$p > 0.10$	No evidence against the null hypothesis.
$0.05 < p < 0.10$	Weak evidence against the null hypothesis
$0.01 < p < 0.05$	Moderate evidence against the null hypothesis
$0.05 < p < 0.001$	Good evidence against null hypothesis.
$0.001 < p < 0.01$	Strong evidence against the null hypothesis
$p < 0.001$	Very strong evidence against the null hypothesis

- Most authors refer to **statistically significant** as $P < 0.05$ and **statistically highly significant** as $P < 0.001$ (less than one in a thousand chance of being wrong).
- $P < 0.1$ is considered a **trend** but not statistically significant

Event Coding

- Common in observational studies, usability tests, etc
- Two or more human **coders** (not the researchers) label the data independently
- Consistency/reliability of the coding must be confirmed by **Cronbach's alpha** measure before reporting the results
- Higher alpha values are better, with 0.7 considered the lower bound for reliability

Cronbach's alpha	Internal consistency
$\alpha \geq 0.9$	Excellent
$0.9 > \alpha \geq 0.8$	Good
$0.8 > \alpha \geq 0.7$	Acceptable
$0.7 > \alpha \geq 0.6$	Questionable
$0.6 > \alpha \geq 0.5$	Poor
$0.5 > \alpha$	Unacceptable

Other Study Considerations...

Legal issues and safety

- All human subject studies require approval from the Institutional Review Board (IRB)
- The IRB evaluates proposals for issues of
 - Data privacy
 - Protection of minors and vulnerable adults
 - Mental or emotional stress or humiliation
 - Physical safety

Motivations for IRB Creation

- Nazi Experiments
- Tuskegee Syphilis Study
- Milgram Study
- Stanford Prison Study

Nazi Experiments 1940s

- Medical experiments
 - Gunshot wounds
 - Traumatic amputations
 - Chemical & biological warfare agents
- Problems
 - Physical harm
 - No informed consent

Tuskegee Syphilis Study

1932-1972

- Medical study
 - Clinical study conducted between 1932 and 1972 in Tuskegee, Alabama by the U.S. Public Health Service to study the natural progression of untreated syphilis in rural black men who thought they were receiving free health care from the U.S. government.
- Problems
 - Physical harm
 - No informed consent

Milgram Study

- 1960's
- Psychological study
 - series of social psychology experiments conducted by Yale University psychologist Stanley Milgram, which measured the willingness of study participants to obey an authority figure who instructed them to perform acts that conflicted with their personal conscience.
- Problems
 - Psychological stress
 - Deception used – no informed consent

<https://www.youtube.com/watch?v=yr5cjyokVUs>

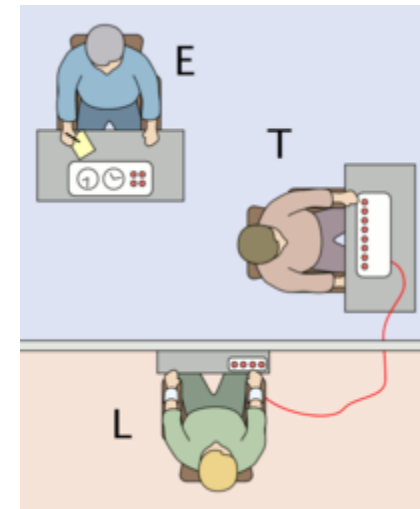
Public Announcement

**WE WILL PAY YOU \$4.00 FOR
ONE HOUR OF YOUR TIME**

Persons Needed for a Study of Memory

*We will pay five hundred New Haven men to help us complete a scientific study of memory and learning. The study is being done at Yale University.

*Each person who participates will be paid \$4.00 (plus 50¢ carfare) for approximately 1 hour's time. We need you for only one hour; there are no further obligations. You may choose the time you would like to come (evening, weekdays, or weekends).



Stanford Prison Study

- 1971
- Psychology study
 - a study of the psychological effects of becoming a prisoner or prison guard.
 - Twenty-four male students were selected to take on randomly assigned roles of prisoners and guards in a mock prison situated in the basement of the Stanford psychology building.
- Caused harm to participants



Prisoner #416 was newly admitted as one of our stand-by prisoners. Unlike the other prisoners, who had experienced a gradual escalation of harassment, this prisoner's horror was full-blown when he arrived. The "old timer" prisoners told him that quitting was impossible, that it was a real prison.

Prisoner #416 coped by going on a hunger strike to force his release. After several unsuccessful attempts to get #416 to eat, the guards threw him into solitary confinement for three hours, even though their own rules stated that one hour was the limit. Still, #416 refused.

At this point #416 should have been a hero to the other prisoners. But instead, the others saw him as a troublemaker. The head guard then exploited this feeling by giving prisoners a choice. They could have #416 come out of solitary if they were willing to give up their blanket, or they could leave #416 in solitary all night.

Most elected to keep their blanket and let their fellow prisoner suffer in solitary all night.

IRB Creation

- Protect human participants from harm:
 - Physical
 - Psychological
 - Social
 - Legal
 - Economic
 - Dignatory (reputation)

How this affects HRI research and you

- All US government funded projects involving human subjects require IRB review prior to funding being awarded
 - If your research is funded by NSF, NIH, etc, your project may already have completed an IRB review
 - If you are working independently, you may need IRB review for a study
- All user study participants must be given information about the study protocol, must consent to the study, and must be debriefed following the study
 - Special forms and guidelines for research involving minors
 - Deception in the protocol is possible, but highly restricted
 - Collection of data such as audio or video recording also requires consent
 - Participants may be compensated for their time with a small payment, but money may not be used to “unduly influence potential human subjects to participate in research activities”
 - Compensation must not be contingent upon completion of study procedures

Abstract: Robots are being introduced in our society but their social status is still unclear. A critical issue is if the robot's exhibition of intelligent life-like behavior leads to the users' perception of animacy. The ultimate test for the life-likeness of a robot is to kill it. We therefore conducted an experiment in which the robot's intelligence and the participants' gender were the independent variables and the users' destructive behavior of the robot the dependent variables. Several practical and methodological problems compromised the acquired data, but we can conclude that the robot's intelligence had a significant influence on the users' destructive behavior. We discuss the encountered problems and the possible application of this animacy measuring method.

Excerpt:

“A qualitative video analysis showed that almost all participants giggled or laughed during the last phase of the experiment. Their reactions are to some degree similar to the behavior participants showed during Milgram's famous obedience to authority experiments [9]. In his experiments, participants were instructed to use electric shocks at increasing levels to motivate a student in a learning task. The experimenter would urge the participant to continue the experiment if he or she should be in doubt. Similar to this study, the participants were confronted with a dilemma and to release some of the pressure, they resorted to laughter.”

This type of study would likely not be allowed in the US by the IRB