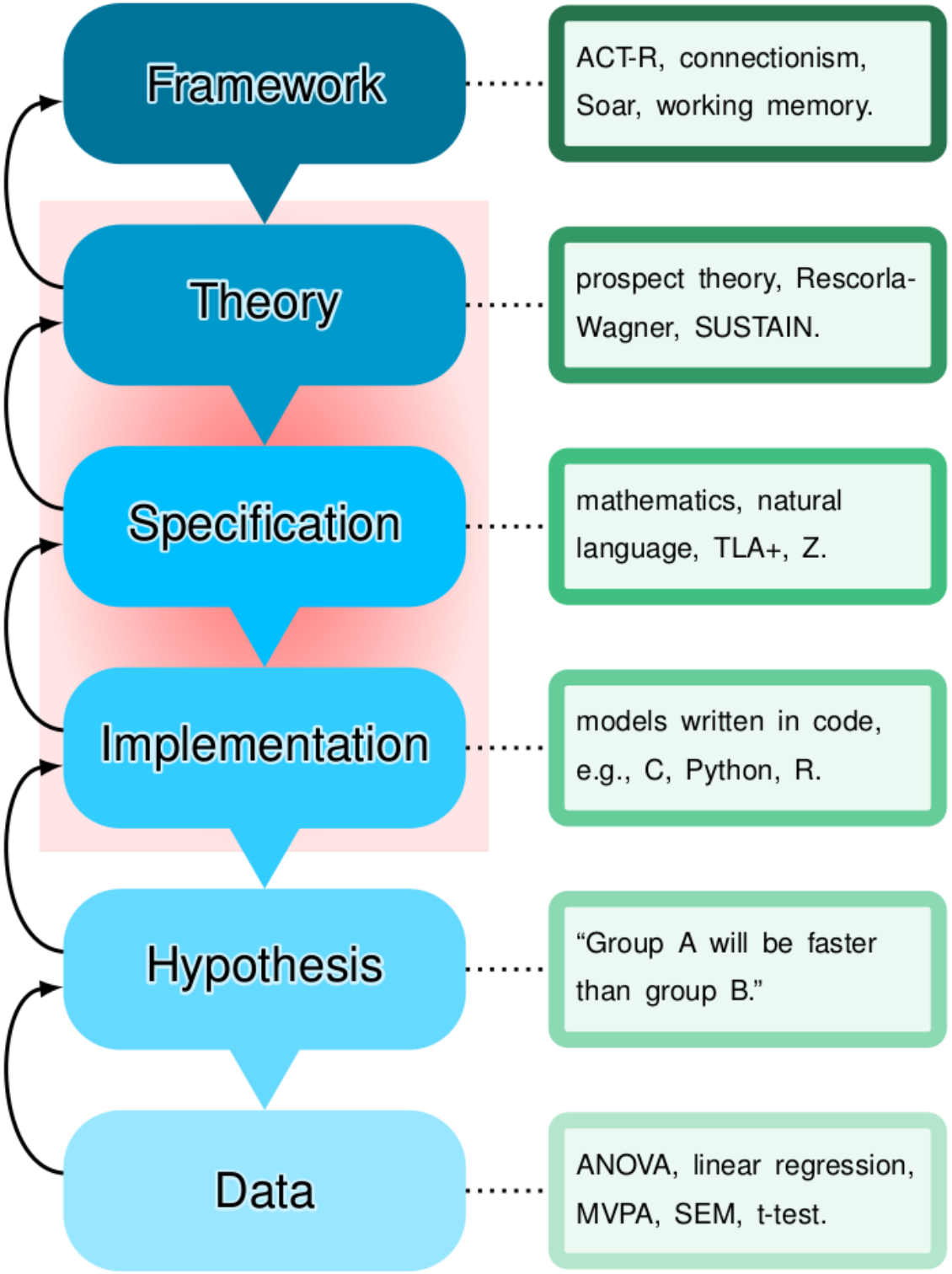
Model Evaluation

Since all models are wrong the scientist must be alert to what is importantly wrong. It is inappropriate to be concerned about mice when there are tigers abroad.

- George Box (1976)



Guest & Martin (2020)

Framework

- Connectionism
  - Parallel Processing
  - Distributed Representations
- Bayes Cognitive Models
  - Hypothesis Space
  - Prior Beliefs
  - Bayesian Updating

Theory

- Describing, Explaining, Predicting
  - Formal Language Theory
  - Set Theory
  - Words and Rules
  - Prospect Theory
  - Survival Analysis
  - Optimality Theory
  - Control Theory

Specification

- Core Assumptions of the Model
  - If the implementation does NOT meet these requirements, then it CANNOT be considered a valid implementation of our theory.
- Special Case: Linking hypothesis
  - How do the model predictions relate to the empirical behavior?

Implementation

- Auxilliary Assumptions
  - Existential Scope
    - Mutable commitments (e.g., noise is Gaussian)
  - Universal Scope
    - Insignificant commitments (e.g., C++)

Hypothesis

- Prevalence
  - X occurs greater than chance.
- Relationship
  - X is associated with Y.
- Causality
  - X precedes Y.
  - X is sufficient to cause Y.
  - X is necessary to cause Y.

Data

- Observation
- Simulation
- Experimentation

- Data are not theory neutral by default.

High Level Model Evaluation

van Rooij, I., & Blokpoel, M. (2020). Formalizing verbal theories. Social Psychology.

https://doi.org/10.1027/1864-9335/a000428

Low Level Model Evaluation

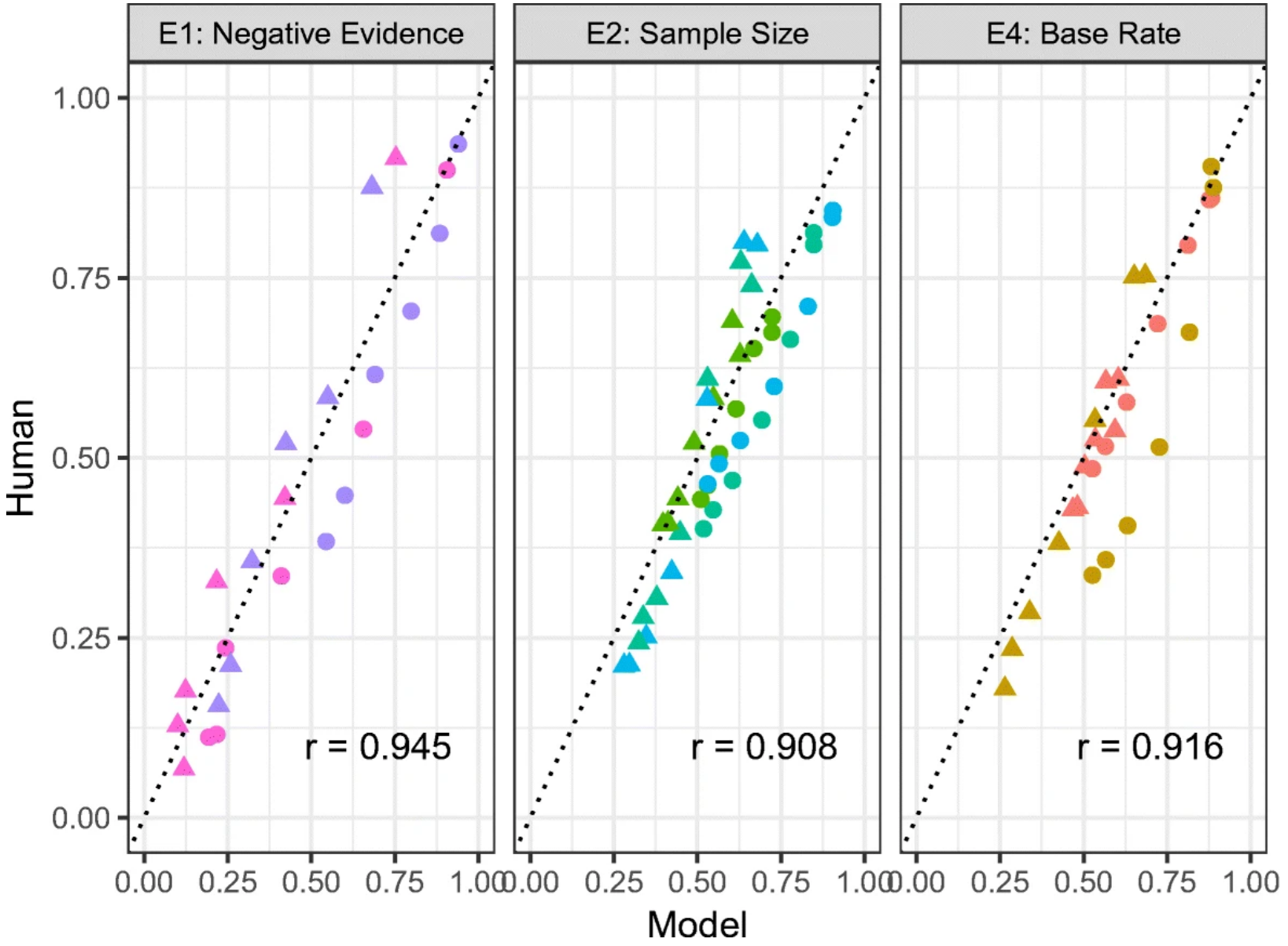
Statistics

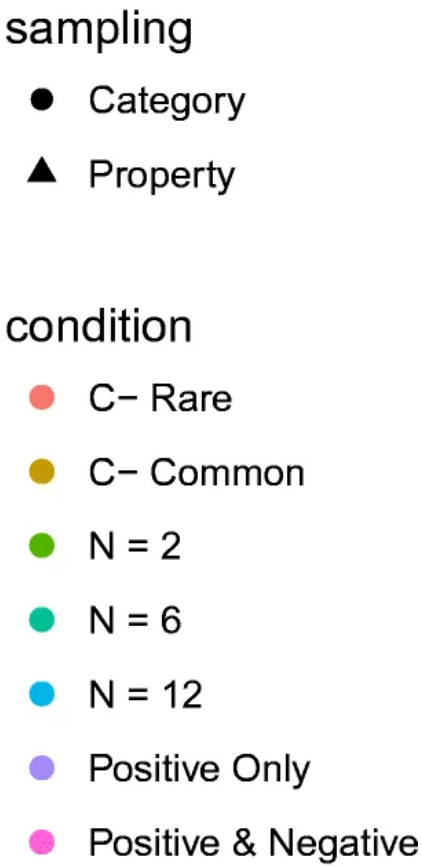
- Null Hypothesis Significance Test
- Bayes Factor
- AIC/BID
- MDL
- ...

Prediction

- Weak generalizaion
  - The training set and the evaluation set are drawn from the same generative model for the same task.
- Strong generalization
  - Out-of-distribution
  - Cross-task

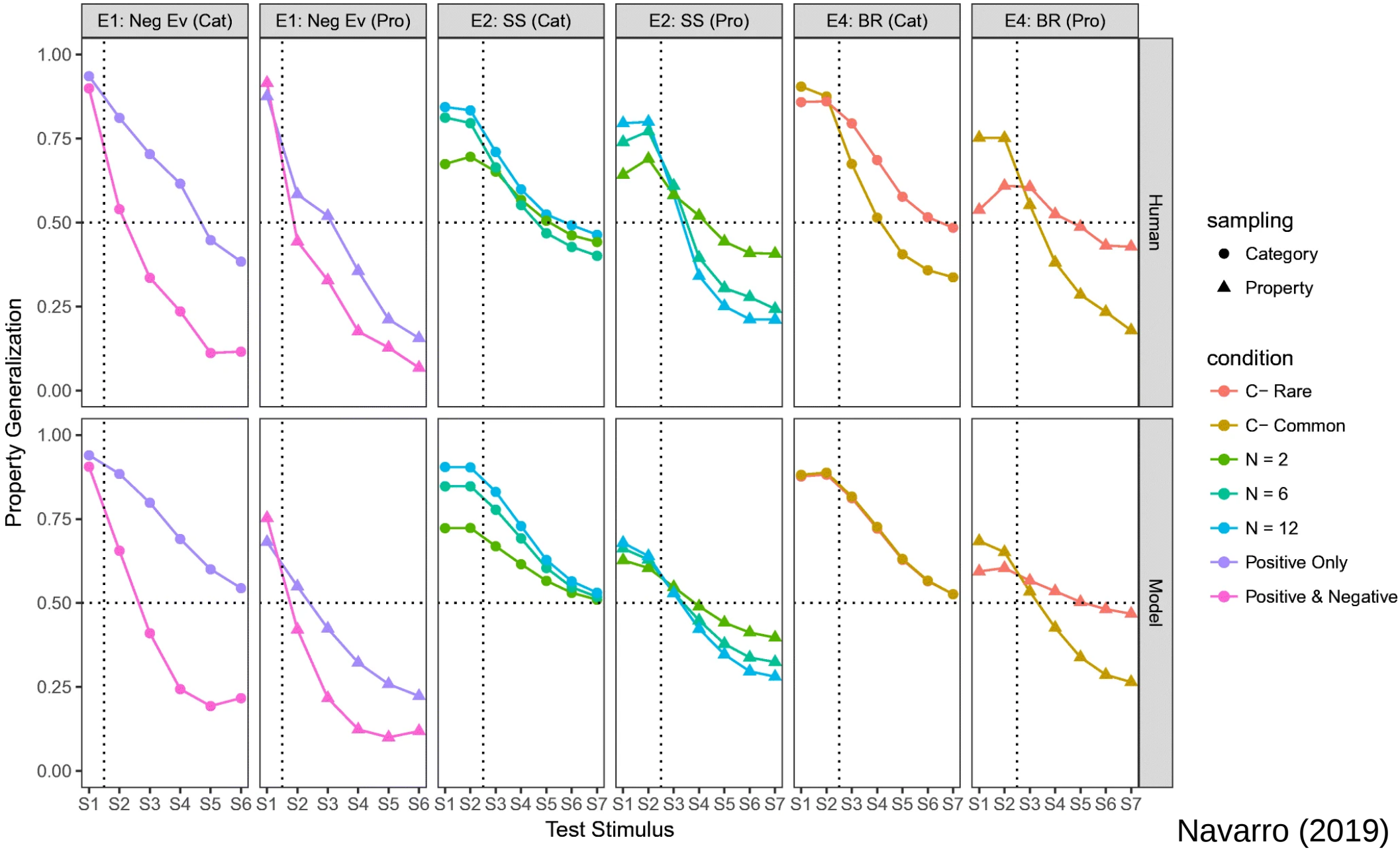
Explanation/Description





Navarro (2019)

Explanation/Description



Navarro (2019)

Take-homes

- Computational Modelling is a plurality of paths
- Model evaluation can occur at all points on the path---not just at the level of data.
- Model evaluation against data is very often not what we care about---yet sociologically it’s what we do.