Humans of AI

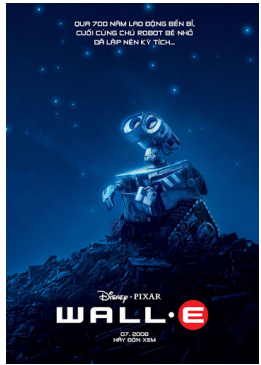
Modeling Humans in Collaborative AI Systems

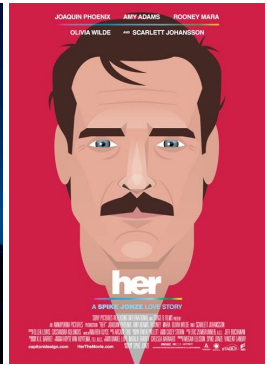
Shiwali Mohan

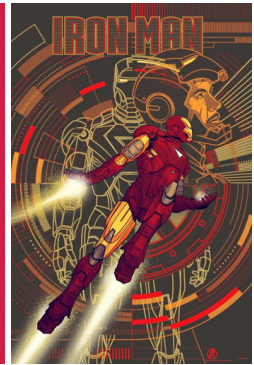
July 10, 2020

Senior Member of Research Staff, Palo Alto Research Center

Namaste!

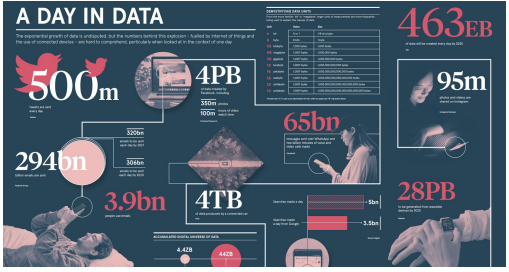


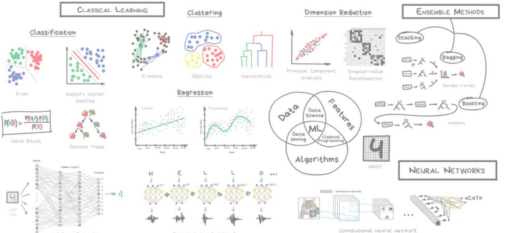




Intelligent collaborators:
independent, long-living entities with
goal-driven, problem-solving behavior who
interact and communicate with humans
learning from their experience

AI Research Now

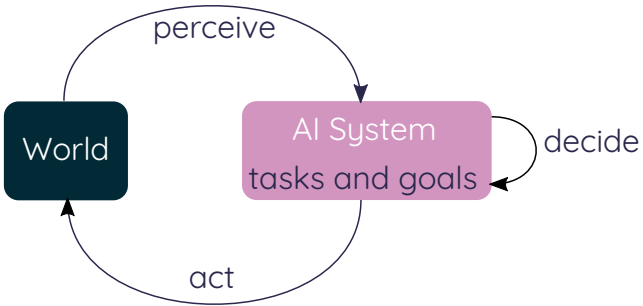




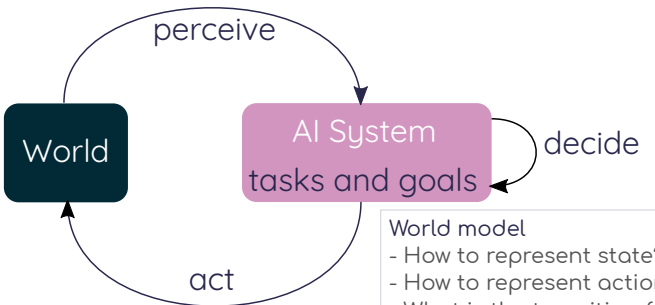
Show your favorite algorithm is better than SOTA on a task-agnostic metric. Fin.

Will algorithmic research by itself result in methods for designing intelligent collaborators?

'AI as a System' View

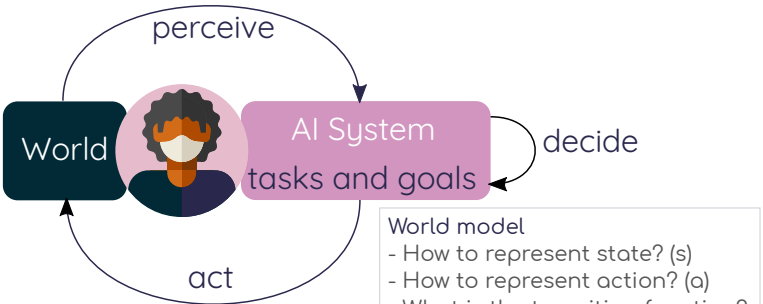


'AI as a System' View



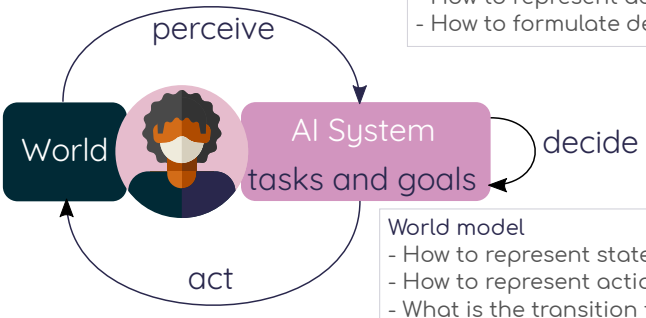
World model
- How to represent state? (s)
- How to represent action? (a)
- What is the transition function?
- How are decisions evaluated?

'AI as a System' View



World model
- How to represent state? (s)
- How to represent action? (a)
- What is the transition function?
- How are decisions evaluated?

'AI as a System' View



Human model?
- How to represent state?
- How to represent action?
- How to formulate decisions?

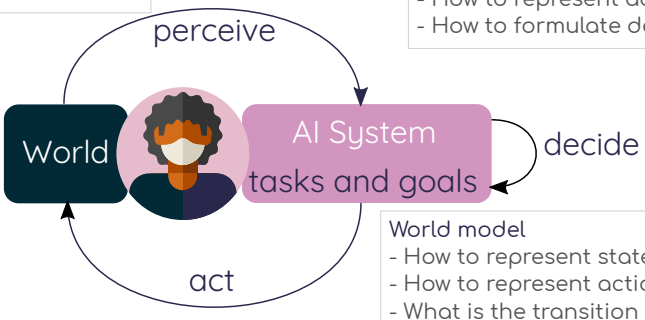
World model
- How to represent state? (s)
- How to represent action? (a)
- What is the transition function?
- How are decisions evaluated?

'AI as a System' View

Artificial Intelligence Journal 2018

Albrecht, S. V., & Stone, P. 2018. *Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems.* Artificial Intelligence, 258, 66-95.

Human model?
- How to represent state?
- How to represent action?
- How to formulate decisions?



World model
- How to represent state? (s)
- How to represent action? (a)
- What is the transition function?
- How are decisions evaluated?

AAAI Presidential Address 2018

Khambhampat, S. 2018. *Challenges of Human-Aware AI Systems*

How do we model the human collaborator?

- that is computational and explicit
- that is causal and prescriptive
- enables reasoning and learning about the human

A Constrained Approach

1. Define a problem where AI success crucially depends on modeling the human collaborator
2. Develop human-centered desiderata, metrics, and evaluation scheme
3. Adopt methods from humanist sciences, design prescriptive models
4. Design interactive AI algorithms, embody in end-to-end systems

Outline

1. Why model humans?
2. Interactive task learning
3. Sustainable transportation
4. Health behavior change
5. Humans of AI

1. Interactive Task Learning

The Problem: Deployment in Dynamic Environments

- AI designers cannot predict all deployment usecases
- AI should be designed to enable non-expert human programming

Interactive Task Learning: A New Field of Enquiry

Ernst Strungmann Forum 2017

Machine learning: Tom Mitchell
Cognitive architectures: Shiwali Mohan, John Laird, Ken Forbus, Christian LeBiere, Paul Rosenbloom
Robotics: Andrea Thomaz, Julie Shah, Maya Cakmak, Peter Stone, Matthias Scheutz
Psychology: Ken Koedinger, Andrea Stocco, Peter Pirolli
NLP: Joyce Chai, Parisa Kordjamshidi

Modeling the Human Teacher's Expectations

Approach: Cognitive science, HRI

Classical machine learning

- Batch: dataset -> model
- Phased: training -> testing
- Passive: learn when asked
- Big data
- Data: confounding

Interactive learning

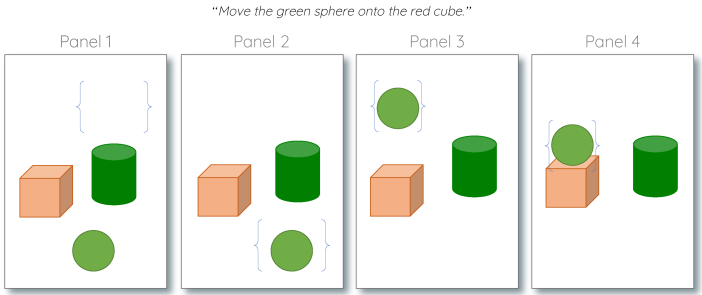
- Incremental: experience -> knowledge
- Online
- Active: learn when failed
- Small data
- Teacher: benevolent

Embodied Language Learning

Approach: Psycholinguistics

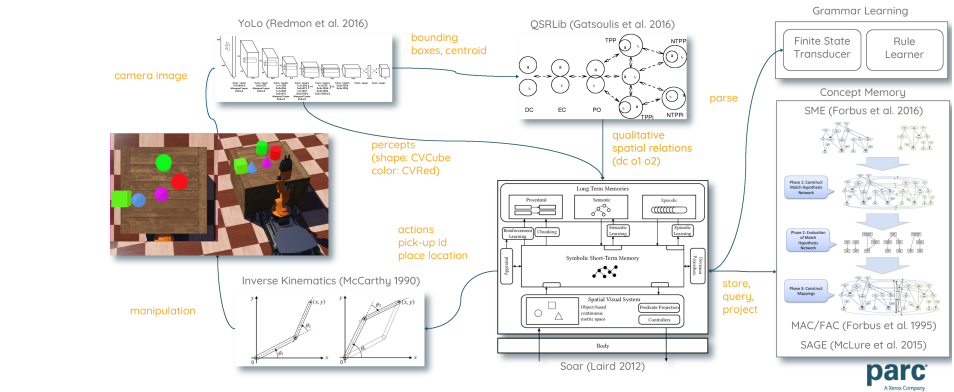
DARPA GAILA: Where does meaning come from?

- NLP (BERT, GPT) derives meaning from statistical patterns in word usage
- GLP (VQA task) derives meaning from paired visual/linguistic stimuli
- The Indexical Hypothesis (Glenberg and Robertson 1999): Humans use language referentially in service of collaborative action (**Mohan** et al. 2013)



AILEEN

Approach: Advanced Cognitive Systems



Shiwali Mohan, Matt Klenk, Matthew Shreve, Kent Evans, Aaron Ang, John Maxwell. *Characterizing an Analogical Concept Memory for Newellian Cognitive Architectures.* To appear in the Proceedings of the Eigth Annual Conference on Advances in Cognitive Systems (ACS). 2020

Interactive Concept Learning

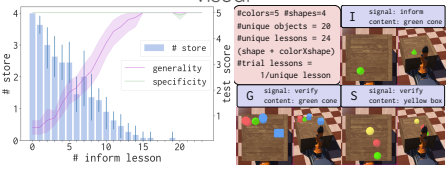
Demo!

Evaluating Interactive Concept Learning

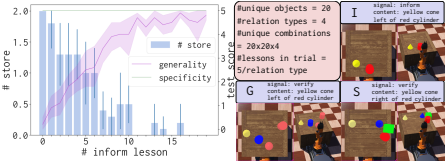
A New Evaluation Scheme

- Not just accuracy!
- What can be learned?
- Is learning active?
- How quick is the learning?
- How general is the acquired knowledge?
- How correct is the acquired knowledge?

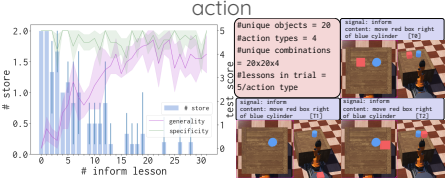
visual



spatial



action



Quick Look

1. Define a human-model-based AI problem
   - How do humans teach?
   - What constraints does human teaching put on a learning system?

2. Define human-centered metrics
   - Quick, active learning
   - Generalization and correctness

3. Adopt methods from humanist sciences
   - Psycholinguistics - the Indexical Hypothesis (Glenberg and Robertson 1999)
   - Cognitive science - analogical reasoning and generalization (Gentner, 2003)

4. Design interactive AI systems
   - Hybrid-AI, end-to-end systems
   - Deep learning, cognitive architectures

Publications

1. Preeti Ramaraj, Matt Klenk, **Shiwali Mohan**. *Understanding Intentions in Human Teaching to Design Interactive Task Learning Robots.* To appear in Robotics Science and Systems Workshops. 2020
2. **Shiwali Mohan**, Matt Klenk, Matthew Shreve, Kent Evans, Aaron Ang, John Maxwell. *Characterizing an Analogical Concept Memory for Newellian Cognitive Architectures.* To appear in the Proceedings of the Eigth Annual Conference on Advances in Cognitive Systems (ACS). 2020
3. John Laird and **Shiwali Mohan**. *Learning Fast and Slow: Levels of Learning in General Autonomous Intelligent Agents.* In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI/Blue Sky). 2018.
4. **Shiwali Mohan** and John Laird. *Learning Goal-Oriented Tasks from Situated Interactive Instruction* In Proceedings of the Twenty Eigth AAAI Conference on Artificial Intelligence (AAAI). 2014.

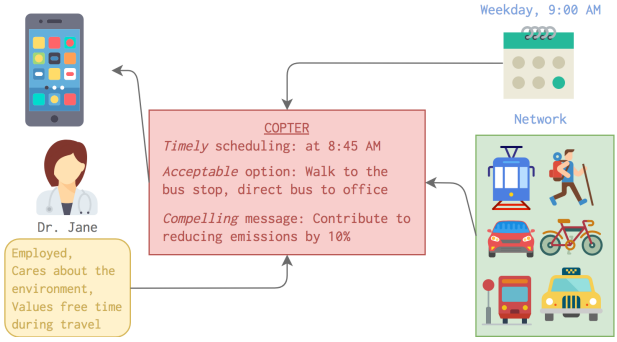
Michigan's Rosie can learn over 60 games and several manipulation and navigational tasks from natural instruction! http://soargroup.github.io/rosie/

2. Sustainable Transportation

The Problem: Energy Consumption in Transportation

- Transportation is one of the largest consumers of energy - 29% of energy in US in 2016
- It is far from efficient - both under and over utilization of networks
- Congestion wastes 6.1 billion hours and 3.1 billion gallons fuel per year (Schrank et al. 2015)
- Energy efficient transportation is an important technology and policy problem - ARPA-E TransNet 2013-2018

The Influence Problem



Shiwali Mohan, Hesham Rakha, and Matt Klenk. *Acceptable Planning: Influencing Individual Behavior to Reduce Transportation Energy Expenditure of a City.* Journal of Artificial Intelligence Research 66 (2019)

Understanding Acceptability

Choice theory from behavioral economics: Tversky and Kahneman, 1986

- Utility can be expressed in terms of attributes relevant to the decision
  - mode dependent: cost, distance, time
  - person dependent: income, education

$$\text{val}(x_i, p) = \gamma_1 \times x_{i1} + \ldots + \gamma_k \times x_{ik} + \lambda_1 \times f_{p1} + \ldots + \lambda_l \times f_{pl}$$

- Observed choice is determined by portion of utility

$$\Pr(i, p) = \frac{e^{\text{val}(x_i, f_p)}}{\sum_{j \in C} e^{\text{val}(x_j, f_p)}}$$

- Switching cost (-gain) to change route, higher gain, better acceptability

$$\Delta_{r,u} = -\Delta_{u,r} = \ln \frac{\Pr(r, p)}{\Pr(u, p)}$$

Designing An Acceptable Planning System

Approach: Hybrid goal-driven planning

Recommend a plan acceptable to an individual giver their local transportation context

1. Formulate multi-modal planning problem taking into account local transportation network, determine plausible alternatives (Dvorak, **Mohan**, Bellotti, and Klenk 2018)
2. Bias plan recommendation by acceptability. How do we estimate acceptability of an arbitrary route?
3. Evaluate energy saving capacity of such an approach. How do we measure the impact?

Estimating Acceptability

Approach: Machine learning

$$\Delta_{r,u} = -\Delta_{u,r} = \ln \frac{Pr(r,p)}{Pr(u,p)}$$

- Problem: multi-class prediction
- Dataset: trip data (CHTS) from CalTrans 2012 - 2012
- Features: trip related (distance), person-related (demographics), network-related (transit pass, license), experience (bike trips in the past week)
- Hypothesis: Dr. Jane's utility function is close to others' who are similar to

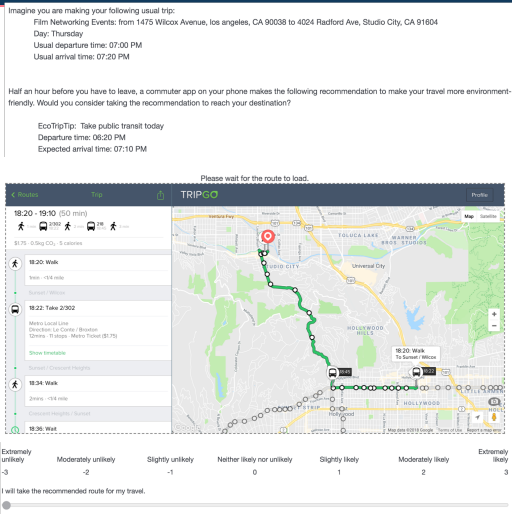
Table 1: F1 scores on 20% test set

Mode	Baseline 1	Baseline 2	RF	MLP
Walk	0.00	0.12	0.82*	0.62
Cycle	0.00	0.00	0.81*	0.28
Bus	0.00	0.02	0.78*	0.38
Subway/train	0.00	0.00	0.58*	0.05
Drive	0.72	0.56	0.93*	0.86
Ride	0.00	0.28	0.84*	0.65
Motorcycle	0.00	0.00	0.80*	0.00
Total	0.68	0.40	0.88*	0.74
Category				
Non-motorized	0.00	0.05	0.83*	0.60
Public transit	0.00	0.14	0.79*	0.43
Motorized	0.90	0.82	0.97*	0.93
Total	0.68	0.70	0.94*	0.86

Evaluating Acceptability

Approach: Behavioral Economics and Decision Theory

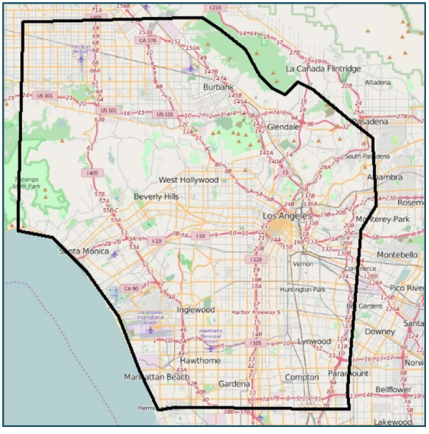
- Population: 49 (27 female, 22 male) regular drivers in LA
- Profiler survey: classifier features, regular weekly trips
- Experiments: 10 choice experiments per participant
- Analyses: mixed-effects linear and logit models
- Conclusions: acceptability impacts adoption



Measuring Impact

Approach: Complexity Science, Transportation Modeling, Agent-based Modeling

- LA transportation network: 170,000 roadway links, 1 million daily transit trips
- State-of-the-art simulation model of the LA region (Elbery et al. 2018), mode adoption model
- Experiment:10% influenced population
- Conclusions: 4% energy and 20% time savings in Los Angeles



Quick Look

1. Define a human-model-based AI problem
   - What underlies a person's choice of transportation mode?
   - Which mode recommendations are acceptable?

2. Define human-centered metrics
   - Acceptability and adoption
   - Choice experiments

3. Adopt methods from humanist sciences
   - Behavioral economics
   - Complexity science
   - Transportation modeling

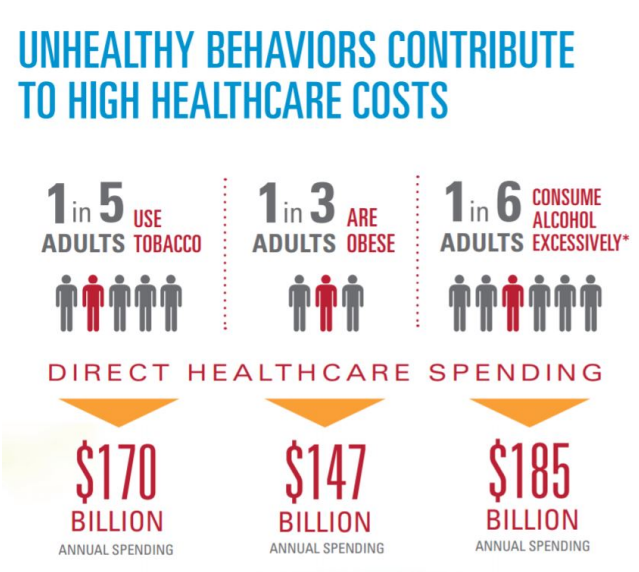
4. Design interactive AI systems
   - Multi-modal planning
   - Machine learning
   - Expected embodiment as an app

Publications

1. **Shiwali Mohan**, Hesham Rakha, and Matt Klenk. *Acceptable Planning: Influencing Individual Behavior to Reduce Transportation Energy Expenditure of a City.* Journal of Artificial Intelligence Research 66 (2019)
2. **Shiwali Mohan**, Matthew Klenk, and Victoria Bellotti. *Exploring How to Personalize Travel Mode Recommendations For Urban Transportation." IUI Workshops* (2019)
3. **Shiwali Mohan**, Frances Yan, Victoria Bellotti, Ahmed Elbery, Hesham Rakha, Matt Klenk *On Influencing Individual Behavior for Reducing Transportation Energy Expenditure in a Large Population.* Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (2019)
4. Filip Dvorak, **Shiwali Mohan**, Victoria Bellotti, Matt Klenk. *ICAPS Proceedings of the 6th Workshop on Distributed and Multi-Agent Systems*

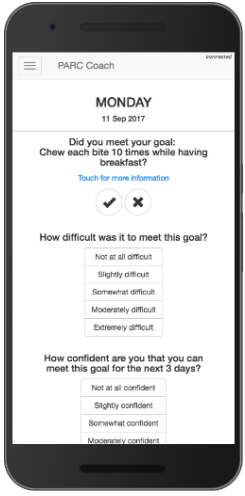
3. Health Behavior Change

The Problem: Healthcare Costs from Unhealthy Behaviors



Individual Coaching for Behavior Change

- Human-human coaching is extremely powerful but costly
- Devices are pervasive
- Can intelligent systems support health behavior change?



NutriWalking

Approach: Interactive AI system

Goal: Support sedentary trainees develop a habit of regular aerobic exercise

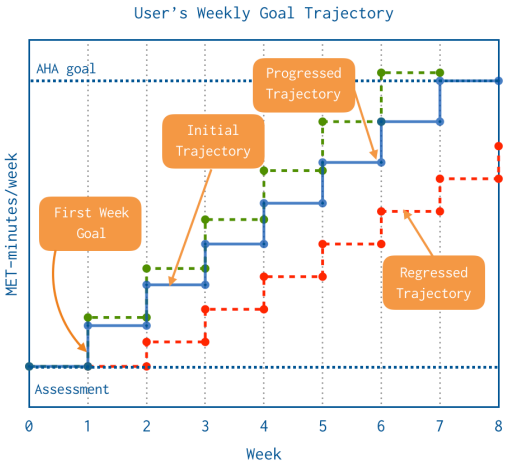
- Exercise: walking; versatile, easy, effective
- Trainees: people managing diabetes
- Embodiment: a smartphone application
- Interaction: assessment, daily reports, weekly goal setting
- AI: adaptive goal setting, scheduling and planning

Shiwali Mohan, Anusha Venkatakrishnan, Andrea Hartzler. *Designing an AI Health Coach and Studying its Utility in Promoting Regular Aerobic Exercise.* In ACM Transactions on Interactive Intelligent Systems. 2020

Modeling Aerobic Capability Growth

Approach: Physical therapists practitioners model

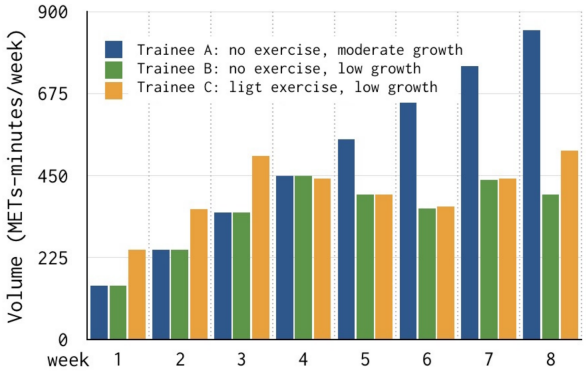
- Goal setting theory (Shilts and Townsend 2000): set goals that are difficult yet attainable
- AHA recommended level of activity - 30 minutes of moderate activity 5 times a week
- Practitioners guidelines
  - Frequency, Intensity, Time, Type to adjust Progress Volume (FITT-VP)
- Adaptive goal setting



Evaluating Adaptive Goal Setting

Approach: four-pronged, multi-disciplinary

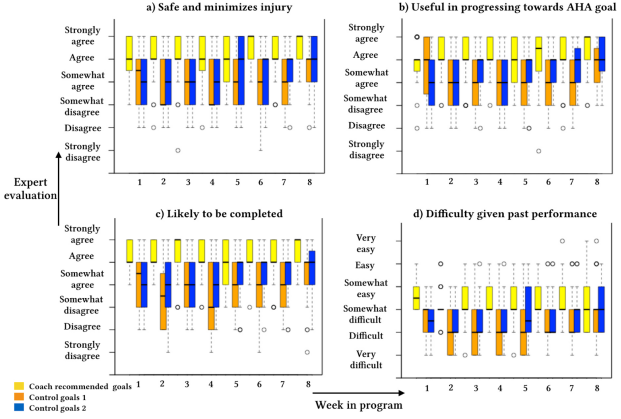
1. Is it adaptive?



Evaluating Adaptive Goal Setting

Approach: four-pronged, multi-disciplinary

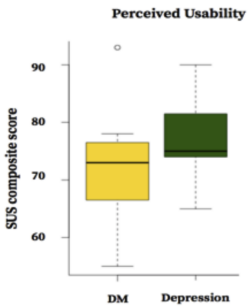
2. Do it experts agree with it?

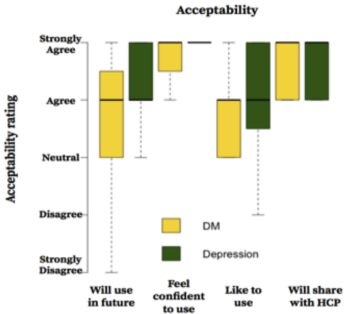


Evaluating Adaptive Goal Setting

Approach: four-pronged, multi-disciplinary

3. Do trainees understand it? 8 participants managing diabetes, 7 managing depression





Evaluating Adaptive Goal Setting

Approach: four-pronged, multi-disciplinary

4. Is it effective?

21 participants managing diabetes used NutriWalking for 6 weeks, ecological contexts!

1. Increased exercise volume [✓]
2. Over-optimistic with self-assessment [X] [✓]
3. Personalized goals + collaborative selection led to more successful completion [✓]
4. Rate of perceived measurement scale provides informative feedback for adaptation [✓]

Quick Look

1. Define a human-model-based AI problem
   - What underlies a person habits?
   - How does a person learn a new behavior?

2. Define human-centered metrics
   - Safety, difficulty
   - Acceptability
   - Behavior change in ecological contexts

3. Adopt methods from humanist sciences
   - Cognitive psychology
   - Physical therapy
   - Human-computer interaction

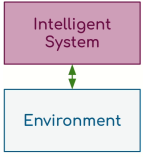
4. Design interactive AI systems
   - Adaptive goal setting: scheduling + measurement
   - Embodiment in an app

Publications

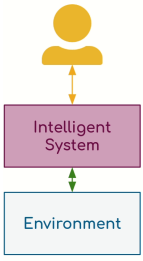
1. **Shiwali Mohan**. *Exploring the Role of Common Model of Cognition in Designing Adaptive Coaching Interactions for Health Behavior Change.* (minor revisions). In ACM Transactions on Interactive Intelligent Systems. 2019.
2. **Shiwali Mohan**, Anusha Venkatakrishnan, Andrea Hartzler. *Designing an AI Health Coach and Studying its Utility in Promoting Regular Aerobic Exercise.* In ACM Transactions on Interactive Intelligent Systems. 2020
3. Aaron Springer, Anusha Venkatakrishnan, **Shiwali Mohan**, Les Nelson, Michael Silva, Peter Pirolli. *Leveraging Self-Affirmation to Increase mHealth Behavior Change.* In Journal of Medical Information Research. 2018
4. Peter Pirolli, **Shiwali Mohan**, Anusha Venkatakrishnan, Les Nelson, Michael Silva, Aaron Springer. *Implementation Intention and Reminder Effects on Behavior Change in a Mobile Health System: A Predictive Cognitive Model.* In Journal of Medical Information Research. 2017
5. **Shiwali Mohan**, Anusha Venkatakrishnan, Michael Silva, and Peter Pirolli. *On Designing a Social Coach to Promote Regular Aerobic Exercise.* In the Proceedings of the 29th IAAI/AAAI. 2017
6. Andrea Hartzler*, Anusha Venkatakrishnan*, **Shiwali Mohan**, Paula Lozano, James D Ralston, ...,Les Nelson, Peter Pirolli. *Acceptability of a Team-Based Mobile Health Application for Lifestyle Self-Management in Individuals in Chronic Illnesses* In 38th Annual International Conference of the Engineering in Medicine and Biology Society. 2016

Humans of AI

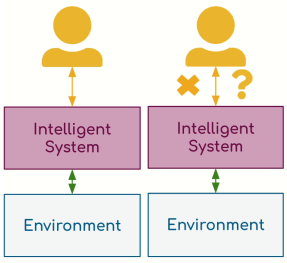
Modeling Humans in AI Systems



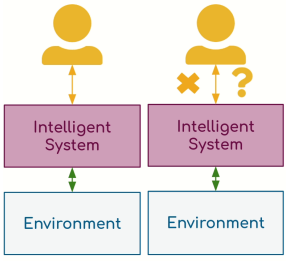
Modeling Humans in AI Systems

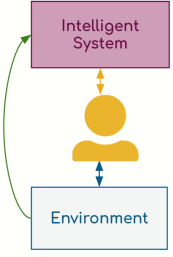


Modeling Humans in AI Systems

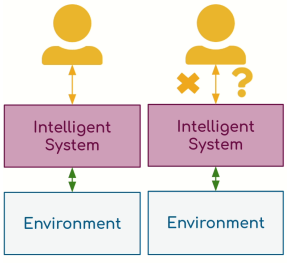


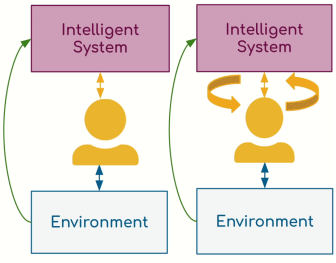
Modeling Humans in AI Systems



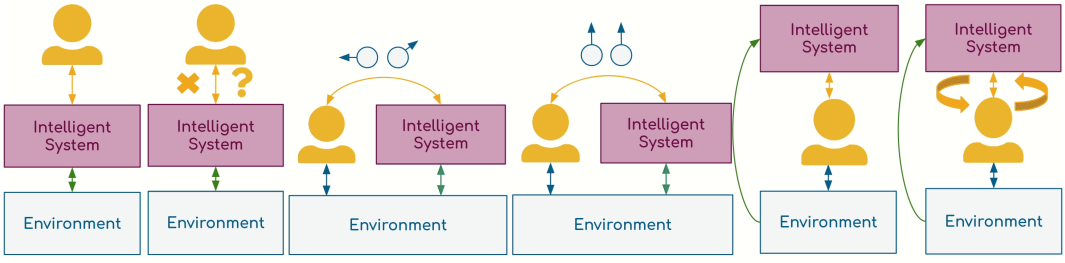


Modeling Humans in AI Systems

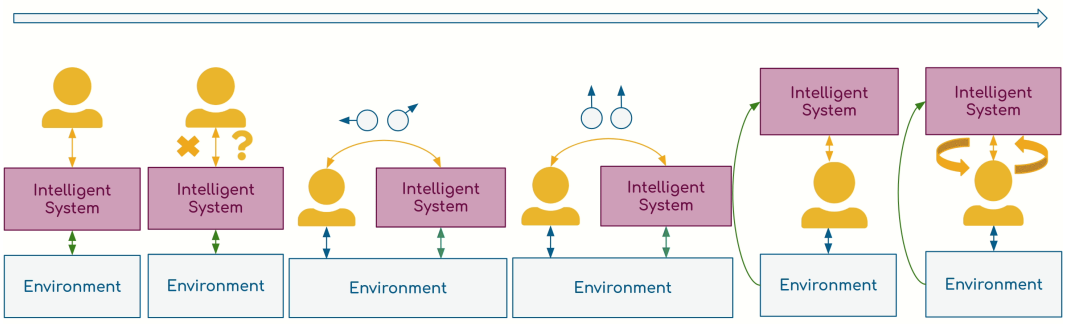




Modeling Humans in AI Systems



Modeling Humans in AI Systems



Acknowledgements

Colleagues:
Victoria Bellotti, John Laird, Peter Pirolli, Matt Klenk, Anusha Venkatakrishnan, Matthew Shreve, John Maxwell, Aaron Ang, Kent Evans, Preeti Ramaraj, Charlie Ortiz, James Kirk, Aaron Mininger, Kyle Dent, Bob Price and many others

Funders:

Humans of AI: Takeaways

- AI and ML is not just CV, NLP, Games
  - Truly hard and extremely interesting AI problems exist out of these traditions
  - Look beyond prediction, to intelligent, long-term, sequential decision making
  - Improve problem domain equity: maternal healthcare in India, drought-robust agricultural policy in India, non-traditional education & training in US
- Humans of AI are important - even critical - and cannot be an afterthought
- Adoption of AI in human societies requires a comprehensive understanding of human-AI ecosystems

@shiwalimohan
shiwali.mohan@parc.com, shiwali.mohan@gmail.com