W911NF-19-1-0217

ARO MURI: Foundations of Decision Making with Behavioral and Computational Constraints

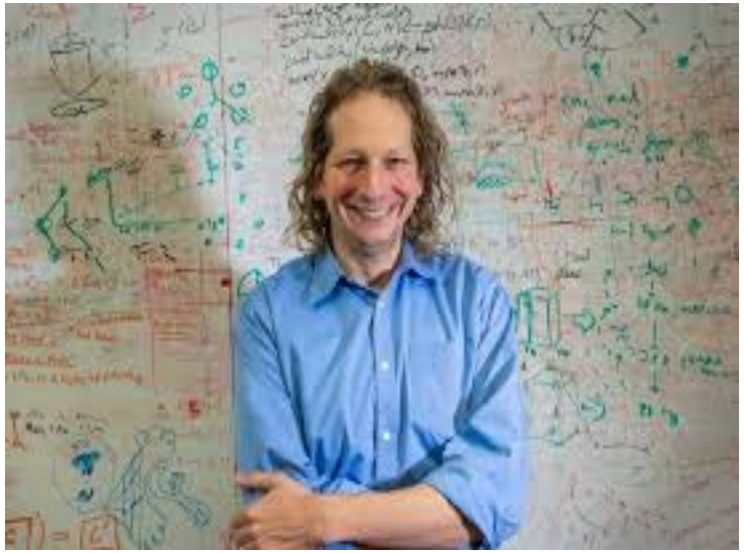
Austin Benson[♭], Ali Jadbabaie[♮], Joe Halpern[♭],
Jon Kleinberg[♭], Elchanan Mossel[♮], Josh Tenenbaum[♮]
[♮] Massachusetts Institute of Technology
[♭] Cornell University

Lead PI: Ali Jadbabaie, JR East Professor of Engineering, MIT

The Team

Ali Jadbabaie
Lead PI

Elchanan Mossel

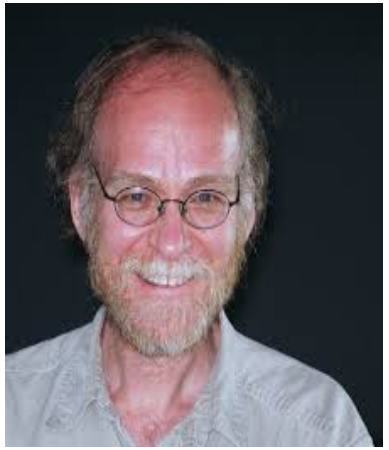


Josh Tenenbaum

Massachusetts Institute of Technology

3 Vannevar Bush Fellows
2 McArthur Fellows
2 NAE Members
3 AAAS Members

Austin Benson



Joe Halpern

Jon Kleinberg

Cornell University

New awards/recognitions 2019-2020

- Jon Kleinberg: 2019 Vannevar Bush Fellow
- Josh Tenenbaum: 2019 Mac Arthur Fellow, AAAS member
- Elchanan Mossel: 2020 Vannevar Bush Fellow
- Joe Halpern: Member, NAE class of 2019
- Austin Benson: JP Morgan Chase Faculty Fellowship in AI

Motivation and Overview (I)

- Rational Decision Theory [Savage 54, Friedman and Savage 48] imposes unrealistic cognitive and computational demands on individuals

- Growing body of evidence has lead to a behavioral approach, informed by studies in psychology, sociology, and cognitive sciences, to include behavioral biases and heuristics in decision making [Tversky and Kahneman 74, 75, Thaler and Sunstein 2008, Rabin 98,…]
- These two distinct directions have evolved separately for the past 50 years

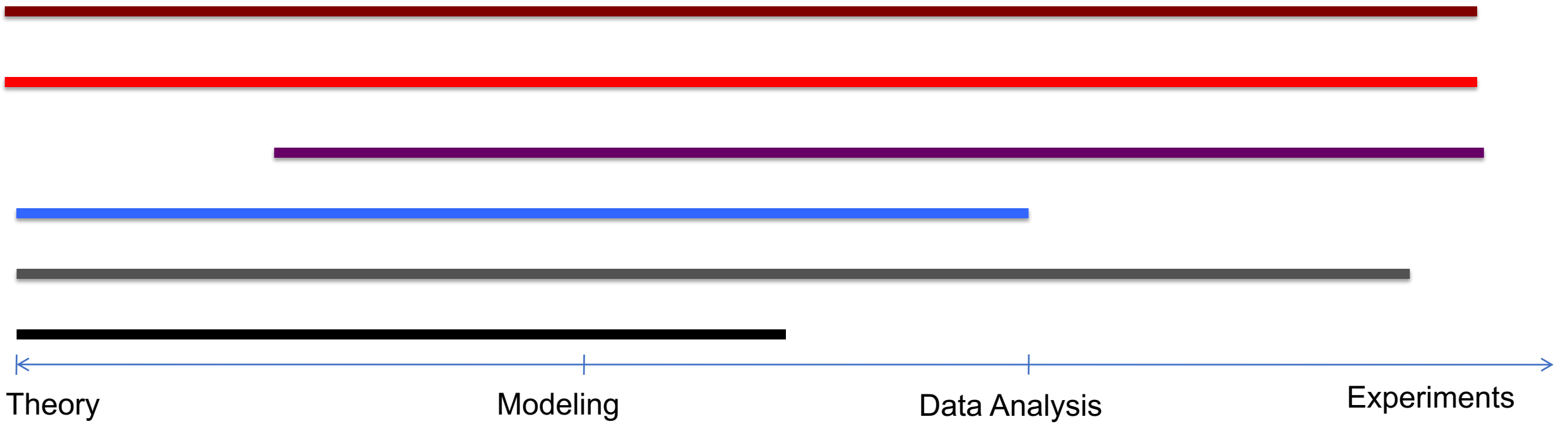
Motivation and Overview (II)

- A theory that *interpolates* between the two approaches?
- Our goal: To create a rigorous unified theory for behavioral and rational decision making
  - for individuals
  - and in groups

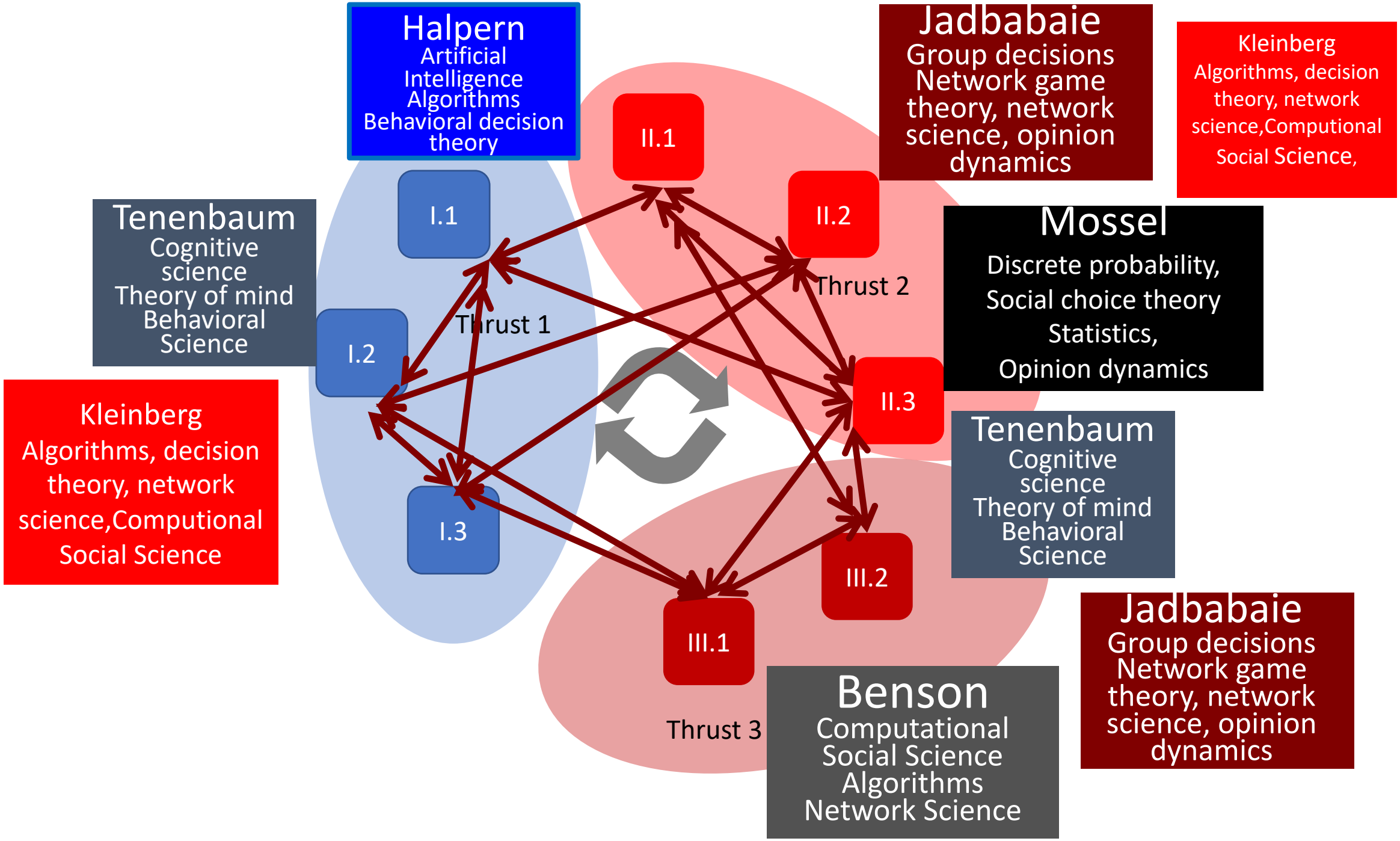
- Can we *rationalize* behavioral approaches by incorporating biases or deriving them as a result of rationality under resource-constraints
  - Computational rationality [Halpern and Pass 2011]
  - Resource-rational decision making [Griffiths, Lieder, Goodman 2015; Vul, Goodman, Griffiths, Tenenbaum 2015

Team Expertise and Connections to Proposal Thrusts

Jadbabaie	Kleinberg	Tenenbaum	Halpern	Benson	Mossel
Group decisions Network game theory, network science, opinion dynamics	Algorithms, decision theory, networks Computational Social Science	Cognitive science Theory of mind Behavioral Science	Artificial Intelligence Algorithms Behavioral decision theory	Computational Social Science Algorithms Network Science	Discrete probability, Social choice theory Statistics, Opinion dynamics



Thrust 1: Individual decision Making Under Computational and Cognitive Constraints	Thrust II: A unified Theory of Group Decision Making	Thrust 3: Modeling and Experimental Investigations
[I1] Computationally-constrained Bayesian decisions	[II1] An axiomatic foundation for group opinion exchange	III1] Data-driven study of cognitive biases
[I2] Choice-set Effects	[II2] Group decision making with resource constraints	[III2] Experiments on Amazon Mechanical Turk
[I3] Planning and decision making with biases	[II3] Polarization	


Halpern
Artificial Intelligence
Algorithms
Behavioral decision theory
Jadbabaie
Group decisions
Network game theory, network science, opinion dynamics
Kleinberg
Algorithms, decision theory, network science,Computional Social Science,
Tenenbaum
Cognitive science
Theory of mind
Behavioral Science
Mossel
Discrete probability,
Social choice theory
Statistics,
Opinion dynamics
Kleinberg
Algorithms, decision theory, network science,Computional Social Science
Tenenbaum
Cognitive science
Theory of mind
Behavioral Science
Jadbabaie
Group decisions
Network game theory, network science, opinion dynamics
Benson
Computational Social Science
Algorithms
Network Science
I.1
I.2
I.3
Thrust 1
II.1
II.2
II.3
Thrust 2
III.1
III.2
Thrust 3

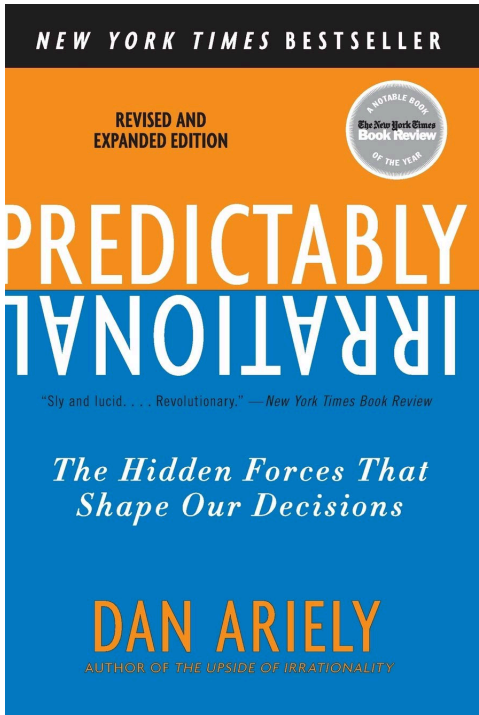
ARO MURI First Year Review
Foundations of Decision Making with Behavioral and Computational Constraints
Thursday, October 22, 2020

2:30-3:00 pm	**MURI Overview** *Ali Jadbabaie*
3:00-3:55 am	**Thrust 1: Individual decision making under computational and cognitive constraints** *Joe Halpern, Josh Tenenbaum, and Jon Kleinberg*
3:55-4:00 pm	**Break**
4:00-4:55 pm	**Thrust 2: A unified theory of group decision making** *Ali Jadbabaie, Josh Tenenbaum, Elchanan Mossel*
4:55-5:00 pm	**Break**
5:00-5:50 pm	**Thrust 3: Data-driven modeling and experimental evaluations : Choice-set effects and habit formation** *Austin Benson, Jon Kleinberg, and Ali Jadbabaie*
5:50-6:10 pm	**Wrap-Up and feedback**
6:10pm	**Adjourn**

Thrust I: Individual decision making under computational and cognitive constraints

- Behavioral scientists posit that human decision makers do not seem to act “rationally” but act in a predictively irrational way
- Our goal: Explain this predictive irrationality as a consequence of computational and cognitive resource constraints

- Optimism, satisficing and negativity bias
- Universalization, as a resource-efficient tool for moral judgement in absence or rules, laws, and norms
- Stereotype formation: limited time or information
- Later (thrust 3): Habit formation



Universalization as a tool for making moral decisions

Universalization reasoning guides moral judgment

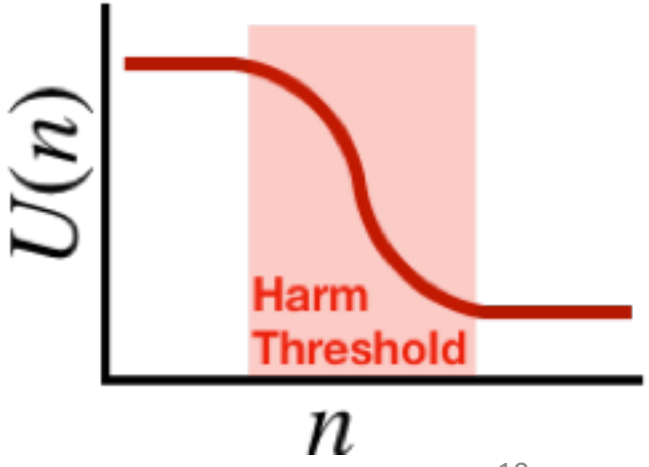
Sydney Levine[a,b,1], Max Kleiman-Weiner[a,b], Laura Schulz[a], Joshua Tenenbaum[a,1], and Fiery Cushman[b,1]

[a]Harvard University; [b]Massachusetts Institute of Technology

This manuscript was compiled on February 13, 2020

collective action problems called “threshold problems”.

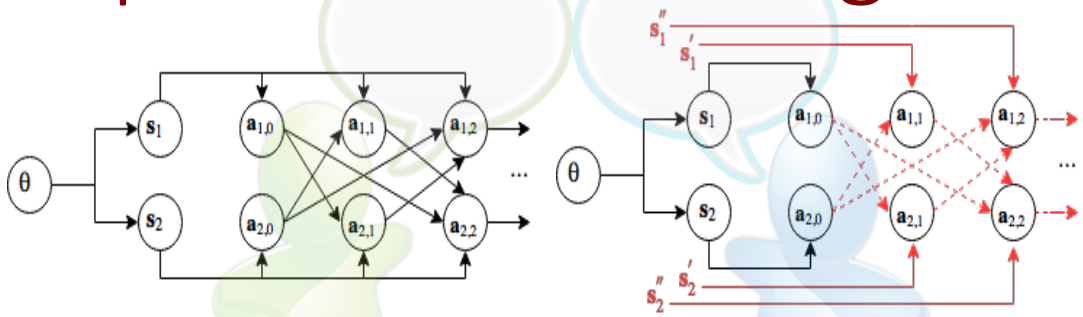
- If one person acts, there is no negative outcome.
- But if many people act, things go badly for everyone.
- Only universalization captures our moral judgments in threshold cases



Models of stereotype formation

- Mechanisms leading to stereotypes Kleinberg-Mullainathan [2020]:
  - In the presence of low information or limited available time, we are more prone to fall back on stereotypes
  - These stereotypes often work to the detriment of groups that are already disadvantaged.
- Can we find a formal basis for these properties?
  - Can such a model suggest useful interventions?

Thrust II: Resource-rational group decision-making



Algorithms & complexity

Biases & heuristics

Hazla, Rahimian, Mossel, Jadbabaie [COLT 2019]
Rahimian, Hazla, Mossel, Jadbabaie [OR, 2020]
Salhab, Ajorlou, Tenenbaum, Jadbabaie, Cog Sci 2020

- Bayesian and non-Bayesian group decisions under resource constraints
  - Group decision-making & inference with single samples
- Effect of Persuasive campaigns on polarization-curse of dimensionality
- Motives for sharing information in social networks
  - Can persuasion as a motive for sharing information/news lead to spread of misinformation?

Extending Resource-Rationality to Group decision-making

REVIEW

Computational rationality: A converging paradigm for intelligence in brains, minds, and machines

Samuel J. Gershman[1,*], Eric J. Horvitz[2,*], Joshua B. Tenenbaum[3,*]
+ See all authors and affiliations

Science 17 Jul 2015:
Vol. 349, Issue 6245, pp. 273-278
DOI: 10.1126/science.aac6076

Article | Figures & Data | Info & Metrics | eLetters | PDI

Abstract

After growing up together, and mostly growing apart in the second half of the 20th century, the fields of artificial intelligence (AI), cognitive science, and neuroscience are reconverging on a shared view of the computational foundations of intelligence that promotes valuable cross-disciplinary exchanges on questions, methods, and results. We chart advances over the past several decades that address challenges of perception and action under uncertainty through the lens of computation. Advances include the development of representations and inferential procedures for large-scale probabilistic inference and machinery for enabling reflection and decisions about tradeoffs in effort, precision, and timeliness of computations. These tools are deployed toward the goal of computational rationality: identifying decisions with highest expected utility, while taking into consideration the costs of computation in complex real-world problems in which most relevant calculations can only be approximated. We highlight key concepts with examples that show the potential for interchange between computer science, cognitive science, and neuroscience.

Volume 43 2020 , e1

Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources

Falk Lieder (a1) and Thomas L. Griffiths (a2)

DOI: https://doi.org/10.1017/S0140525X1900061X Published online by Cambridge University Press: 04 February 2019

Related commentaries (25) Author response

COGNITIVE SCIENCE
A Multidisciplinary Journal

Original Article | Free Access

One and Done? Optimal Decisions From Very Few Samples

Edward Vul, Noah Goodman, Thomas L. Griffiths, Joshua B. Tenenbaum

First published: 28 January 2014 | https://doi.org/10.1111/cogs.12101 | Citations: 113

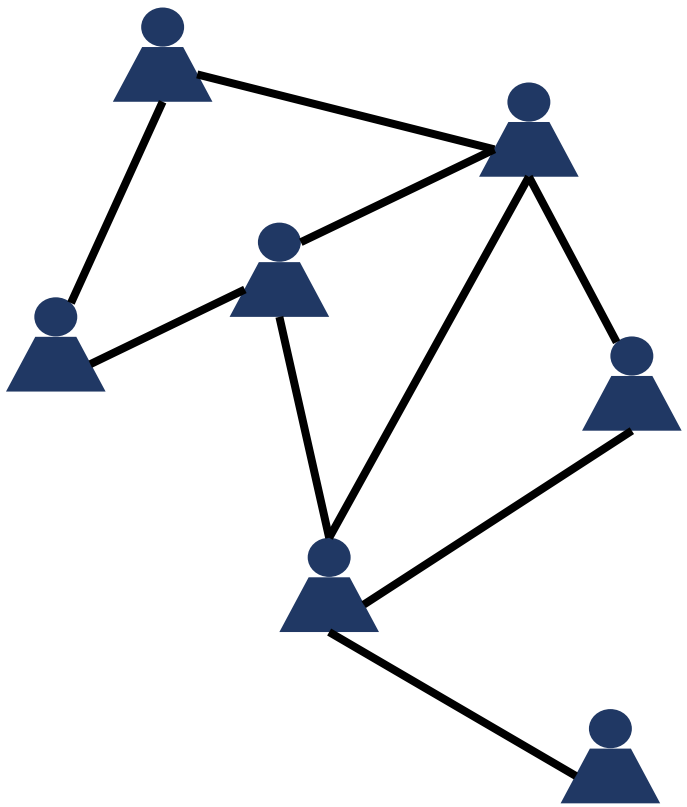
- ❑People often appear to make decisions based on just one or a few samples from the appropriate posterior probability distribution, rather than using the full distribution
- ❑Acting based on a few samples can be easily reconciled with optimal Bayesian inference
- ❑Can groups learn if they only communicate samples?

Bayesian Social Learning with Sampling

- We modify the social learning setting to the framework of Vul, Goodman, Griffiths, Tenenbaum 2015

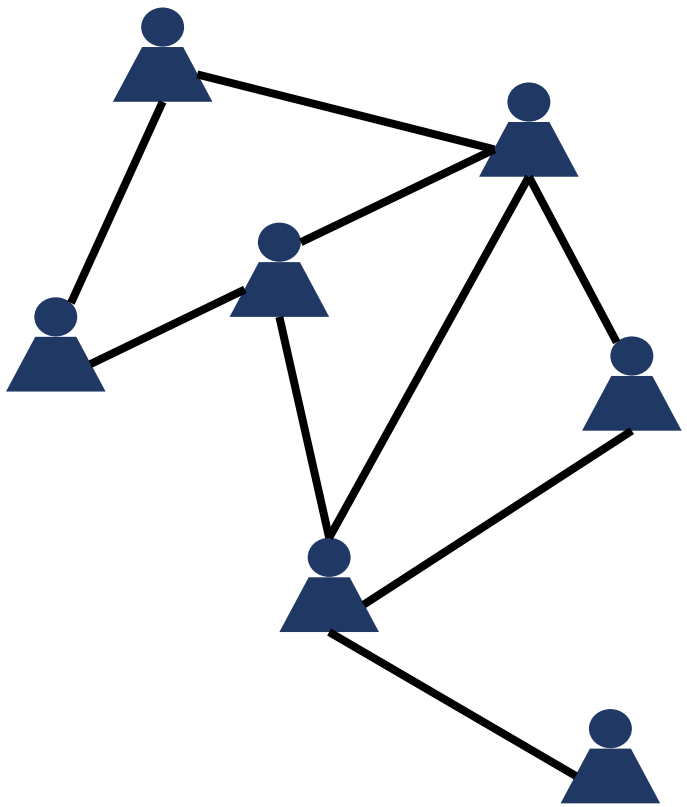
- A group of fully Bayesian agents with private observations communicate a single sample from their belief distribution at each round

- Questions:
  - Do agents aggregate information?
  - If so, how well and how quickly?
  - Is it computationally tractable?



Bayesian Social Learning with belief sampling

- Surprisingly, there is *no informational cost* to sampling from beliefs.
- Bayesian agents aggregate information efficiently: as if all private signals were publicly known



- Sampling beliefs by fully Bayesian agents *does not* cause herding, polarization or confirmation bias. It only potentially slows convergence

- Not true in Non-Bayesian setting

Salhab, Ajorlou, Tenenbaum, Jadbabaie, ***CogSci 2020***

II.3 Polarization and affirmation-seeking agents

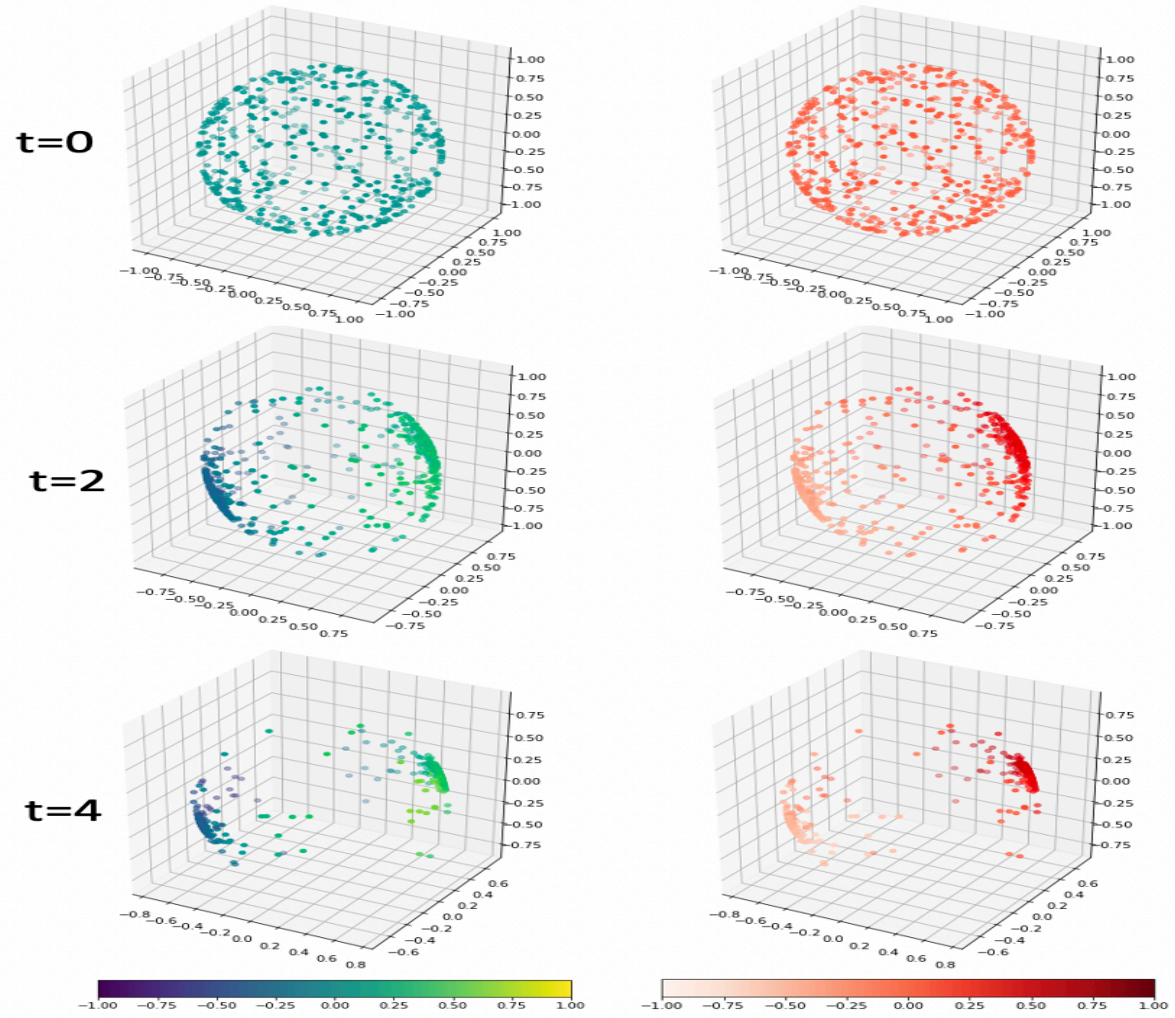
- Persuasive campaigns and polarization?

[Hazla, Jin, Ramyanaran, Mossel, 2020]

- Opinions u are d dimensional unit vectors $\in\ S^{d-1}$
- After intervention v, u updates to a vector proportional to :

$$w = u + \eta \cdot \langle u, v \rangle \cdot v, \qquad \eta > 0,$$

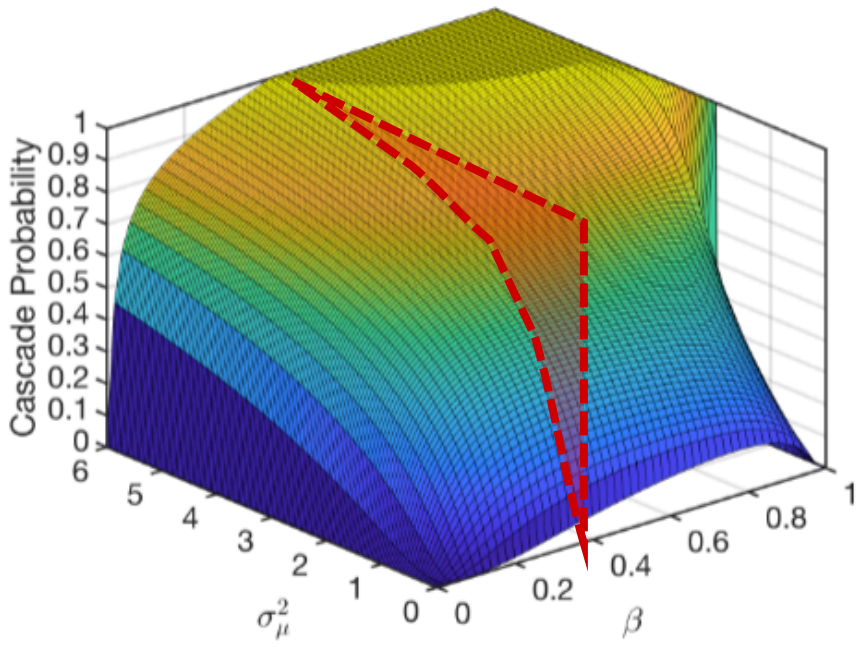
- Main results: Campaigns lead to polarization.
- No peer effect, network structure, etc.



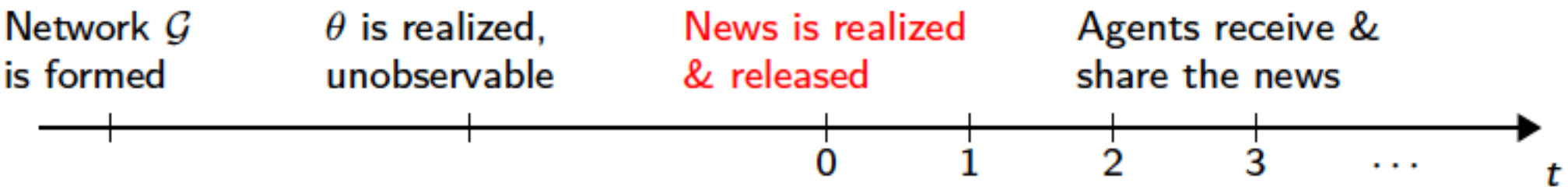
Persuasion as a motive for sharing on social media

- Individuals use Bayes rule to update their perspectives when receiving news with finite precision. Will decide to share so as to persuade followers with different perspectives

- How does precision of the news, structure of the network, and diversity of perspectives impact virality
- If population perspectives are very diverse, news with low credibility goes viral!
-

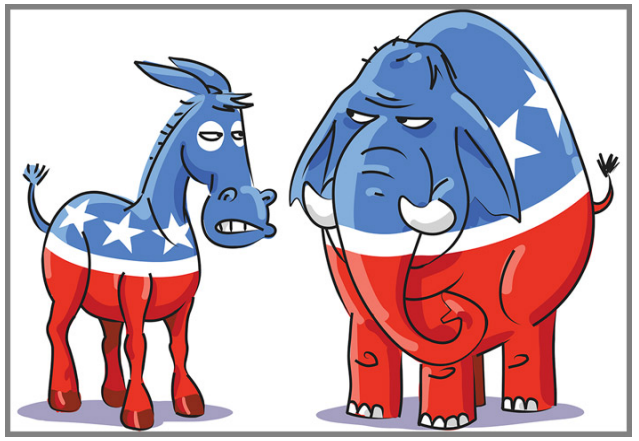


Cascade likelihood versus the credibility β and the diversity in perspectives σ_μ^2.



News subscription on social media

- Countless sources of information/limited attention span
- Sources are **biased** and are motivated to persuade the public opinion:
  - promote news that can sway their subscribers towards their own political agenda or conceal the unfavorable one.
- Main questions:
  - How do people choose their news sources? what can they infer when their sources do not disclose?
  - Sources: what news to disclose or conceal?
- Joint work with Muhamet Yildiz



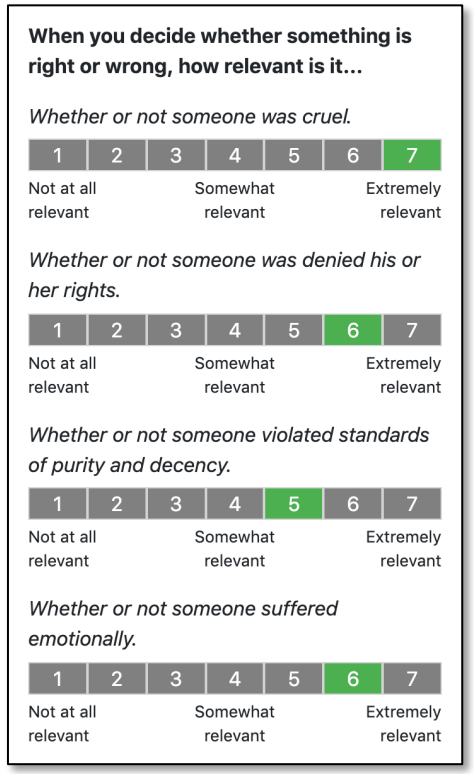
Credit: kbeis/Getty Images. Stanford News.

Paper on SSRN, ID: 3689990

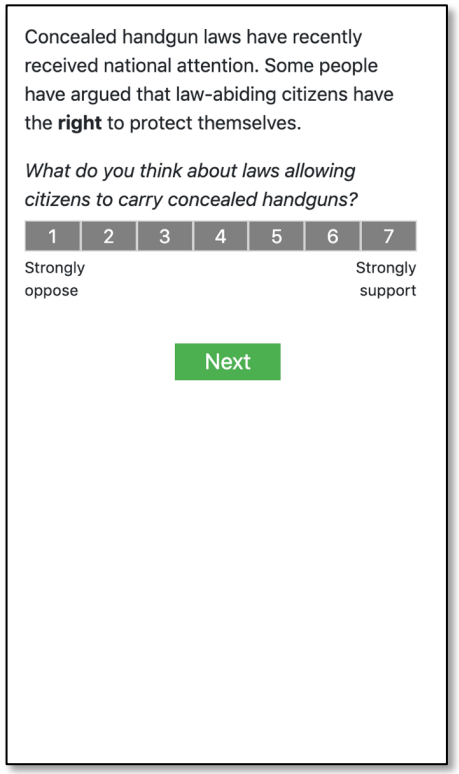
Identifying heterogeneity in persuasion

Question: What psychological/moral characteristics affect whether somebody finds an argument persuasive?
Method: Online randomized control trial, with adaptive experiment design

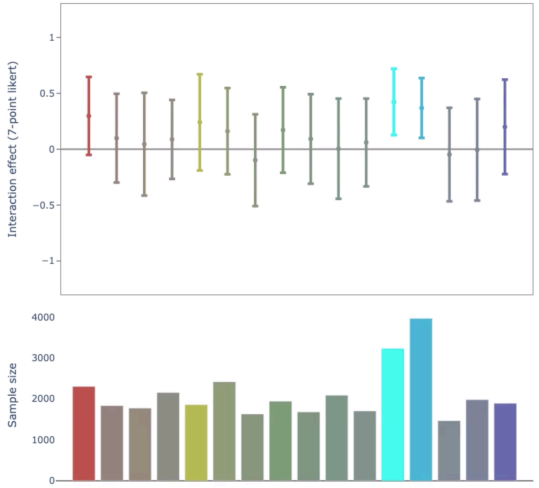
1. Pre-survey
(Moral foundations, etc)



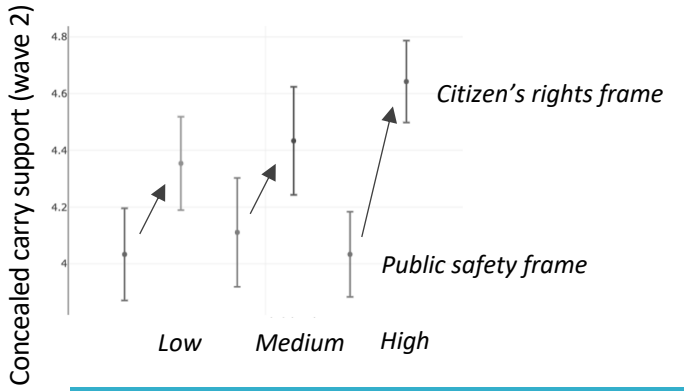
2. Treatment + Outcome
(Tweet, vignette, etc.)



3. Optimize questions in-the-loop



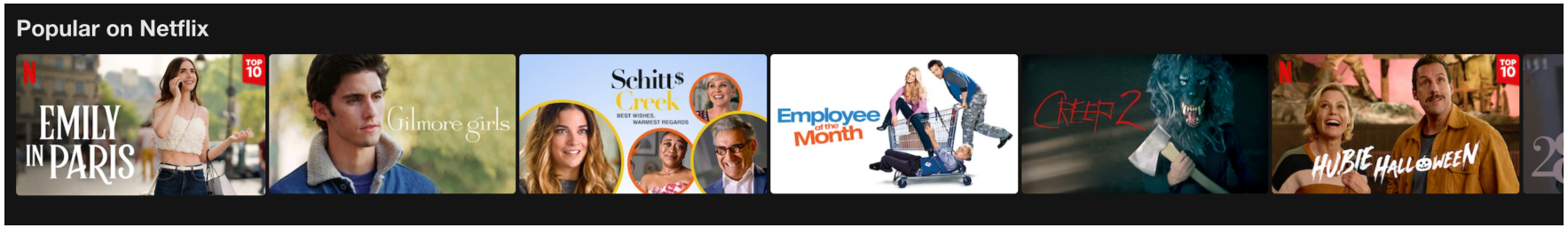
4. Identify strongest
predictors of persuasion



Strongest predictor question:

"When you decide whether something is right or wrong, how relevant is it ***whether or not someone was denied his or her rights****?"*

Discrete choice models provide a framework for reasoning about decision making and biases.



Universe U of options (movies, foods, health policies)

$\longrightarrow$ subset C of discrete alternatives (what's available)

$\longrightarrow$ choose item x from C (decision)

- **Key Questions**
  - How does changing the "choice set" C alter decisions or cause bias?
  - How can we learn the biases from data ? Can we optimize choice sets to steer behavior?
  - How do changes in choice sets let us understand habits?
  - How do comparisons reveal latent habits or skills?

Choice-sets and behavioral biases: habit disruption

- Last year we talked about repeat-behavior bias (Benson 2016,18)
- Habits: ***automatic processes*** that are learned from ***repeated responses*** and are triggered through various ***contextual or mental cues***
- Resource-constrained rationality: By forming habits we free up our mental resources for more important deliberative tasks
- *Habits are persistent*, even when the outcome structure is changed, and the behavior is no longer optimal
- Defining habit through state dependence (inertia)
- This year: Rigorously measure effect of disruptions on habits [Tohidi et al. in prep.]
- Word orderings are a widespread form of repeated behavior in language; can study disruptions to these patterns in text data [VanKoevering-Benson-Kleinberg 2020]
- Formalize notions of habit and state-dependence

ARO MURI First Year Review
Foundations of Decision Making with Behavioral and Computational Constraints
Thursday, October 22, 2020

2:30-3:00 pm	**MURI Overview** *Ali Jadbabaie*
3:00-3:55 am	**Thrust 1: Individual decision making under computational and cognitive constraints** *Joe Halpern, Josh Tenenbaum, and Jon Kleinberg*
3:55-4:00 pm	**Break**
4:00-4:55 pm	**Thrust 2: A unified theory of group decision making** *Ali Jadbabaie, Josh Tenenbaum, Elchanan Mossel*
4:55-5:00 pm	**Break**
5:00-5:50 pm	**Thrust 3: Data-driven modeling and experimental evaluations : Choice-set effects and habit formation** *Austin Benson, Jon Kleinberg, and Ali Jadbabaie*
5:50-6:10 pm	**Wrap-Up and feedback**
6:10pm	**Adjourn**