Cognitive Modeling of the Behavioral Effectiveness of Non-Pharmaceutical Interventions

Christian Lebiere, Carnegie Mellon University, (cl@cmu.edu) in collaboration with:

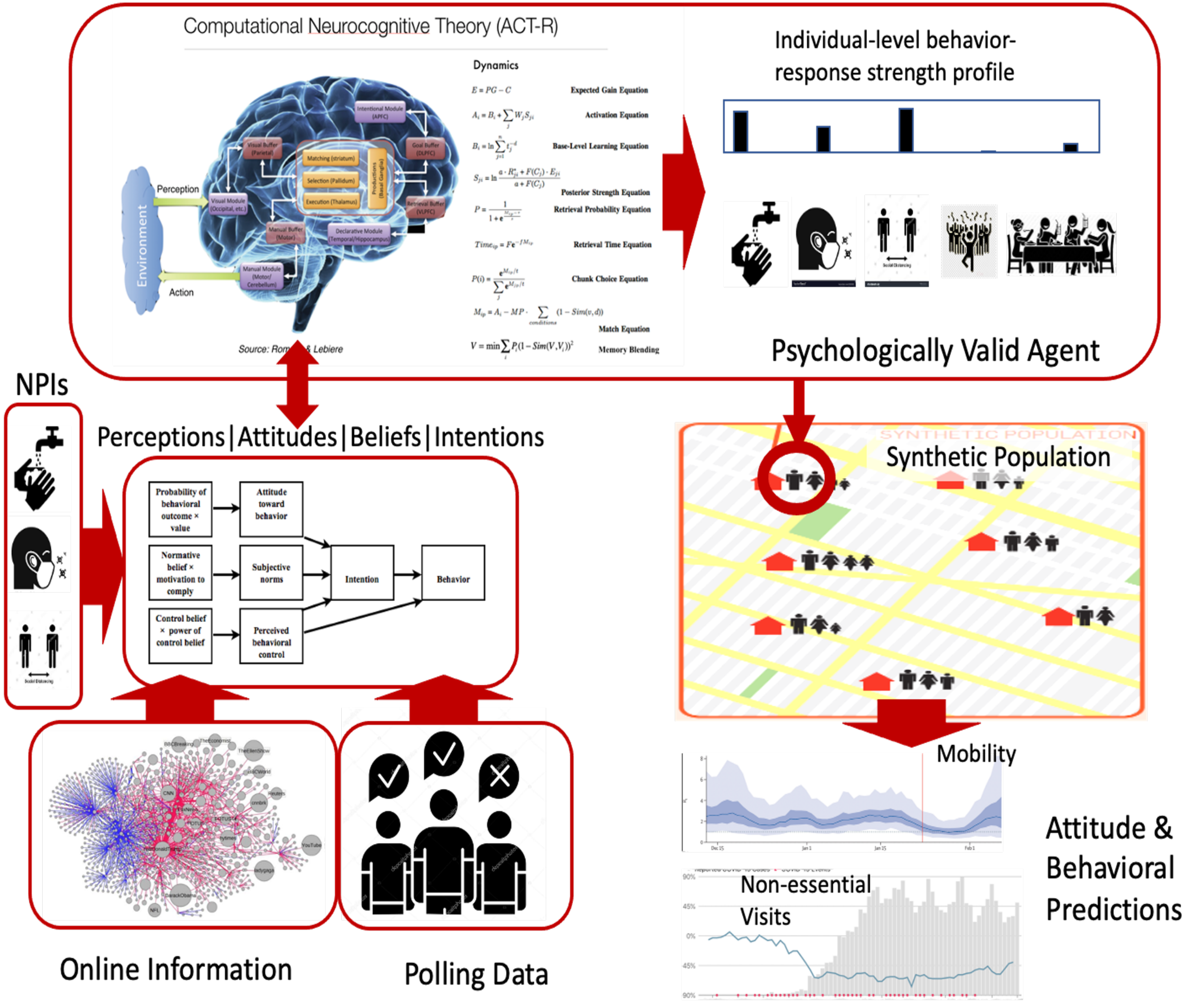
Kathleen Carley, Michael Martin & Konstantinos Mitsopoulos, Carnegie Mellon

Peter Pirolli, Bonnie Dorr, Tomek Strzalkowski, Adam Dalton,
Brodie Mather & Archna Bathia, Institute for Human and Machine Cognition

Mark Orr & Stephen Eubank, University of Virginia Biocomplexity Institute

From Information To Behavior

- Build on:
  - Theories of individual health psychology
  - Predictive computational cognitive models of behavior change
- **Psychologically Valid Agent**-based epidemiological models
  - More accurately predict the dynamics of behavior change
  - Response to NPIs government messaging, mass media, social media, information framing, etc



Computational Neurocognitive Theory (ACT-R)

Source: Romero & Lebiere

Individual-level behavior-response strength profile

Psychologically Valid Agent

NPIs

Perceptions | Attitudes | Beliefs | Intentions

Synthetic Population

Mobility

Non-essential Visits

Attitude & Behavioral Predictions

Online Information

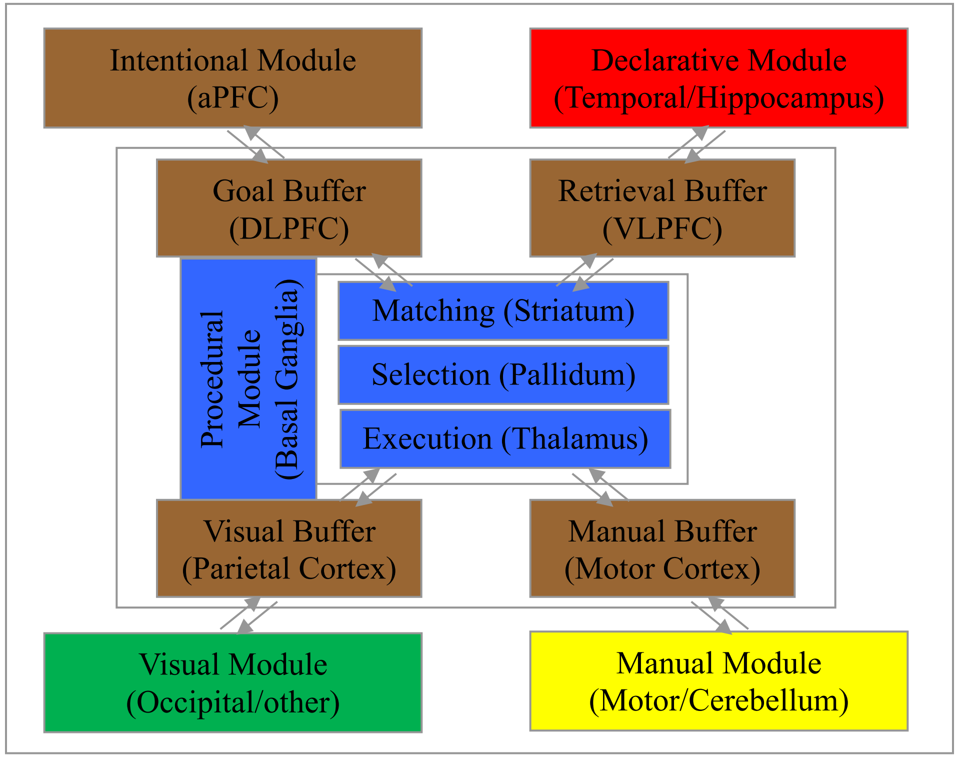
Polling Data

Environment
Perception
Action
Visual Module (Occipital, etc.)
Visual Buffer (Parietal)
Matching (striatum)
Selection (Pallidum)
Execution (Thalamus)
Productions (Basal Ganglia)
Intentional Module (APFC)
Goal Buffer (DLPFC)
Retrieval Buffer (VLPFC)
Manual Buffer (Motor)
Declarative Module (Temporal/Hippocampus)
Manual Module (Motor/ Cerebellum)
Dynamics
$E = PG - C$ Expected Gain Equation
$A_i = B_i + \sum_j W_j S_{ji}$ Activation Equation
$B_i = \ln \sum_{j=1}^{n} t_j^{-d}$ Base-Level Learning Equation
$S_{ji} = \ln \frac{a \cdot R_{ji}^* + F(C_j) \cdot E_{ji}}{a + F(C_j)}$ Posterior Strength Equation
$P = \frac{1}{1 + e^{-\frac{M_{ip} - \tau}{s}}}$ Retrieval Probability Equation
$Time_{ip} = Fe^{-fM_{ip}}$ Retrieval Time Equation
$P(i) = \frac{e^{M_{ip}/t}}{\sum_j e^{M_{jp}/t}}$ Chunk Choice Equation
$M_{ip} = A_i - MP \cdot \sum_{conditions} (1 - Sim(v,d))$ Match Equation
$V = \min \sum_i P_i (1 - Sim(V, V_i))^2$ Memory Blending
Social Distancing
Probability of behavioral outcome × value
Attitude toward behavior
Normative belief × motivation to comply
Subjective norms
Control belief × power of control belief
Perceived behavioral control
Intention
Behavior
SYNTHETIC POPULATION
Dec 15
Jan 1
Jan 15
Feb 1
90%
45%
0%
-45%
-90%

Cognitive Architectures and Human Behavior

- ACT-R provides constrained, principled framework to model complex human behavior
- Modules for knowledge, action selection, working memory, perception and motor actions
- Integrates symbolic knowledge and statistical adaptivity
- Reflects individual differences and emergent cognitive biases
- Accounts for training effects by modeling learning processes

ACT-R Cognitive Architecture


Intentional Module (aPFC)
Declarative Module (Temporal/Hippocampus)
Goal Buffer (DLPFC)
Retrieval Buffer (VLPFC)
Procedural Module (Basal Ganglia)
Matching (Striatum)
Selection (Pallidum)
Execution (Thalamus)
Visual Buffer (Parietal Cortex)
Manual Buffer (Motor Cortex)
Visual Module (Occipital/other)
Manual Module (Motor/Cerebellum)

Brown: Working Memory; Red: Declarative Memory; Blue: Procedural Memory; Green: Perception; Yellow: Action

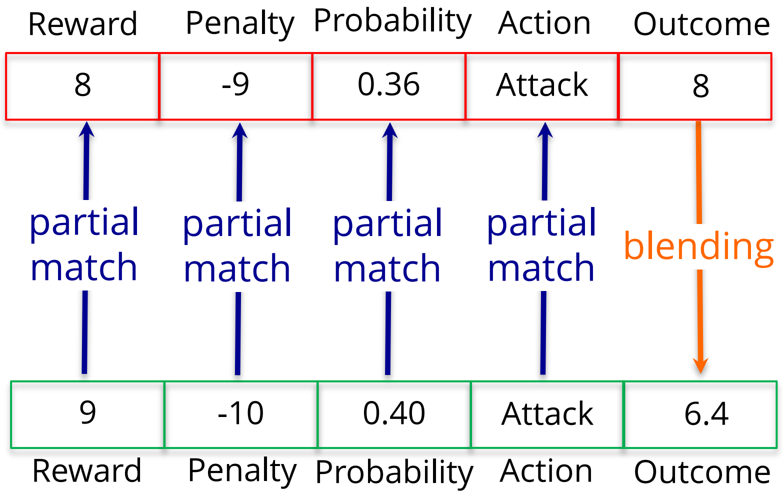
Instance-Based Learning (IBL)

- Instance-Based Learning (IBL) models make decisions by generalizing previous situations
- Leverage cognitive mechanisms for memory (decay, rehearsal, priming) and pattern matching (partial matching, blending)
- Individual biases reflect distinct experience history (knowledge)
- Additional biases from cognitive mechanisms (e.g., recency bias)
- Predict training effectiveness by diversity, size of instance base

$$A_i = \ln \sum_{j=1}^{n} t_j^{-d} + MP * \sum_{k} Sim(v_k, c_k) + \varepsilon_i$$

$$P_i = \frac{e^{A_i/s}}{\sum_j e^{A_j/s}} \qquad V = \mathrm{argmin} \sum_{i} P_i \times \left(1 - Sim(V, V_i)\right)^2$$

Memory Instance (one of many)

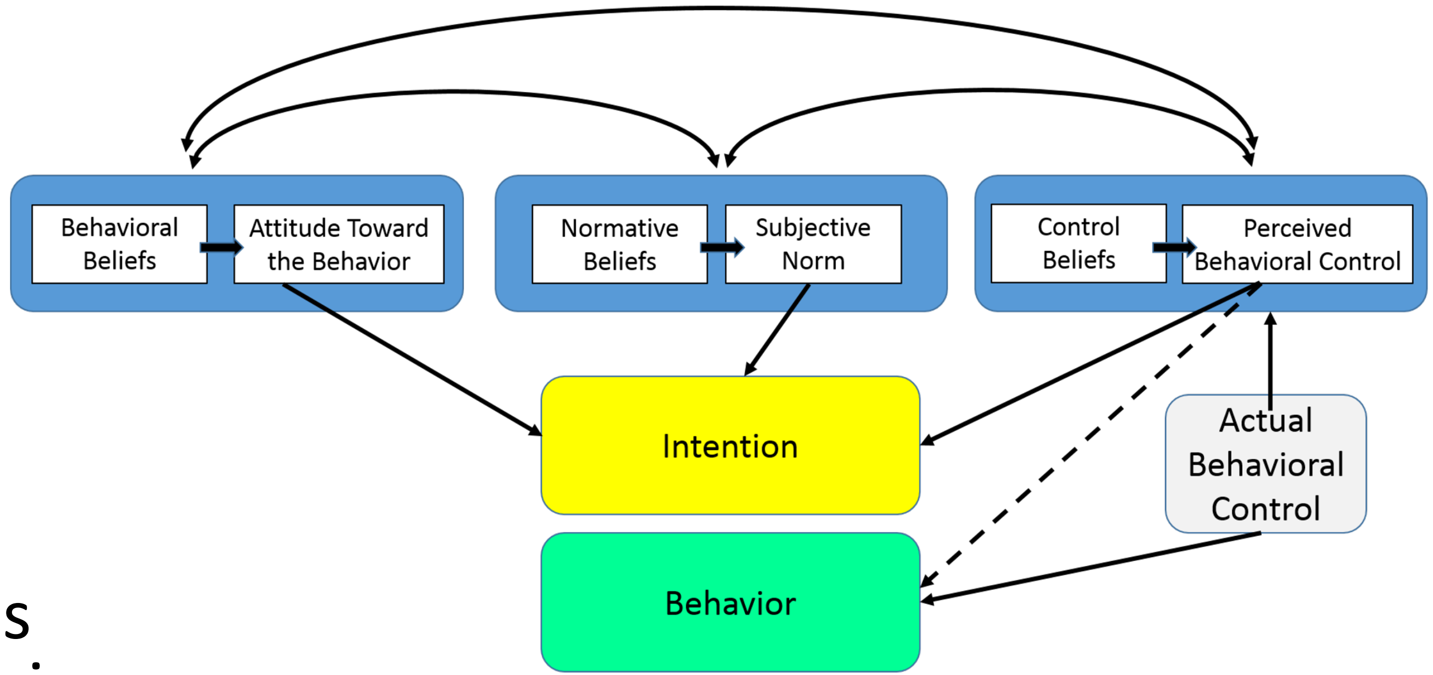


Reward	Penalty	Probability	Action	Outcome
8	-9	0.36	Attack	8
partial match	partial match	partial match	partial match	blending
9	-10	0.40	Attack	6.4
Reward	Penalty	Probability	Action	Outcome

Event Representation (Context)

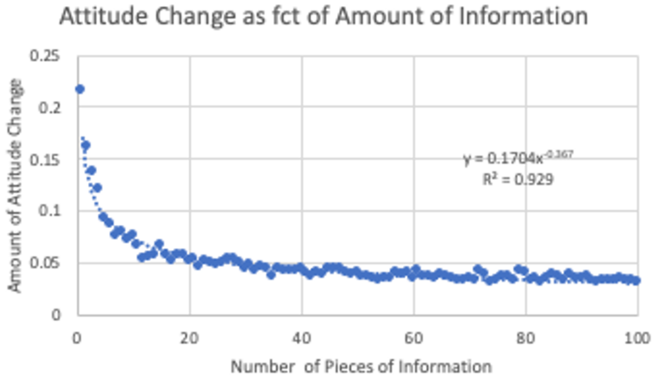
Theory of Planned Behavior

- predict and explain wide range of health behaviors and intentions
- Social norms reflect normative beliefs toward desired behavior
- Attitudes reflect intentions toward execution of behavior
- Control beliefs reflect perception of ability to control behavior

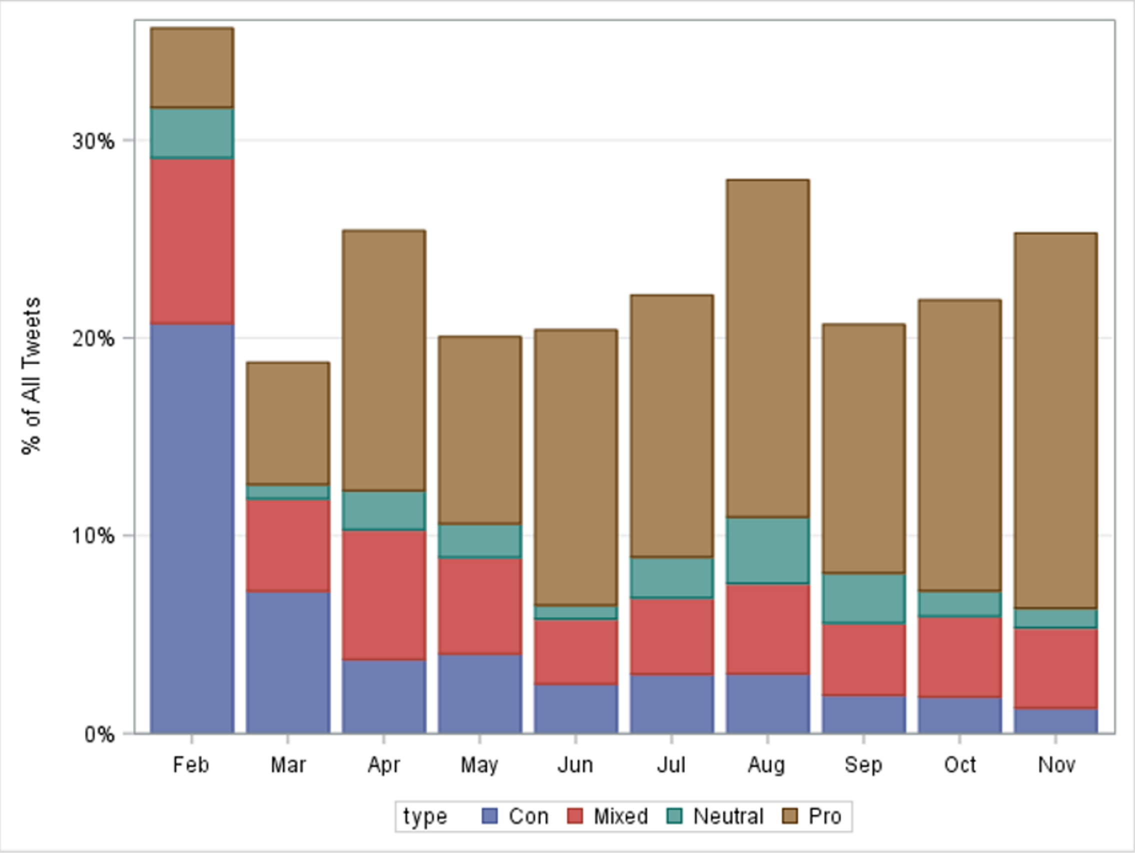


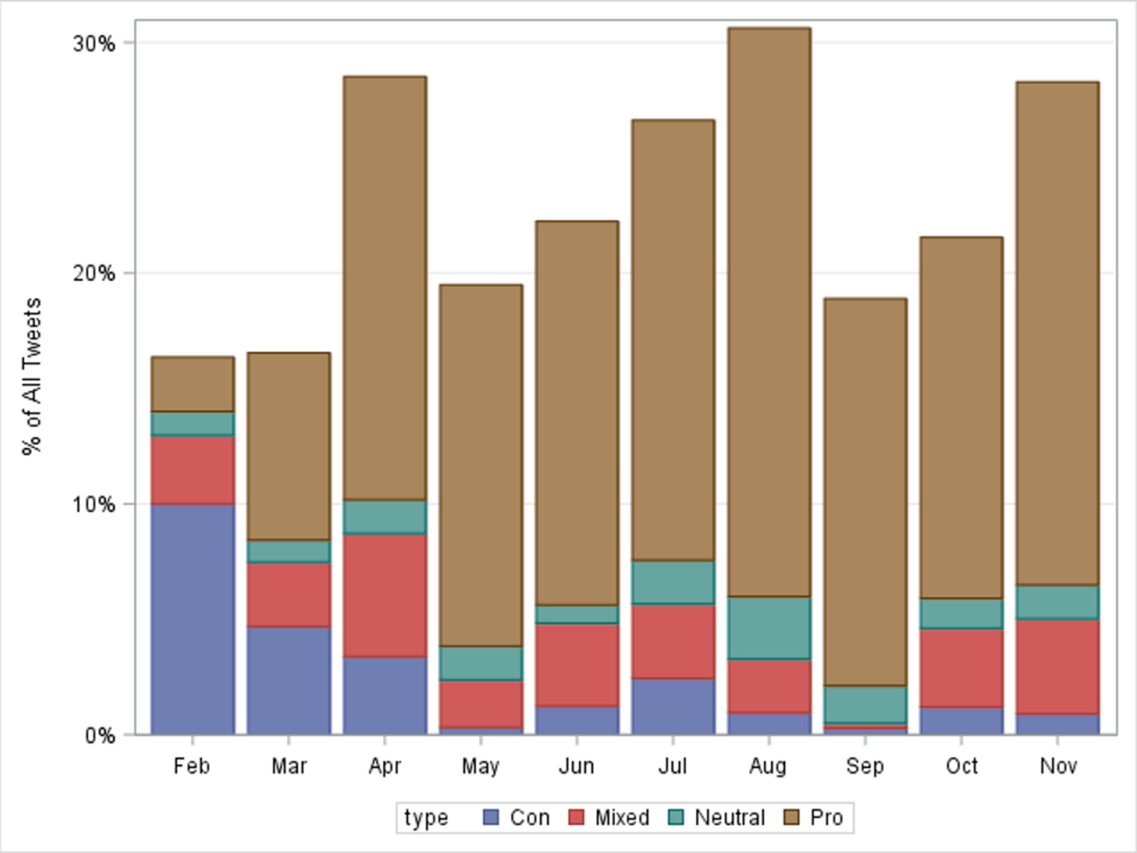
Social Norms

- Attitudes reflect input from agent’s global network
  - Social/political affiliation and social network polarization
  - Transfer from other norms e.g. drunk driving
- Attitudes are modulated by local context
  - Prevalence of behavior in local environment
  - Immediate peer pressure from direct contacts
- Formalized as chunks encoding competing actions
  - Wear-mask vs not-wear-mask actions
  - Competing actions rather than action-valence better fits IBL approach
  - Competing base-level activation of alternative action chunks



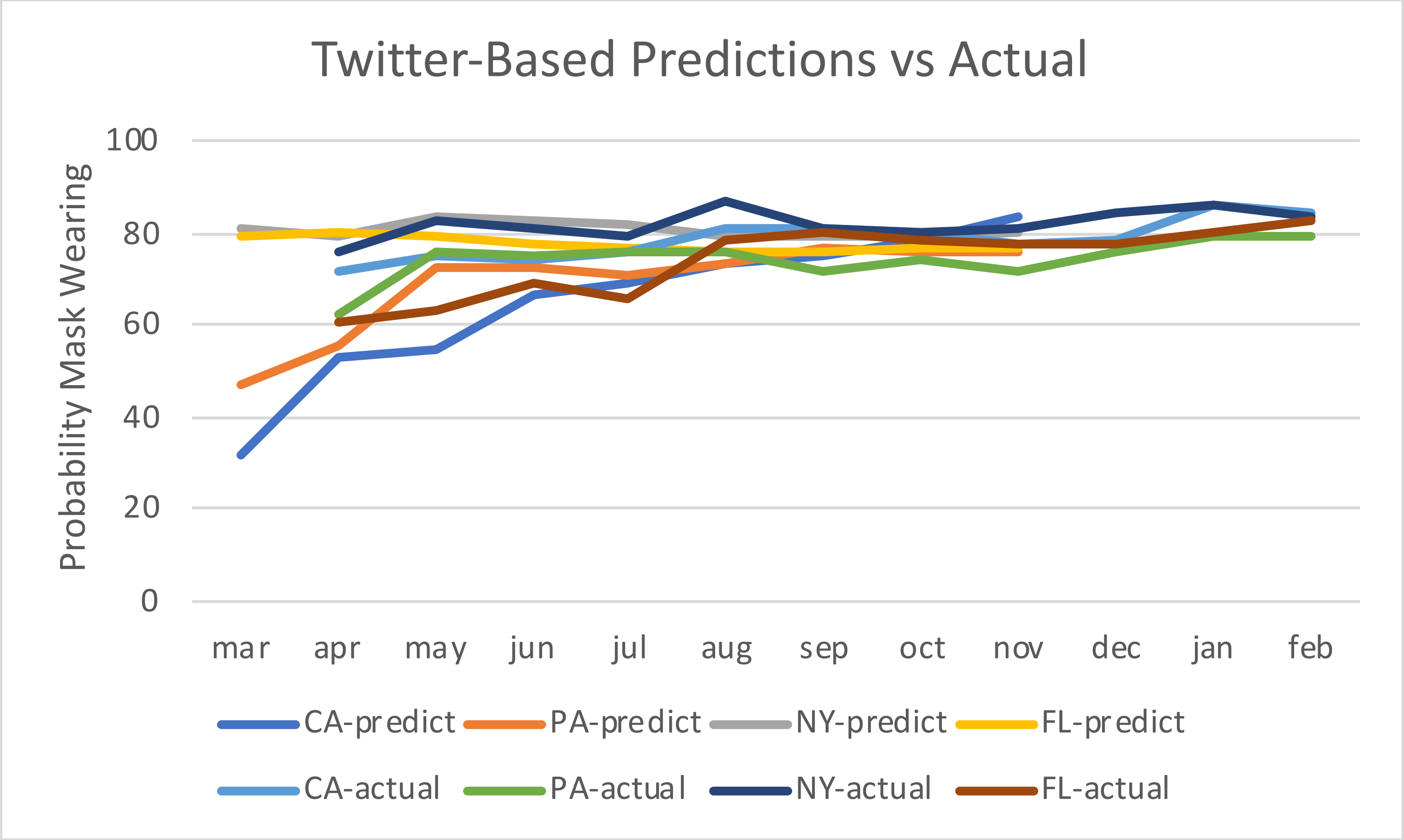
Evolving Social (Media) Norms in CA and PA





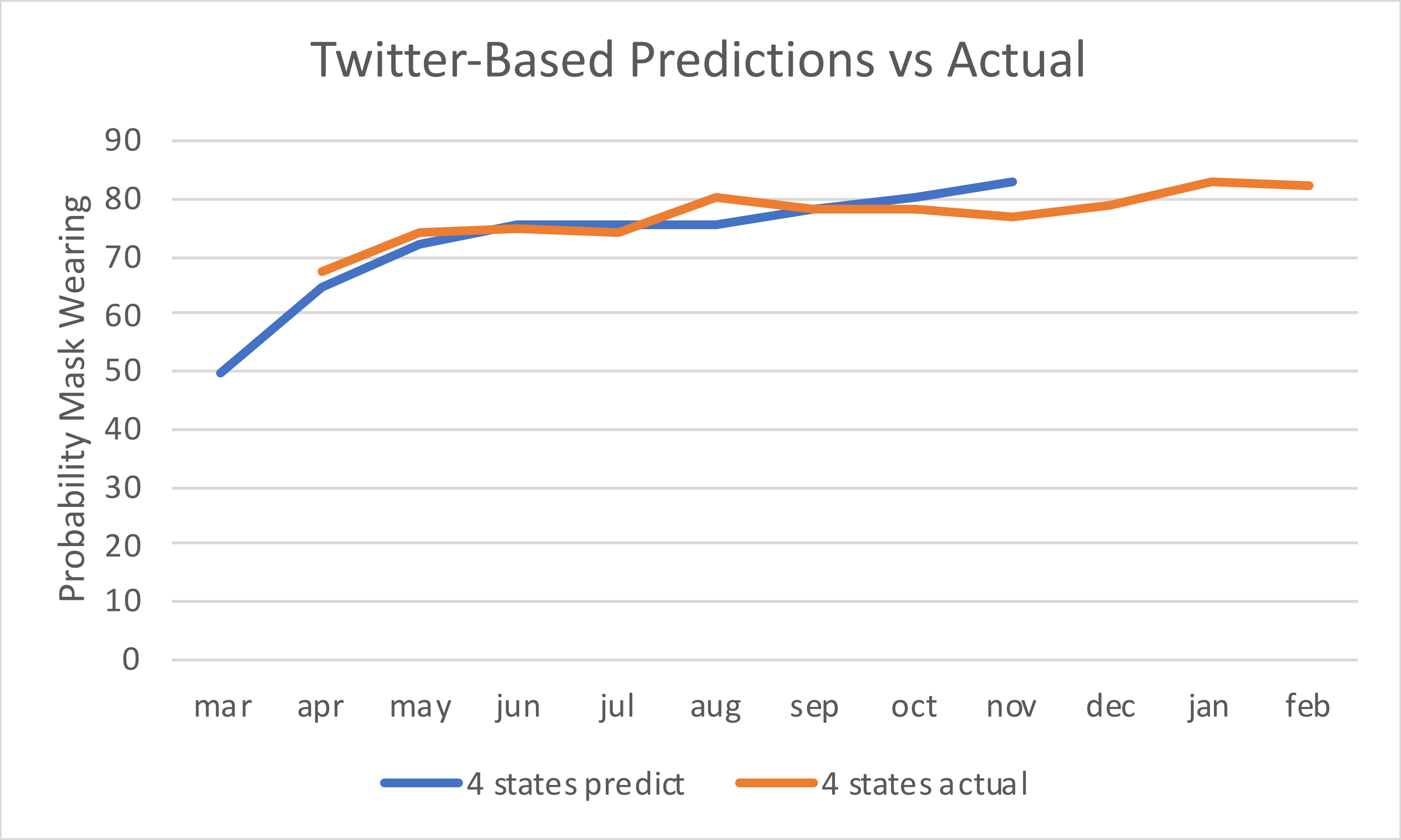
Mask Wearing Predictions in 4 Large States

- Memory strength of norms abstracted from ratio of pro- and anti-mask Twitter hashtags
- Reflects power laws of recency and frequency
- 1 parameter estimated: blending temperature
  - suggests stronger social cohesion in NY and CA
- Validated against CovidStates data



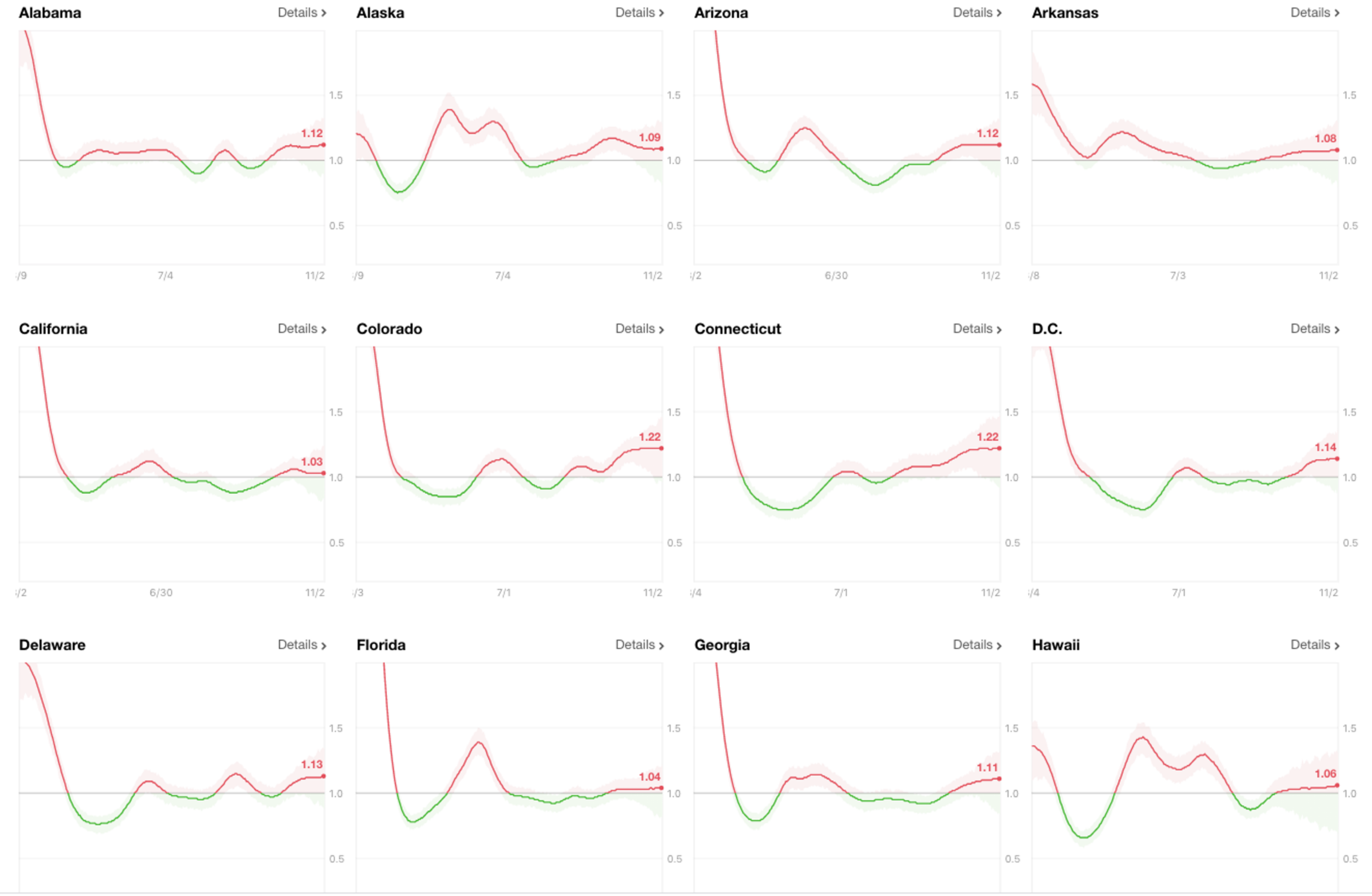
Mask Wearing Aggregate Validation

- Aggregate Twitter data and unweighted mask wearing probabilities
- Temperature parameter estimated near neutral value (1.2)
- Better fit than single states
  - Limitations of noisy data
  - National nature of twitter
  - Unrepresentative sample
  - Other behavior factors



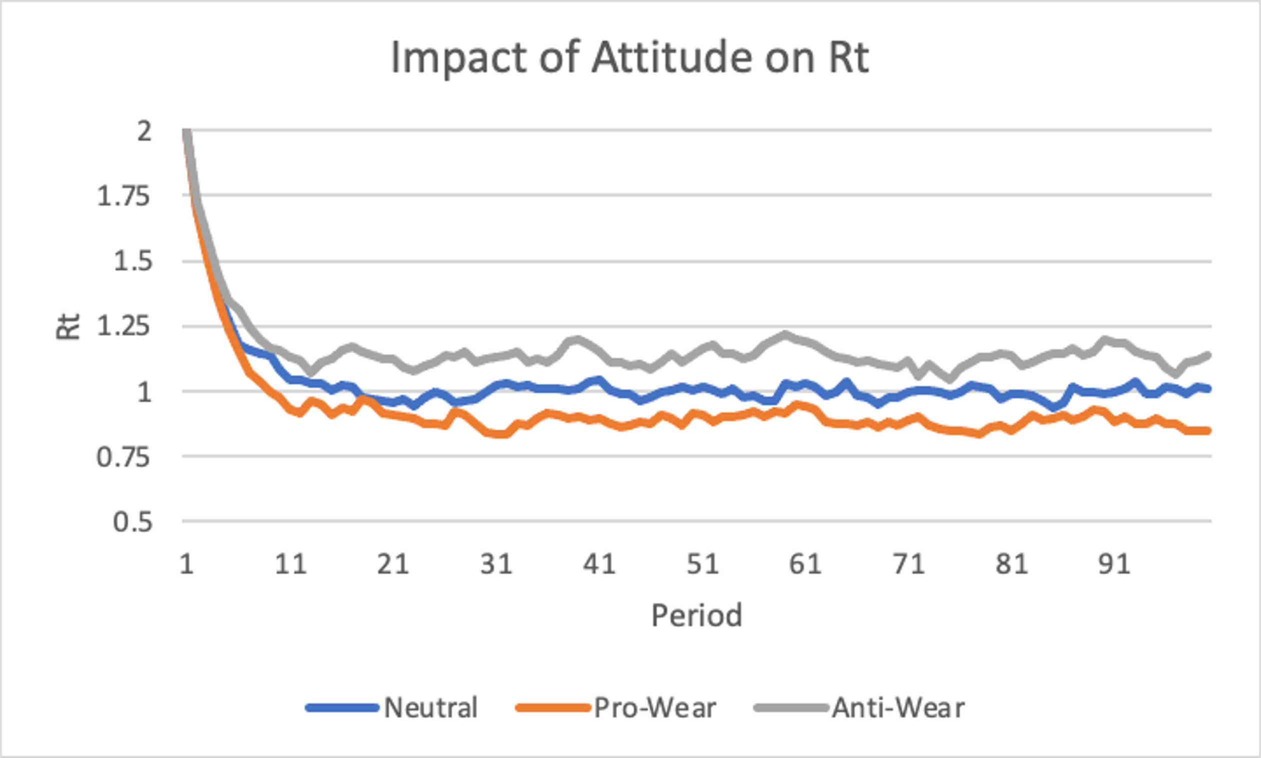
Control: Damped Oscillations of Infection Rate

- Damped oscillations in Rt observed at all levels
- Attitude Representation
    - Context-free norm chunks
        - Wear vs don’t wear
    - Contextual (Rt) reactions
        - “Fear” vs “hope”
    - Chunk activation strength
    - Context estimated from temporal aggregation
- Match context -> action
    - Norm damping effect
    - Contextual oscillations



Influence of Attitude on Rt

- Strengths of pro- and anti-mask attitudes represented as distinct activation of corresponding beliefs
- Stronger impact on asymptote of infection rate than initial response
  - Due to effect of natural reactions ("hope" and "fear") that are triggered regardless of attitude by the current state of the pandemic
- Population as heterogeneous mix of attitudes activated at different points, moving between curves

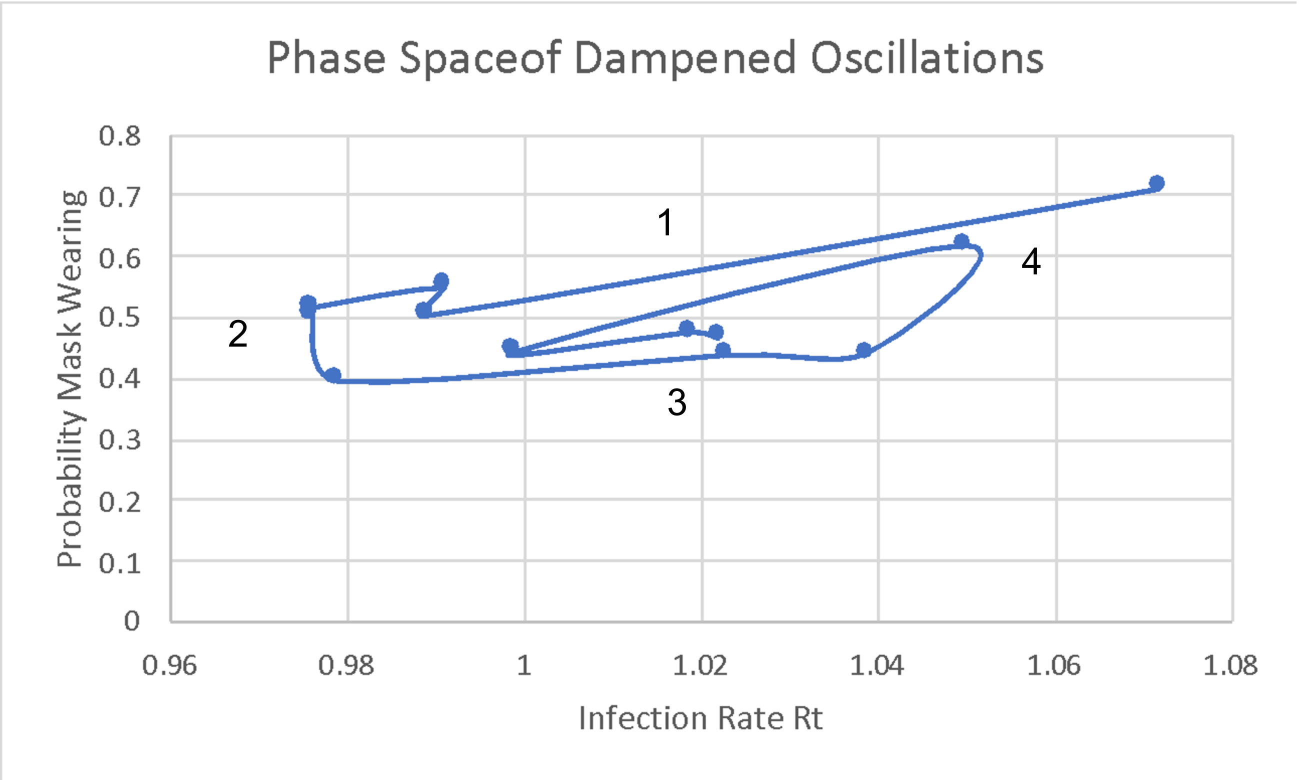


Dynamics of Attitudes and Infections

- Feedback delays in dynamical system control induce oscillations

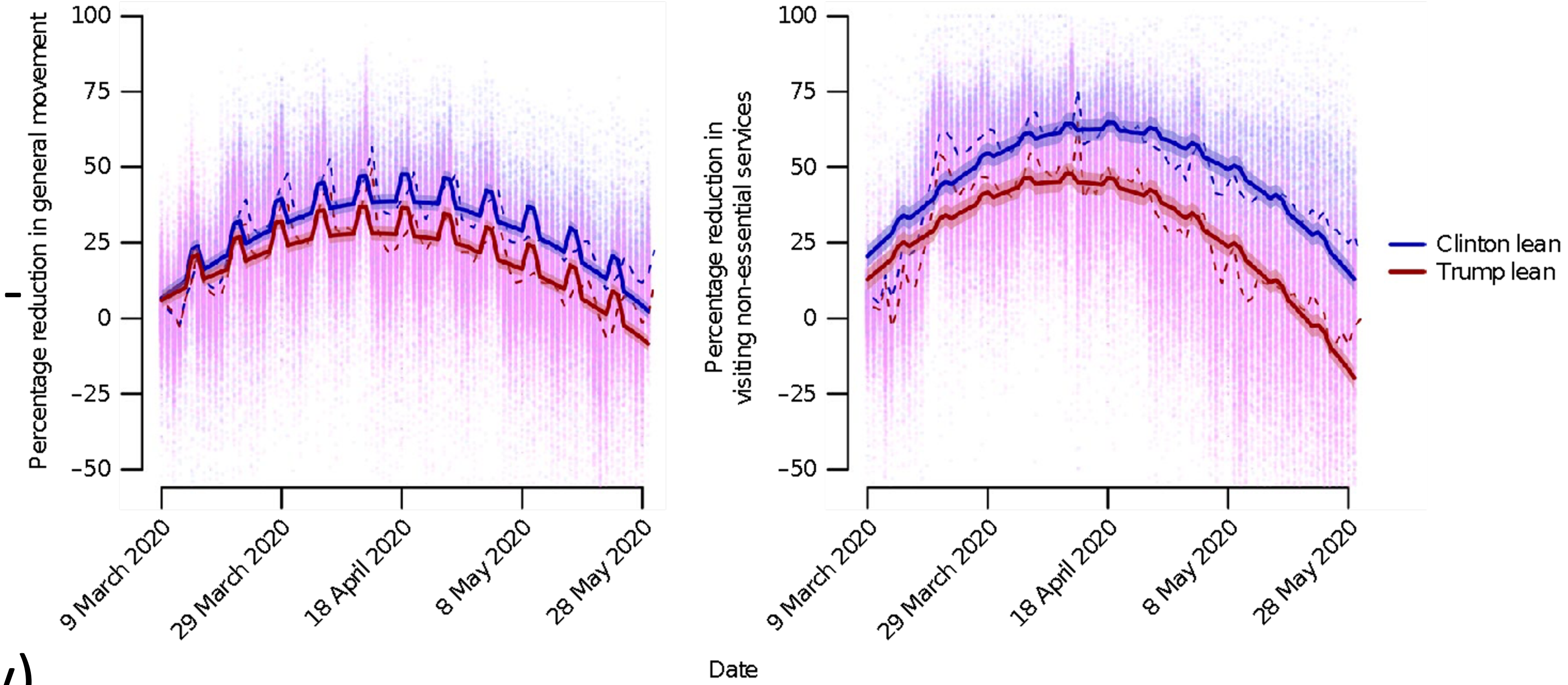
1. High mask wearing compliance reduce initial infection rate
2. Lower infection rate leads to relaxing in mask wearing
3. Lower mask wearing probability leads to rising infection rate
4. Rising infection rates leads to increase in mask wearing

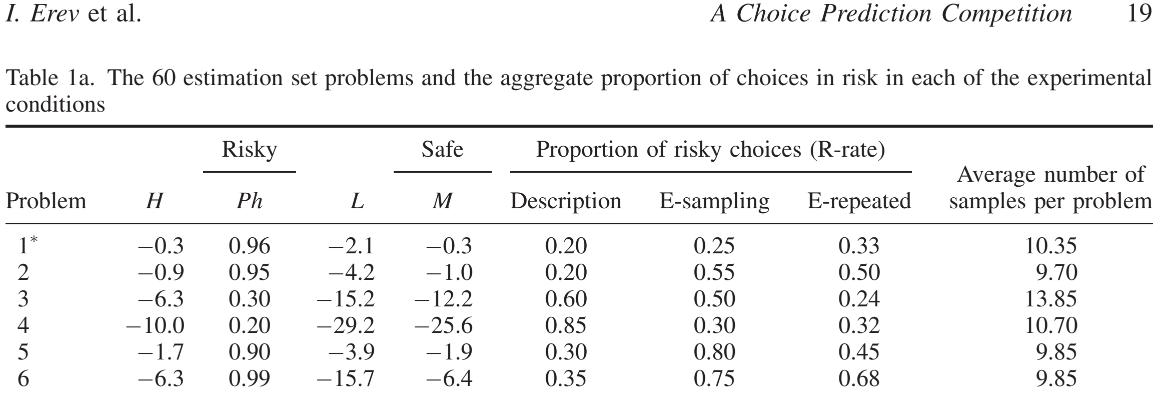
- Amplitude of oscillations decrease over time but can rise with external shocks (mandates, weather, events)



Behavioral Attitudes: Increasing Polarization

- Mobility reduction decreases over time
- Differences tied to socio-political orientation amplifies over time
- Fundamental task is decision making under uncertainty (safe vs risky)
- Well-studied problem can lead to dynamic effects such as cognitive biases





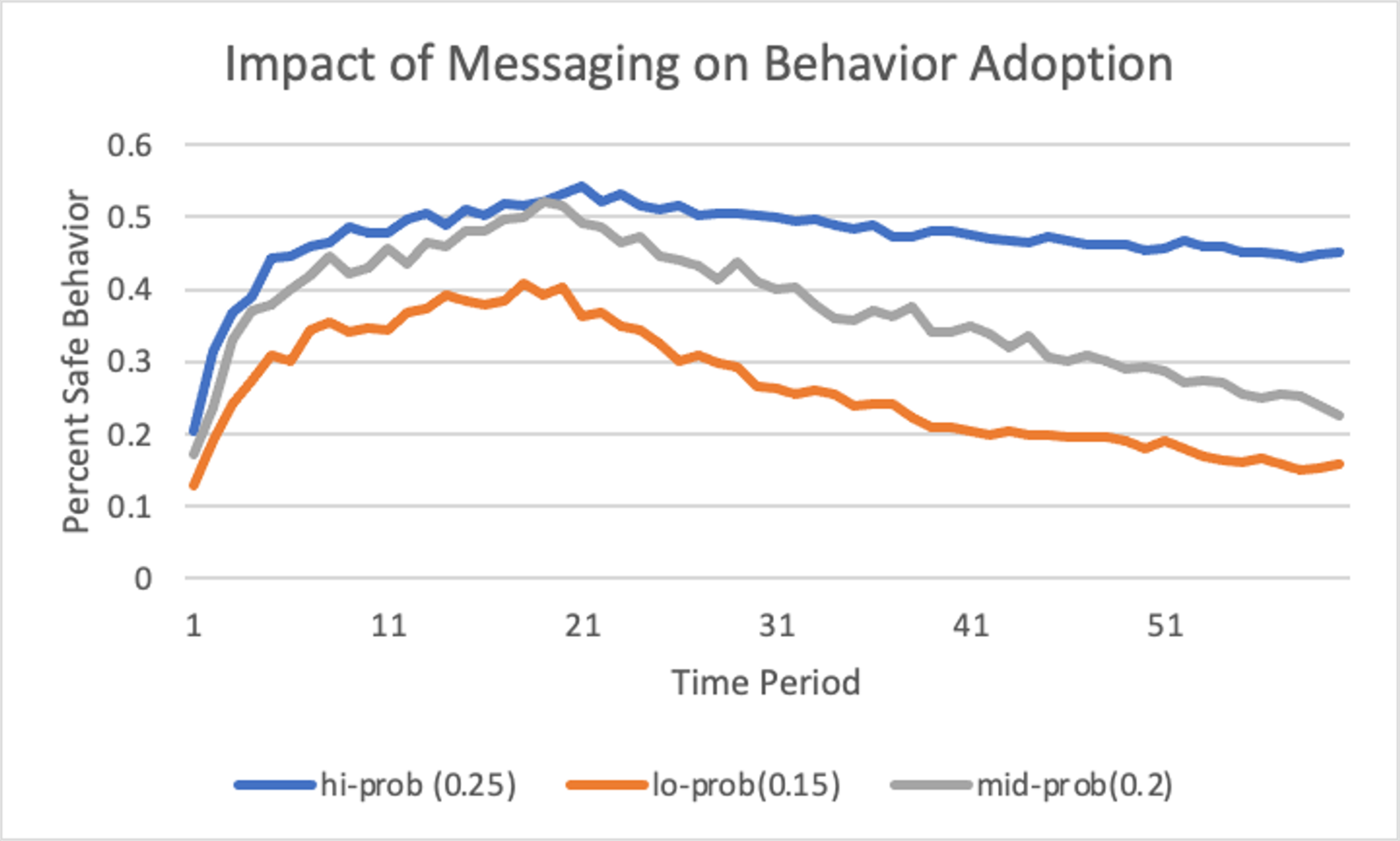
I. Erev et al. *A Choice Prediction Competition* 19

Table 1a. The 60 estimation set problems and the aggregate proportion of choices in risk in each of the experimental conditions

		Risky		Safe	Proportion of risky choices (R-rate)			
Problem	*H*	*Ph*	*L*	*M*	Description	E-sampling	E-repeated	Average number of samples per problem
1*	−0.3	0.96	−2.1	−0.3	0.20	0.25	0.33	10.35
2	−0.9	0.95	−4.2	−1.0	0.20	0.55	0.50	9.70
3	−6.3	0.30	−15.2	−12.2	0.60	0.50	0.24	13.85
4	−10.0	0.20	−29.2	−25.6	0.85	0.30	0.32	10.70
5	−1.7	0.90	−3.9	−1.9	0.30	0.80	0.45	9.85
6	−6.3	0.99	−15.7	−6.4	0.35	0.75	0.68	9.85

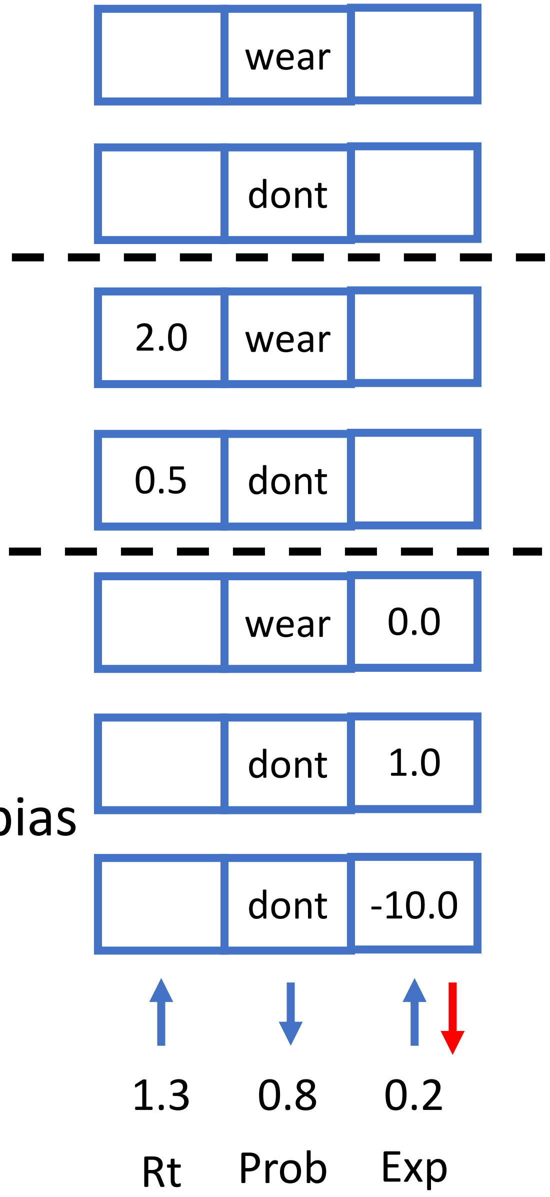
Impact of Messaging on Behavior Adoption

- Previous behavior presented as risky
- Alternative behavior slightly less desirable
- Messaging competes with experience
- No confirmation bias storage of expectation
- Standard parameters



Integrating 3 Levels of Modeling

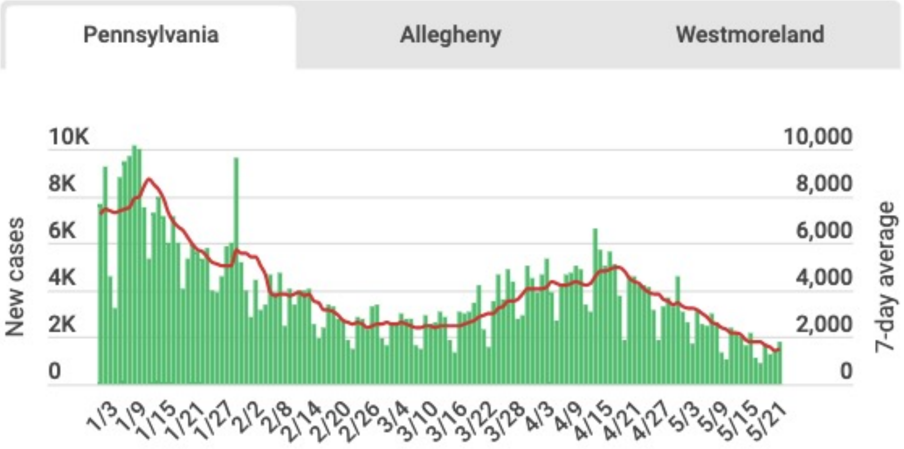
- Pro- and anti-mask wearing norms
  - Driven by prevalence of social norms but context-free
- Attitude-driven reactions
  - Context is estimate of pandemic expansion or contraction (~Rt)
  - Driven by affective reactions such as hope and fear
- Decision-making under risk and uncertainty
  - Driven by combination of risk messaging and direct experience
  - Reflects sampling dynamics, e.g., risk aversion and confirmation bias
- Combine reactive, affective and reflective decision making
  - Generate expectation -> capture habituation effects
  - Generate action to meet context and expectation
  - Activation strength of each structure modulates its effect
    - Reflect individual dispositions and exposure to information environment

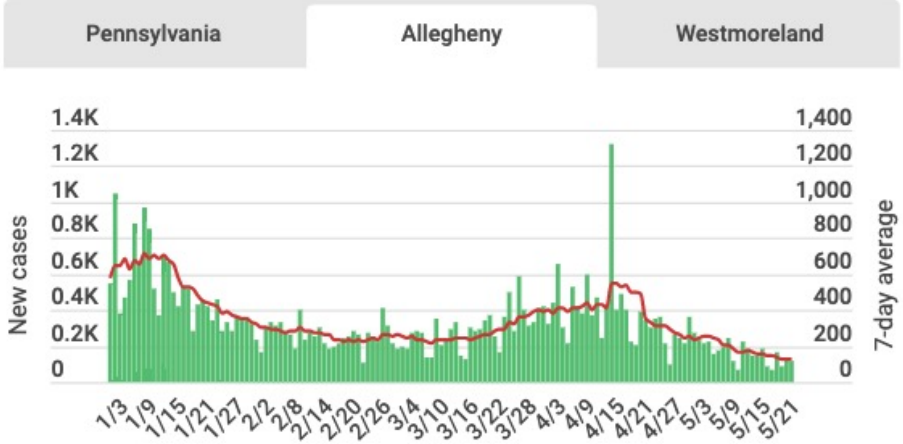


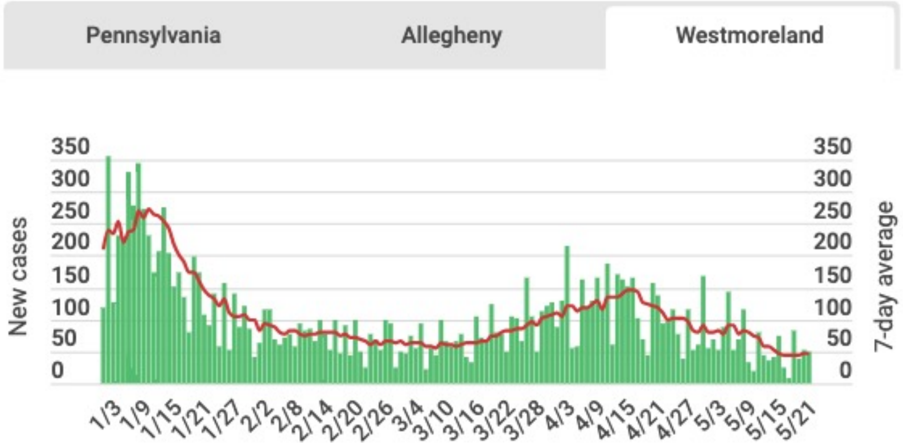
	wear	
	dont	
2.0	wear	
0.5	dont	
	wear	0.0
	dont	1.0
	dont	-10.0
1.3	0.8	0.2
Rt	Prob	Exp

Integrating Multiple Factors

- Modeling previously unseen pattern:
  - starting from a stable plateau (Rt~1) of February-March
  - gradual (quasi-linear, not exponential) increase in late March
  - plateau about 40-50% above original level in early April
  - decrease symmetric to original increase in late April through May
- Pattern reproduced at national, state and local levels
  - Largely independent of local characteristics
- Pattern present in cases, hospitalizations and deaths with usual time lag

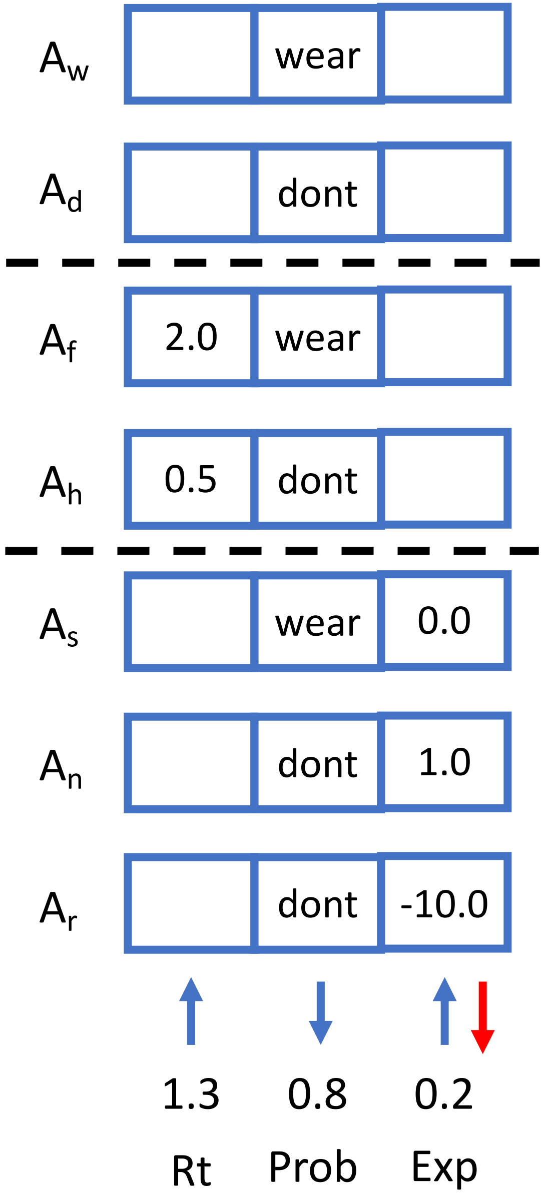






Cognitive and Biological Mechanisms

- Using four concurrent biological and cognitive mechanisms
- Two external (pharmaceutical/biological) mechanisms
  - gradual introduction of vaccines that proportionally reduces Rt from a roughly linear decrease in the Susceptible population
  - sigmoid increase of factor ~0.4 in R0 (and consequently Rt) resulting from the spread of the more contagious British variant
- Two internal (cognitive/behavioral) mechanisms:
  - fairly sudden relaxation of behavior triggered by the introduction of vaccines and seasonal/official relaxation
    - Expressed in first level of cognitive model, i.e. norms, in terms of proportional increase of A_d activation level
  - gradual tightening in behavior resulting from the original increase in cases
    - Expressed at the second level of the cognitive model, i.e., higher A_f activation resulting from closer match to increasing Rt

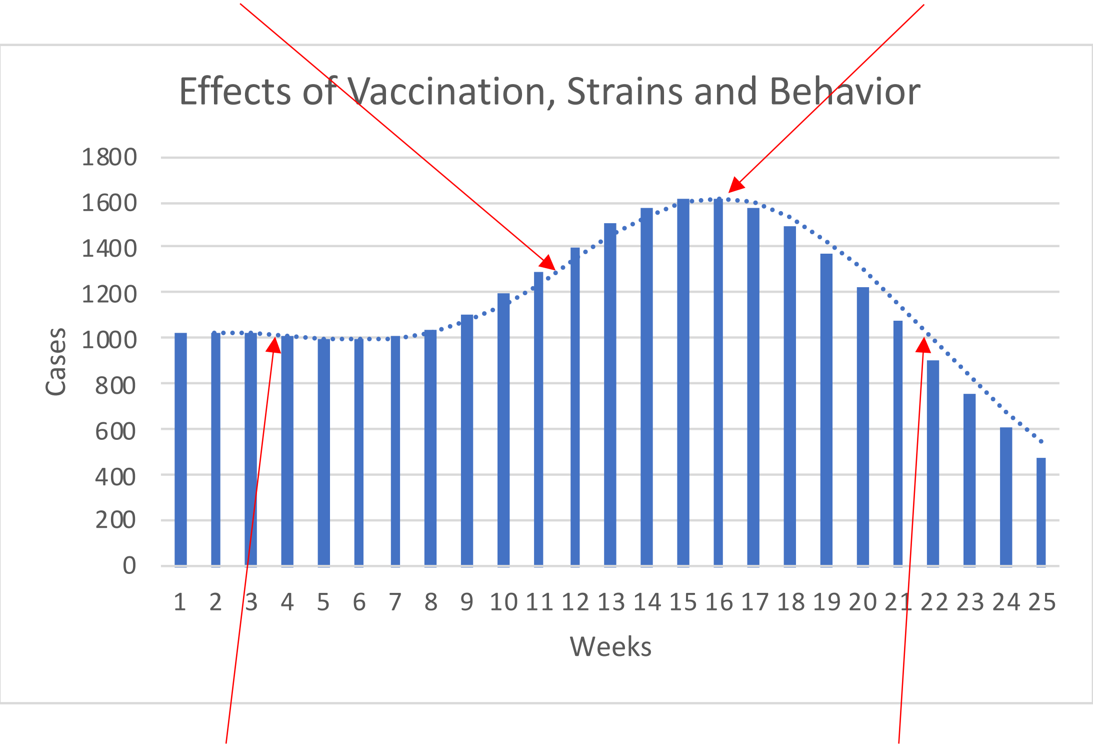


A_w		wear	
A_d		dont	
A_f	2.0	wear	
A_h	0.5	dont	
A_s		wear	0.0
A_n		dont	1.0
A_r		dont	-10.0
	↑	↓	↑↓
	1.3	0.8	0.2
	Rt	Prob	Exp

Phases of Model Behavior

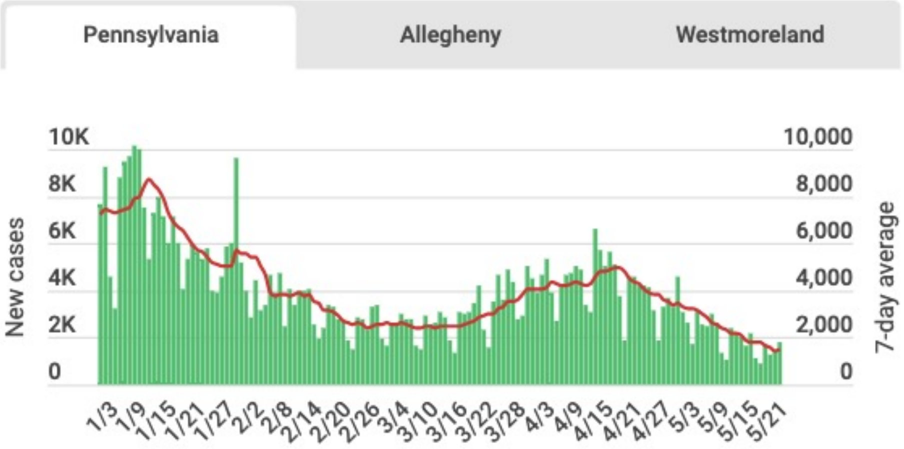
Near-exponential explosion of British variant and behavior relaxation overcomes gradual vaccination

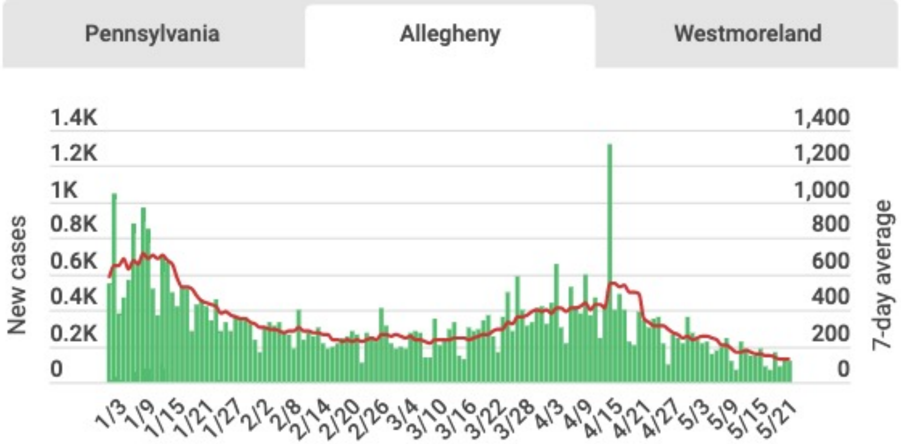
Behavior tightening and increasing vaccinations stabilize against British variant

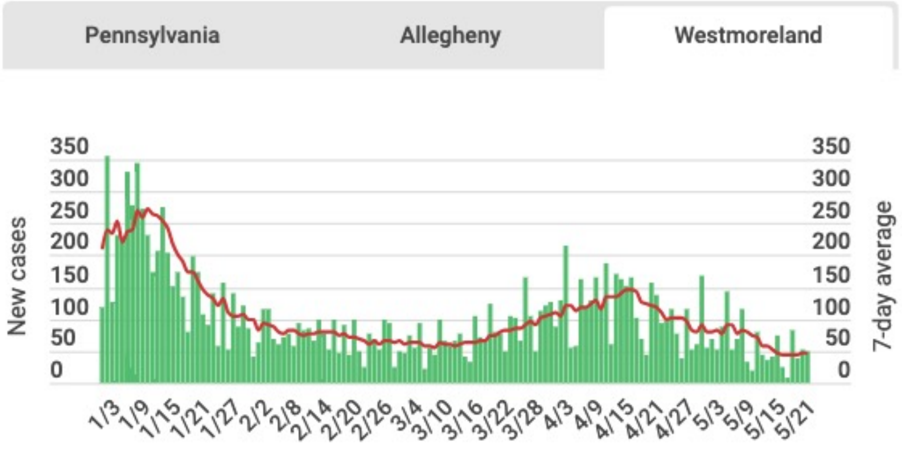


Early, slow increase in vaccinations and British variant, stabilized by cognitive dynamics

Vaccinations approaching herd immunity and behavior tightening overcome British variant saturation

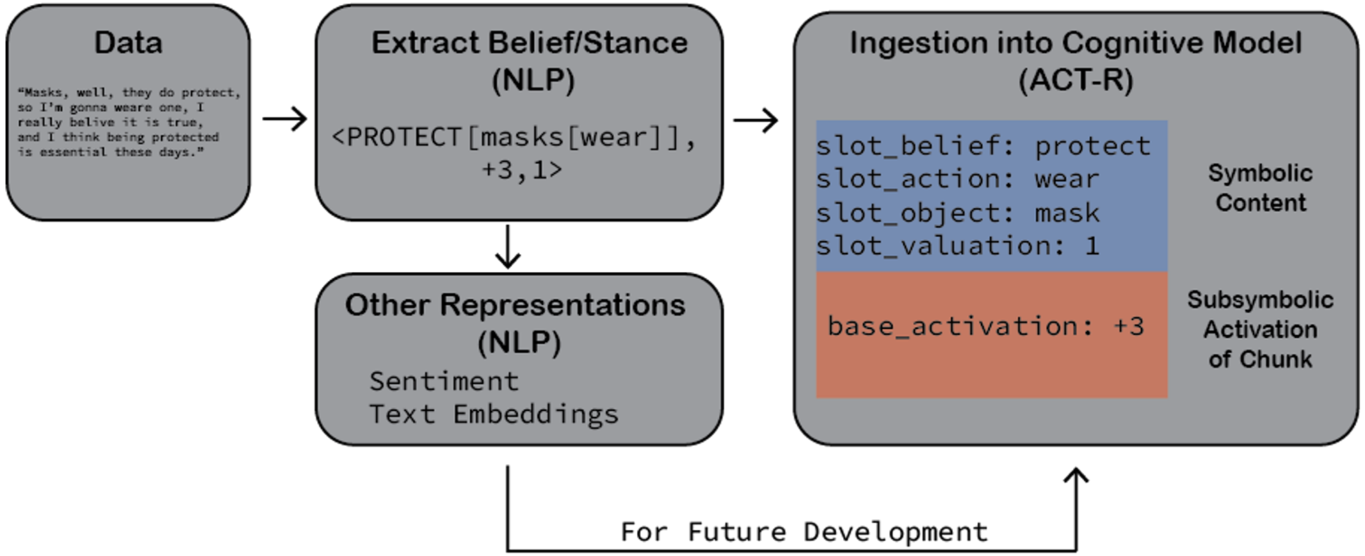






Future Work

- Integrate 3 model levels & time scales of dynamics
- Improve fidelity of belief stance representations
- Integrate effect of diverse stances in model dynamics
- Integrate external events in control dynamics
- Extend model to account for vaccination decisions
- Integrate behavior model into large-scale ABM



Data

"Masks, well, they do protect, so I'm gonna weare one, I really belive it is true, and I think being protected is essential these days."

Extract Belief/Stance (NLP)

<PROTECT[masks[wear]], +3,1>

Other Representations (NLP)

Sentiment
Text Embeddings

Ingestion into Cognitive Model (ACT-R)

```
slot_belief: protect
slot_action: wear
slot_object: mask
slot_valuation: 1
```

Symbolic Content

```
base_activation: +3
```

Subsymbolic Activation of Chunk

For Future Development

Acknowledgements

This Work is supported by the National Science Foundation Grant No. 2033390 and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via 2020-20092500001.