Caltech

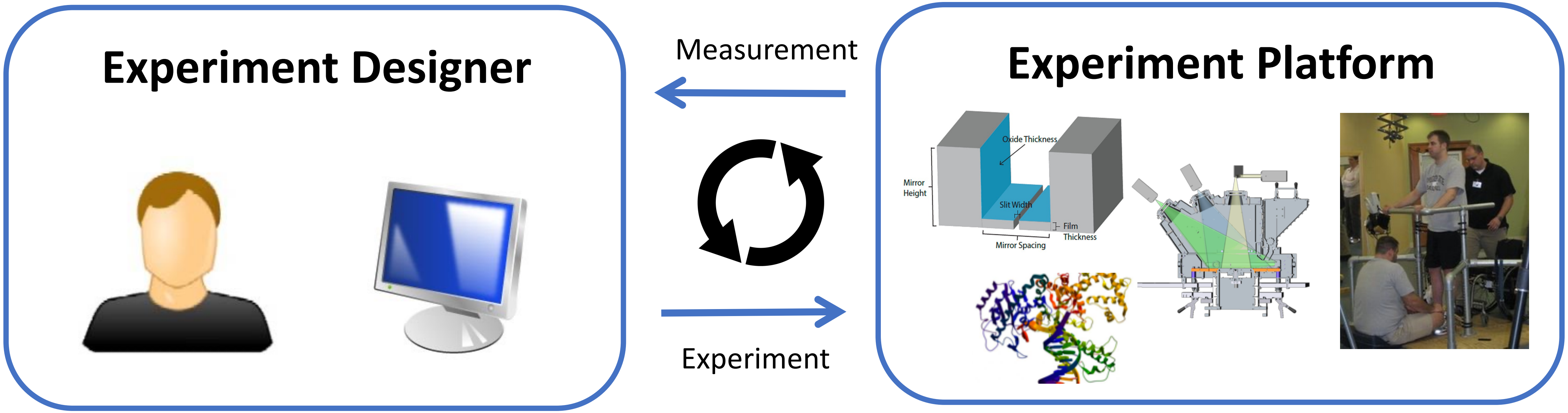
Personalized Preference Learning

from Spinal Cord Stimulation to Exoskeletons

Yisong Yue

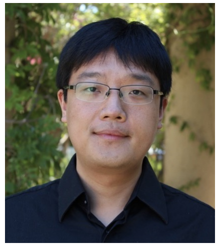
Goal-Oriented Experiment Design

(aka "Interactive Learning”)



- Iterative & adaptive
- Utility maximizing (for a suitable definition of utility)

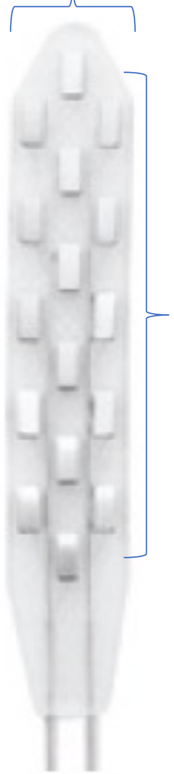
Personalized Therapies



Yanan Sui

Medtronic human array

10 mm



49 mm

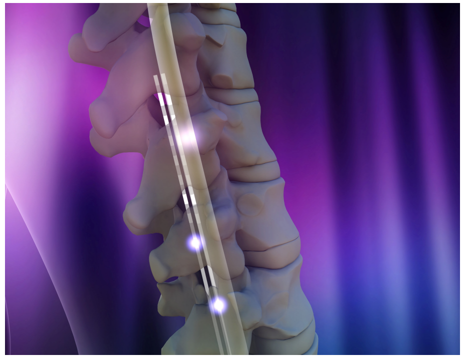


Image source: williamcapicottomd.com

SCI Patient

Each patient is unique

10^9 possible configurations!

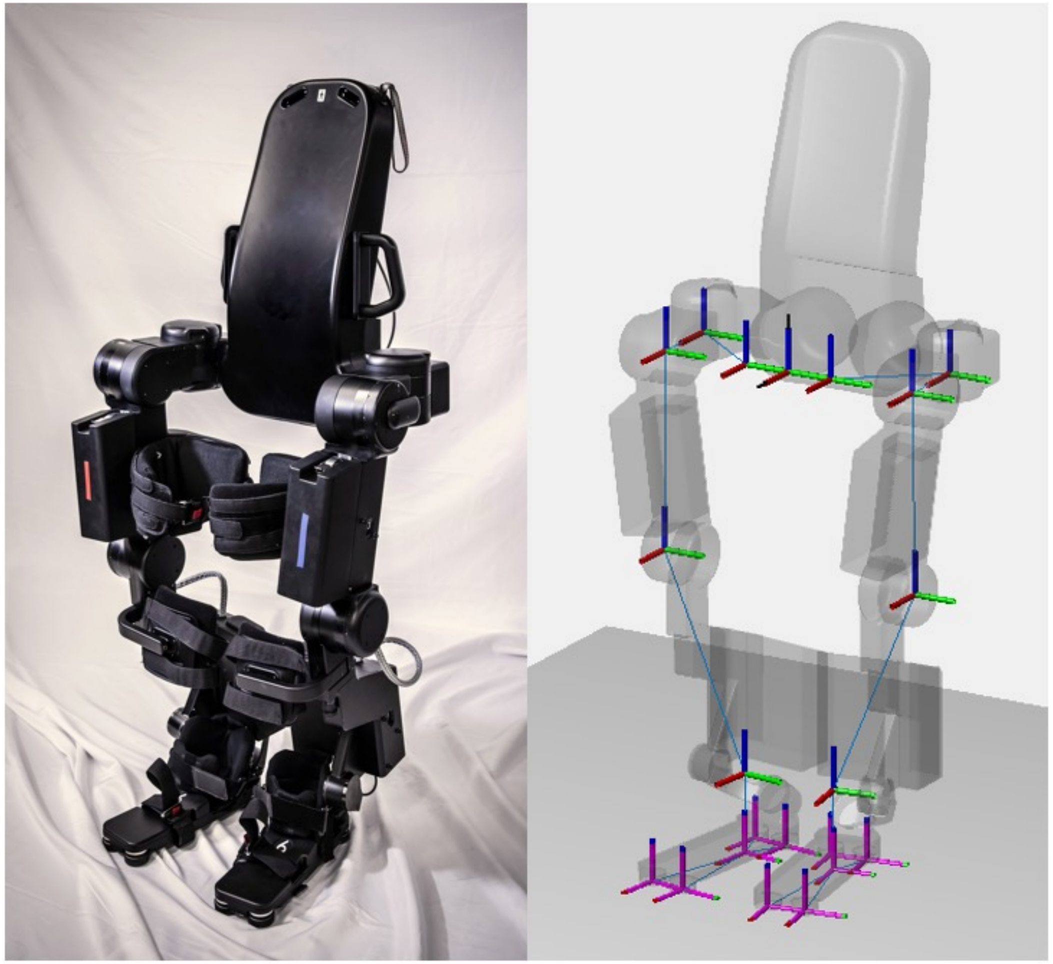
Personalized Therapies

Maegan Tucker

Ellen Novoseller

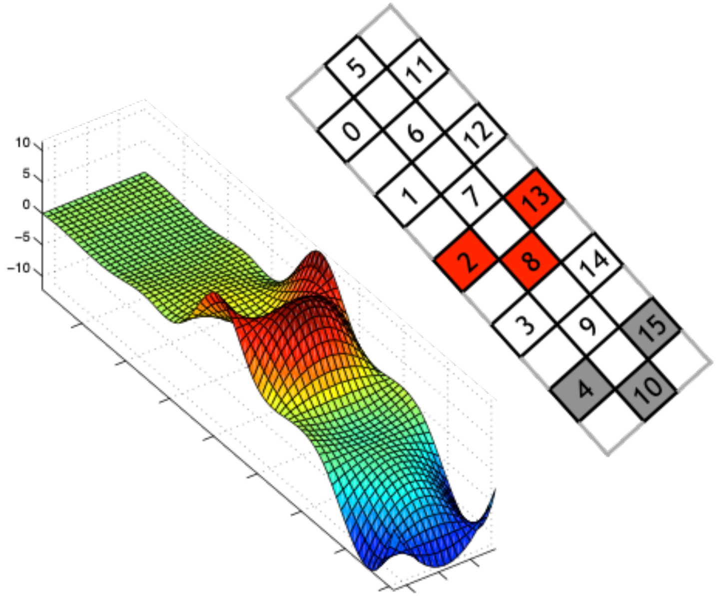
Amy Li

Erdem Biyik



Atalante Exoskeleton by Wandercraft

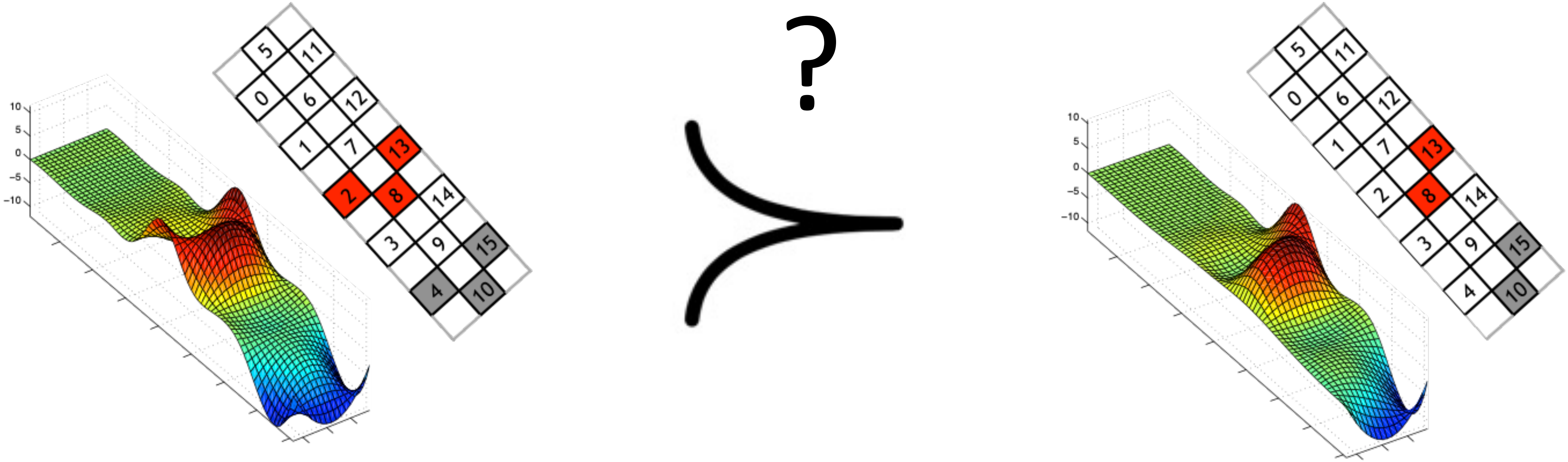
Key Question: How to Elicit Feedback?



Absolute Feedback: “That felt good, 4/5 rating.”

Challenge: humans are not consistent in providing absolute feedback.

Our Approach: Preference Feedback



Multi-dueling Bandits with Dependent Arms, Sui, Zhuang, Burdick & Yue, UAI 2017
Correlational Dueling Bandits with Application to Clinical Treatment in Large Decision Spaces, Sui, Yue & Burdick, IJCAI 2017
Preference-Based Learning for Exoskeleton Gait Optimization, Tucker, Novoseller, et al., ICRA 2020 *(Best Paper Award)*
Human Preference-Based Learning for High-dimensional Optimization of Exoskeleton Walking Gaits, Tucker et al., IROS 2020
ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes, Li, Tucker, et al., ICRA 2021

Preference-Based Interactive Learning

Core Algorithms & Theory

The K-armed Dueling Bandits Problem

Yisong Yue
Dept. of Computer Science
Cornell University
Ithaca, NY 14853
yyue@cs.cornell.edu

Josef Broder
Center for Applied Math
Cornell University
Ithaca, NY 14853
jbroder@cam.cornell.edu

Robert Kleinberg
Dept. of Computer Science
Cornell University
Ithaca, NY 14853
rdk@cs.cornell.edu

Thorsten Joachims
Dept. of Computer Science
Cornell University
Ithaca, NY 14853
tj@cs.cornell.edu

Bayesian Active Learning for Classification and Preference Learning

Neil Houlsby, Ferenc Huszár, Zoubin Ghahramani, Máté Lengyel
Computational and Biological Learning Laboratory
University of Cambridge

Towards Preference-Based Reinforcement Learning

Johannes Fürnkranz, Eyke Hüllermeier,
Weiwei Cheng, Sang-Hyeun Park

Applications to Computer Systems

In Proceedings of UIST 2005, Seattle, WA, USA

Preference Elicitation for Interface Optimization

Krzysztof Gajos and Daniel S. Weld
University of Washington
Seattle, WA 98195, USA
{kgajos,weld}@cs.washington.edu

Active Exploration for Learning Rankings from Clickthrough Data

Filip
Department
Co
Ith
filip@

Applications to Autonomous Systems

Learning Preferences for Manipulation Tasks from Online Coactive Feedback

Ashesh Jain, Shikhar Sharma, Thorsten Joachims, Ashutosh Saxena
Department of Computer Science, Cornell University, USA.
{ashesh, ss2986, tj, asaxena}@cs.cornell.edu

Active Preference-Based Learning of Reward Functions

Dorsa Sadigh, Anca D. Dragan, Shankar Sastry, and Sanjit A. Seshia
University of California, Berkeley, {dsadigh, anca, sastry, sseshia}@eecs.berkeley.edu

Remainder of Talk

Sparring Family of Algorithms

Mainly these two papers

- Sparring [Ailon, Karnin, Joachims, ICML 2014]
- SelfSparring [Sui, Zhuang, Burdick, Yue, UAI 2017]
- CoSpar [Tucker, Novoseller, Kann, Sui, Yue, Burdick, Ames, ICRA 2020] *(Best Paper Award)*
- LineCoSpar [Tucker, Cheng, Novoseller, Cheng, Yue, Burdick, Ames, IROS 2020]

Briefly Survey Other Considerations

Measurements are relative rather than absolute!

(Basic) Interaction Protocol

1. Choose two actions (x_1,x_2)
2. Compare x_1 vs x_2 *(stochastic)*
3. Feedback $y=1$ if x_1 wins ($y=0$ o.w.)
4. Repeat

Example Preference Matrix

	A	B	C	D	E	F
A	0	0.03	0.04	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	0.03
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Values are Pr(row > col) – 0.5

The K-armed Dueling Bandits Problem, Yue, Broder, Kleinberg, Joachims, COLT 2009

Measurements are relative rather than absolute!

(Basic) Interaction Protocol

1. Choose two actions (x_1,x_2)
2. Compare x_1 vs x_2 *(stochastic)*
3. Feedback y=1 if x_1 wins (y=0 o.w.)
4. Repeat

Compare E & F: P(E > F) = 0.53

Example Preference Matrix

	A	B	C	D	E	F
A	0	0.03	0.04	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	**0.03**
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Values are Pr(row > col) – 0.5

The K-armed Dueling Bandits Problem, Yue, Broder, Kleinberg, Joachims, COLT 2009

Measurements are relative rather than absolute!

(Basic) Interaction Protocol

1. Choose two actions (x_1,x_2)
2. Compare x_1 vs x_2 *(stochastic)*
3. Feedback y=1 if x_1 wins (y=0 o.w.)
4. Repeat

Compare A & C: P(A > C) = 0.54

Example Preference Matrix

	A	B	C	D	E	F
A	0	0.03	**0.04**	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	0.03
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Values are Pr(row > col) – 0.5

The K-armed Dueling Bandits Problem, Yue, Broder, Kleinberg, Joachims, COLT 2009

Measurements are relative rather than absolute!

(Basic) Interaction Protocol

1. Choose two actions (x_1,x_2)
2. Compare x_1 vs x_2 *(stochastic)*
3. Feedback y=1 if x_1 wins (y=0 o.w.)
4. Repeat

Active Learning: Identify Entire Matrix

Optimization/Bandits: Converge to Best Action

Our Main Focus!

Compare A & A: P(A > A) = 0.50

Example Preference Matrix

	A	B	C	D	E	F
A	**0**	0.03	0.04	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	0.03
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Values are Pr(row > col) – 0.5

The K-armed Dueling Bandits Problem, Yue, Broder, Kleinberg, Joachims, COLT 2009

(Prior Work) Sparring

Standard Absolute Feedback

- Instantiate 2 Bandit algorithms: P_1 & P_2
- For t = 1, …
  - P_1 chooses x_1
  - P_2 chooses x_2
  - Compare x_1 vs x_2
  - Provide feedback
    - P_1 observes if x_1 won
    - P_2 observes if x_2 won

Player 1 (columns) / Player 2 (rows)

	A	B	C	D	E	F
A	0	0.03	0.04	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	0.03
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Reducing Dueling Bandits to Cardinal Bandits, Ailon, Karnin & Joachims, ICML 2014

Intuition

- Reduction to standard Bandit settings
  - Each player selfishly maximizes own reward

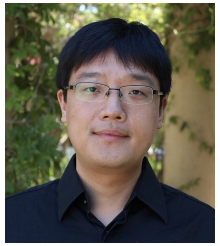
- Instantiate P_1
- For t = 1, …
  - P_1 chooses x_1
  - Plays x_1
  - Observes feedback

- Instantiate P_2
- For t = 1, …
  - P_2 chooses x_2
  - Plays x_2
  - Observes feedback

Challenges:

- Hard to analyze
- Inefficient (throws away information)

SelfSparring



Yanan
Sui

- Instantiate 1 Bandit algorithm P
- For t = 1, …
  - P chooses x_1
  - P chooses x_2
  - Compare x_1 vs x_2
  - Provide feedback
    - P observes both sides

Probabilistic Bandit Algorithm (Thompson Sampling)

Player 1

Player 2	A	B	C	D	E	F
A	0	0.03	0.04	0.06	0.10	0.11
B	-0.03	0	0.03	0.05	0.08	0.11
C	-0.04	-0.03	0	0.04	0.07	0.09
D	-0.06	-0.05	-0.04	0	0.05	0.07
E	-0.10	-0.08	-0.07	-0.05	0	0.03
F	-0.11	-0.11	-0.09	-0.07	-0.03	0

Multi-dueling Bandits with Dependent Arms, Sui, Zhuang, Burdick & Yue, UAI 2017

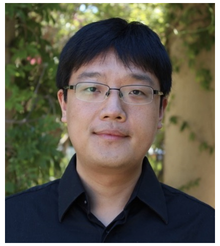
Benefits of SelfSparring

- Rigorous convergence analysis (regret bound)

- Performs very well in practice

- **Easily extendable to new settings**

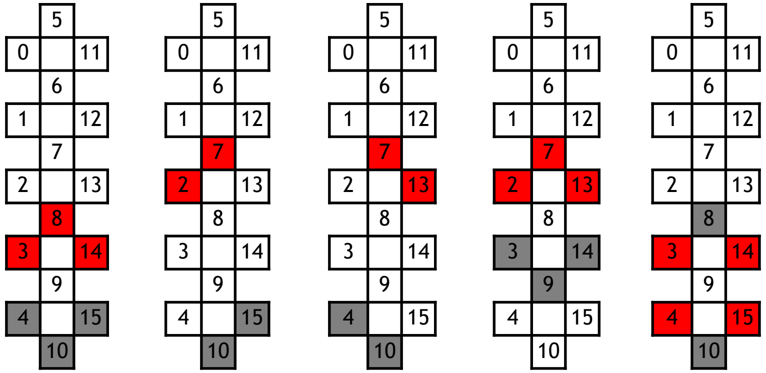
Extension: Comparing Multiple Actions



Yanan Sui

- Need to consider multiple actions before giving feedback

- E.g., rank a batch of 5 electro-stimulation stimuli



- **SelfSparring trivially generalizes!**

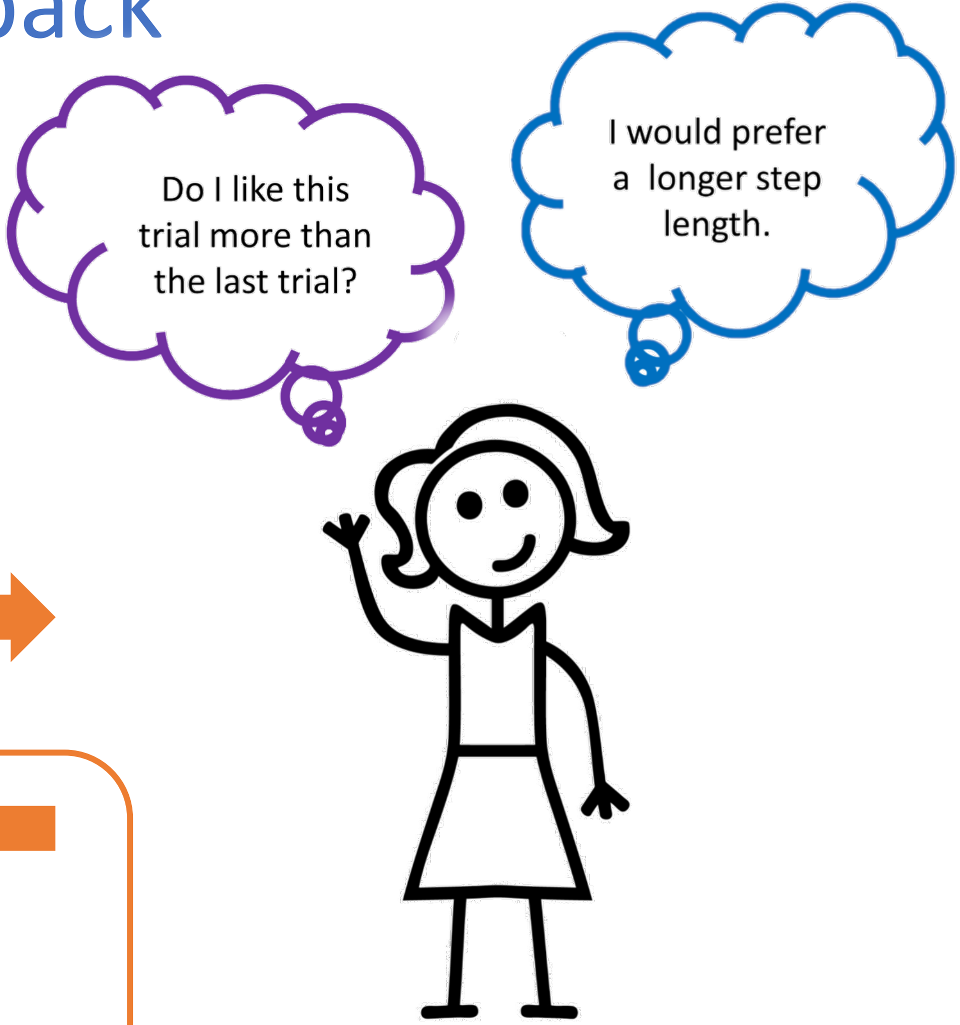
Multi-dueling Bandits with Dependent Arms, Sui, Zhuang, Burdick & Yue, UAI 2017

Clinical Experiments

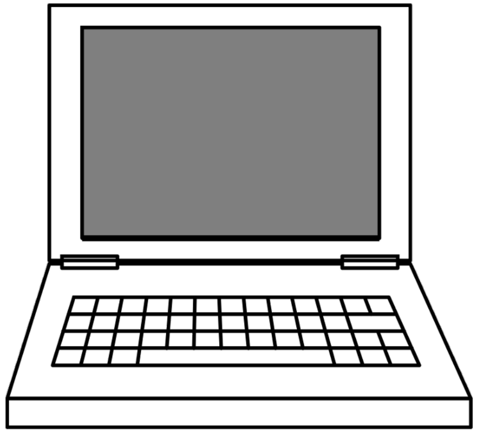
Extension: Mixed-Initiative Feedback

Maegan Tucker

Ellen Novoseller



Exoskeleton gait



CoSpar Algorithm

Agent

User feedback:

- Preferences
- Coactive feedback

Mixed-Initiative Feedback

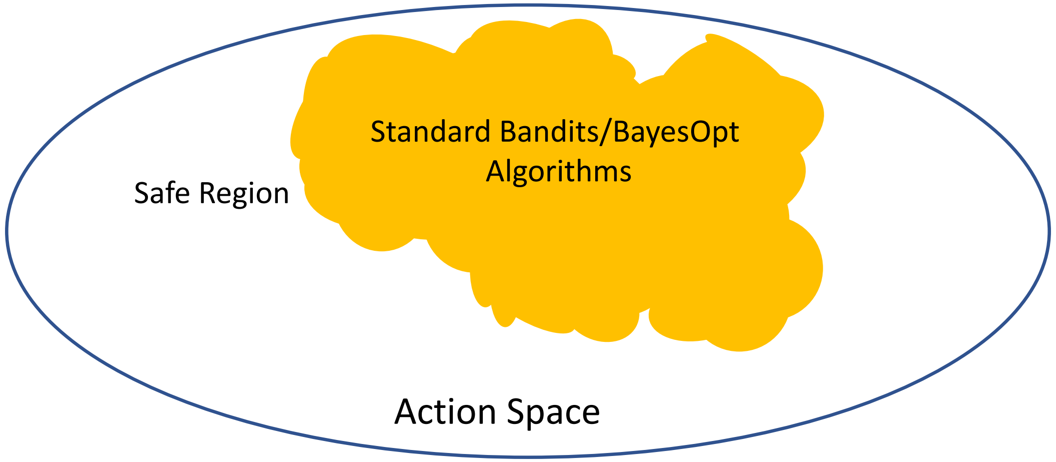
Exoskeleton User

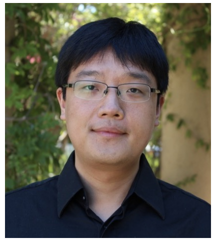
Environment

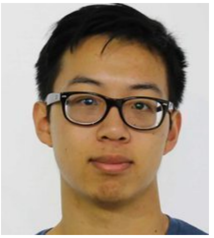
Preference-Based Learning for Exoskeleton Gait Optimization, Maegan Tucker, Ellen Novoseller, et al., ICRA 2020 *(Best Paper Award)*

Unsafe Actions

- Avoiding Unsafe or Undesirable Regions
- Requires Absolute Feedback from Users
  - **"This feels uncomfortable / painful."**




Yanan Sui


Vincent Zhuang


Amy Li


Maegan Tucker


Erdem Biyik


Ellen Novoseller

ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes, Amy Kejun Li et al., ICRA 2021
Stagewise Safe Bayesian Optimization with Gaussian Processes, Yanan Sui et al., ICML 2018

High Dimensions


Maegan Tucker

Myra Cheng



Step Length   Step Duration   Step Width   Step Height   Pelvis Roll   Pelvis Pitch

LineCoSpar: Optimize series of 1-D subspaces

(builds on LineBO alg.)

Human Preference-Based Learning for High-dimensional Optimization of Exoskeleton Walking Gaits, Maegan Tucker et al., IROS 2020

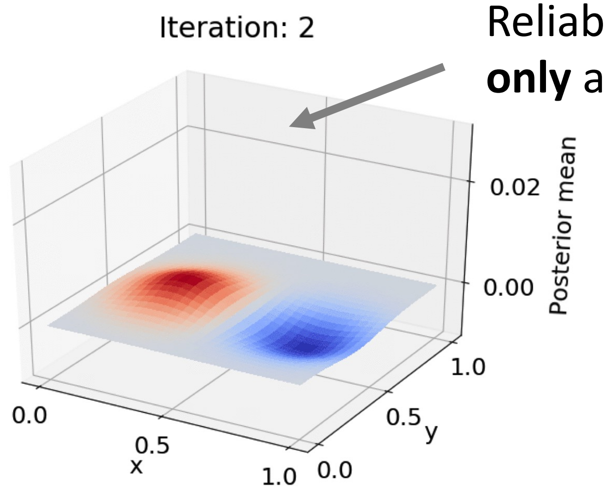
Optimization vs Active Learning

Amy Li

Maegan Tucker

Erdem Biyik

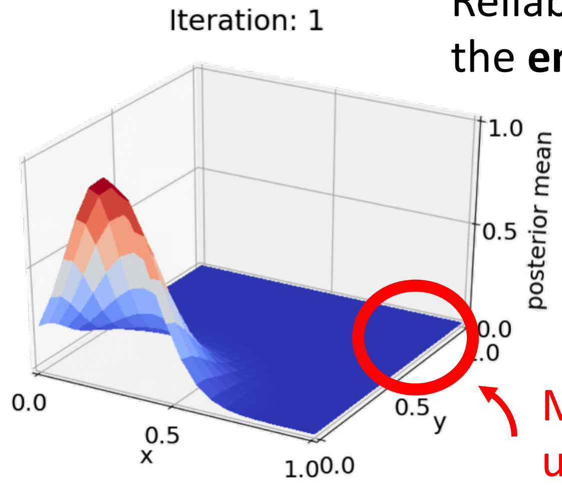
Optimization (CoSpar)



Reliable information **only** around the optimum

Identify optimal gaits for specific users

Active Learning (ROIAL)



Reliable information for the **entire** landscape

More likely to give users "bad" actions

Better understand the science of walking

ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes, Amy Kejun Li et al., ICRA 2021

RoAMS Initiative About Research People Publications Get Involved

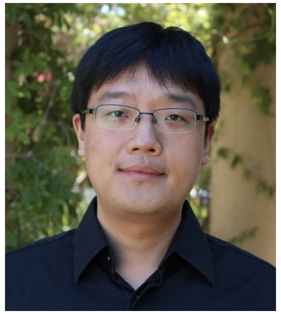
The mission of the Robotic Assisted Mobility Science (RoAMS) initiative is to bring together research frontiers across multiple labs to restore mobility to individuals with a condition that limits or prevents stable locomotion. These research frontiers include mechanical design, machine learning, and formal control algorithm synthesis. This important research has the potential to offer advancements in robotic assistive devices such as prostheses, exoskeleton, exosuits, and spinal cord implants.

Aaron Ames

Joel Burdick

http://roams.caltech.edu/

Thank You!


Yanan Sui

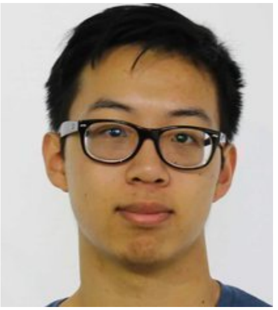

Maegan Tucker


Ellen Novoseller

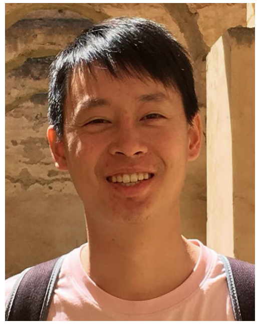

Amy Kejun Li


Erdem Biyik


Myra Cheng


Vincent Zhuang


Claudia Kann


Richard Cheng


Dorsa Sadigh


Joel Burdick


Aaron Ames

Multi-dueling Bandits with Dependent Arms, Yanan Sui et al., UAI 2017
Correlational Dueling Bandits with Application to Clinical Treatment in Large Decision Spaces, Yanan Sui et al., IJCAI 2017
Preference-Based Learning for Exoskeleton Gait Optimization, Maegan Tucker, Ellen Novoseller et al., ICRA 2020 *(Best Paper Award)*
Human Preference-Based Learning for High-dimensional Optimization of Exoskeleton Walking Gaits, Maegan Tucker et al., IROS 2020
ROIAL: Region of Interest Active Learning for Characterizing Exoskeleton Gait Preference Landscapes, Amy Kejun Li et al., ICRA 2021
Stagewise Safe Bayesian Optimization with Gaussian Processes, Yanan Sui et al., ICML 2018