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

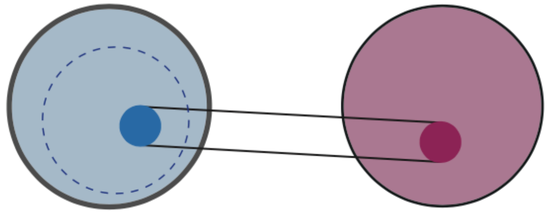
Topics

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

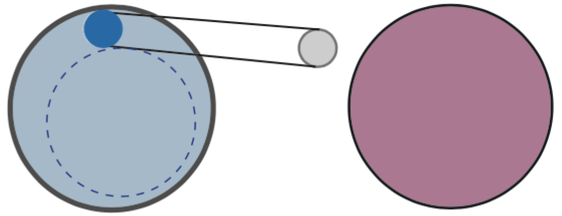
Ways of understanding behavior

Multiple realizations of brain → behavior

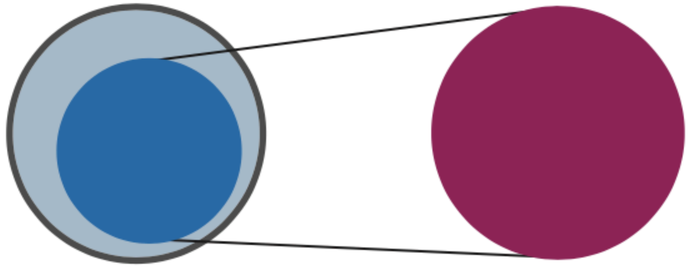
Neural activity patterns Natural behaviors



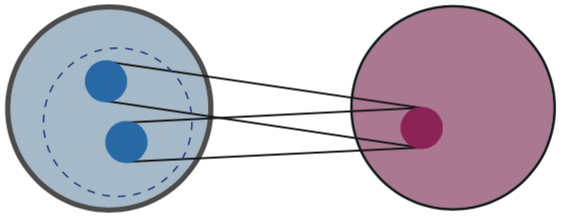
Smaller subset of activity patterns Subset of natural behaviors



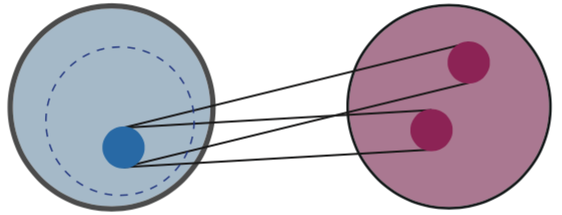
Unnatural activity pattern Behavior outside natural repertoire



Subset of possible activity patterns All natural behaviors



Multiple possible patterns of activity Single natural behavior



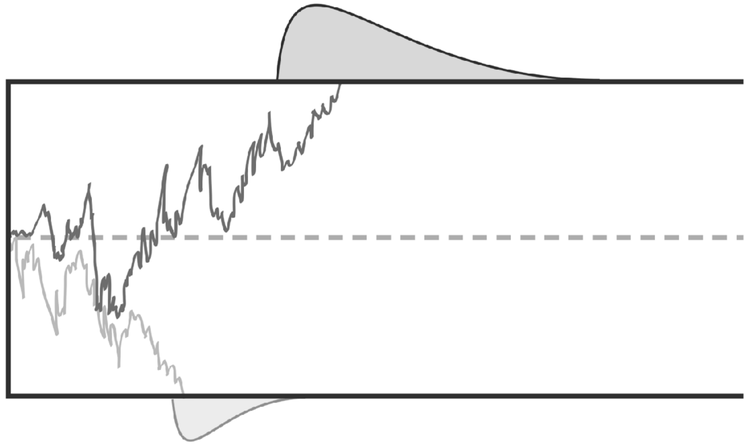
Single pattern of activity Multiple natural behaviors

(Krakauer et al. 2017)

Levels of analysis: intelligence

Behavioral Goal

Cognition



Neural Substrate

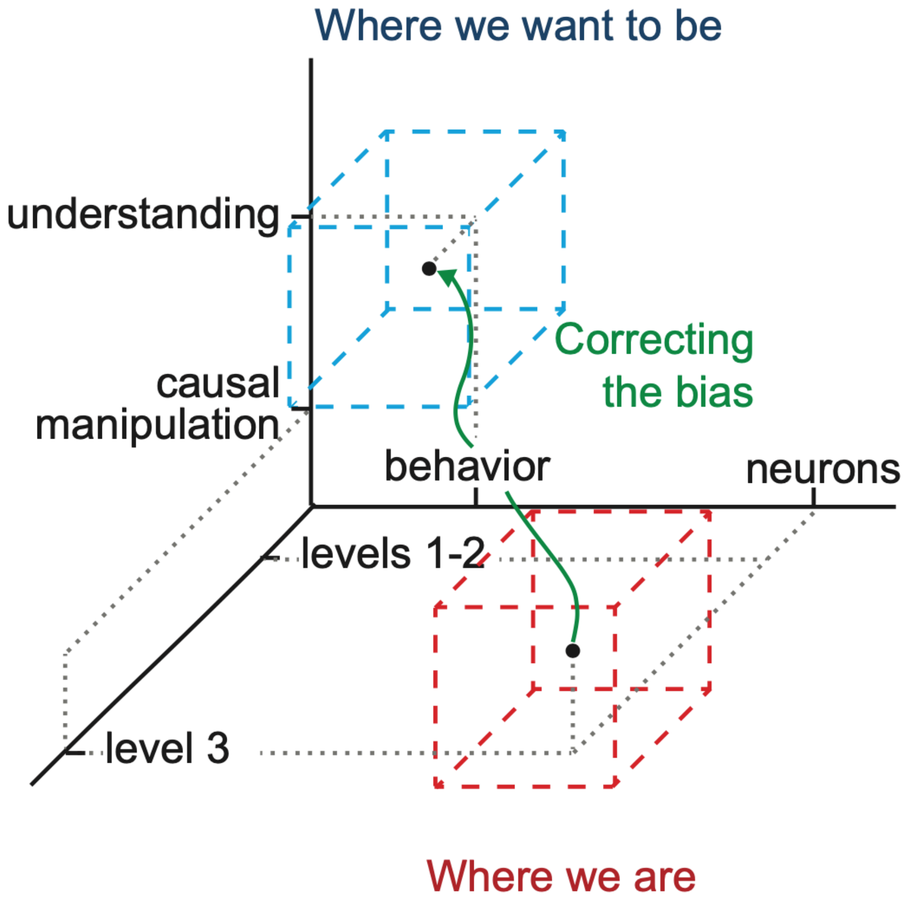
Computation 1

Algorithm 2

Implementation 3

(Krakauer et al. 2017)

Where we need to be



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

→ Theory

(Krakauer et al. 2017)

Structure a theory (of behavior)

The theories we have

Schemata Theory

Prospect Theory

Psychodynamic Theory

Reinforcement Learning

Embodied cognition

Social contamination

Affordances

Associative Learning

Psychosocial stages

Connectionism

What is a theory?*

* in psychology & neuroscience

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

Set of abilities

Primary explananda (things to be explained)

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

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

1) Plausibility constraints

2) Theoretical cycle

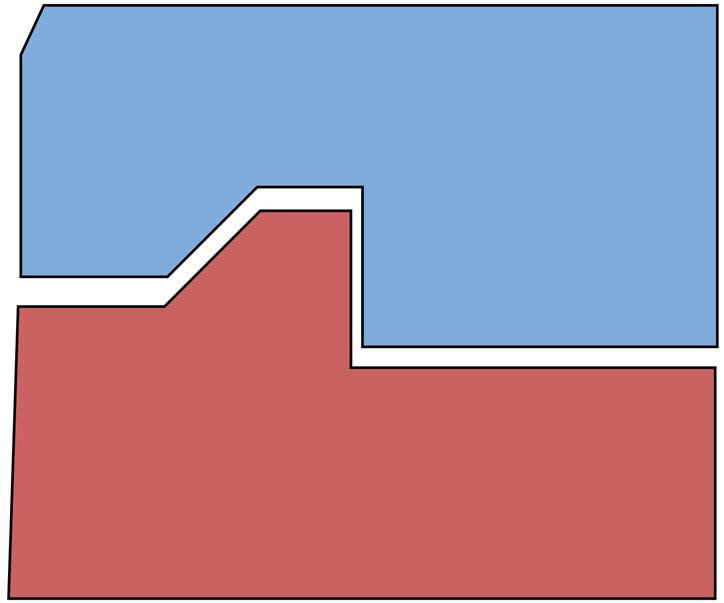
(van Rooij & Baggio 2020)

Plausibility constraints

Assumptions:

1. Theory must provide a means for making rigorous tests possible.

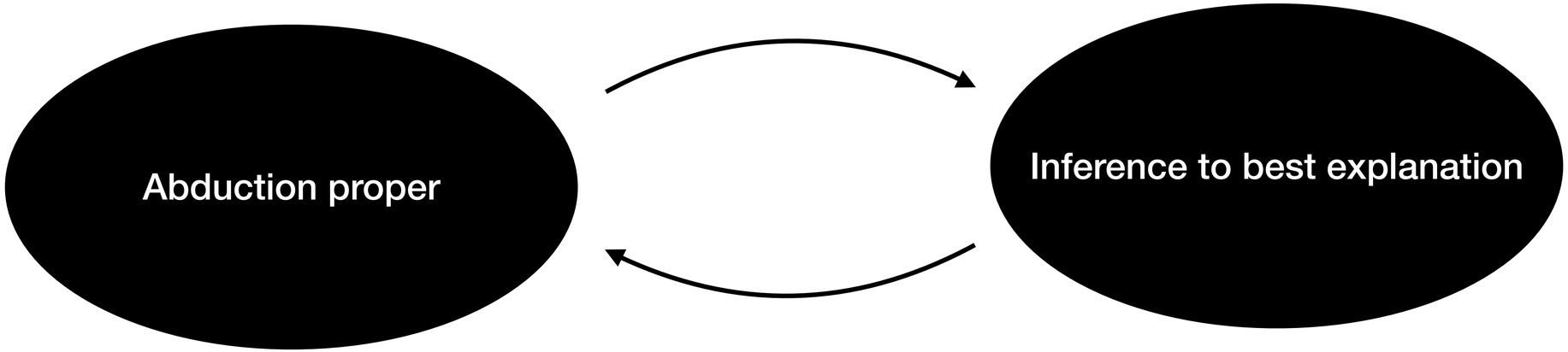
2. Should restrict the number and types of theories/hypotheses considered for testing.



(van Rooij & Baggio 2020)

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

Abduction: Reasoning from observations to generate possible explanations.

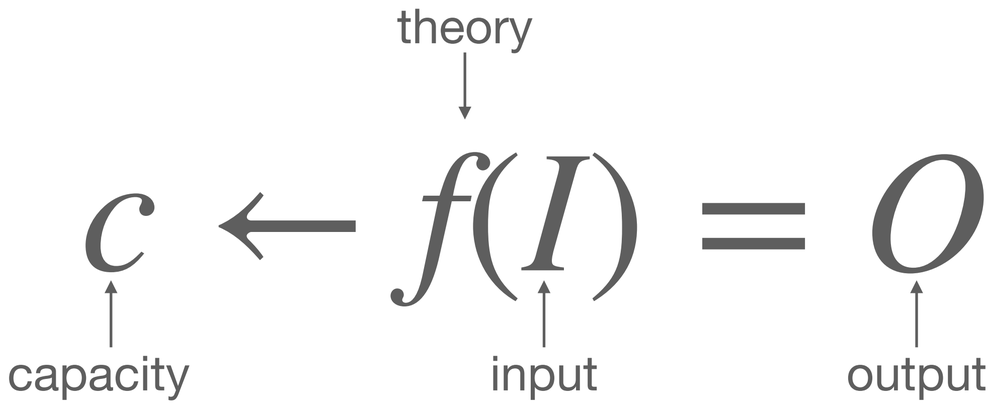


Generate candidate hypotheses

Selecting the “best” explanatory hypothesis

(van Rooij & Baggio 2020)

Structural form of a theory



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

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

(van Rooij & Baggio 2020)

Example

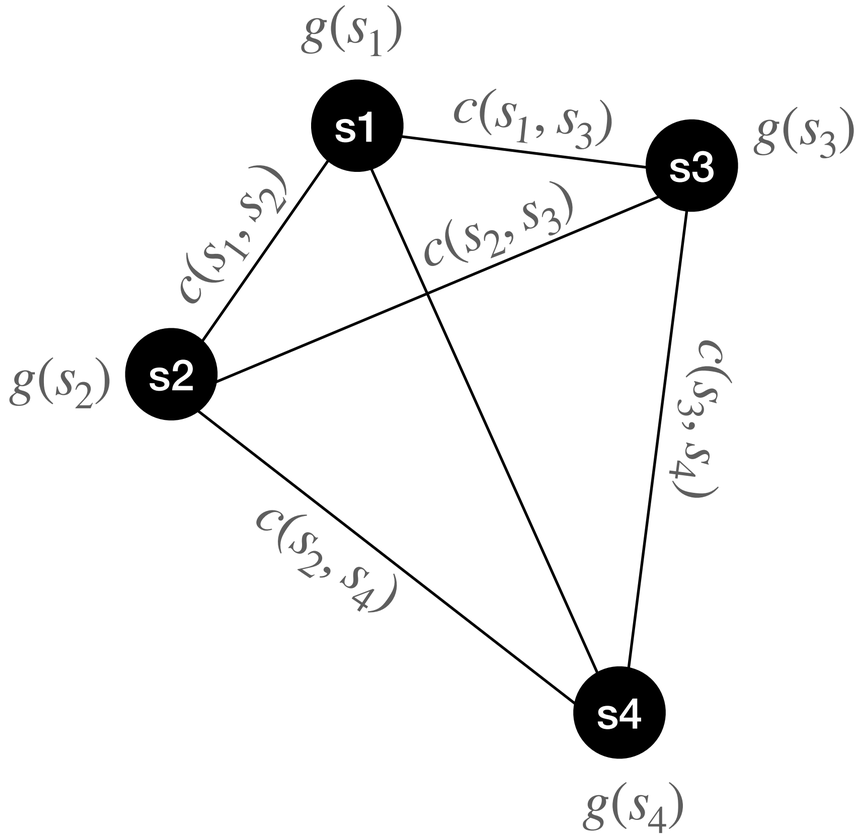
The traveling salesman problem

Foraging *f*

Input: A set of sites $S = \{s_0, s_1, s_2, \ldots, s_n\}$, 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 $\pi(S) = [s^0, s^1, \ldots, s^n, s^0]$ of the elements in S such that $s^0 = s_0$ and the sum of foods collected at $s^1, \ldots, s^n$ exceeds the total cost of the travel, i.e.,

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



(van Rooij & Baggio 2020)

Where does this leave experimentation?

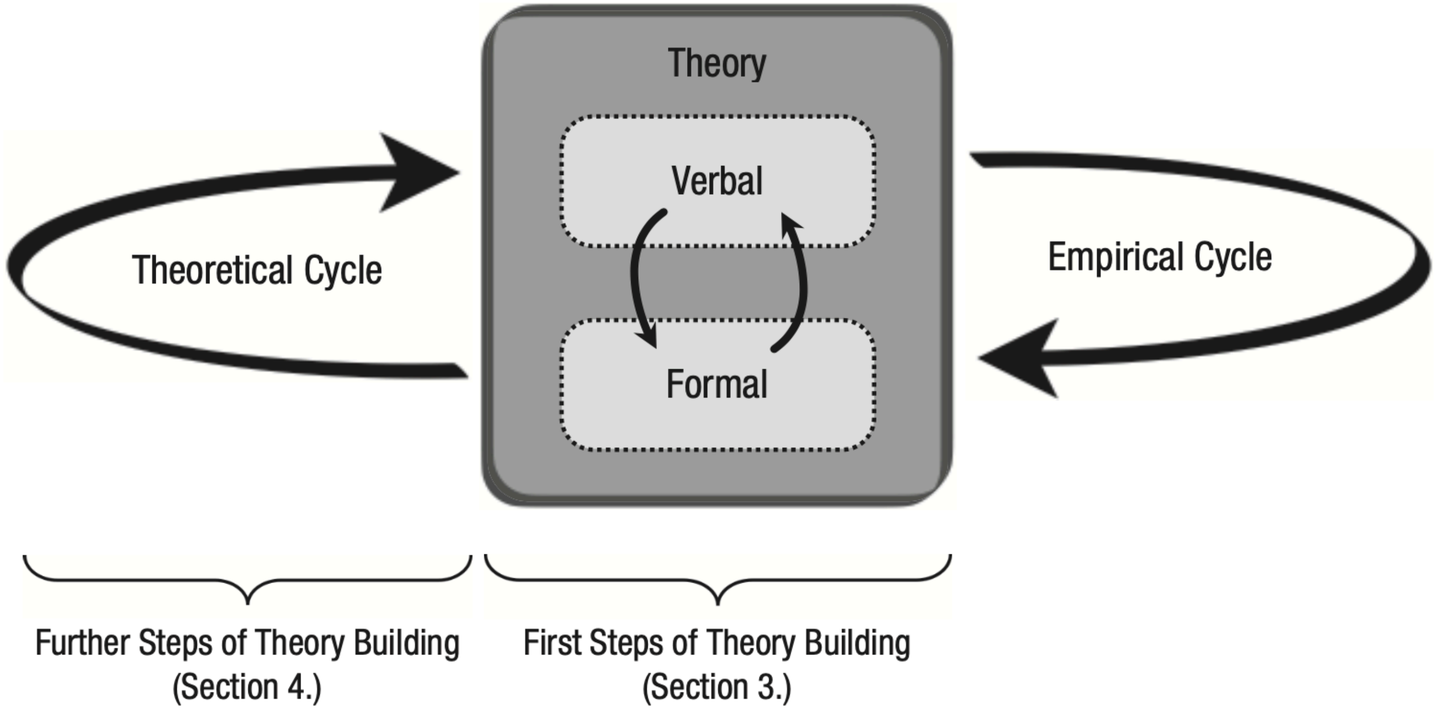
Experimentation: Should be pursued in the context of:

1. Explanatory multilevel theories of capacities.
2. Close continuity with theory development

"[T]he main goal of empirical research, including experimentation (e.g., testing for effects) and computational modeling or simulation, is to further narrow down the space of possible function *after* relevant theoretical constraints have been applied."

(van Rooij & Baggio 2020)

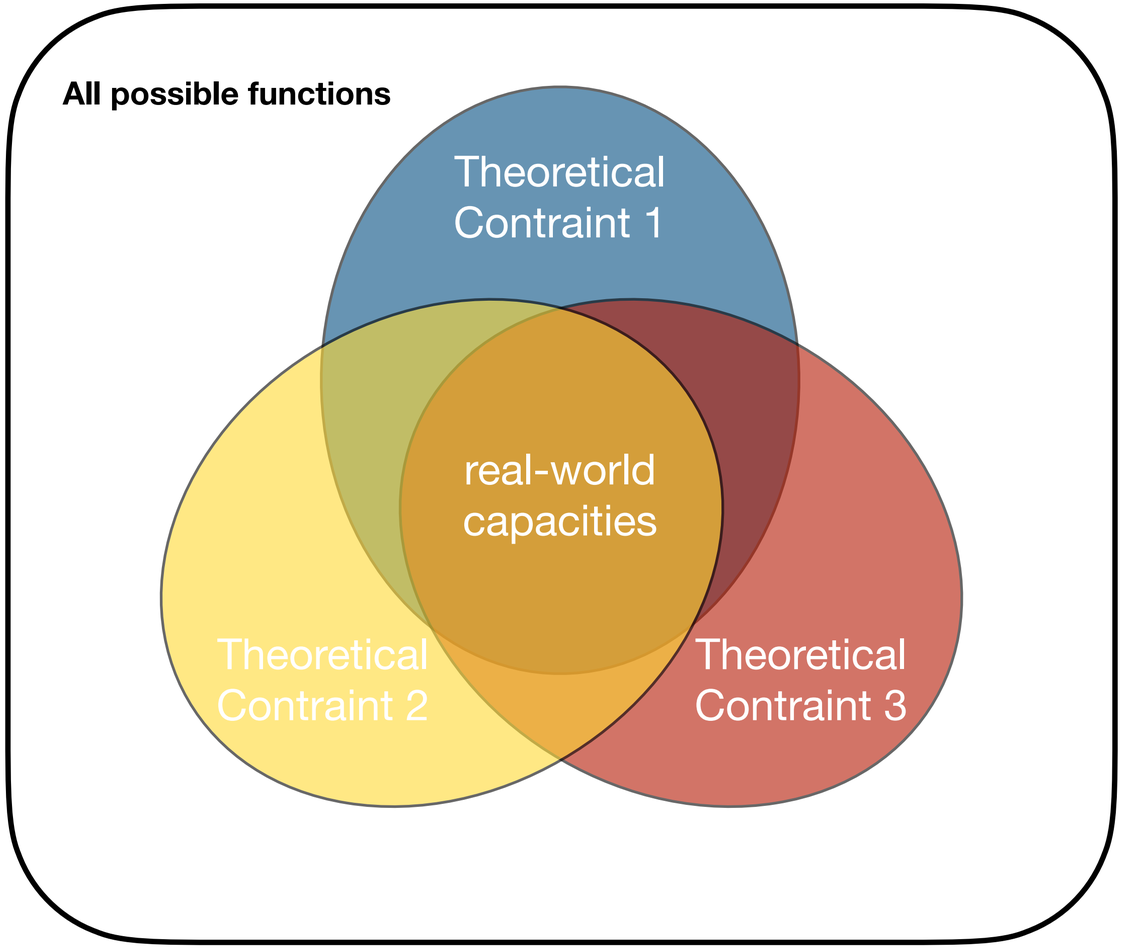
Evolution of a theory



- 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.

(van Rooij & Baggio 2020)

Reducing the space of possible theories



(van Rooij & Baggio 2020)

Take home message

- We need to understand intelligent behaviors at multiple levels: from computation to neurons and back again.
- Our theories that describe these behaviors should be focused on specific sets of capacities that describe narrow input-output relationships.
- Logical constraints define the scope of your theory and empirical observations validate those constraints.

Break out group discussions

Task: Following the approach outlined by van Rooij & Baggio, come up with a *computational* level theory of how college students choose which classes to take.

- Describe the theoretical/plausibility constraints used to construct your theory.
- Describe a way that you could test your theory.
- Define each of the following: *c, f, I, O*