EXPERIMENTAL HRI

(original presentation by P. Baxter at the 2nd Summer School on Social HRI on Åland Islands, Finland)

Gergely Magyar, PhD.
Center for Intelligent Technologies
Department of Cybernetics and Artificial Intelligence
Technical University of Košice

Department of Cybernetics and Artificial Intelligence

Center for Intelligent Technologies

1. What does the word robot mean and where was it born?
2. What is the uncanny valley?

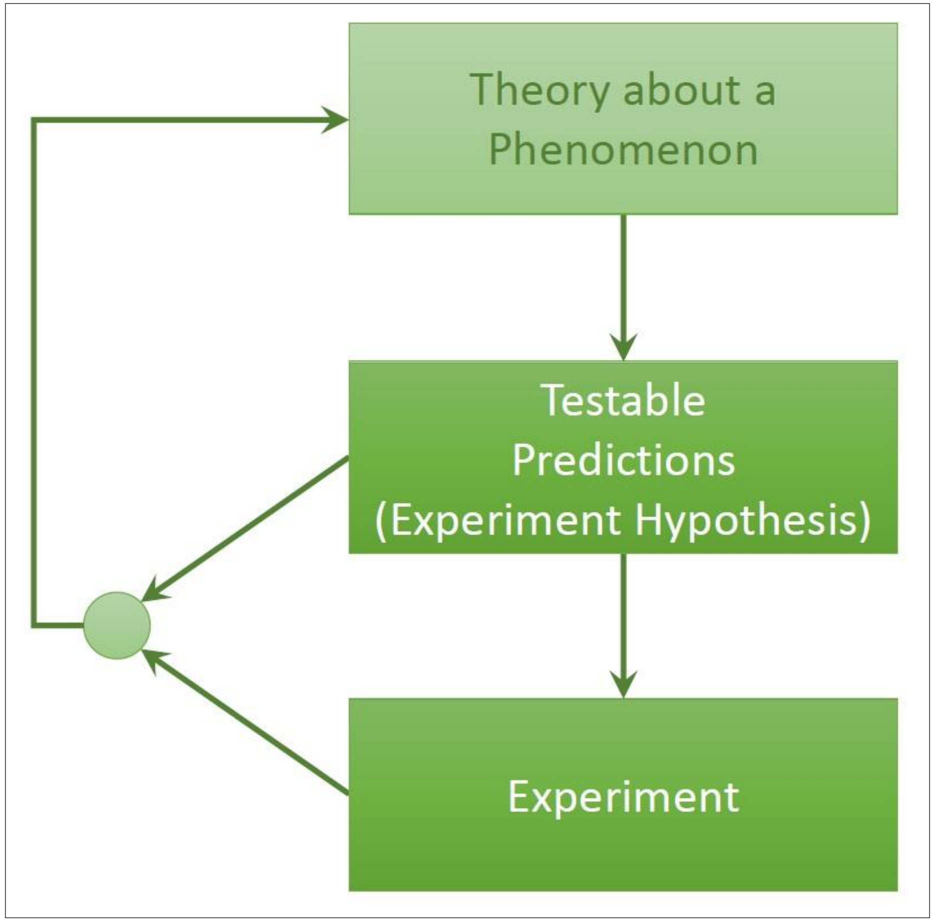
(Good) experiments in social HRI are hard to do!

WHY?

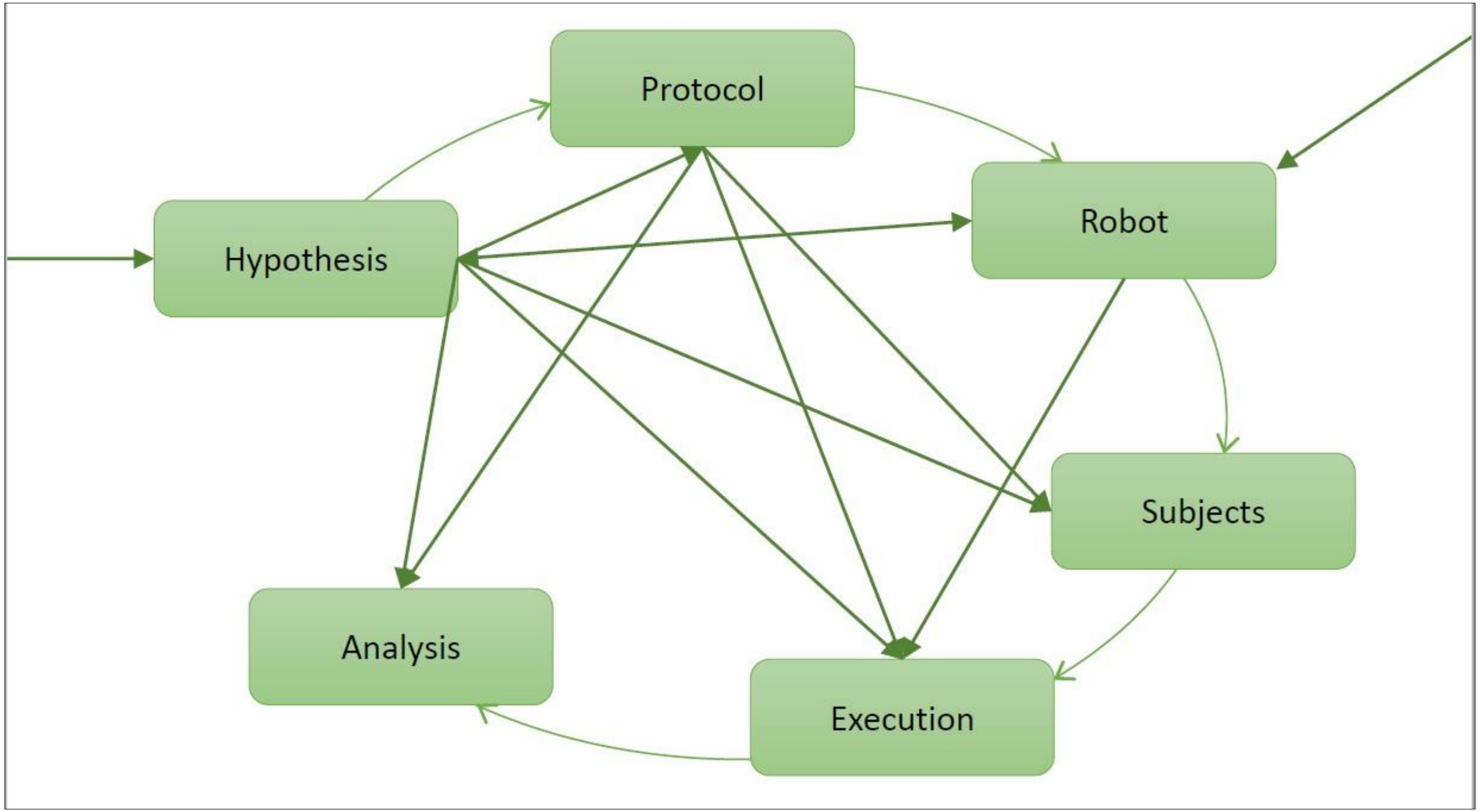
- Technical systems (“Battery low … knock knock”)
- The ‘real world’ (“I can’t connect to the network”)
- People -.-

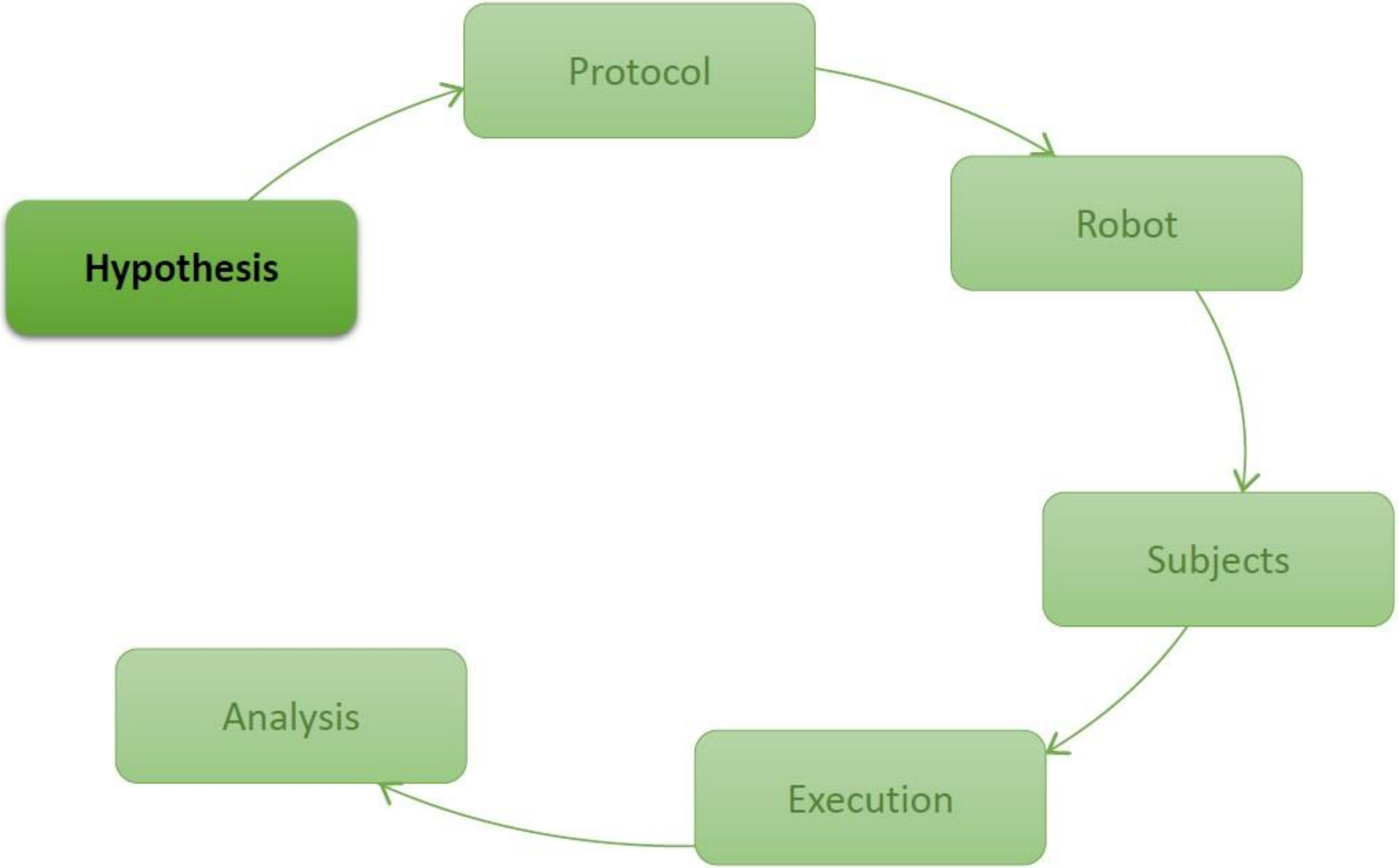
But!!! There are many exciting opportunities and things to learn, so well worth doing, and worth doing well.

THE SCIENTIFIC METHOD



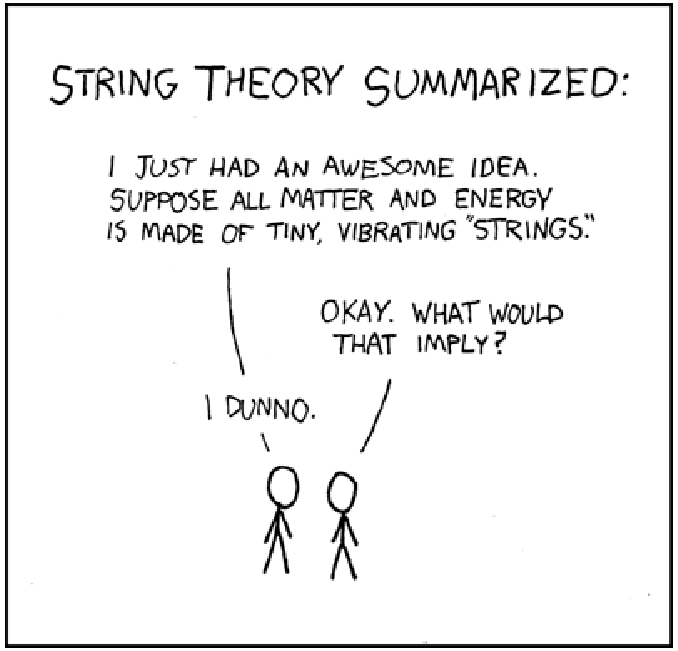
THE “SCIENTIFIC” METHOD




Protocol
Robot
Hypothesis
Subjects
Analysis
Execution

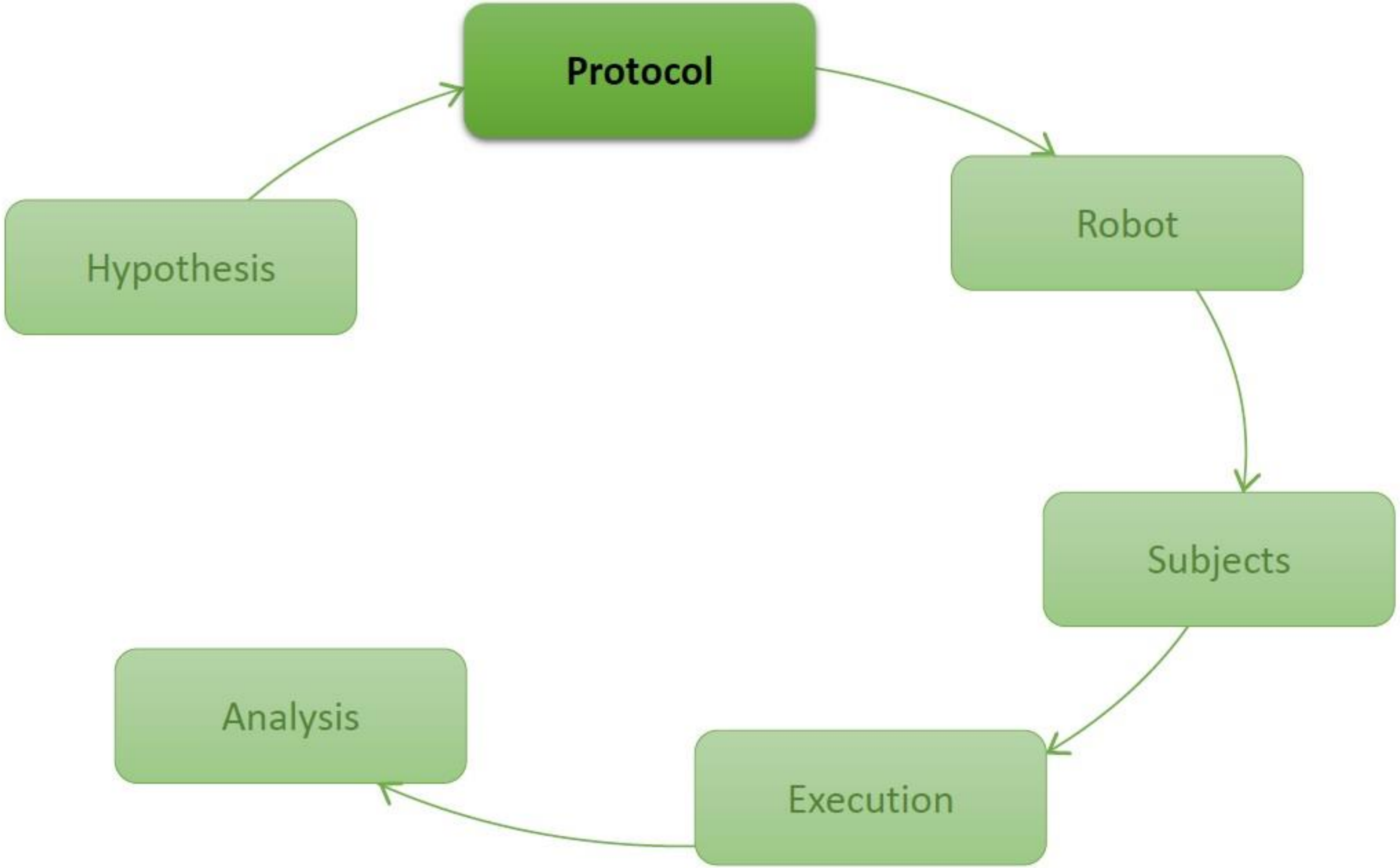
TESTABLE PREDICTIONS

- Let's assume you have a theory about a phenomenon or effect…
- A good experimental hypothesis will make testable prediction
  - Which can be supported or rejected
- The purpose and structure of the experiment will be to test this



EXPLORATORY STUDIES

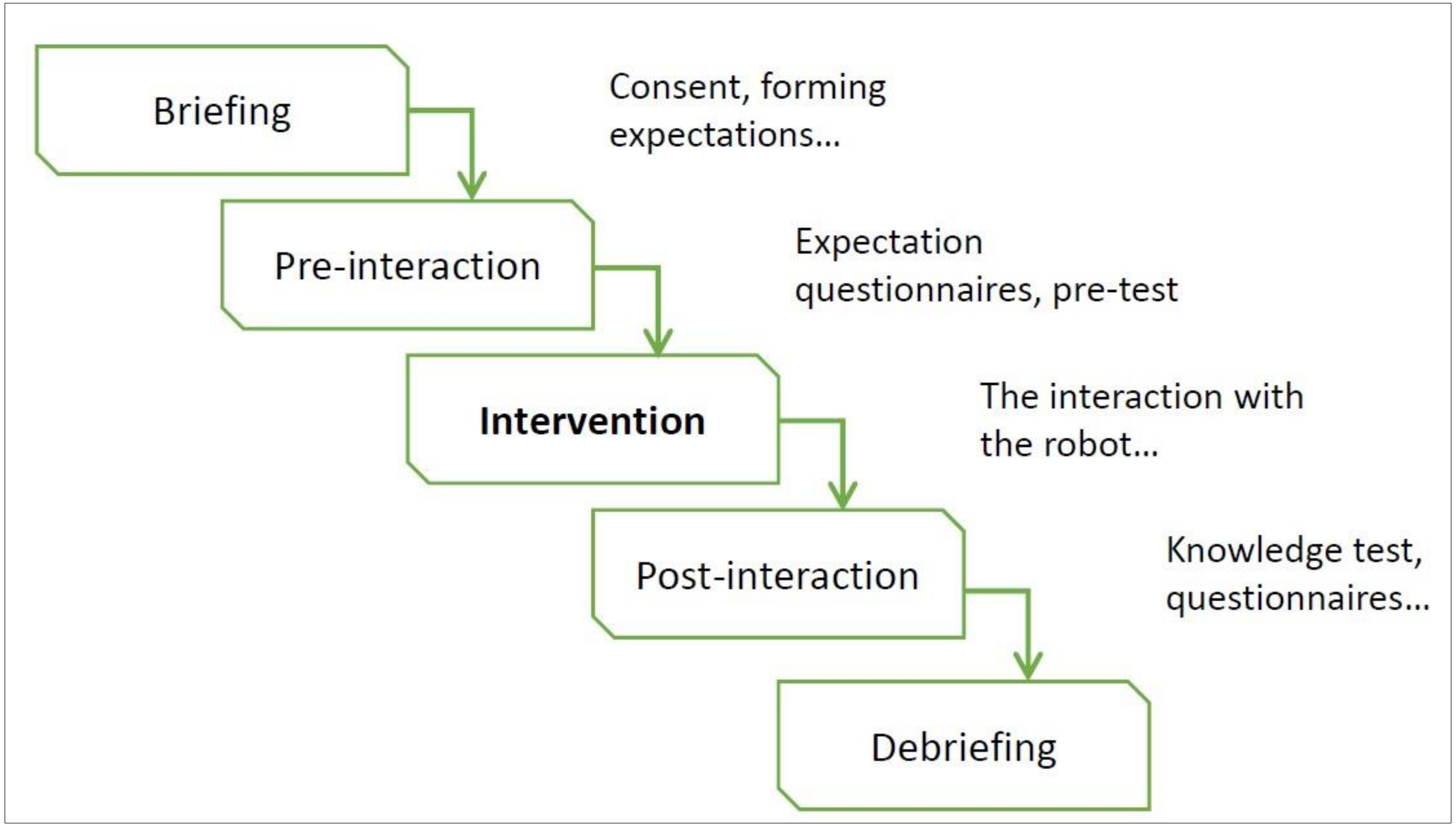
- The preceding argument does not suggest that exploratory studies are useless…
  - …although the standard scientific method suggests that these would still be driven by some expectation or theory
- Includes studies with perhaps only a single experimental condition
- Some degree of control and rigour are still required to draw meaningful observations
  - Reduce the incidence of confounds, or even discover what the confounds are
  - Finding data from which hypotheses can be formed


Protocol
Robot
Subjects
Execution
Analysis
Hypothesis

THE EXPERIMENT PROTOCOL

- Structure of the interaction
  - Things to consider …
- Confounds
- Pilot studies

THE STRUCTURE OF THE INTERACTION



CONSIDERATIONS

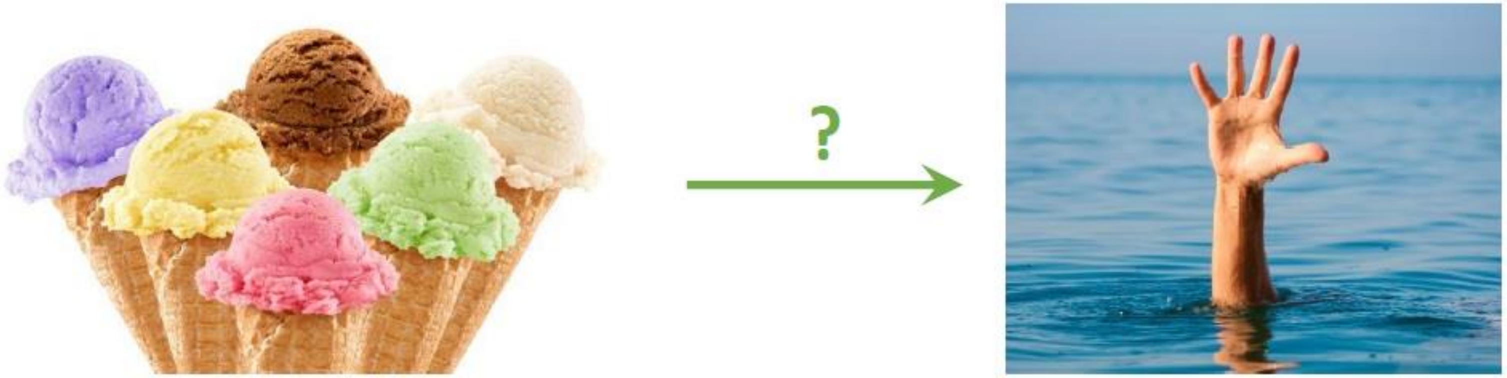
- What role the robot will play?
  - Peer, teacher, assistant, remote presence …
  - Is social all it is made out to be?
- In the pre- and post-interaction parts, is some form of knowledge/skill assessment going to take place?
  - Such as a test of prior knowledge/questionnaires
  - Enables some gain due to the intervention (within some limits)

CONSIDERATIONS (2)

- What is the desired structure of the intervention?
  - Rigid structure (e.g. strict turn-based)
  - Unstructured
  - Trade-offs of potential richness of interaction versus complexity of implementation and analysis
  - If in context of social HRI, then how to decide what is best?
  - A happy middle ground in certain games?

CONFOUNDS

- To mix up …
- Correlates with both the dependent (metric) and independent (e.g. condition) variable
  - Bad example: association between ice-cream consumption and deaths from drowning



CONFOUNDS, CONFOUNDS, AND … CONFOUNDS

- Prior experience
  - Do some subjects have some experience in some other domain that will give them an advantage?
- Existing competence
  - If task is given to participants, are some better than others even before the task starts?
- Biases, etc.
  - Unconscious effects
  - Color …

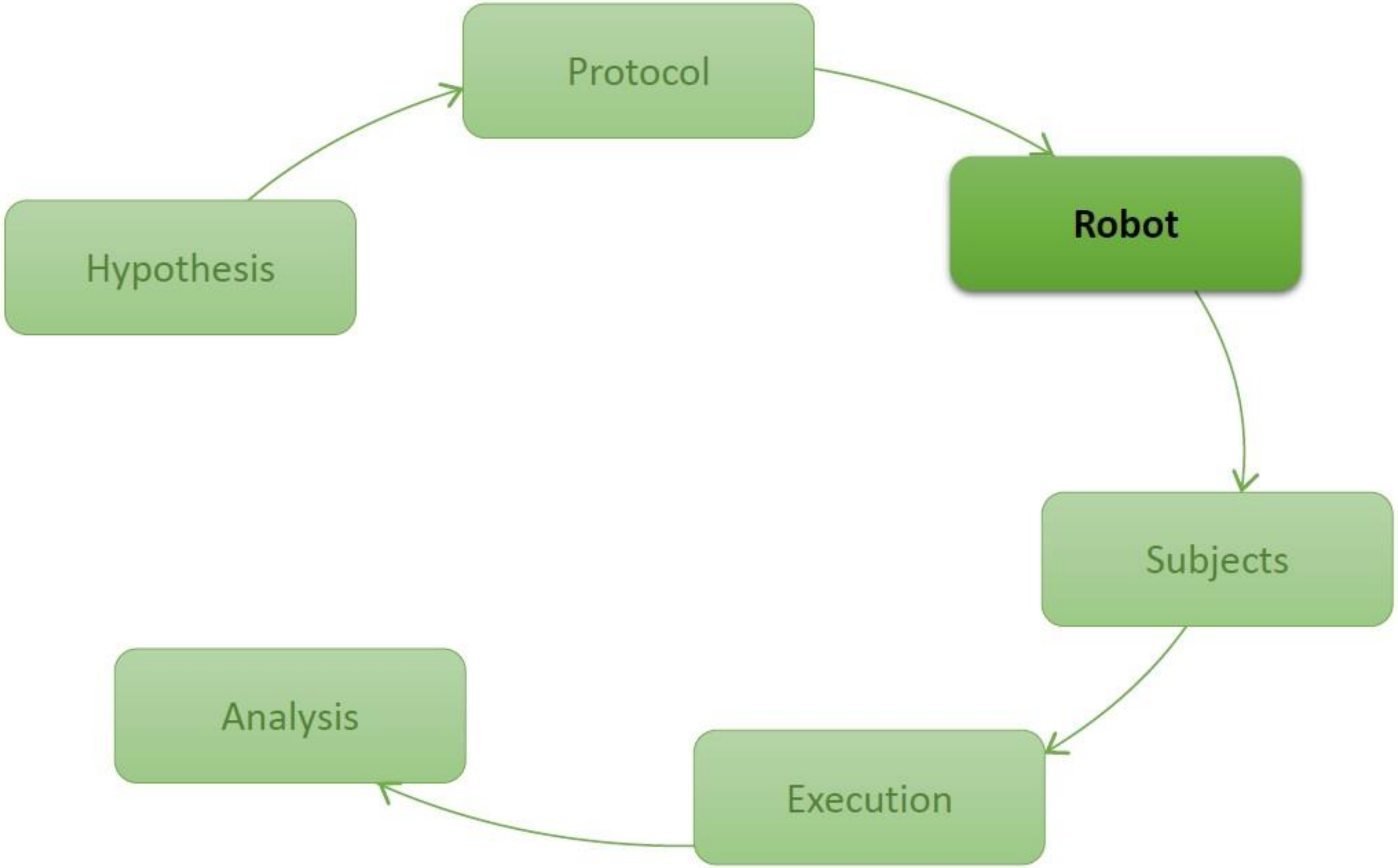
THE NOVELTY EFFECT: YOUR PROBLEMATIC FRIEND?

DEALING WITH CONFOUNDS

- Balancing participants to ensure that the different abilities are equally represented
- Implication that you know what the potential confounds actually are
  - That you have verified this in some way, e.g. in a pilot study
- One potential way of dealing with this is to balance for common potential confounds:
  - Age
  - Gender
  - Teacher-rated ability (in Maths for example)

PILOT STUDIES

- To refine research questions
  - Remember the “Exploratory studies” part previously…
- To validate measures/metrics
  - Are you actually measuring what you intend to measure?


Protocol
Robot
Hypothesis
Subjects
Analysis
Execution

DRAWING PEOPLE IN

- Just using a robot brings many advantages in terms of interest and engagement
    - Great benefit in public spaces and for events open to the public to generate participation

PERCEPTION VS EXPECTATION

- The complex/human-like looking robot may draw in more attention …
- …however, this can give rise to subsequently unfulfilled expectations
- Managing expectations
  - Through robot design (e.g. presence of eyes, mouth and ears), but also through pre-experiment briefing

PERCEPTION VS EXPECTATION (2)

- Maintaining the ‘illusion’
  - The illusion of a robot competence greater than what is technically present
  - Adults seem to be particularly sensitive to breaks in the illusion
  - Children may be more forgiving, but there are limits to the forgiveness
    - E.g. robot when interacting with multiple children vs just one child …

ROBOT CONTROL

- In your experiment, how will you control your robot?
  - What is the most appropriate given your protocol?
- Autonomous behavior
  - Do you have enough sensory information available to enable this?
  - Is adaptivity required, and how would this be implemented, or is reactive behavior sufficient? Neither is trivial …
- Remote-controlled robots
  - Where the hypothesis demands it (e.g. telepresence)
  - Where technical limitations impose it
  - Prehaps even where time pressure necessitates it …

WIZARD OF OZ

- Human making up for technological shortfalls
  - Notably speech recognition (e.g. for children, or in noisy environments)
- Human-human interaction via a robot
  - E.g. tele/remote-presence devices
- Typically used for natural language processing and non-verbal processing, not typically for manipulation
- Control of behavior production, but not typically of recognition
- WoZ as part of an iterative design process

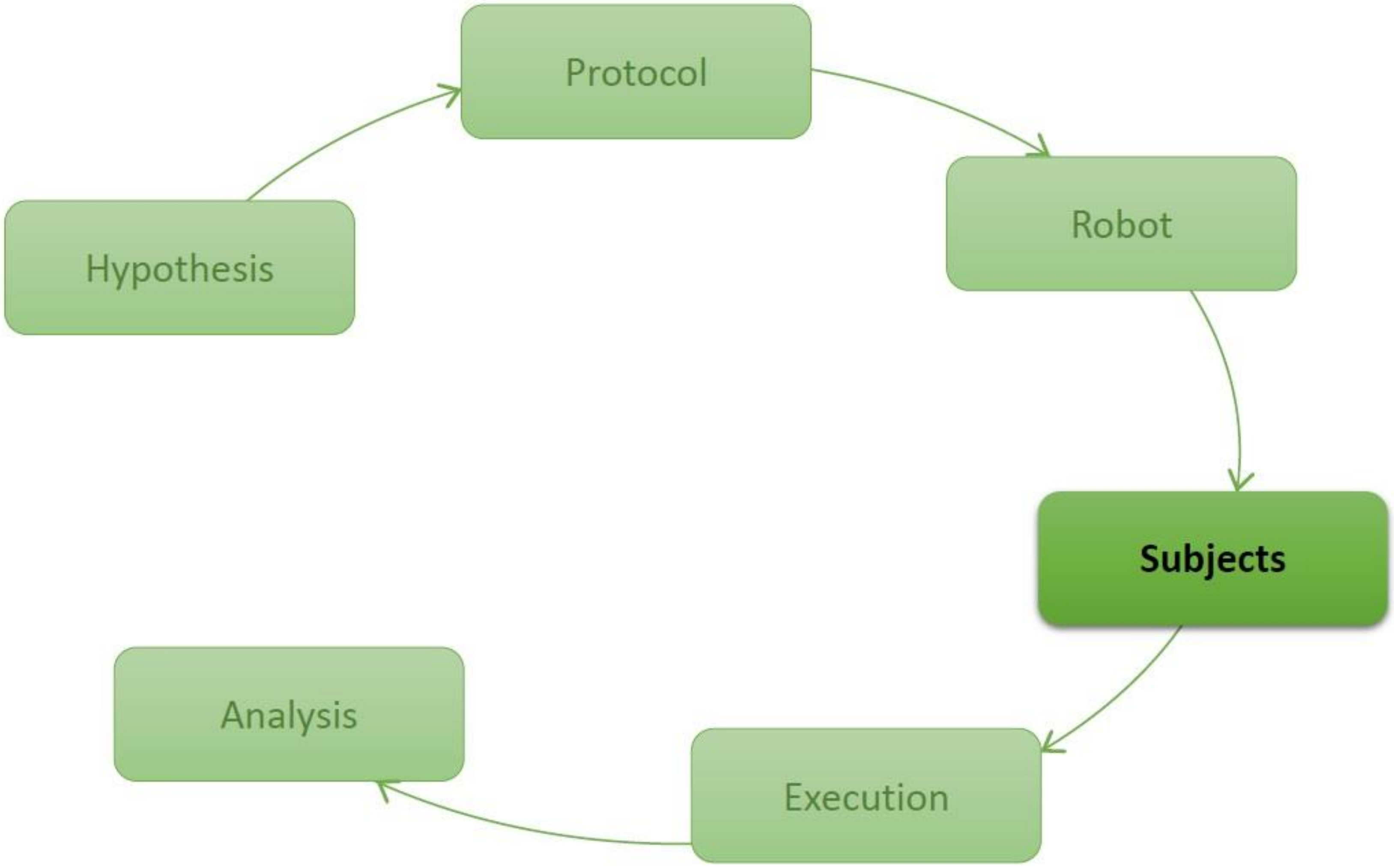
IF WOZ USED AS A CONTROL CONDITION – WHAT ABOUT WIZARD ADAPTIVITY?

TECHNICAL SHORTCUTS

- The ‘Sandtray’
  - A touchscreen interaction mediator between a robot and a person
  - Shortcut for perception
  - Shortcut for complex robot actions

TECHNICAL ROBUSTNESS

- It is guaranteed that the ‘user’ will do things that you had not anticipated!
- Implementing fall-backs and fail-safes
- From experience:
  - Test under the same conditions as the experiment will take place
  - Test with naïve subjects
  - Try and break your system


Protocol
Robot
Hypothesis
Subjects
Analysis
Execution

PRACTICAL RECRUITMENT ISSUES

- Requirements
  - What participants are needed, any particular properties
  - Age range, expertise, mental faculties
  - Special needs concerns: physical and/or mental disabilities
- Availability for the experimental period
  - E.g. in hospitals, need for treatment, and discharge
  - Unless part of the treatment course, sHRI experiments will typically have to give way to medical concerns

TAKING INTO ACCOUNT THE SUBJECT GROUP

- Verifying the task is at the appropriate level; verifying that the metrics/measurements are appropriate
  - E.g. ensuring questionnaires are unambiguous for children (no difficult words)
  - Facilitating ease of completion: e.g. making the task visually interesting and similar to other tasks which they are already familiar (e.g. school work)
- Same ideas apply for all age-ranges and abilities

BEHAVING NATURALLY?

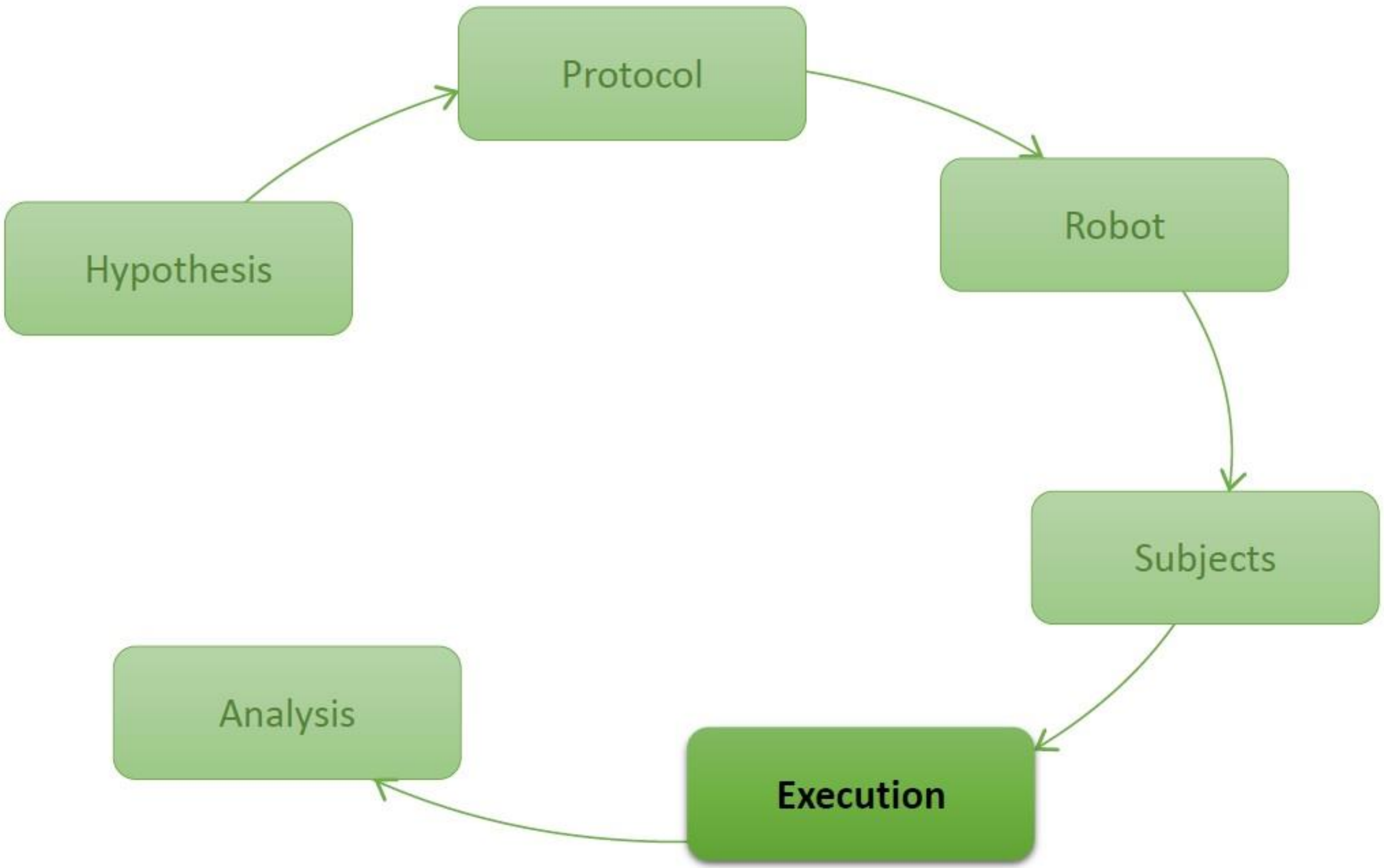
- What does constitute natural behavior?
  - Is this possible in the lab with a robot?
- Trade-off: the control possible in a lab versus the potential natural-ness of behavior
  - The differences between adults and children in this regard

ADAPTIVITY OF BEHAVIOR

- One thing that can be counted on is adaptivity of behavior from your subjects
  - They will adapt whether you want them or not
  - Has consequences for experimental design
- Experience and learning

LEGAL AND ETHICAL ISSUES

- Well defined protocols that need to be followed
  - At national (legislation) and institutional level (ethics committees), e.g. data protection
- Consent
  - From both children and parents, as well as the school


Protocol
Hypothesis
Robot
Subjects
Analysis
Execution

EXECUTION OF EXPERIMENTS

- Problems in the ‘real world’
- What do you need to measure, and how are you going to measure it?
  - Driven by the hypotheses
  - Avoiding/minimizing confounds

PROBLEMS IN THE 'REAL WORLD'

- Obviously many potential technical problems, particularly if running studies of controllable lab conditions
  - People don't do what you expect (especially children …)
  - Variable background noise and light conditions
  - Speech recognition: issues with noise
- The 'human problems'
  - Independence
  - E.g. children copying each other in groups

METRICS

- Performance
- Physiological metrics
- Qualitative analysis
- Questionnaires
- Retrospective behavior analysis

TASK PERFORMANCE

- Does what it says on the tin
- An objective measure (or measures) of how well you participant performs the task:
  - Correct answers
  - Completion
  - …

PHYSIOLOGICAL METRICS

- Reaction time
  - Well established in psychology: memory, attention, priming, biasing, etc.
  - Speed/accuracy trade-offs
- Galvanic skin response, heart rate, pupil dilation
- Brain-computer interfaces

QUALITATIVE METRICS

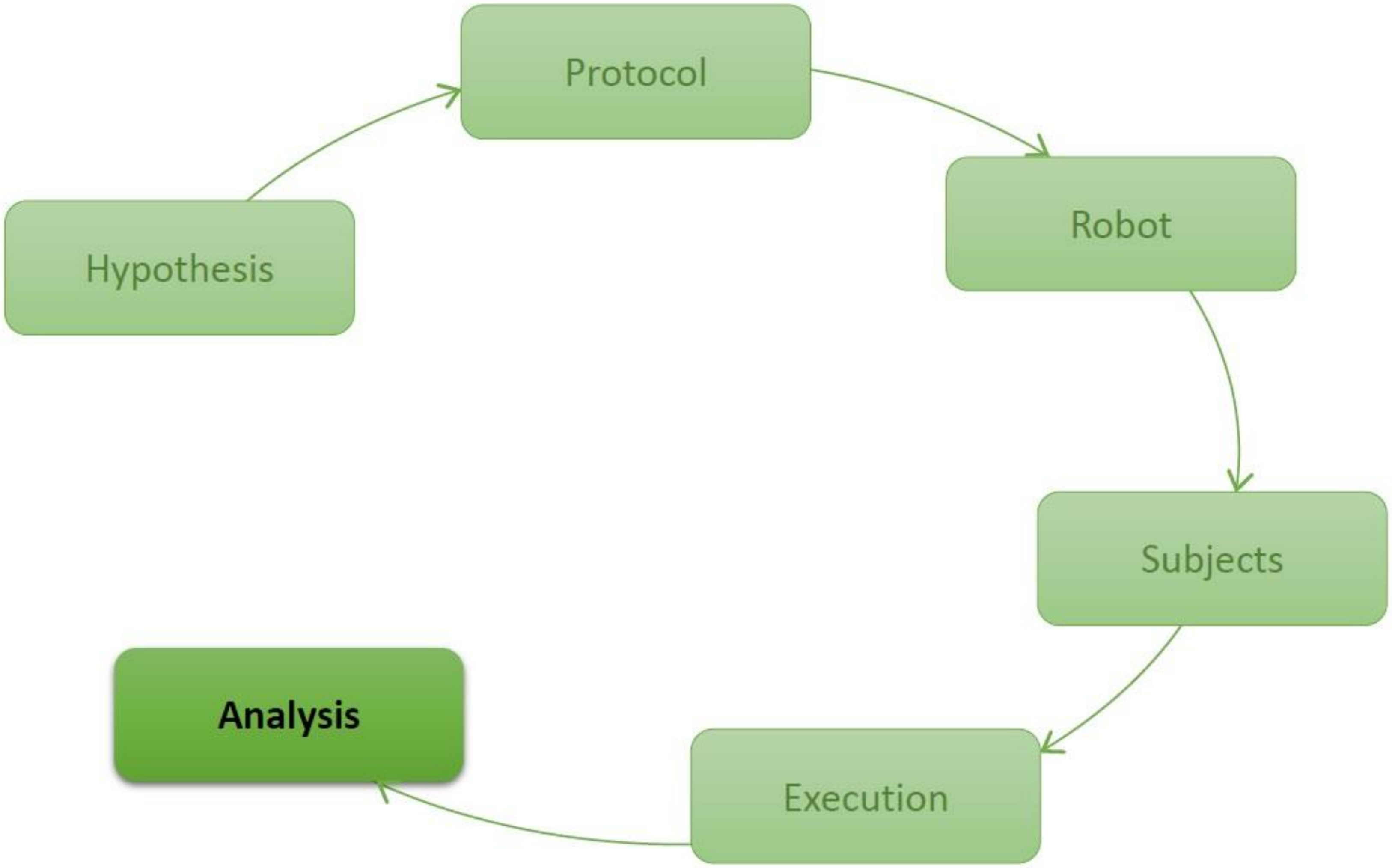
- Ethological techniques
  - Observations of animals as they go about their lives
  - E.g. coding behavior over time
- Conversation analysis
  - Systematic analysis of conversations in interactions: not just what is said, but how and when

QUESTIONNAIRES

- Commonly used, can typically be administered relatively easily
  - Typically to collect attitudes/opinions in a formal way
  - Likert scales, etc.
- Is the questionnaire developed/adapted
  - Valid: does it measure what it intends to measure? Is it appropriate for the target population?
- Problems with administering to children
  - Children seek to please the experimenter, seek to second-guess what the right answer is

RETROSPECTIVE BEHAVIOR ANALYSIS

- Video coding
  - A way of analyzing recorded interactions in depth that is not possible to do in real time, at the time of the experiment
  - Time consuming, but the level of detail obtained can be worth of it
- Subjective vs objective
  - How is the participant feeling over time: coder interpreting psychological state
- Validation
  - If only one person has coded, how do you know that a good job has been done?
  - Second coding, inter-rater agreement (Cohen's Kappa)


Protocol
Hypothesis
Robot
Subjects
Analysis
Execution

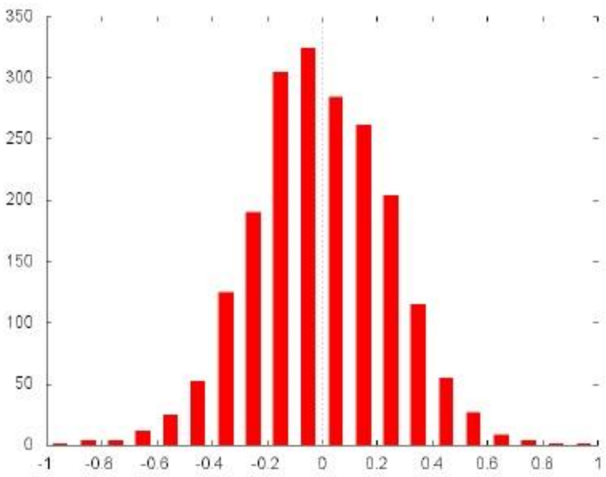
ANALYSIS

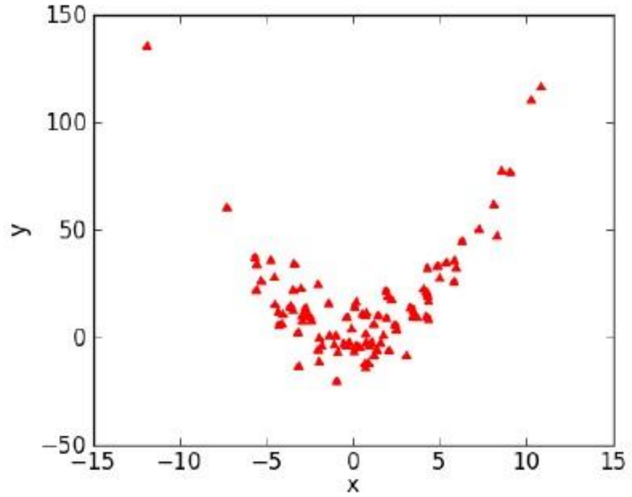
- Descriptive stats and correlation
- Null hypothesis significance testing
- Reporting

GOOD STATISTICS ALONE CAN'T SAVE YOU FROM POOR METHODOLOGY/EXECUTION!

DESCRIPTIVE STATISTICS

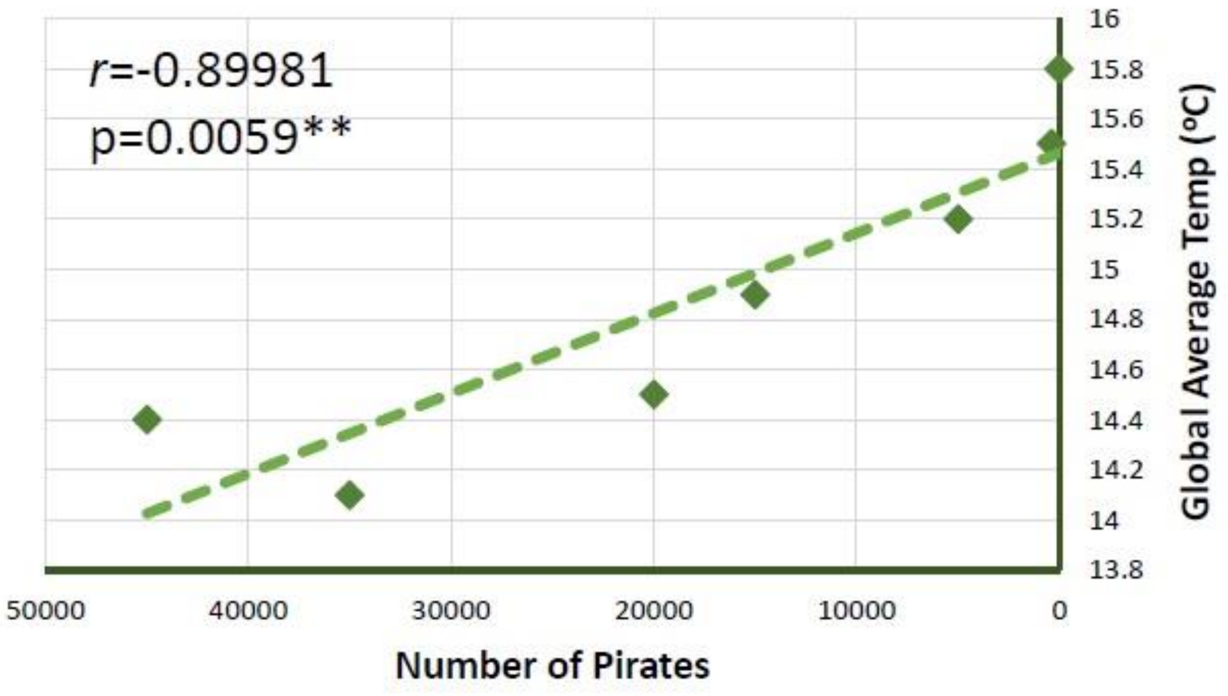
- Descriptive statistics are a very useful tool, and under-rated
  - Just look at your (raw) data visually
  - Scatter plots, frequency plots, etc.





CORRELATION != CAUSATION

- Bradford Hill criteria for causality



SIGNIFICANCE TESTING AND P'S

- The p-value is standard
- Cumming, 2008 analysis of p-value variation
  - 25 experiment simulations
  - 12 significant, 13 not …
  - p-range: <0,001 to 0,759
- Better to use confidence intervals as a descriptive measure of an effect?

REPORTING STATISTICS

Statistic	Purpose	APA Style	Description
Descriptive Statistics			
Mean	To provide an estimate of the population from which the sample was selected.	M = _____	Indicates the center point of the distribution and serves as the reference point for nearly all other statistics.
Standard Deviation	To provide an estimate of the amount of variability/dispersion in the distribution of population scores.	SD = _____	Indicates the variability of scores around their respective mean. Zero indicates no variability.
Measures of Effect Size			
Cohen's d	To provide a standardized measure of an effect (defined as the difference between two means).	d = _____.	Indicates the size of the treatment effect relative to the within-group variability of scores.
Correlation	To provide a measure of the association between two variables measured in a sample.	r(df) = _____	Indicates the strength of the relationship between two variables and can range from −1 to +1.
Eta-Squared	To provide a standardized measure of an effect (defined as the relationship between two variables).	η^2 = _____.	Indicates the proportion of variance in the dependent variable accounted for by the independent variable.
Confidence Intervals			
CI for a Mean	To provide an interval estimate of the population mean. Can be derived from both the z and t distributions.	____% CI [___, ___]	Indicates that there is the given probability that the interval specified covers the true population mean.
CI for a Mean Difference	To provide an interval estimate of the population mean difference. Can be derived from both the z and t distributions.	____% CI [___, ___]	Indicates that there is the given probability that the interval specified covers the true population mean difference.
Significance Tests			
One Sample t Test	To compare a single sample mean to a population mean when the population standard deviation is not known	t(df) = ____, p = ____.	A small probability is obtained when the statistic is sufficiently large, indicating that the two means significantly differ from each other.
Independent Samples t Test	To compare two sample means when the samples are from a single-factor between-subjects design.		
Related Samples t Test	To compare two sample means when the samples are from a single-factor within-subjects design.		
One-Way ANOVA	To compare two or more sample means when the means are from a single-factor between-subjects design.	$F(df_1, df_2)$ = ___, p = ___.	A small probability is obtained when the statistic is sufficiently large, indicating that the set of means differ significantly from each other.
Repeated Measures ANOVA	To compare two or more sample means when the means are from a single-factor within-subjects design.		
Factorial ANOVA	To compare four or more groups defined by a multiple variables in a factorial research design.		

READINGS

- Laurel D. Riek – Wizard of Oz Studies in HRI: A Systematic Review and New Reporting Guidelines
- Aaron Steinfeld et al. – Common Metrics for Human-Robot Interaction
- Robin Murphy, Debra Schreckenghost – Survey of Metrics for Human-Robot Interaction
- Paul Baxter et al. – From Characterising Three Years of HRI to Methodology and Reporting Recommendations
- **Deadline: 15th of October 23:59**

github.com/gergely-magyar/humanoid_technologies

QUESTIONS?