

How do we understand an intelligent system?

Readings for today

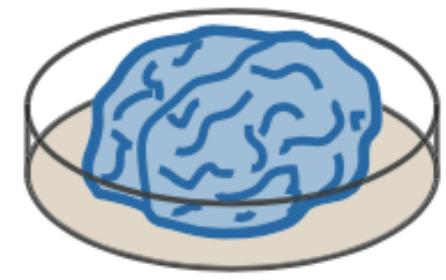
- Krakauer, J. W., Ghazanfar, A. A., Gomez-Marin, A., MacIver, M. A., & Poeppel, D. (2017). Neuroscience needs behavior: correcting a reductionist bias. *Neuron*, 93(3), 480-490.
- van Rooij, I., & Baggio, G. (2020). Theory before the test: How to build high-verisimilitude explanatory theories in psychological science. *PsyArXiv*

Topics

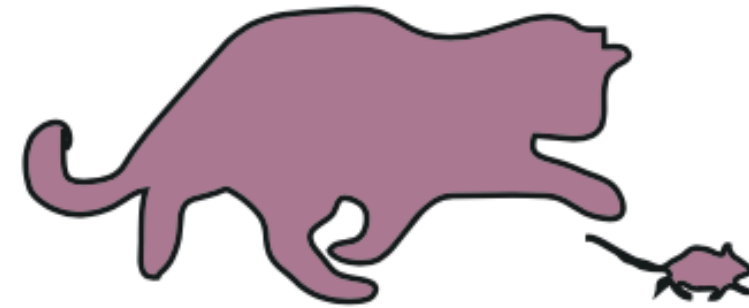
- Ways of understanding behavior
- Structure of a theory (of behavior)

Ways of understanding behavior

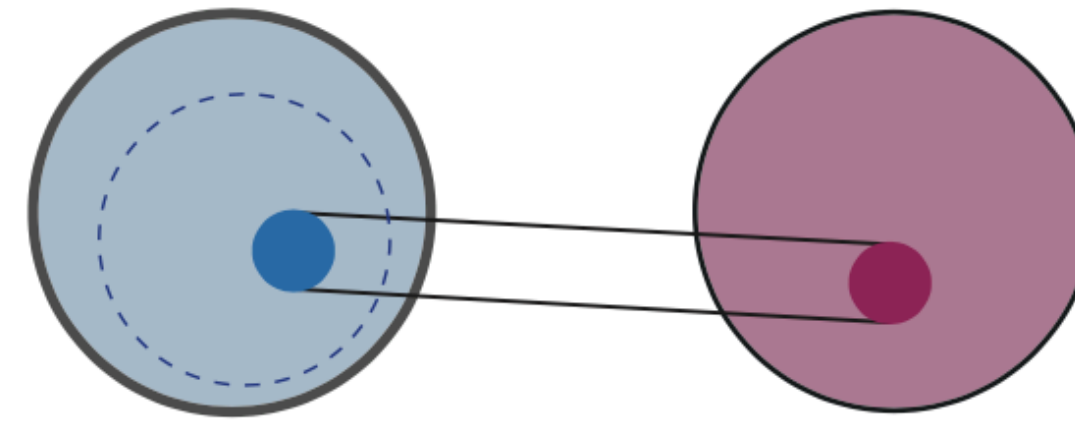
Multiple realizations of brain → behavior



Neural activity patterns

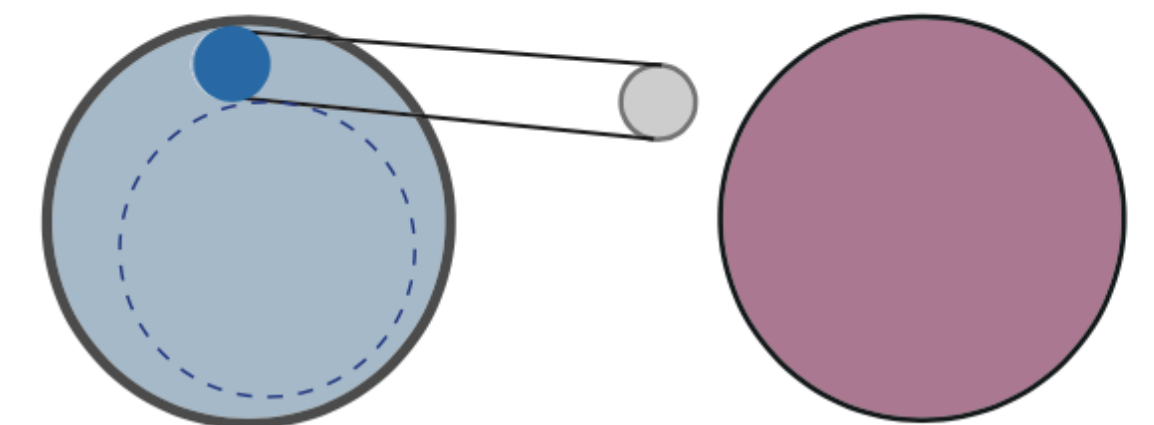


Natural behaviors



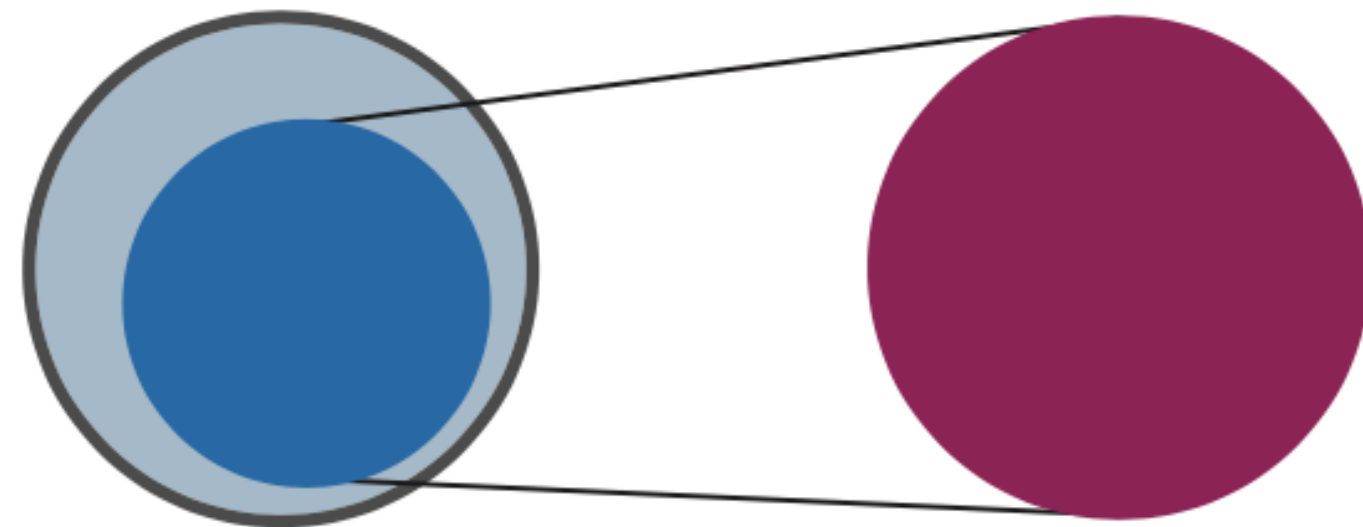
Smaller subset of activity patterns

Subset of natural behaviors



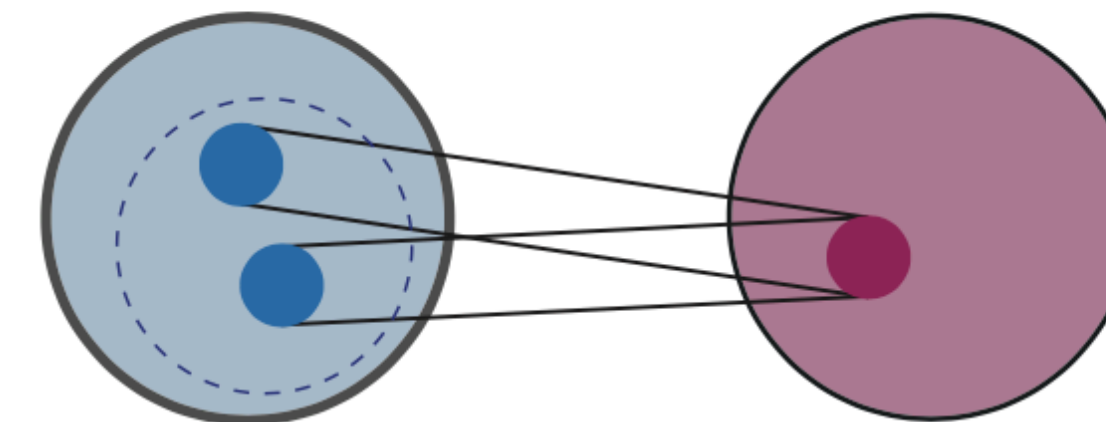
Unnatural activity pattern

Behavior outside natural repertoire



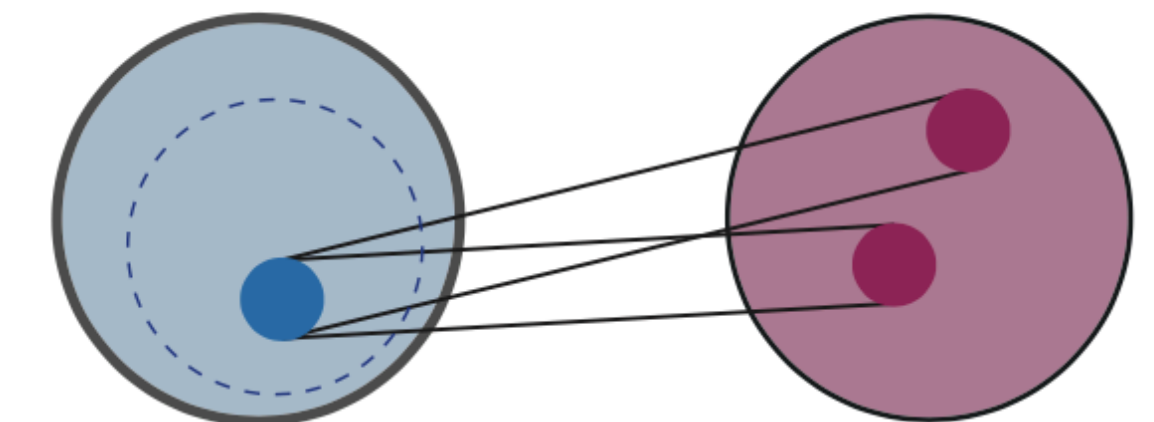
Subset of possible activity patterns

All natural behaviors



Multiple possible patterns of activity

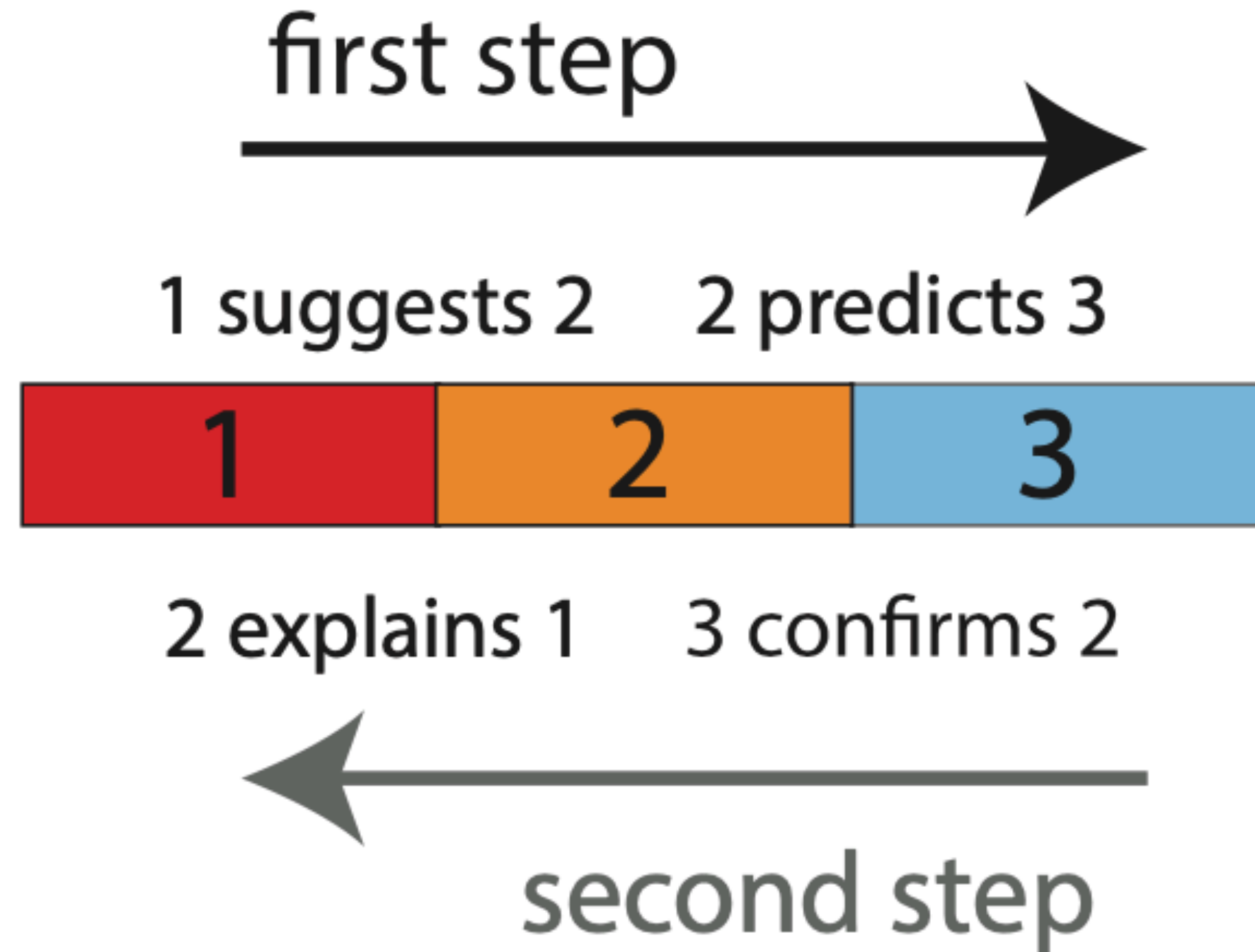
Single natural behavior



Single pattern of activity

Multiple natural behaviors

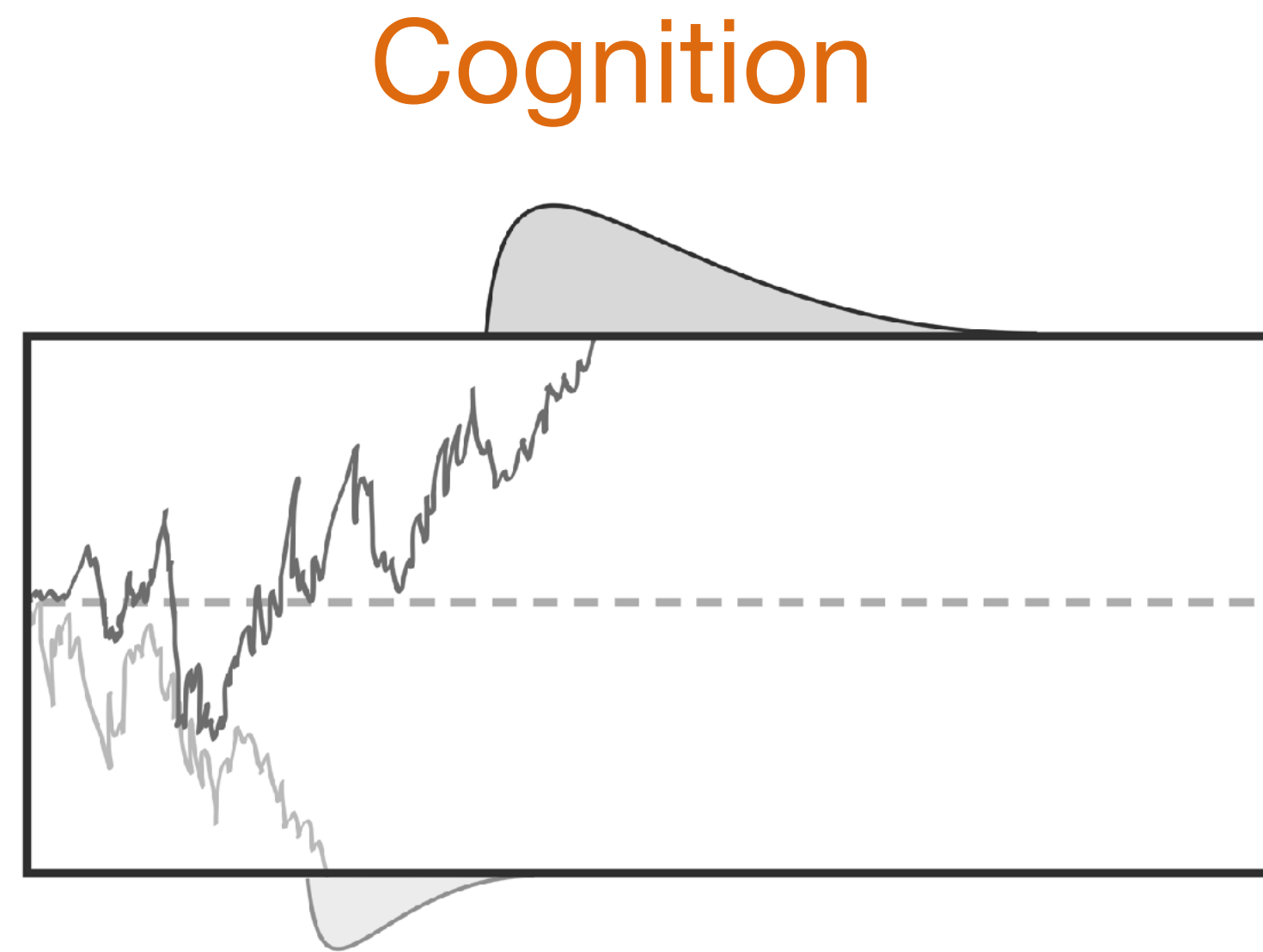
Marr's levels of analysis



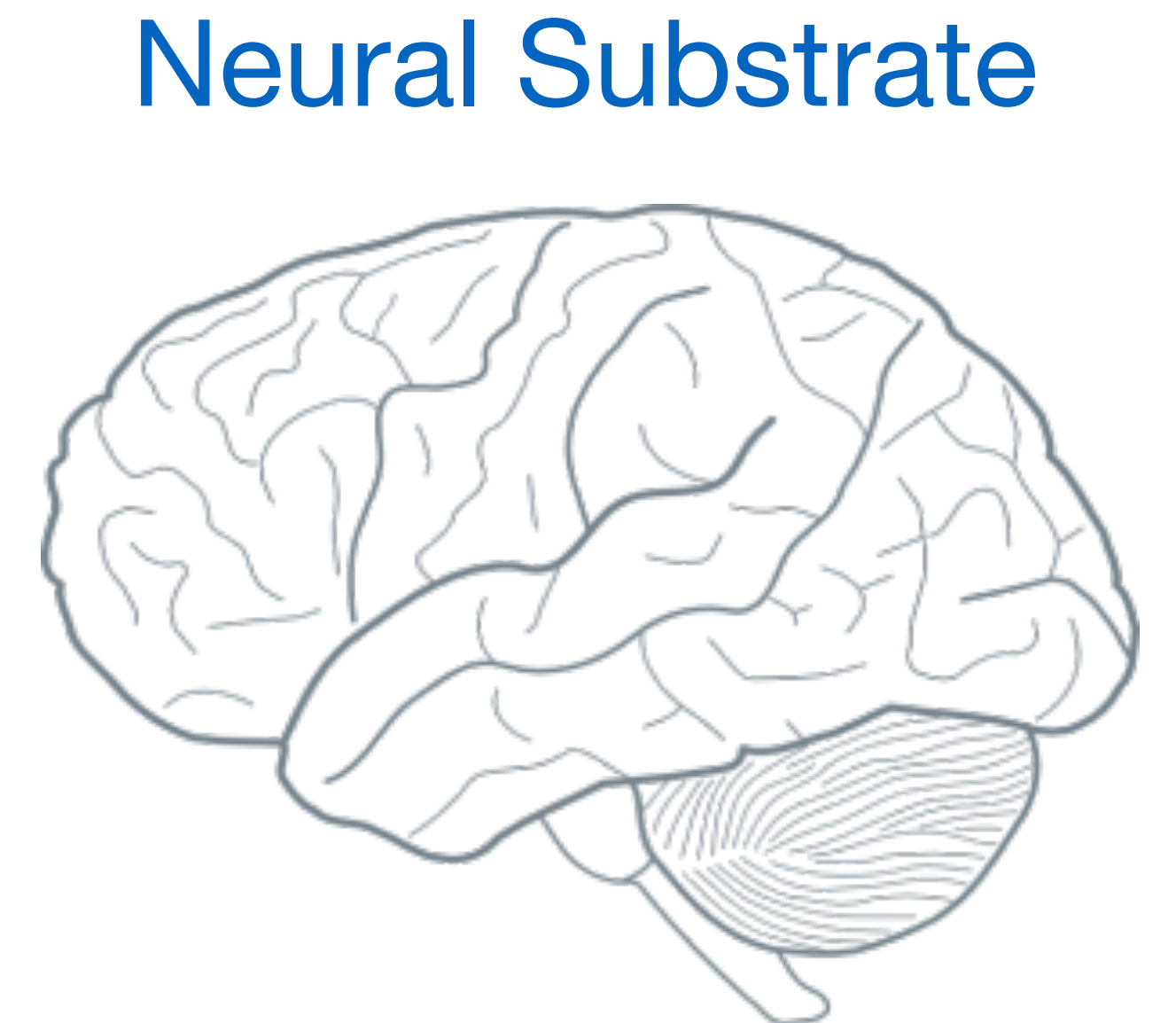
Levels of analysis: intelligence



Computation 1

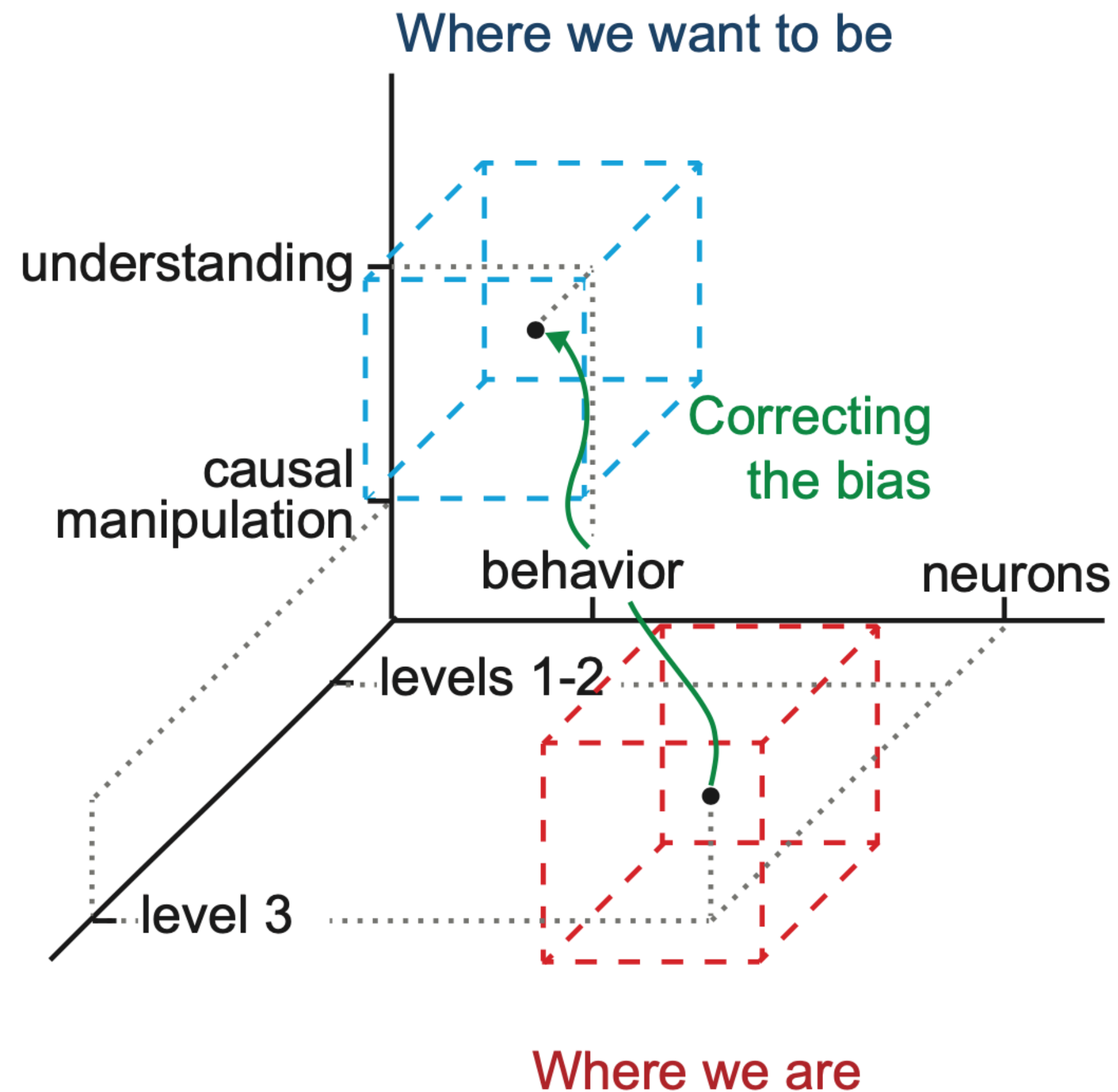


Algorithm 2



Implementation 3

Where we need to be

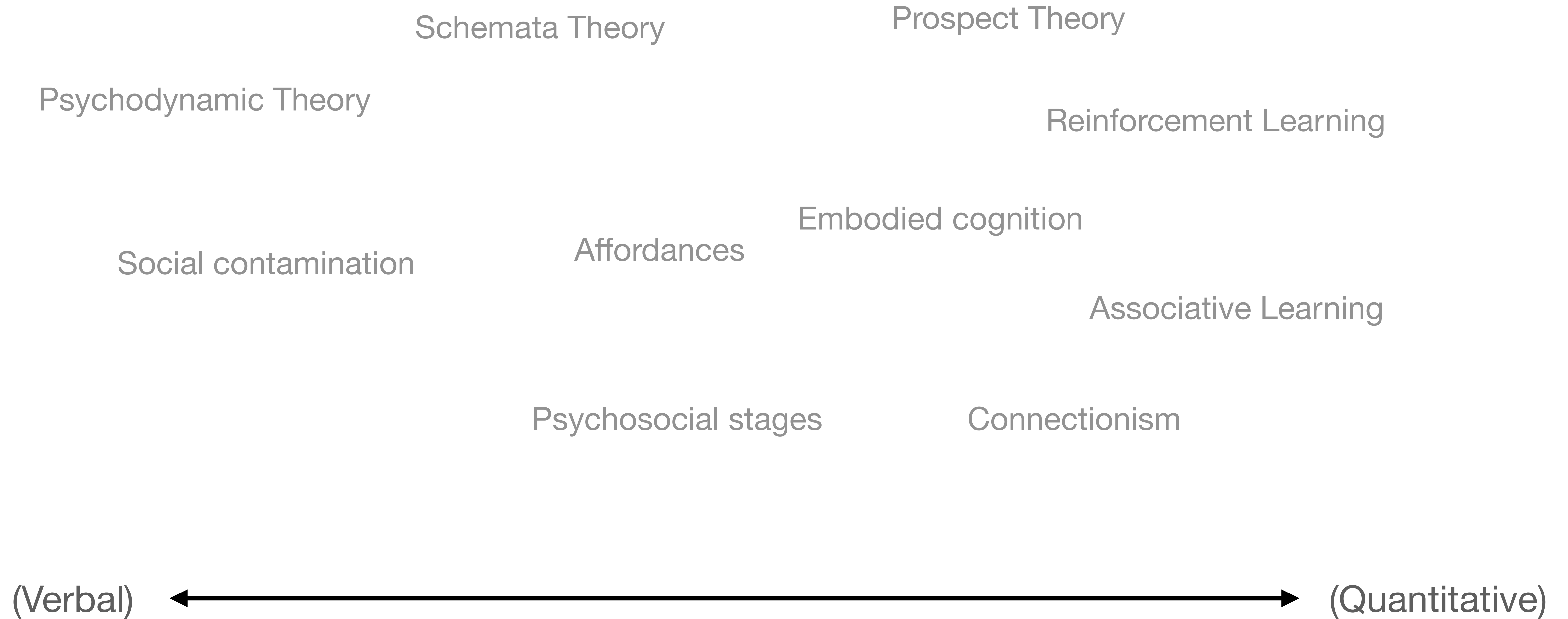


“Understanding something is not the same as just describing it or knowing how to intervene to change it.”

Theory

Structure a theory (of behavior)

The theories we have



What is a theory?*

* in psychology & neuroscience

Theory: A description of a set of *capacities*.


Informal Building a description based on a
theory: collection of observed effects.

Formal Constructing a description using
theory: formal logic *prima facie* via a
constructive strategy.

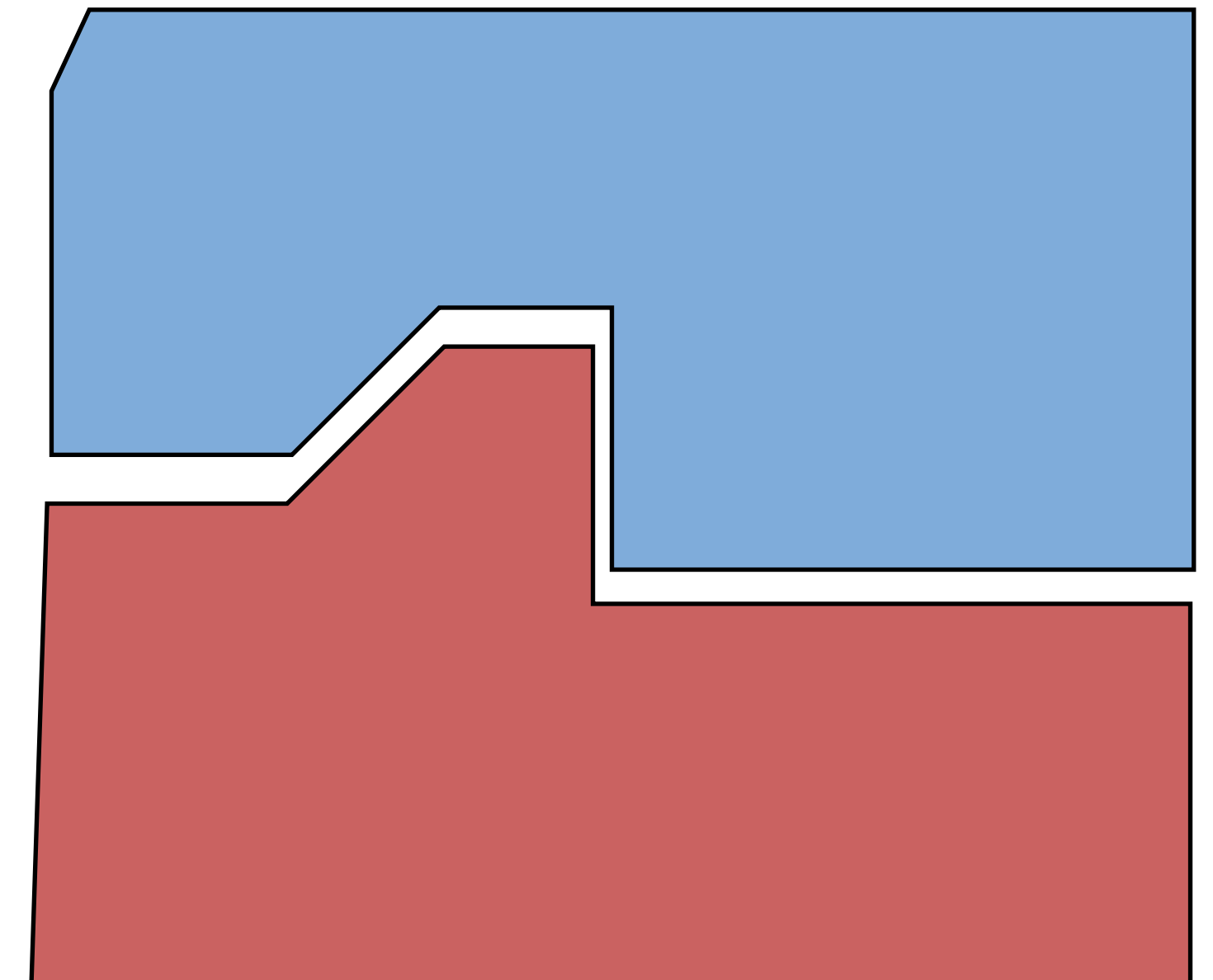
1) Plausibility constraints

2) Theoretical cycle

Plausibility constraints

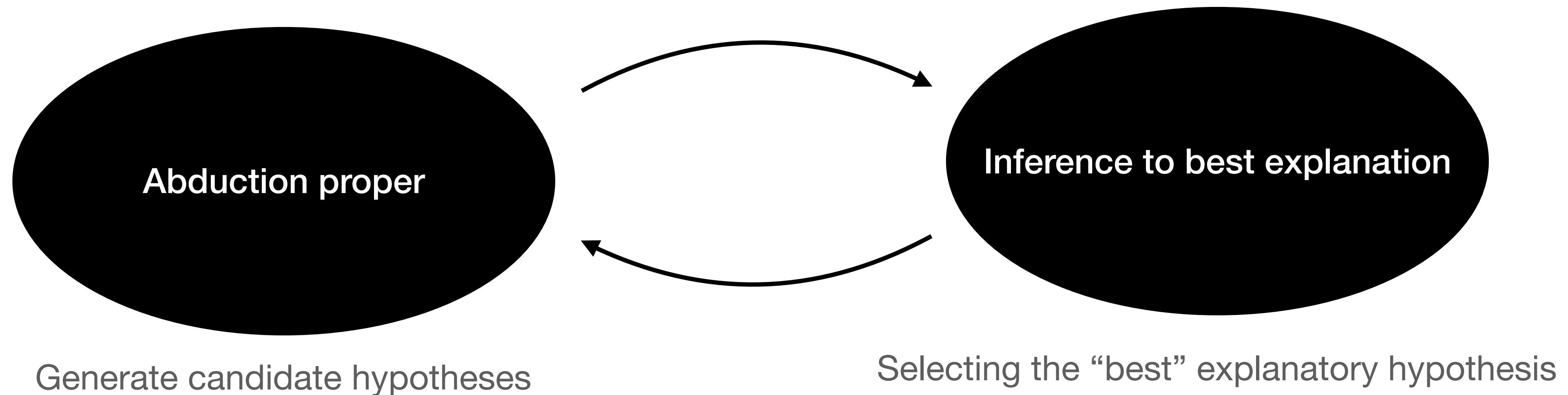
Assumptions:

1. Theory must provide a means for making rigorous tests possible.
2. Should restrict the number and types of theories/hypotheses considered for testing.



How to build a theory, f , of capacities, c ?

Abduction: Reasoning from observations to generate possible explanations.



Structural form of a theory

$$\begin{array}{ccccc} & & \text{theory} & & \\ & & \downarrow & & \\ \underset{\substack{\uparrow \\ \text{capacity}}}{\mathcal{C}} & \leftarrow & \underset{\substack{\uparrow \\ \text{input}}}{f(I)} & = & \underset{\substack{\uparrow \\ \text{output}}}{O} \end{array}$$

e.g. $O = f(I) = \beta_1 I_1 + \beta_2 I_2 + \epsilon$

$$O = f(I) = \beta_1 I_1^2 + g(I_2) + \epsilon$$

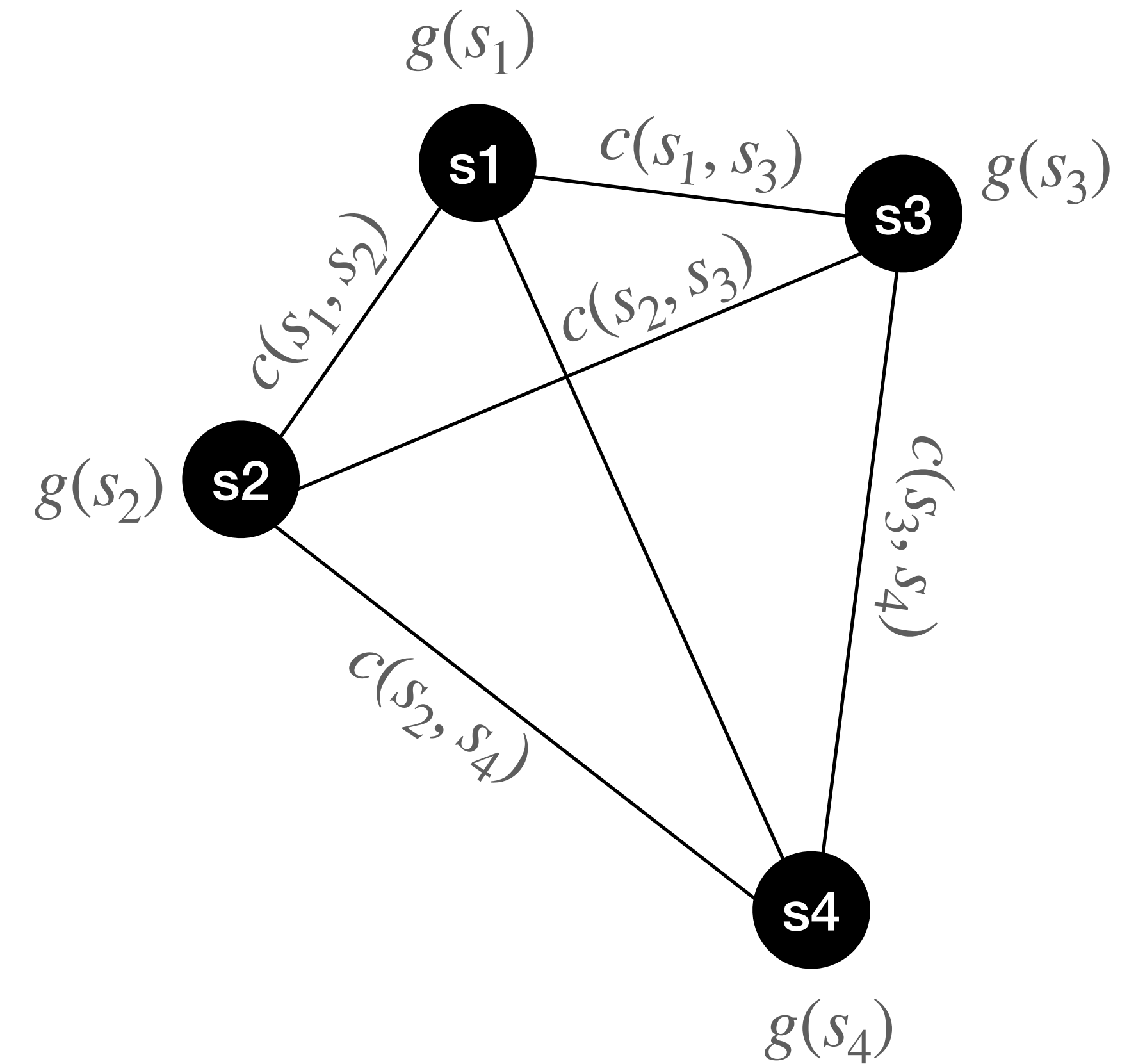
Example

Foraging f

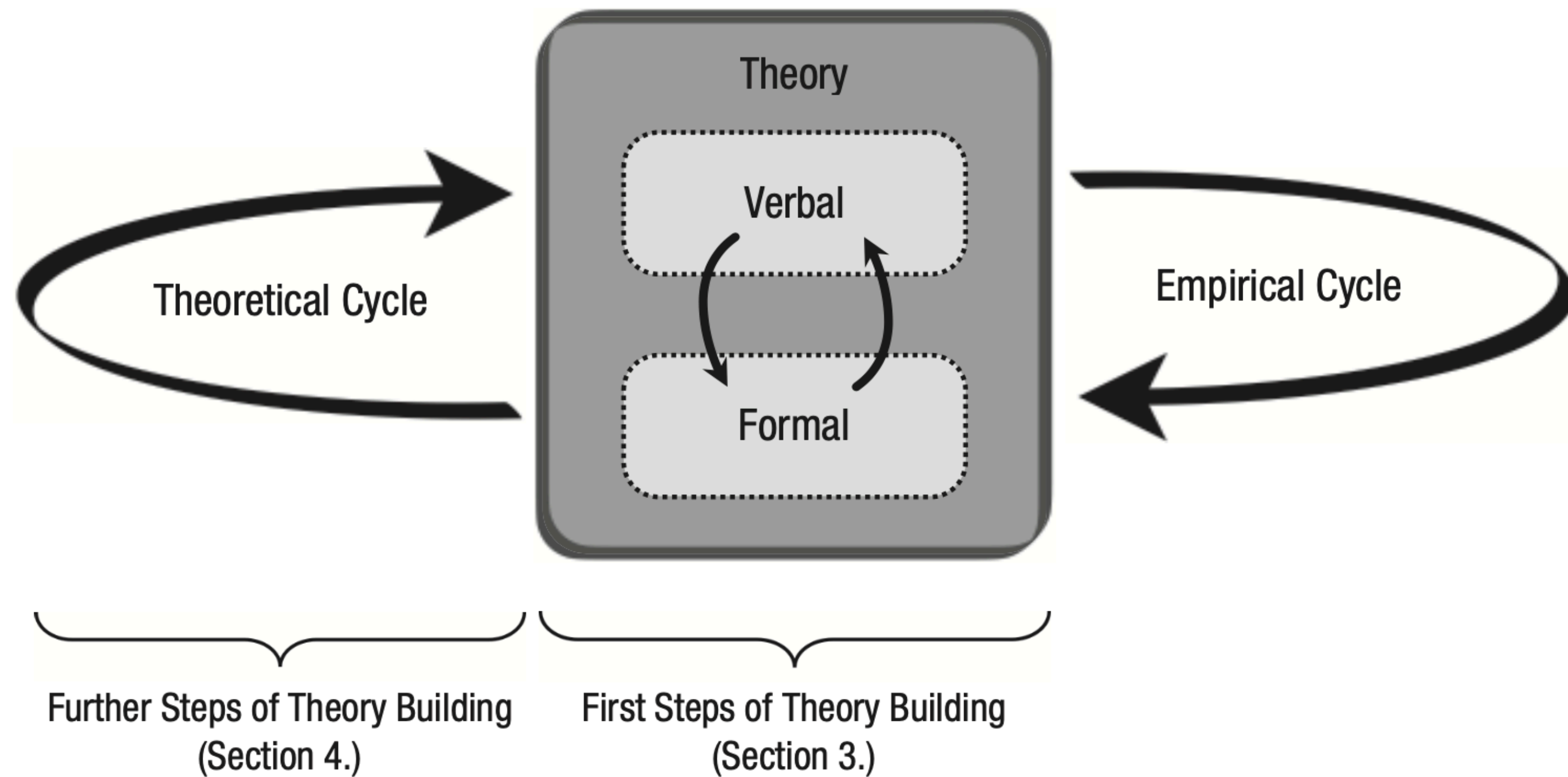
Input: A set of sites $S = \{s_0, s_1, s_2, \dots, s_n\}$, each site $s_i \in S$ with $i > 0$ hosts a particular amount of food $g(s) \in \mathbb{N}$, and for each pair of sites $s_i, s_j \in S$ there is a cost of travel $c(s_i, s_j) \in \mathbb{N}$.

Output: An ordering $\pi(S) = [s^0, s^1, \dots, s^n, s^0]$ of the elements in S such that $s^0 = s_0$ and the sum of foods collected at s^1, \dots, s^n exceeds the total cost of the travel, i.e.,

$$c \leftarrow f(S) = \sum_{s \in S} g(s) \geq c(s^n, s^0) + \sum_{s^i, s^{i+1} \in \pi(S)} c(s^i, s^{i+1})$$

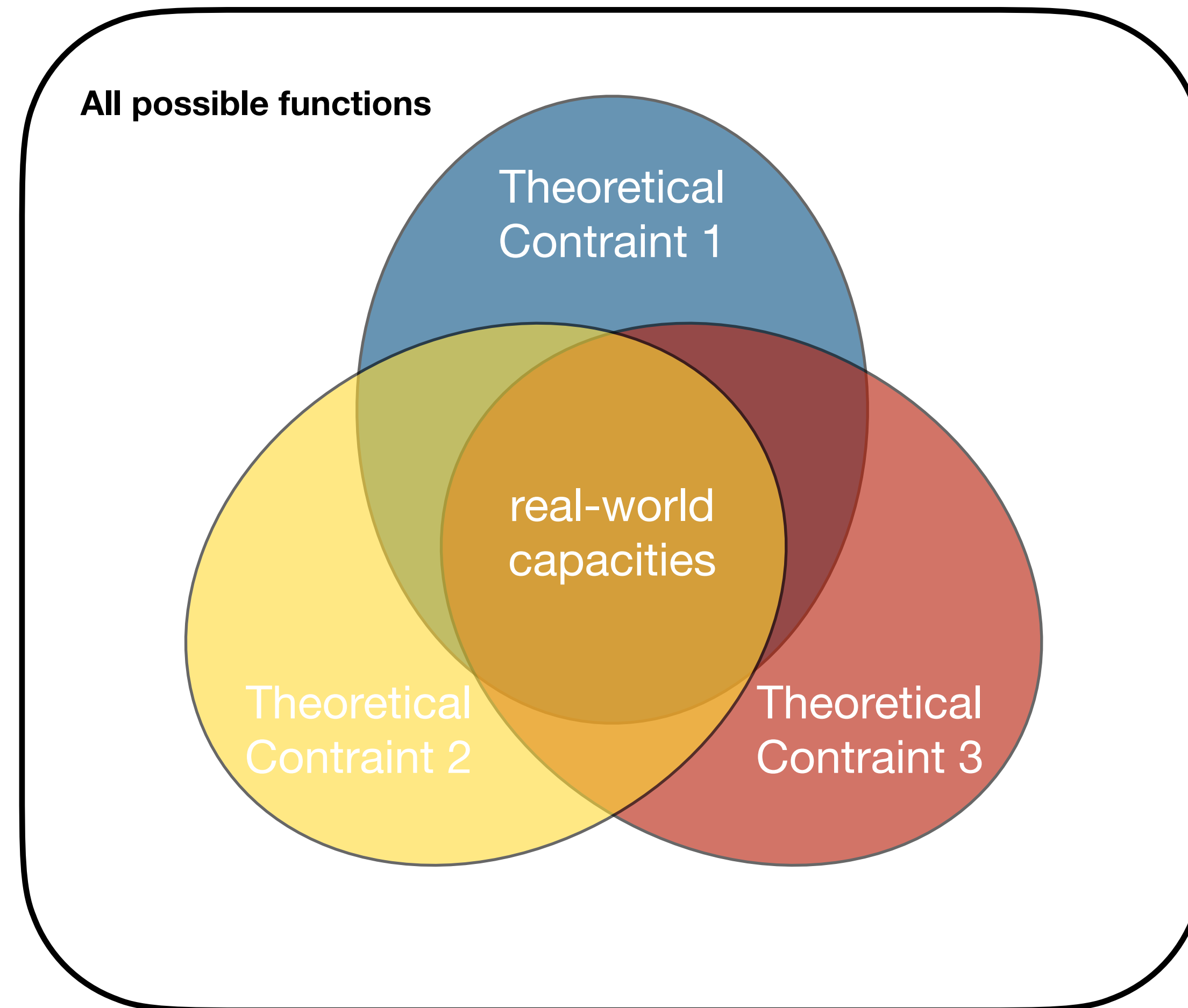


Evolution of a theory



- Start with an informal verbal theory to set conceptual frame.
- Operationalize it to a formal structure to make hypotheses (abduction)
- Design tests to evaluate the hypotheses.
- Use empirical results to refine the form of your theory.

Reducing the space of possible theories



Food for thought

- Does the multiple realizability problem for brain→behavior relationships mean that we can ever come up with a meaningful theory of a behavior?
- How do you see van Rooij & Baggio's constraints on theory structure being informed by Marr's levels of analysis?
- How is Krakauer et al.'s argument for how investigations should progress across levels of analysis compatible or incompatible with van Rooij & Baggio's theory cycle?