Constructing a testable hypothesis

Readings for today

- van Rooij, I., & Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. PsyArXiv

- Platt, J. R. (1964). Strong inference. Science, 146(3642), 347-353.

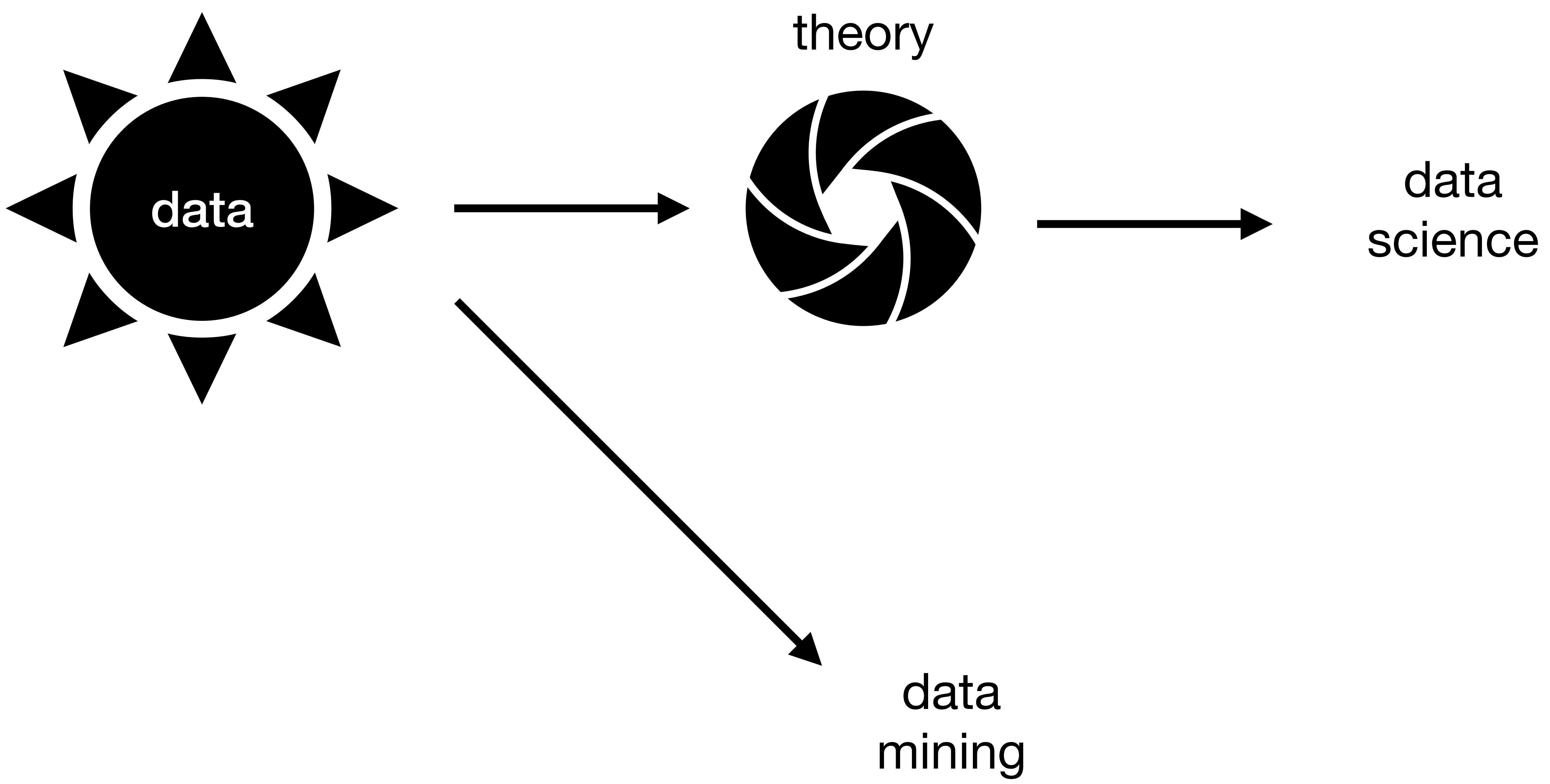
Topics

1. Theory vs. hypothesis
2. Strong inference
3. A *testable* hypothesis

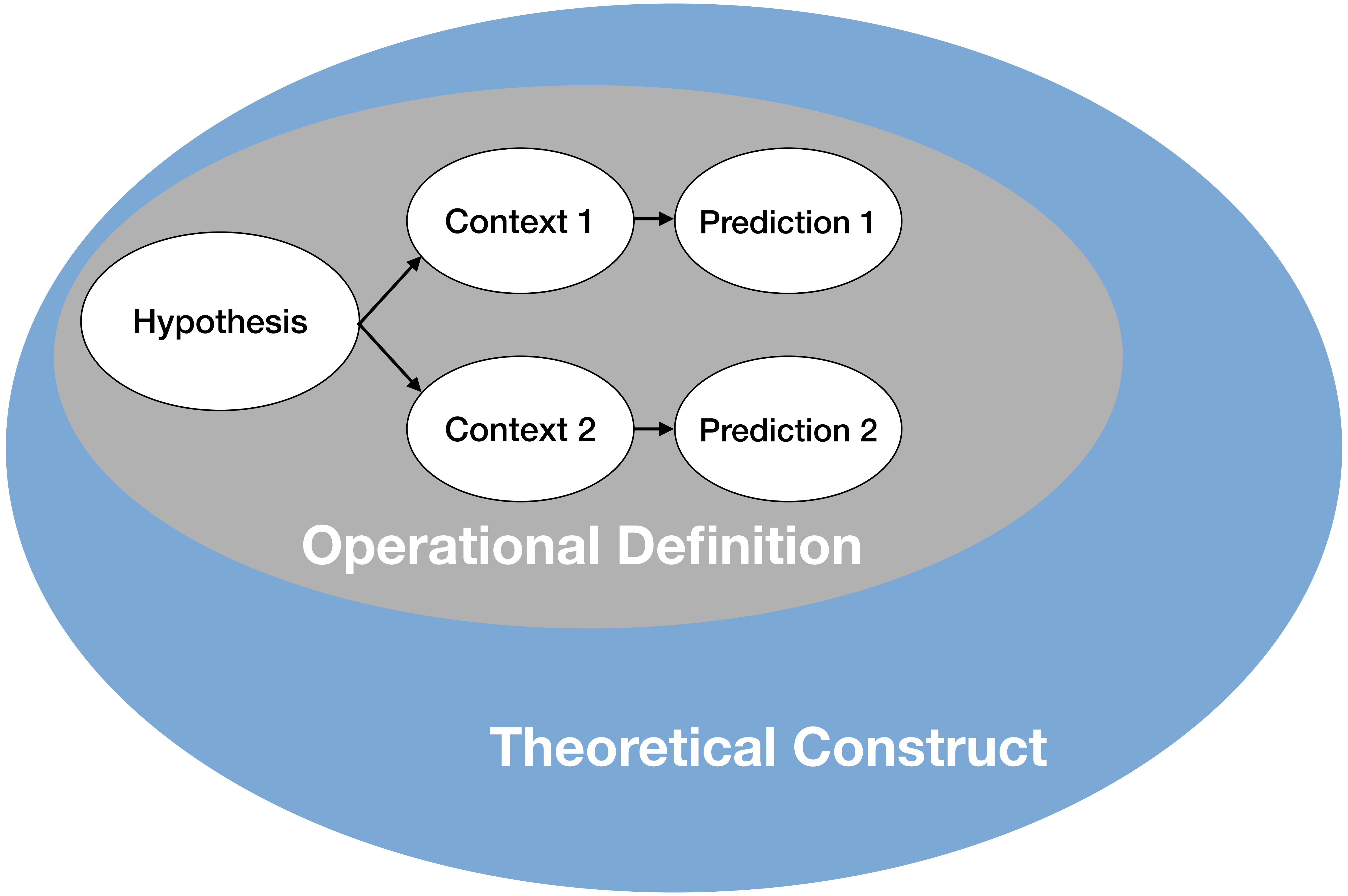
Theory vs. hypothesis

Data science is a *science*

Goal: Develop a clear and veridical understanding of the story *behind* your data by evaluating it from a theoretically driven perspective.



Refining focus from theory to tests



Theoretical Construct: A general description of a process or capacity (e.g., working memory)

Operational Definition: Reformulation of the theory in terms of a process that can be tested (e.g. number of items that can be recalled after a short delay).

Hypothesis: A testable statement about the operational definition as a set of relations (e.g., humans have an upper limit to the number of items they can recall after a 1min delay).

Context: Specific environment that the hypothesis is evaluated in (e.g., digit span task).

Prediction: Specific formulation of the hypothesis in a specific context (e.g., recall errors will increase as digit span increases).

Predictions as projections of theory

Strong inference

Strong inference

A method of inductive inference that is systematic in application

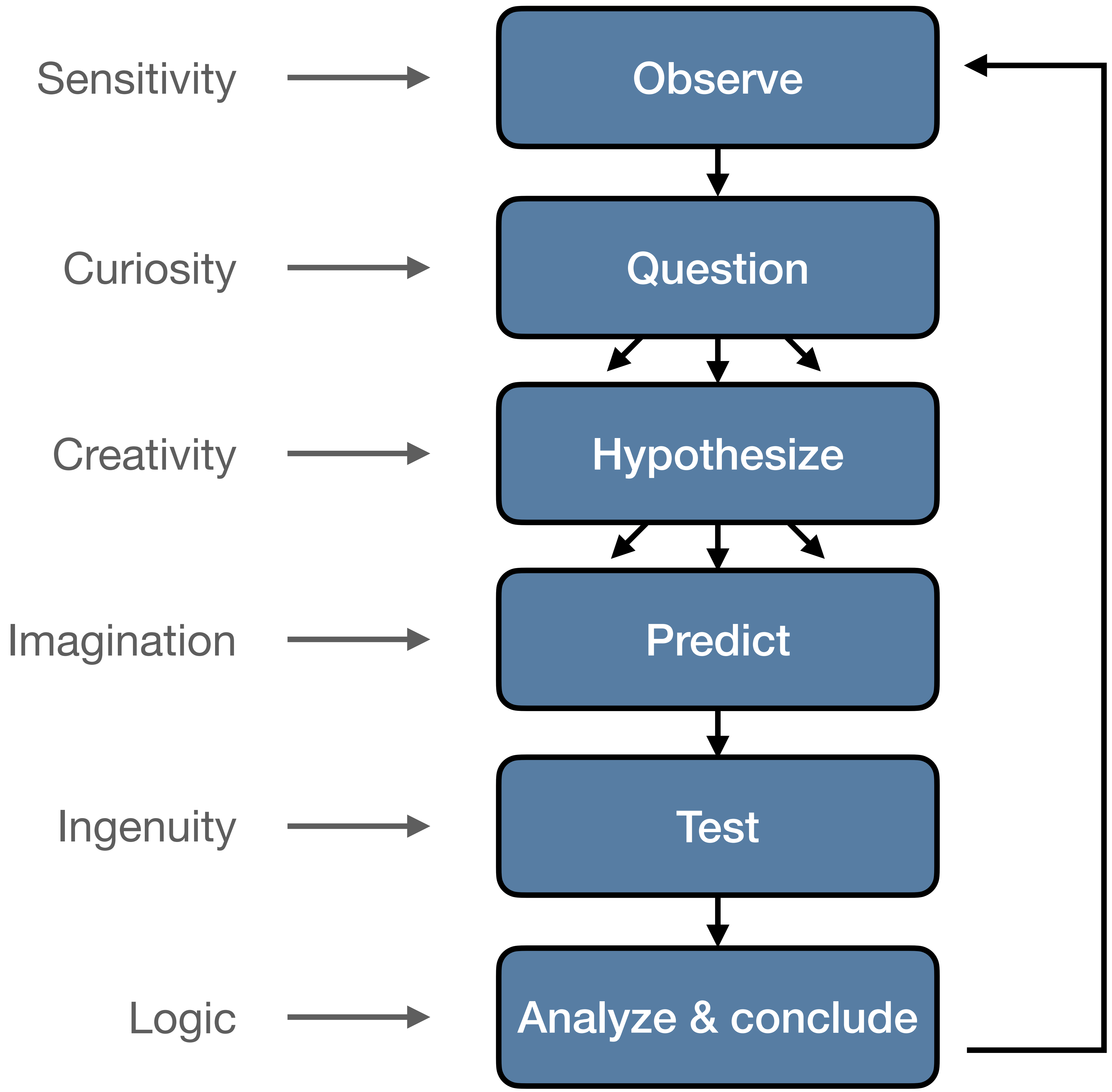
Steps:

1. Devise a ***set*** of hypotheses.
2. Devise an experiment with alternative outcomes, ***each of which can exclude one or more hypotheses***.
3. Carry out a test (e.g., experiment, statistical analysis) that ***can evaluate your full set of hypothesis***.

repeat

(Platt 1964)

Strong inference



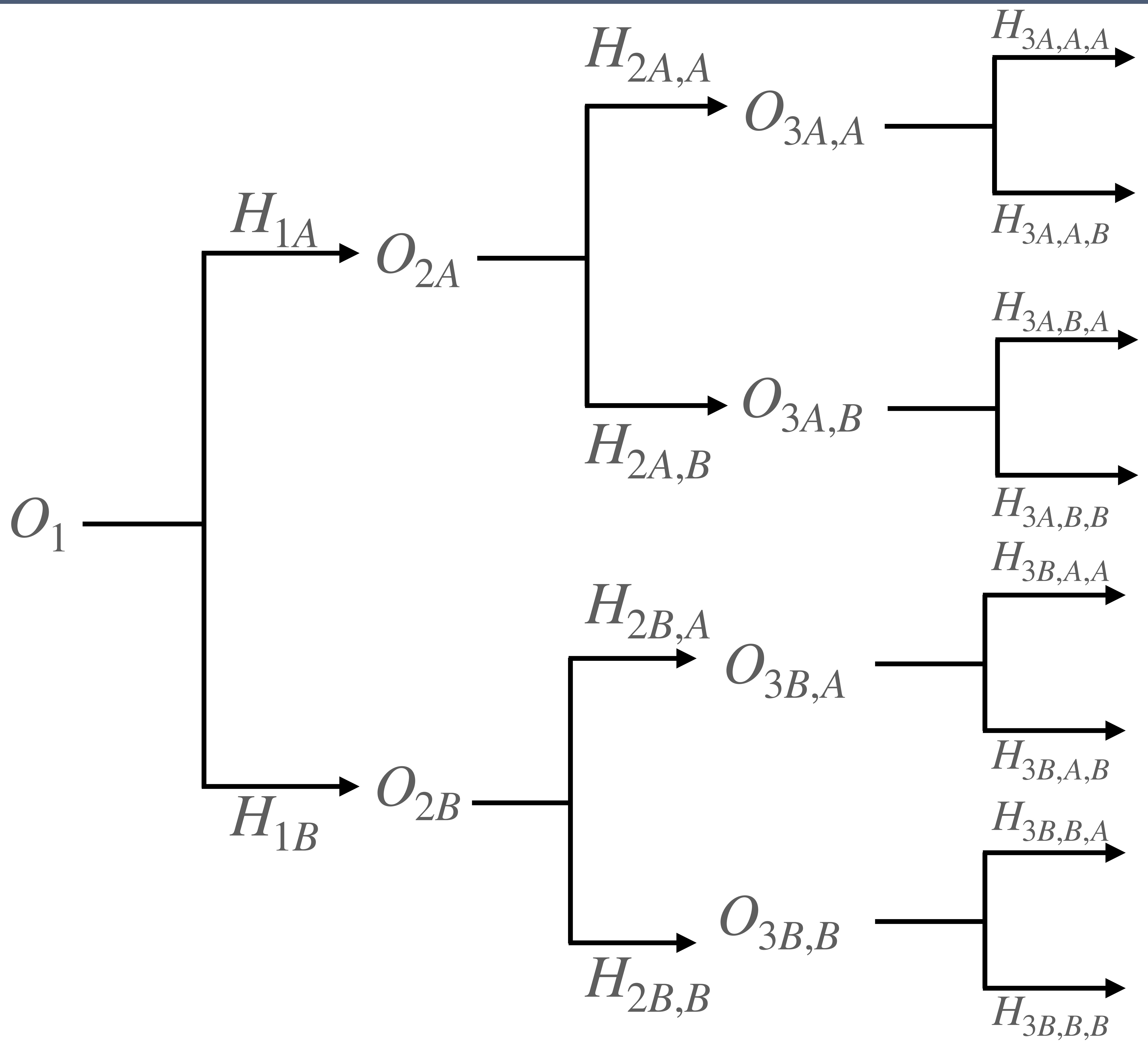
"A theory that cannot be mortally endangered cannot be alive" - Murray Gell-Mann & Yuval Ne'eman

<u>Goal</u>: Data and theory work together to evolve a strict set of axiomatic of statements that provide meaningful explanations & predictions.

<u>Key</u>: Constructing a *testable* hypothesis

(Platt 1964; Fudge 2014)

Conditional inductive trees



<u>Goal</u>: *A priori* thought experiment laying out a set of scenarios of “if/then” statements that lay out a logic for your empirical tests.

(Platt 1964)

A *testable* hypothesis

Popper's falsifiability

The only valid, testable hypotheses are those that are constructed so as to be falsifiable.

Just 1 counter point to disprove.

Falsifiable: All swans are white.

Unfalsifiable: There are black swans.

Exhaustive search required.

(Popper 1959)

Types of hypotheses

Null (H_0) Hypothesis: The hypothesis that needs to be rejected based on your theoretical premise.

- The status quo if your theory is wrong.

Research (H_i) Hypothesis: An alternative to the H_0 that is consistent in form to your theoretical premise

- One of many.

Theory → Hypothesis → Statistical Model

Fundamental form of a theory

$$\underset{\text{capacity}}{c} \leftarrow \overset{\text{theory}}{f}(\underset{\text{input}}{I}) = \underset{\text{output}}{O}$$

(van Rooij & Baggio 2020)

Fundamental form of a statistic

$$\underset{\text{output}}{Y} = \overset{\text{relational function}}{f}(\underset{\text{input}}{X})$$

$$f$$

The form of a statistical test, f, is a quantitative description of a specific hypothesis being evaluated (whether or not a p-value is calculated)

Theory → Hypothesis → Statistical Model

Theory: Hunger impairs working memory.

Operational Definitions:

1. Hunger is the bodily state that occurs when no food is consumed for 4 or more hours.
2. Working memory is how many items that can be recalled after a short period of time in the Digit Span task.

Hypotheses:

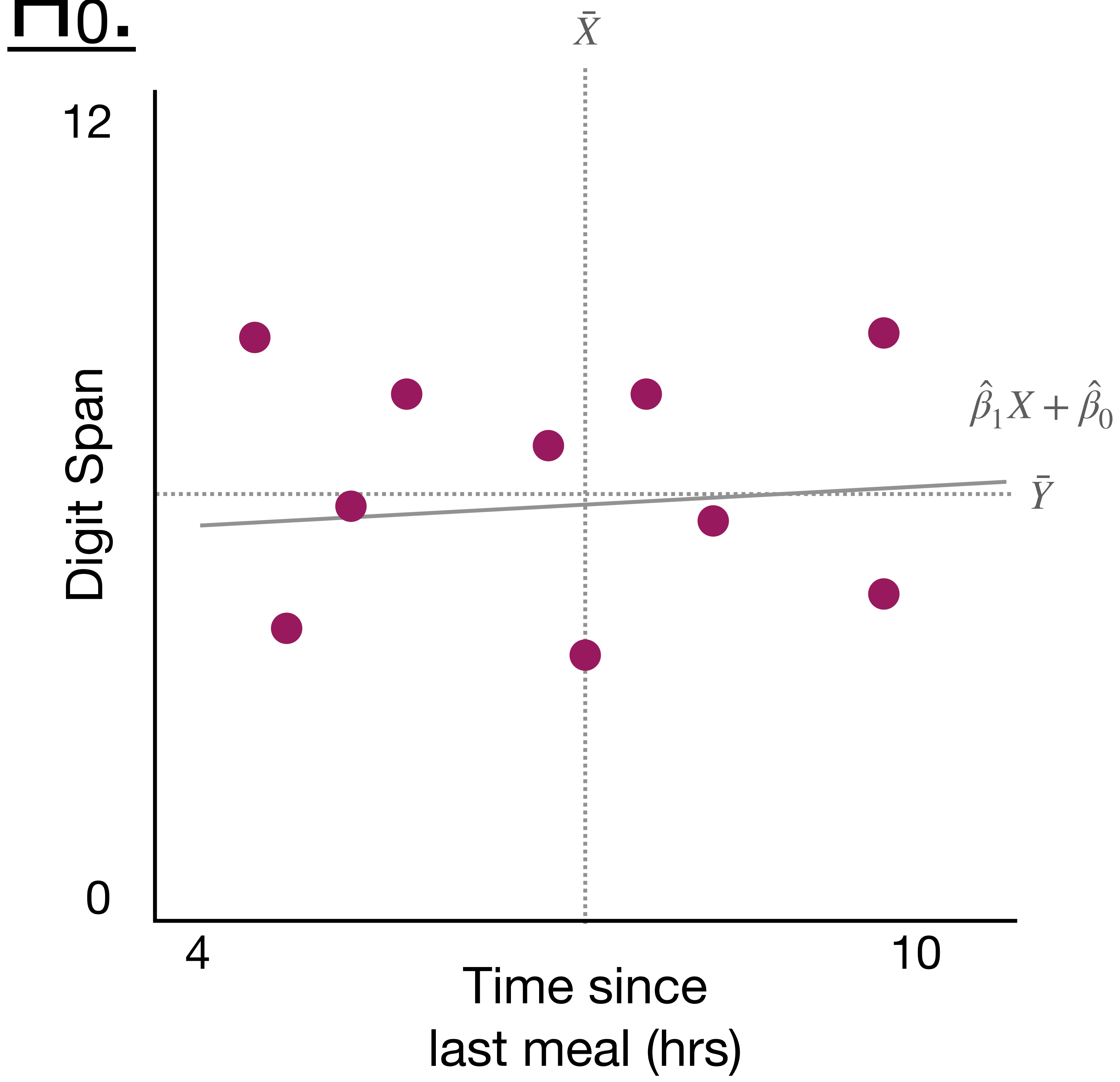
H_0: Digit recall does not change with hours since last meal.

H_A: Digit recall reduces with time since last meal.

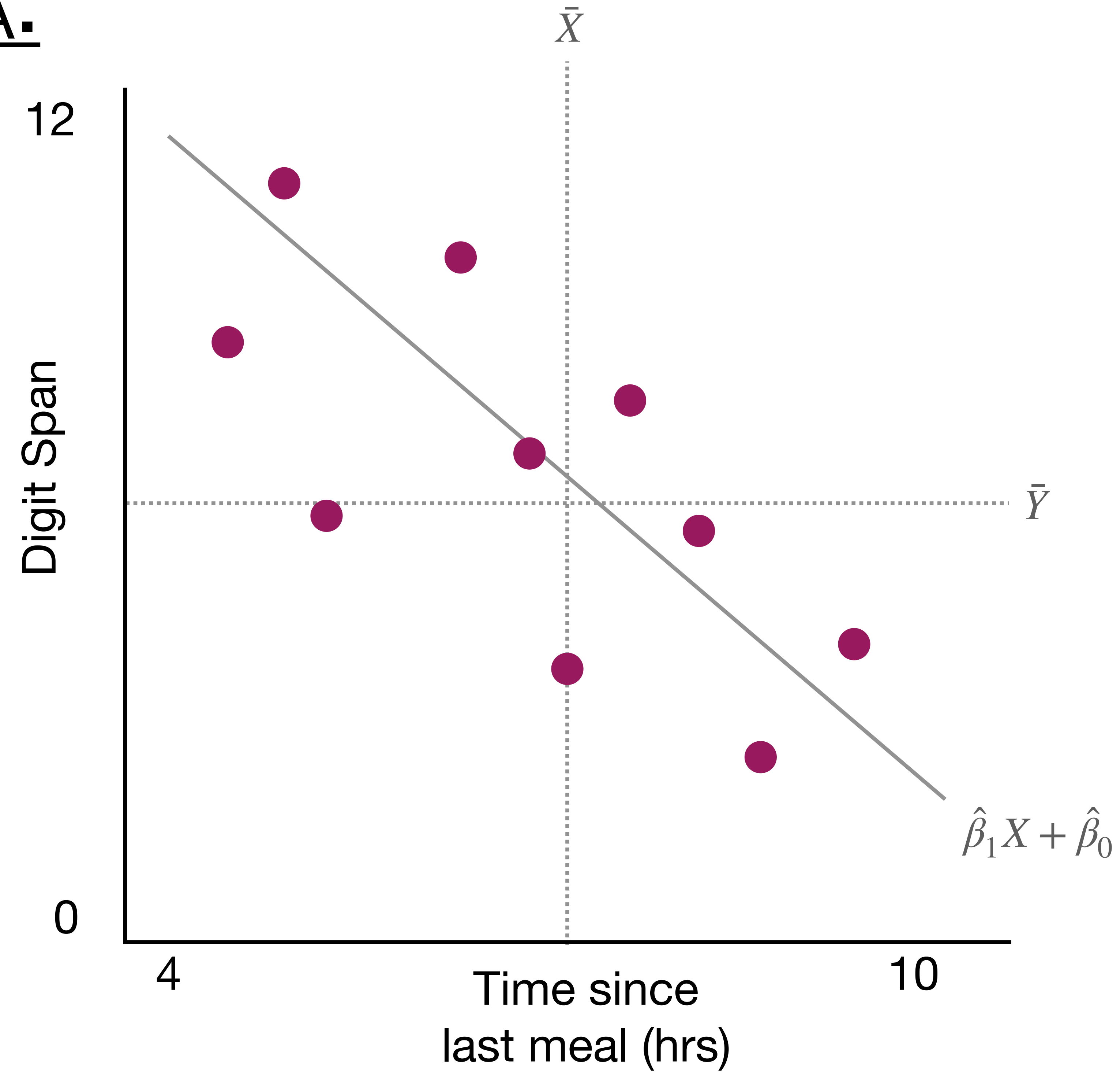
H_B: Digit recall increases with time since last meal.

Theory → Hypothesis → Statistical Model

H0:



HA:



Functional form:

$$Y_{ds} = \beta_1 X_{time} + \hat{\beta}_0$$

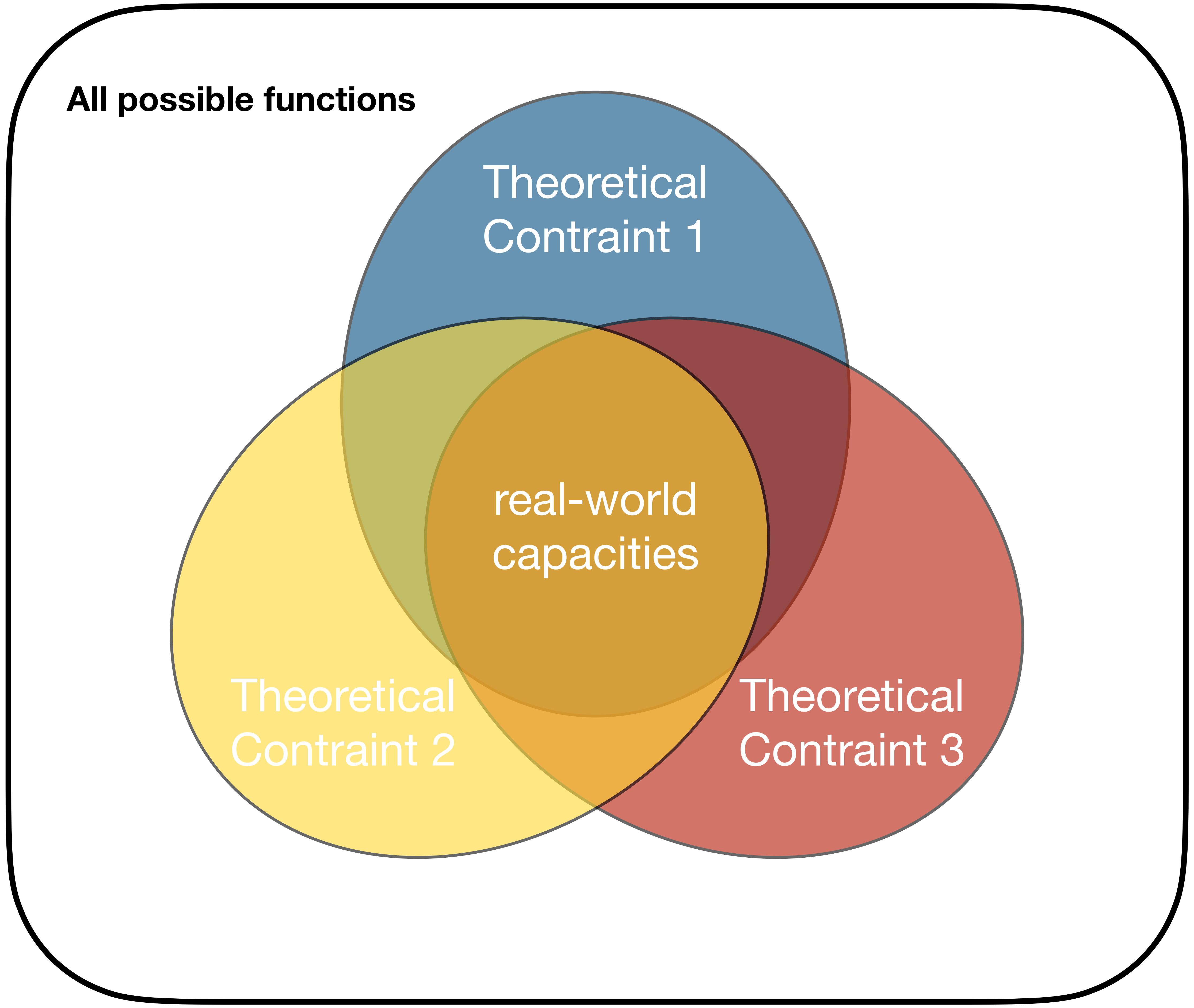
Hypotheses:

H_0: $\beta_1 = 0$

H_A: $\beta_1 < 0$

H_B: $\beta_1 > 0$

Hypotheses as models gives theoretical constraints



Hypothesis:

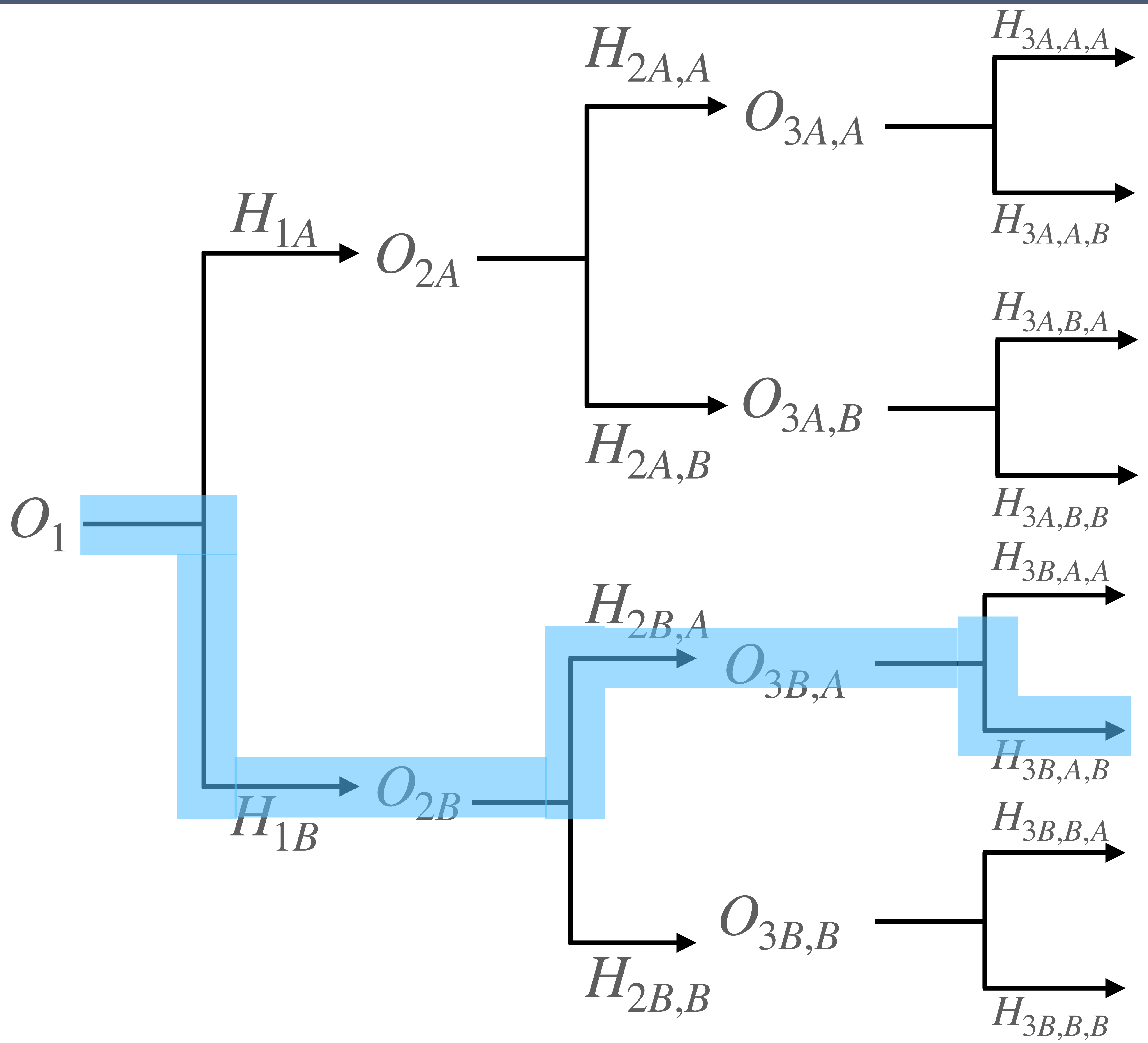
$$Y = \beta_1 X_1 + \beta_2 X_2 + \epsilon$$
$$\beta_1 > 0,\ \beta_2 < 0$$

Constraints:

1. Linear relationship.
2. Learnable (i.e., can be fit).
3. Stationarity of relations.
4. Normality of errors.

(van Rooij & Baggio 2020)

Conditional inductive trees (again)



Final model:

$$Y_{ds} = \beta_1 X_{time} + \beta_2 X_{sleep} + \beta_3 X_{exercise} + \epsilon$$

$$\beta_1 = -0.1,\ \beta_2 = -0.02,\ \beta_3 = 0.3$$

(Platt 1964)

Take home message

- Thinking about your hypotheses in their statistical form gives you the formalism necessary for (1) open theorizing, (2) testability, & (3) strong inference.