

AI+HCI

MICHAEL BERNSTEIN
STANFORD UNIVERSITY



Announcements

- Project fair on Thursday
- Shorter lecture and discussion to earn us back enough time for the project fair
- Jeff Hancock joining next Tuesday while Michael is at CSCW

Course Overview

INTRO	week 1	Intro to Interaction; Intro to Social Computing
	week 2	Intro to Design; Interaction
DEPTH	week 3	Interaction; Social Computing
	week 4	Social Computing
	week 5	Design
BREADTH	week 6	AI+HCI; Media
	week 7	Foundations
	week 8	Access; Programming
	week 9	Collaboration; Visualization
	week 10	Education; Critiques of HCI

People: where the AI hits the road



MIT Personal Robotics Group

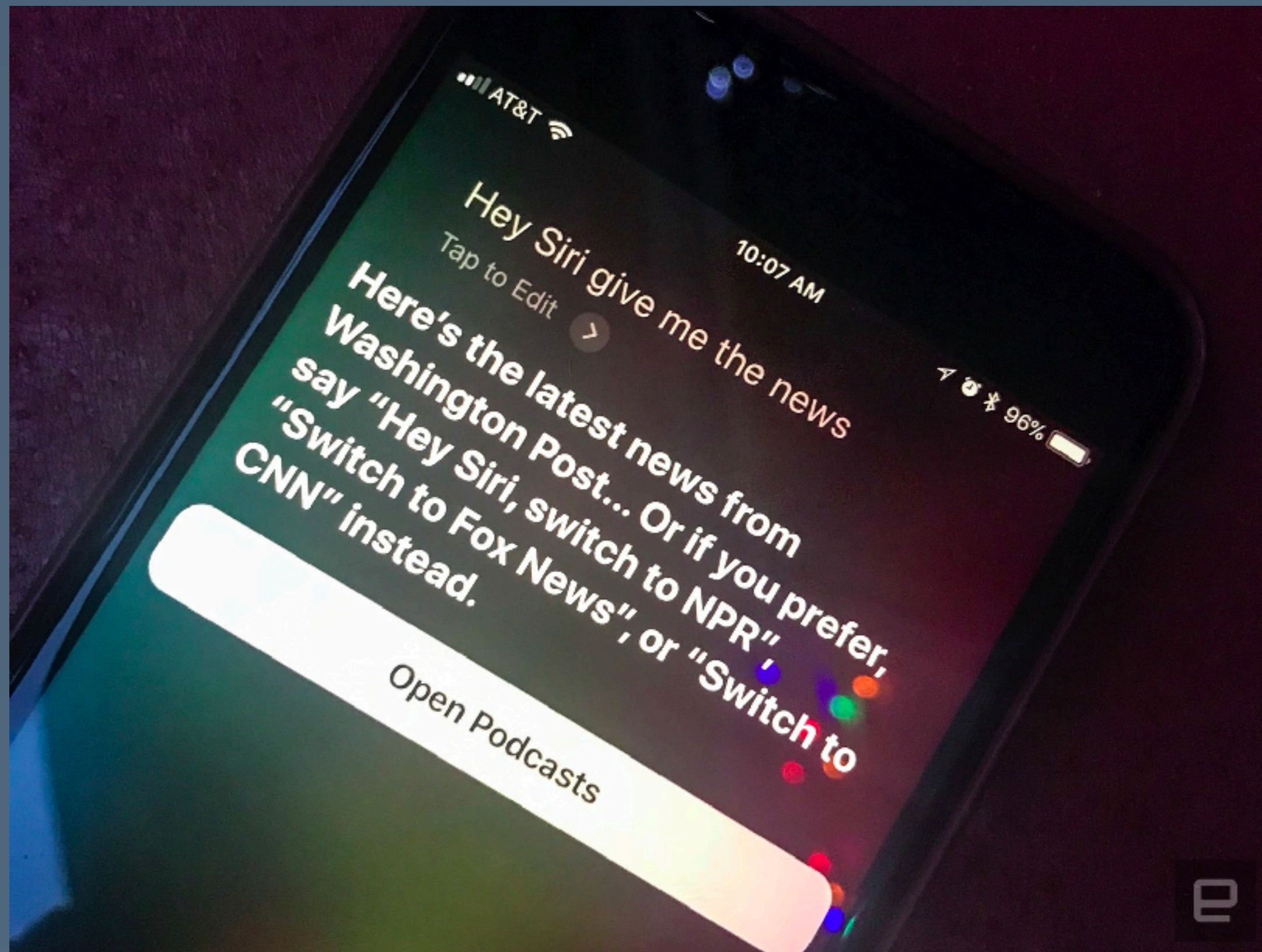


UC Berkeley InterACT laboratory

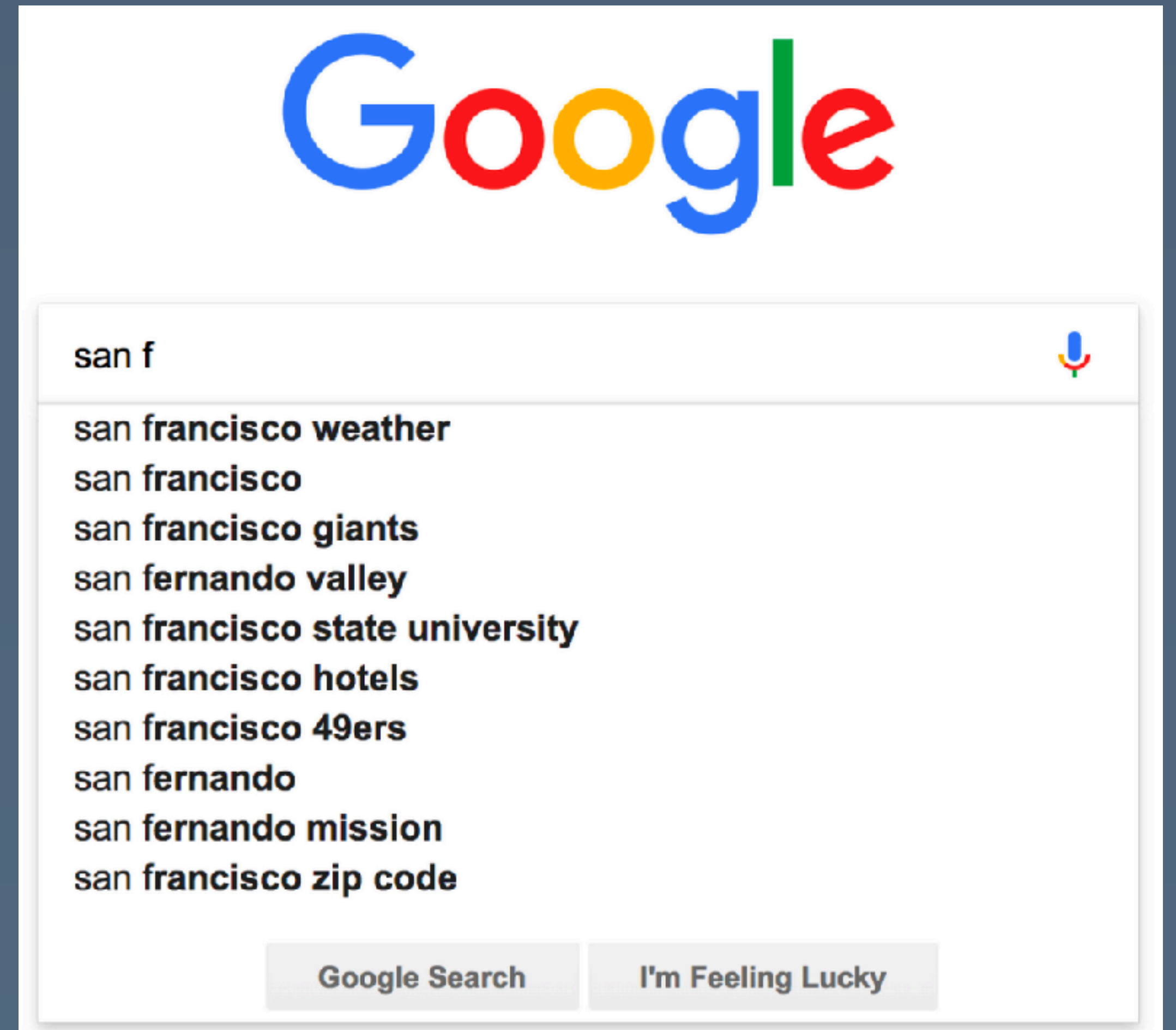


Interactive AI on live TV

Interactive AI in everyday use



Siri, image from Engadget



Google Autocomplete

Automation is not the answer



Intelligence Augmentation

A reaction to:

“AI will take over
human jobs”

INTRODUCTION

OVERALL ABOUT PROGRAM
FILE AS AN *INSTRUMENTA*
CONTROL TECHNIQUES
FILE IMPLEMENTATION
USAGE
ACTIVITIES
CREDITS



AUGMENTING HUMAN INTELLECT: A CONCEPTUAL FRAMEWORK

Prepared for:

DIRECTOR OF INFORMATION SCIENCES
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
WASHINGTON 25, D.C.

CONTRACT AF 49(638)-1024

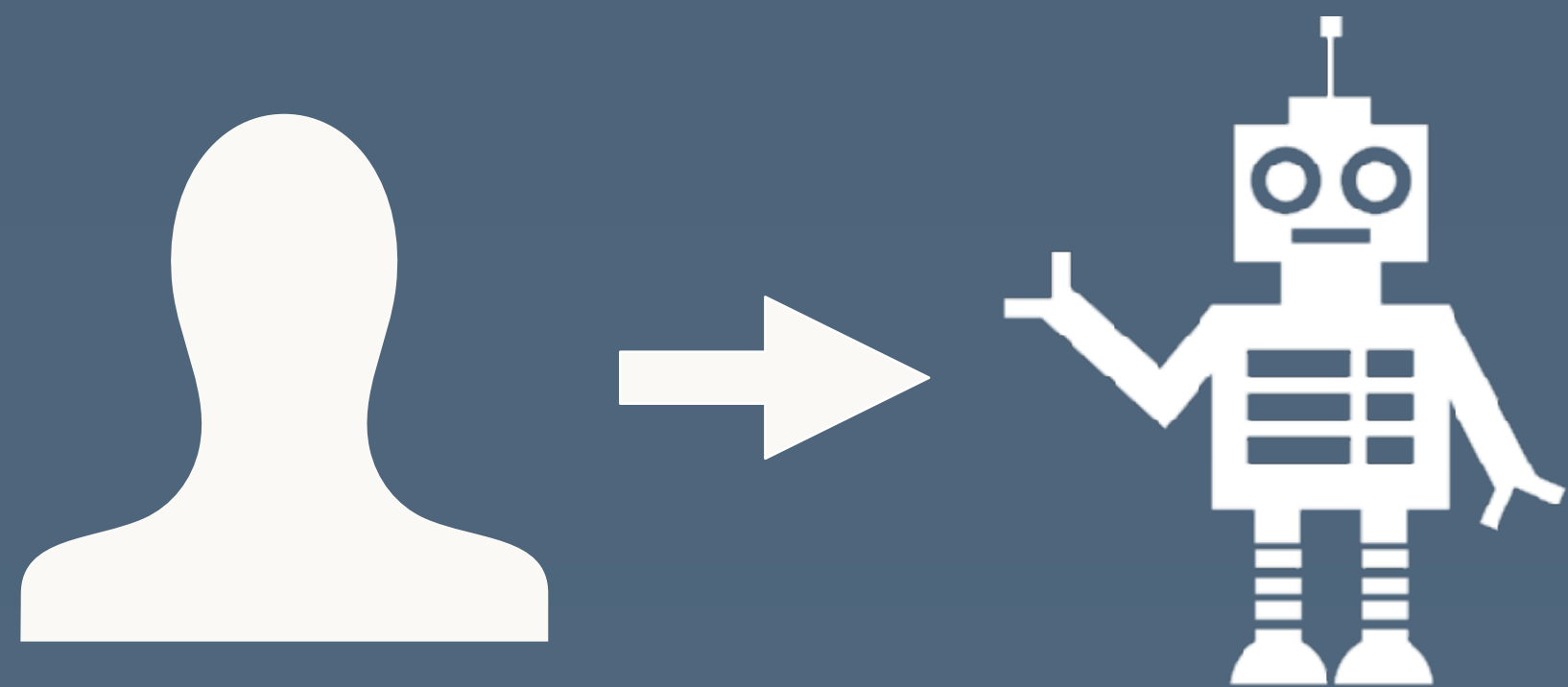
By: D. C. Engelbart

STANFORD RESEARCH INSTITUTE

MENLO PARK, CALIFORNIA

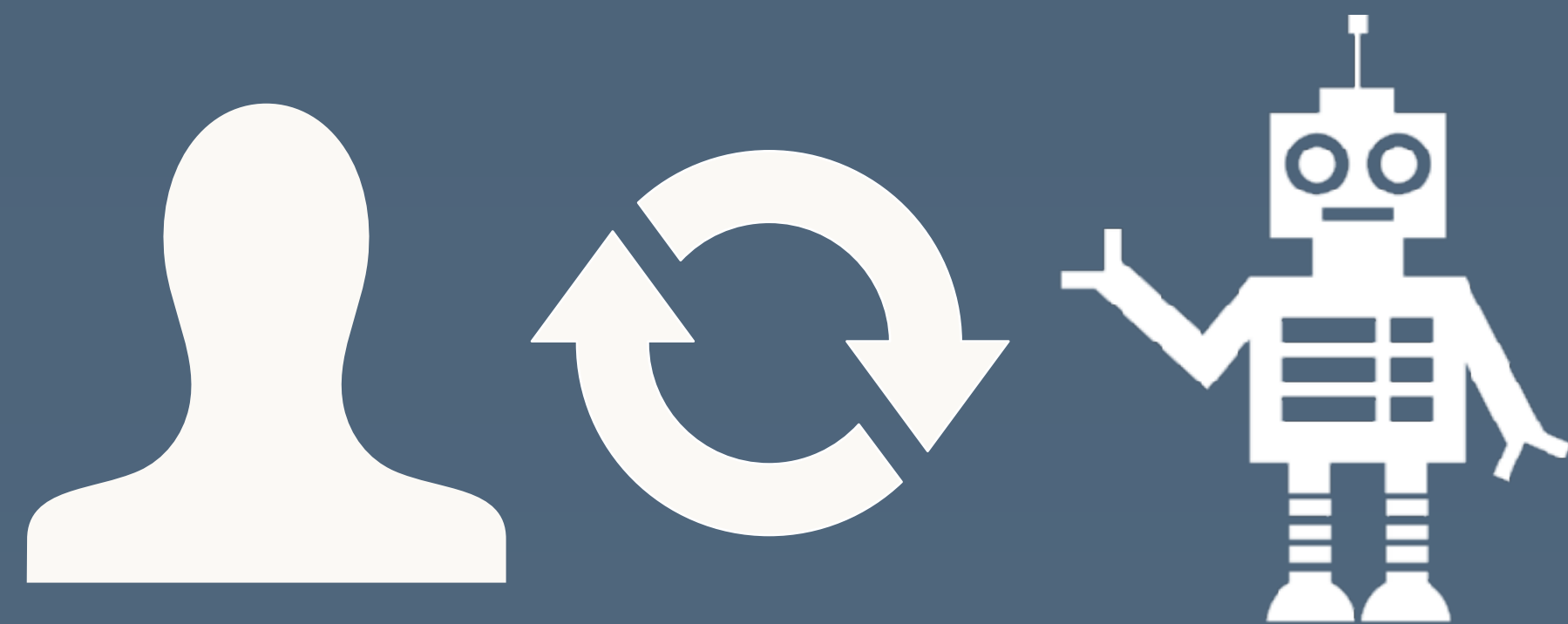


Artificial Intelligence



Replace human intelligence
with artificial intelligence

Intelligence Augmentation



Augment human intelligence
with artificial intelligence

Modeling Uncertainty and Error

Software agents

- Delegate to proactive software and artificial intelligence

Pattie Maes, MIT Media Lab



Direct manipulation

- Users should always have full control

Ben Shneiderman, U. Maryland



A problem has been detected and Windows has been shut down to prevent damage to your computer.

The problem seems to be caused by the following file: kbdhid.sys

MANUALLY_INITIATED_CRASH

If this is the first time you've seen this stop error screen, restart your computer. If this screen appears again, follow these steps:

Check to make sure any new hardware or software is properly installed. If this is a new installation, ask your hardware or software manufacturer for any Windows updates you might need.

If problems continue, disable or remove any newly installed hardware or software. Disable BIOS memory options such as caching or shadowing. If you need to use safe mode to remove or disable components, restart your computer, press F8 to select Advanced Startup Options, and then select Safe Mode.

Technical Information:

*** STOP: 0x000000e2 (0x00000000, 0x00000000, 0x00000000, 0x00000000)

The problem

- Unlike traditional interfaces, introducing an AI into a system creates an element of uncertainty
- Will it understand you correctly? Will it make the correct inferences?
- How do you design a system that can be robust to these kinds of errors?

Mixed-initiative interaction

Mixed-initiative: combine the best of both worlds

[Horvitz CHI '99]

- Utility-based calculation:
 - $u(A,G)$ = (positive) utility of taking an automated action when the goal is correctly guessed
 - $u(A,\neg G)$ = (negative) utility of taking the same action when the goal is incorrectly guessed
 - $u(\neg A,G)$ and $u(\neg A,\neg G)$ similarly

	Desired goal	Not desired goal
Take action	$u(A,G)$	$u(A,\neg G)$
No action	$u(\neg A,G)$	$u(\neg A,\neg G)$

Mixed-initiative: utility calculation

[Horvitz CHI '99]

- What's the expected value of taking action?

$$\cdot (P(G) \cdot u(A, G)) + (P(\neg G) \cdot u(A, \neg G))$$

- What's the expected value of taking no action?

$$\cdot (P(G) \cdot u(\neg A, G)) + (P(\neg G) \cdot u(\neg A, \neg G))$$

	Desired goal	Not desired goal
Take action	$u(A, G)$	$u(A, \neg G)$
No action	$u(\neg A, G)$	$u(\neg A, \neg G)$

Mixed initiative: visually

$$u(\neg A, \neg G)$$

$$u(A, G)$$

Expected
value

If it's
unlikely that
the
user has the
given goal

If it's likely
that the
user has the
given goal

$$u(A, \neg G)$$

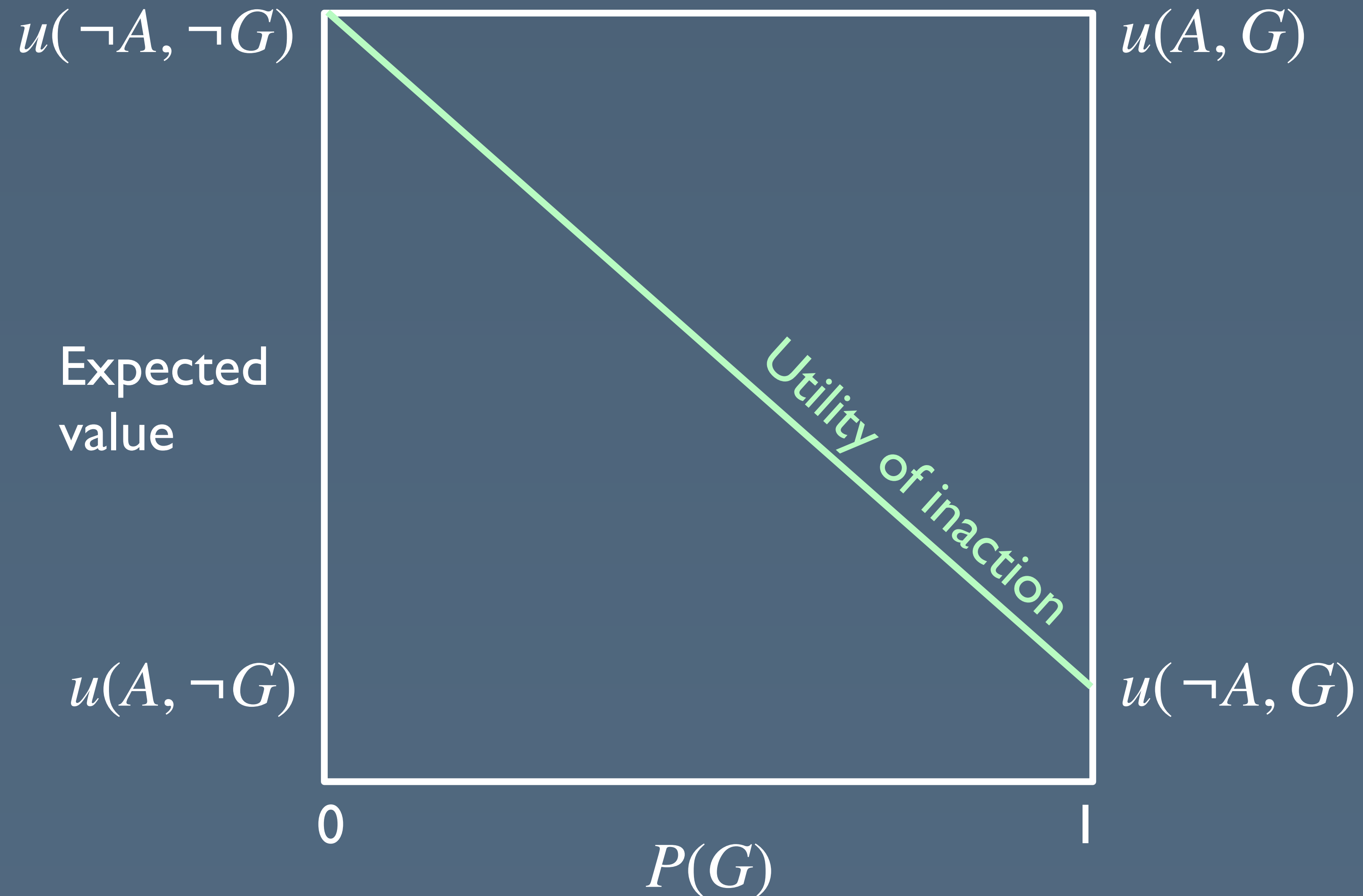
$$u(\neg A, G)$$

0

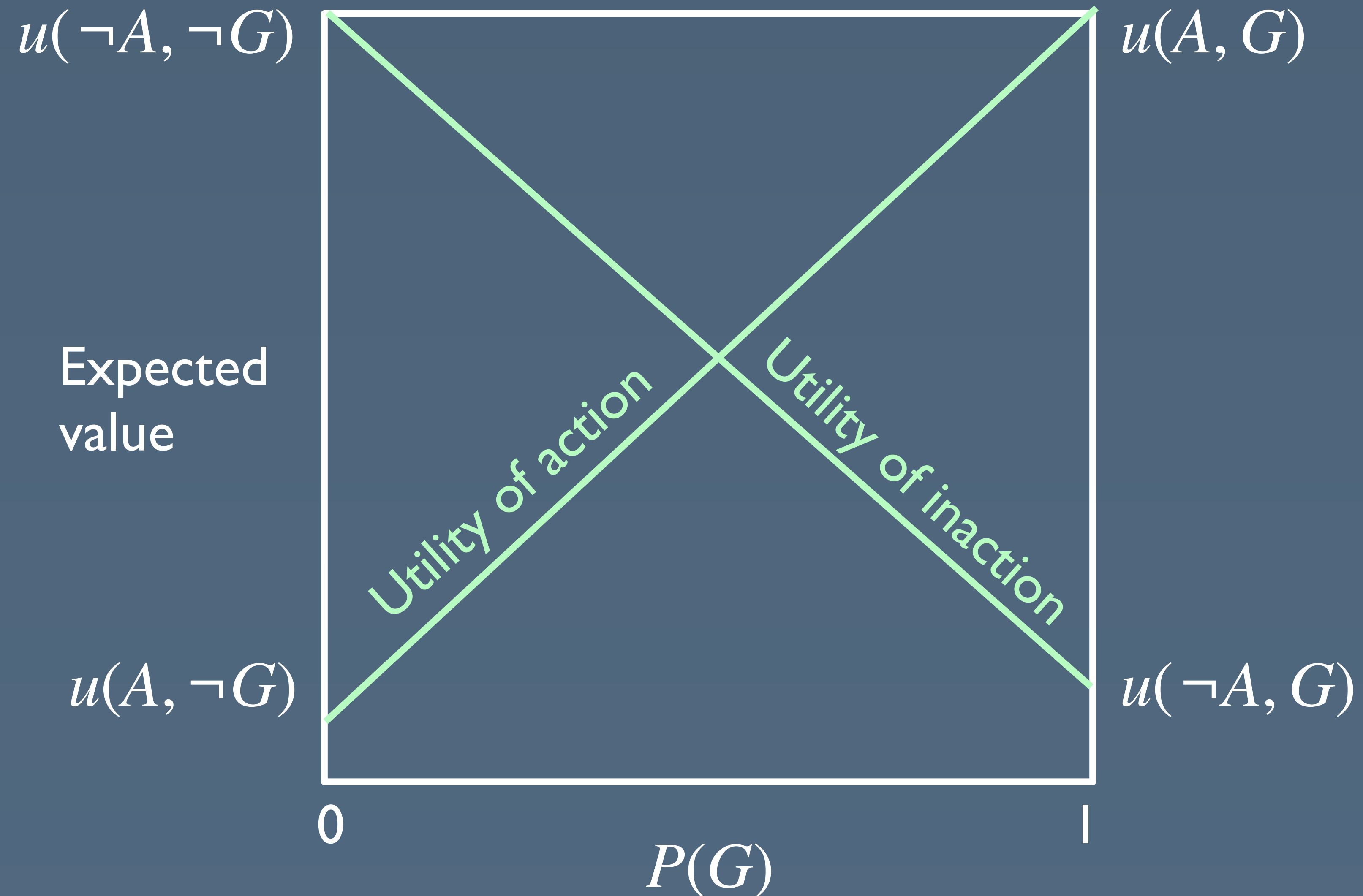
$P(G)$

1

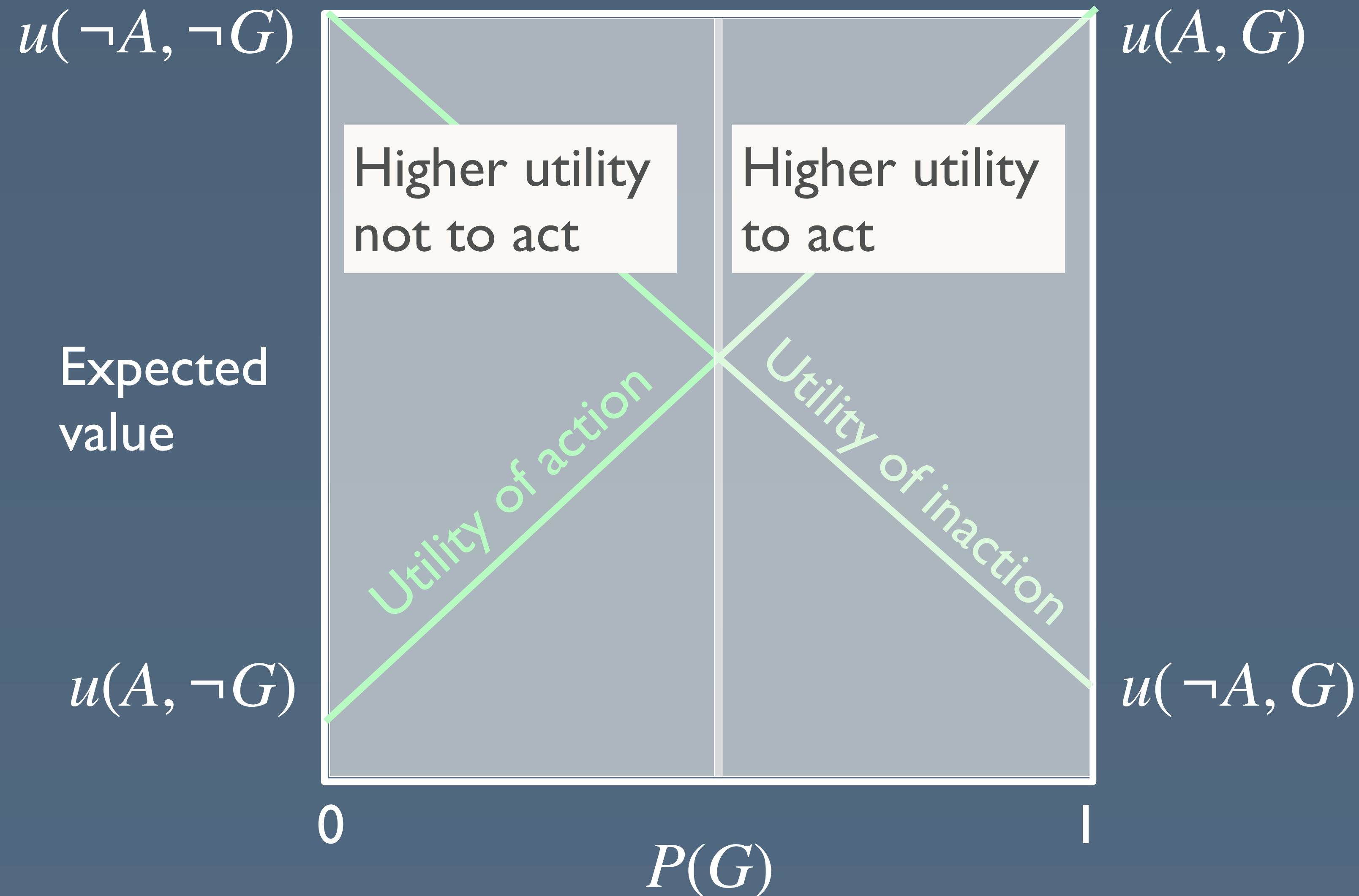
Mixed initiative: visually



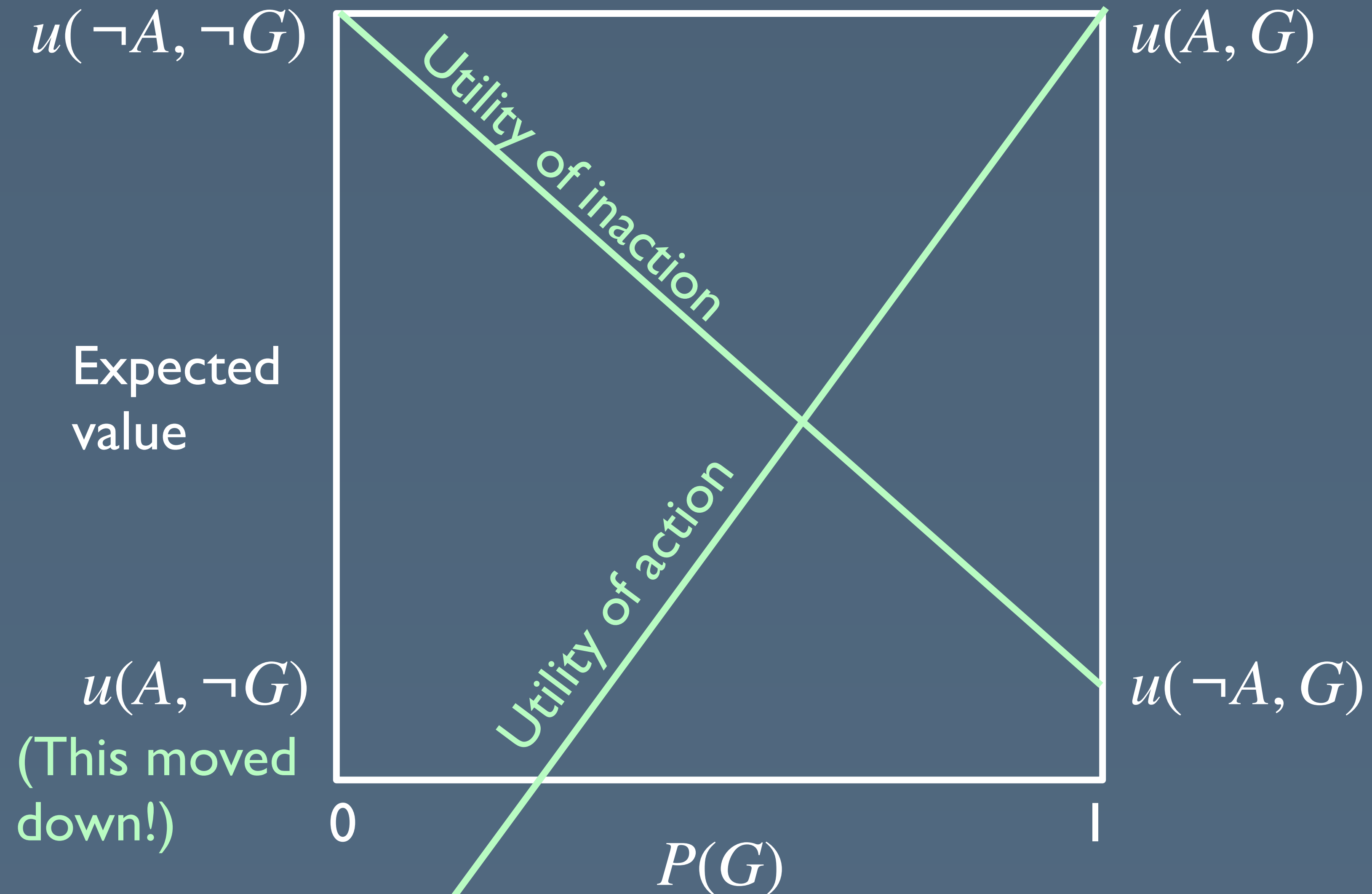
Mixed initiative: visually



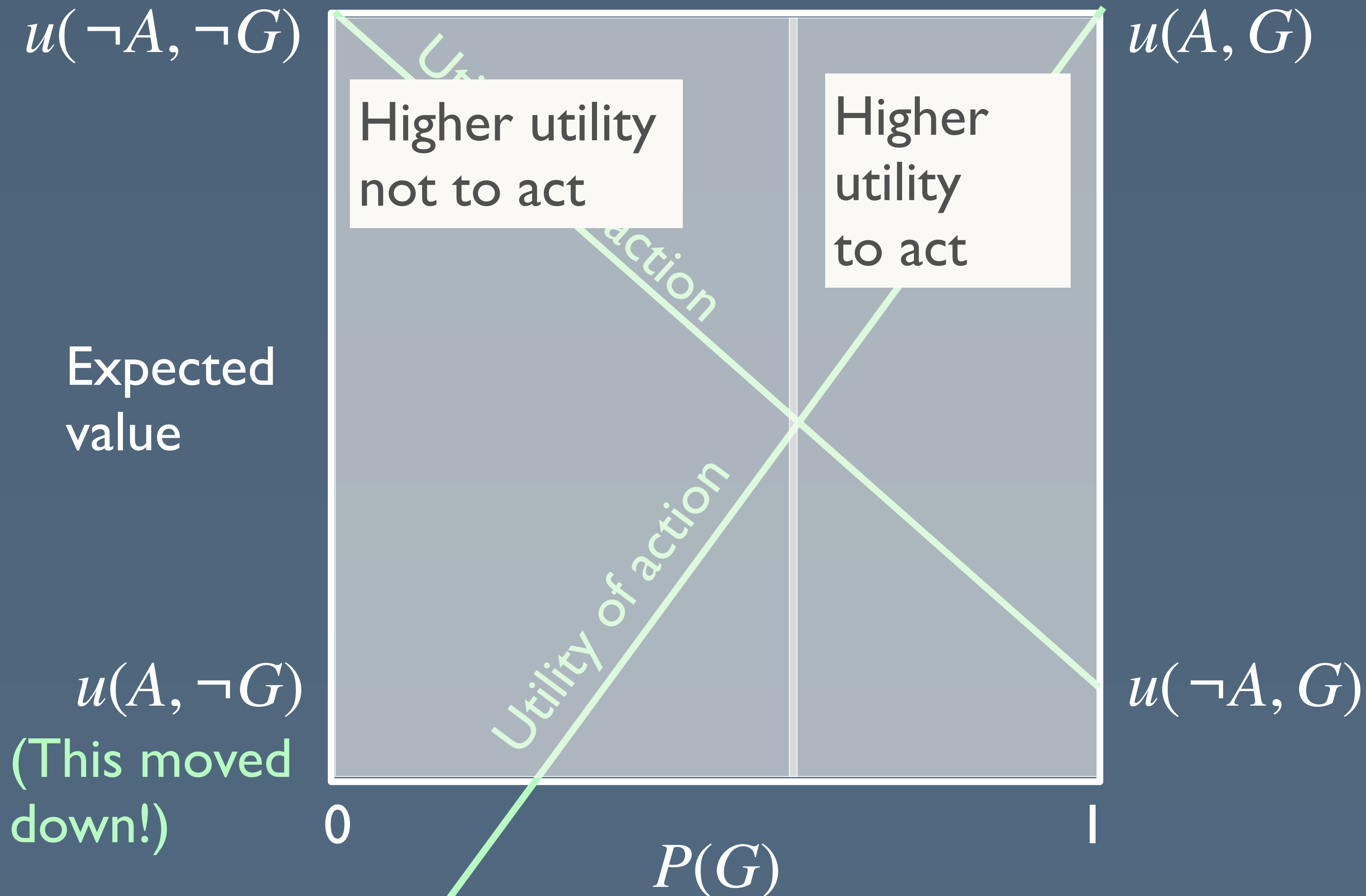
Mixed initiative: visually



What if making an error is costly?



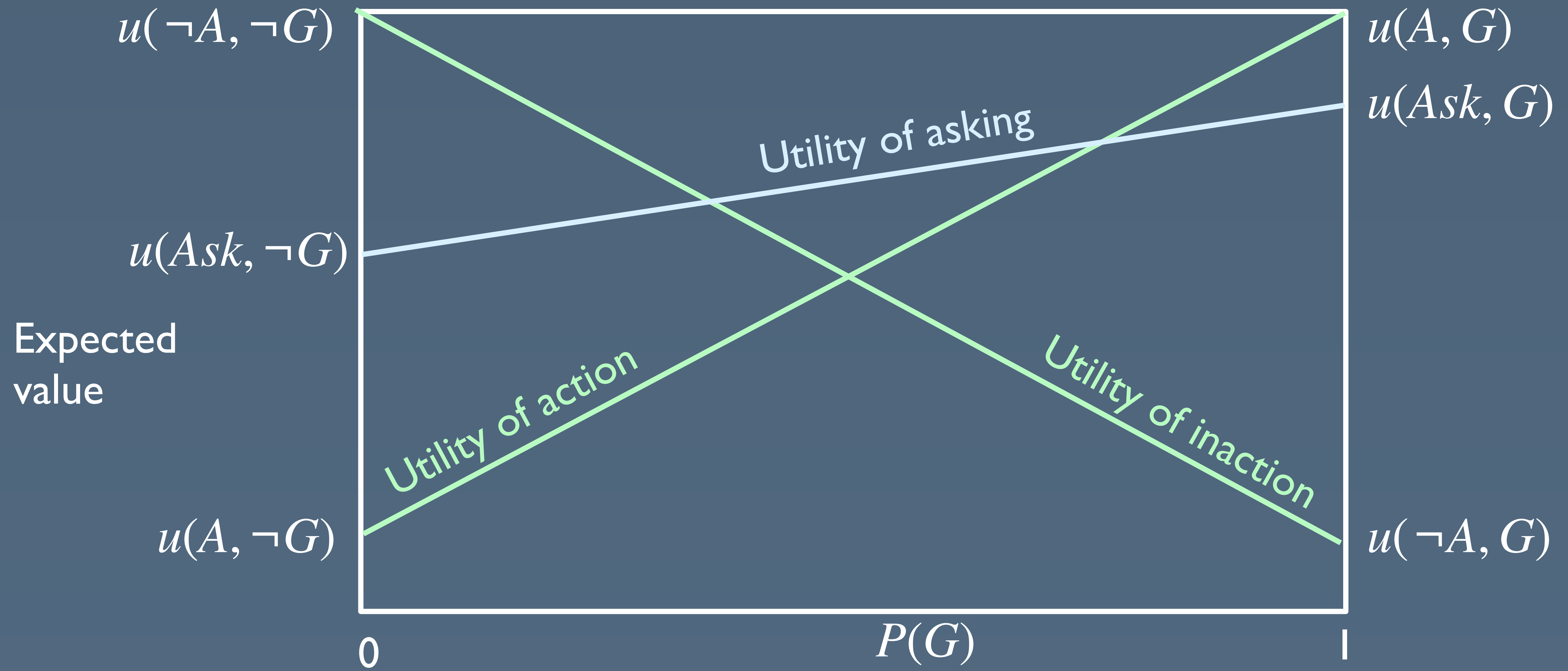
What if making an error is costly?



Now we only take action if we are even more certain that we correctly estimated the user's goal

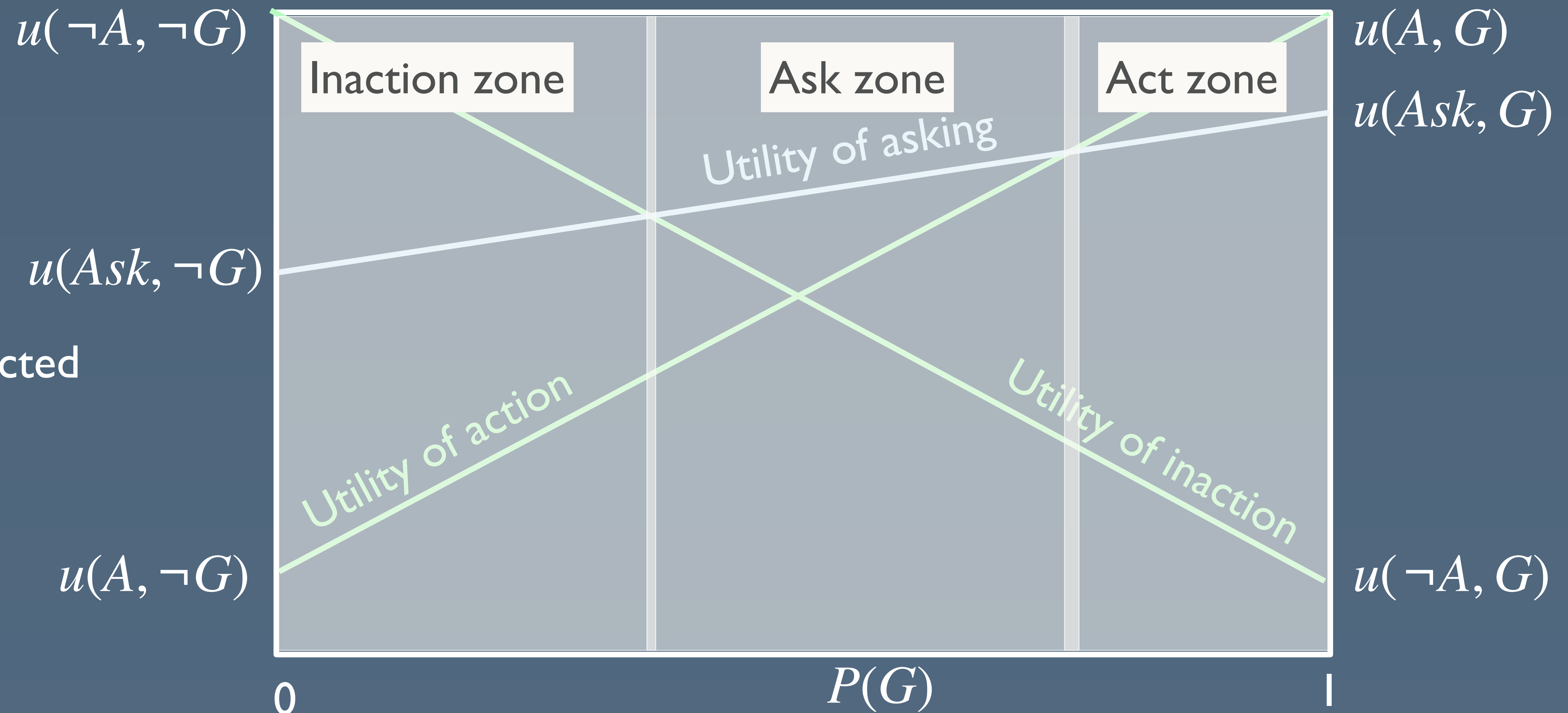
What if we ask the user?

Asking often carries lower risk, but also lower utility



What if we ask the user?

Asking often carries lower risk, but also lower utility

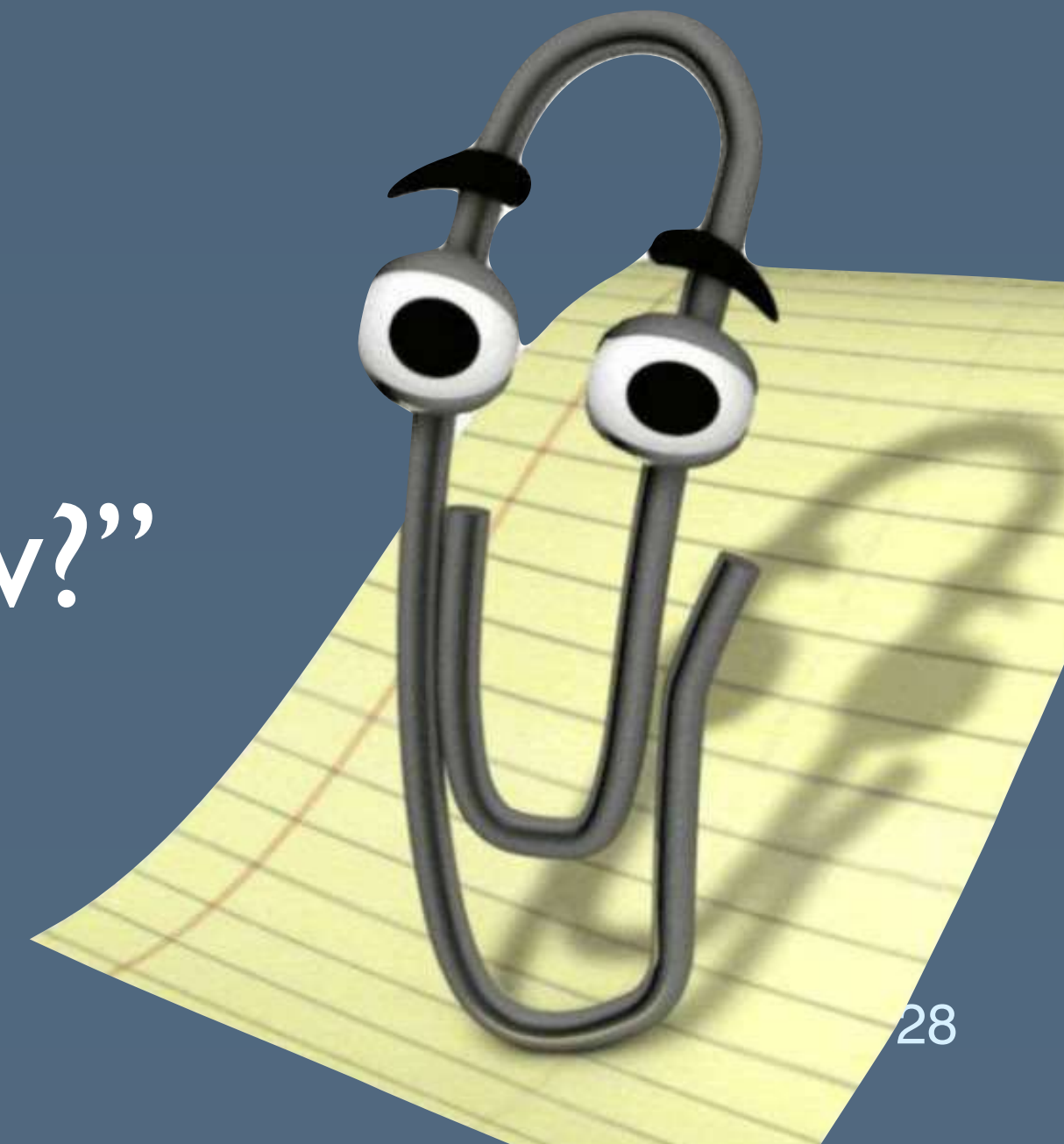


So, when does this screw up?

- When the system cannot accurately assess the probability of the user having the goal $P(G)$

or

- When the utilities are not correctly estimated e.g., too high a utility for asking if the user doesn't have the goal G . "Are you writing a letter right now?"



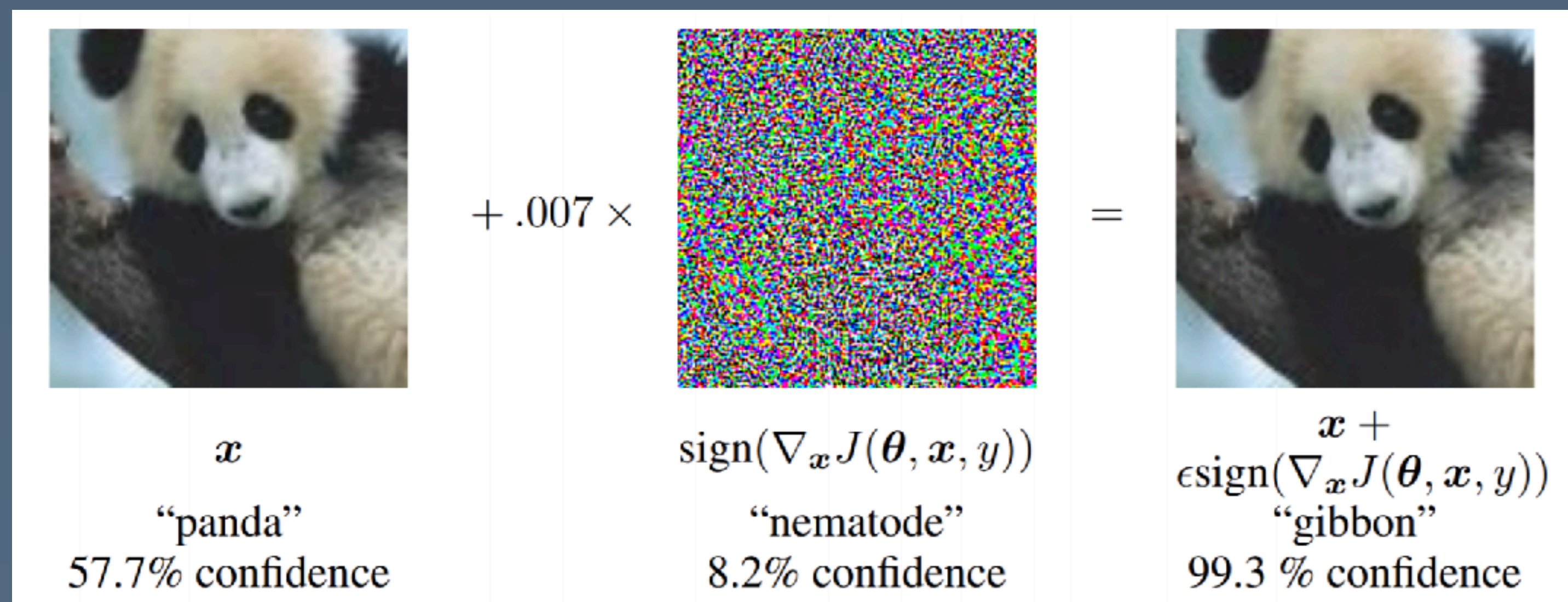
Interactive Machine Learning

Now: diving into the ML models

- AI systems are built on the back of machine learning models
- What lessons can we apply to make these models more powerful and effective?
- Assumed here: a basic knowledge of machine learning

What is your black box learning?

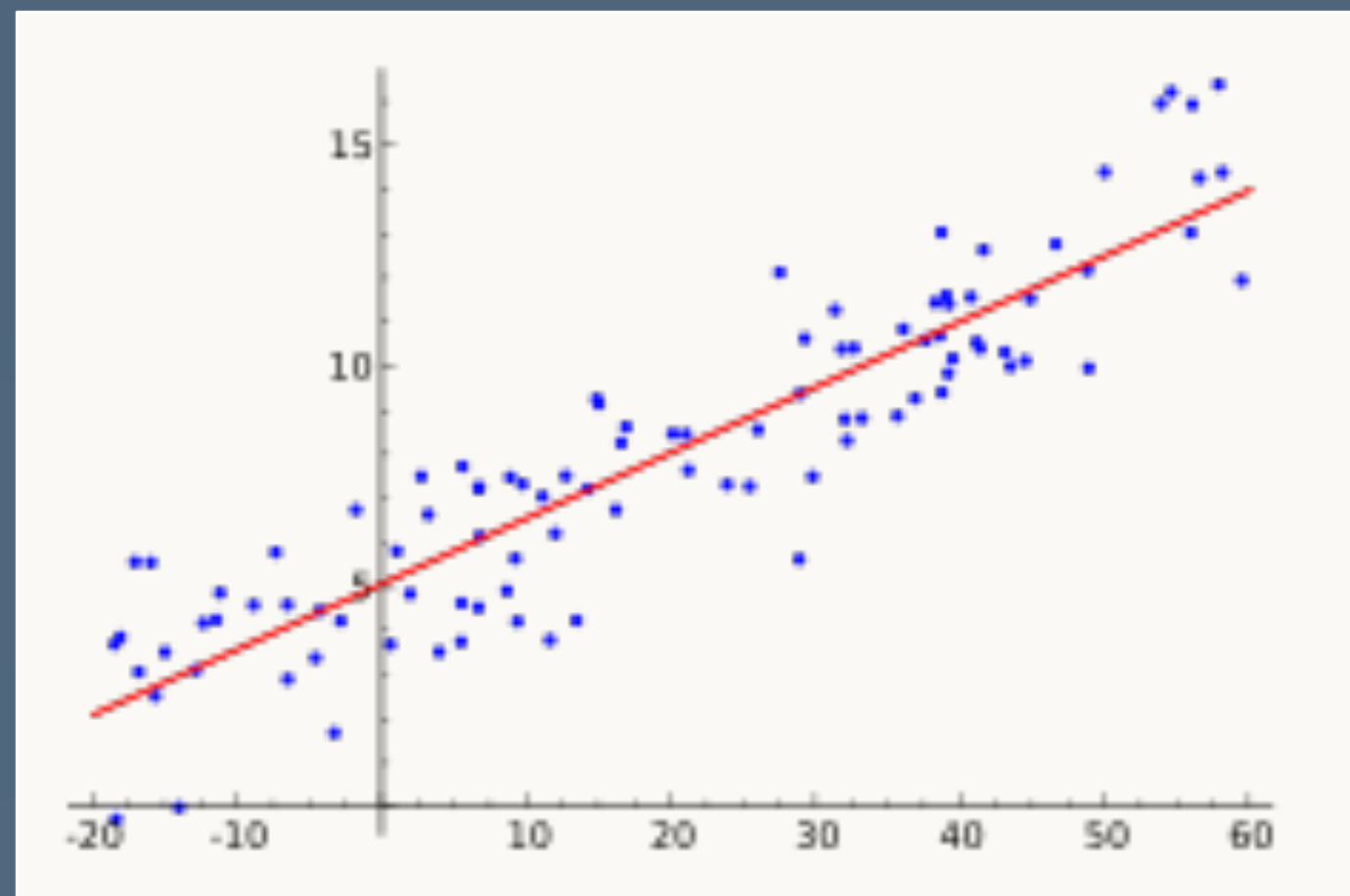
- Machine learning models are powerful, but opaque and unintelligible
- Difficult to predict, design, and debug
- This produces nonintuitive behavior:



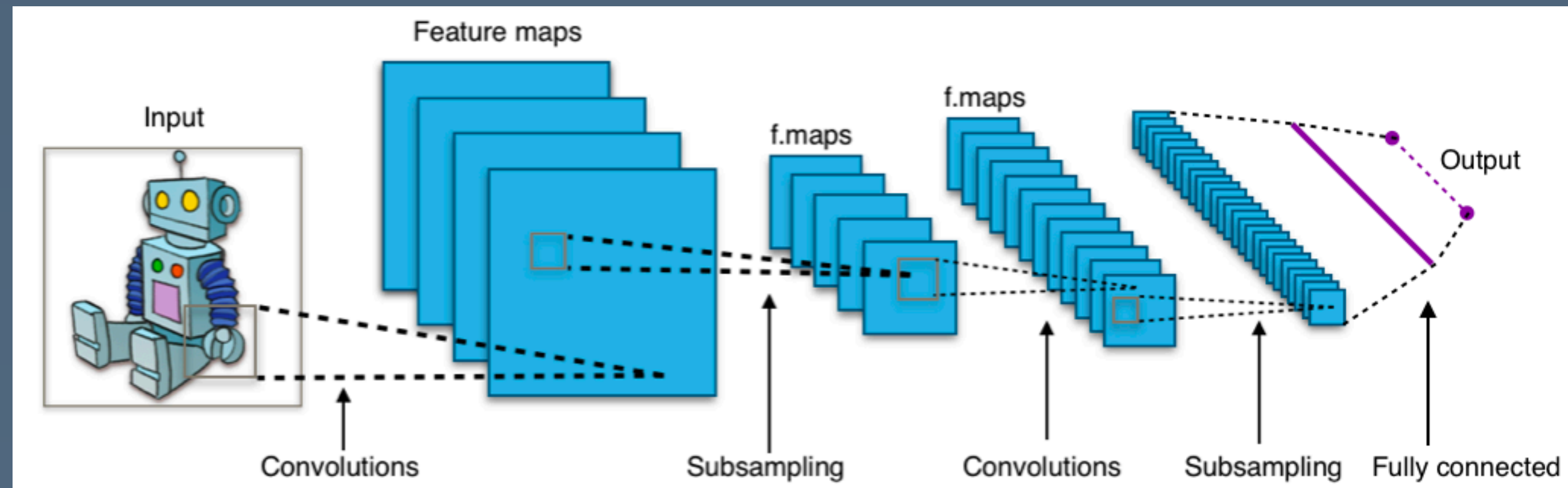
[Goodfellow, Shlens, Szegedy 2014]

Goal: intelligibility in ML models

- A model is *intelligible* to the extent that a human can predict how a change to model's inputs will change its output.
[Weld 2018]



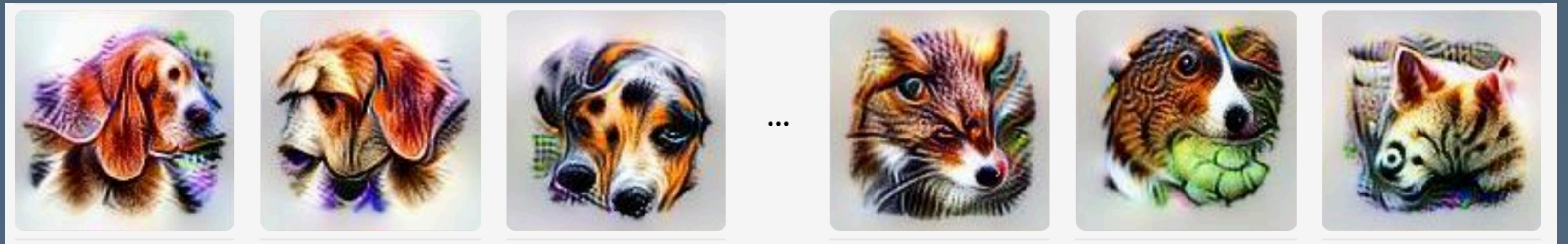
Linear relationship: intelligible
[image from Wikipedia]



Convolutional neural networks: far less intelligible
[image from Wikipedia]

What does the model “see”?

[Olah et al. 2018]



- Above: labrador retriever (left) vs. tiger cat (right)
- If it requires a person to predict its behavior, ML systems that are intelligibility require that people can “see what they see” to a certain extent

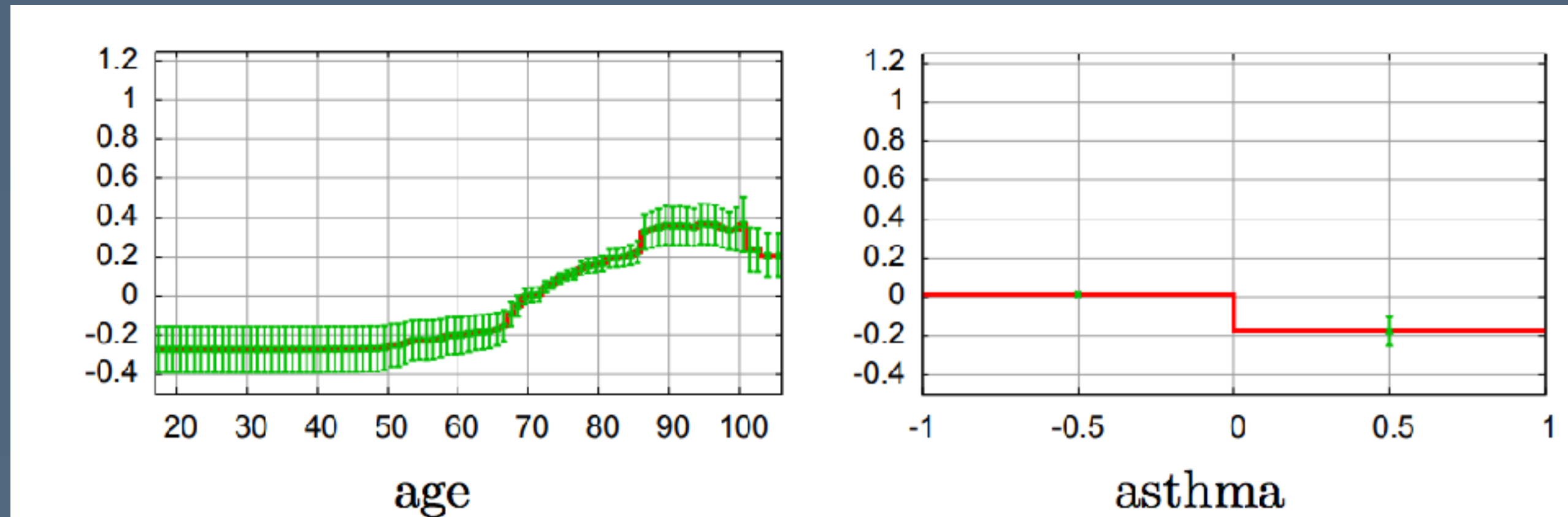
One approach: simplify the model

- Sometimes you can get most of the performance with far higher intelligibility

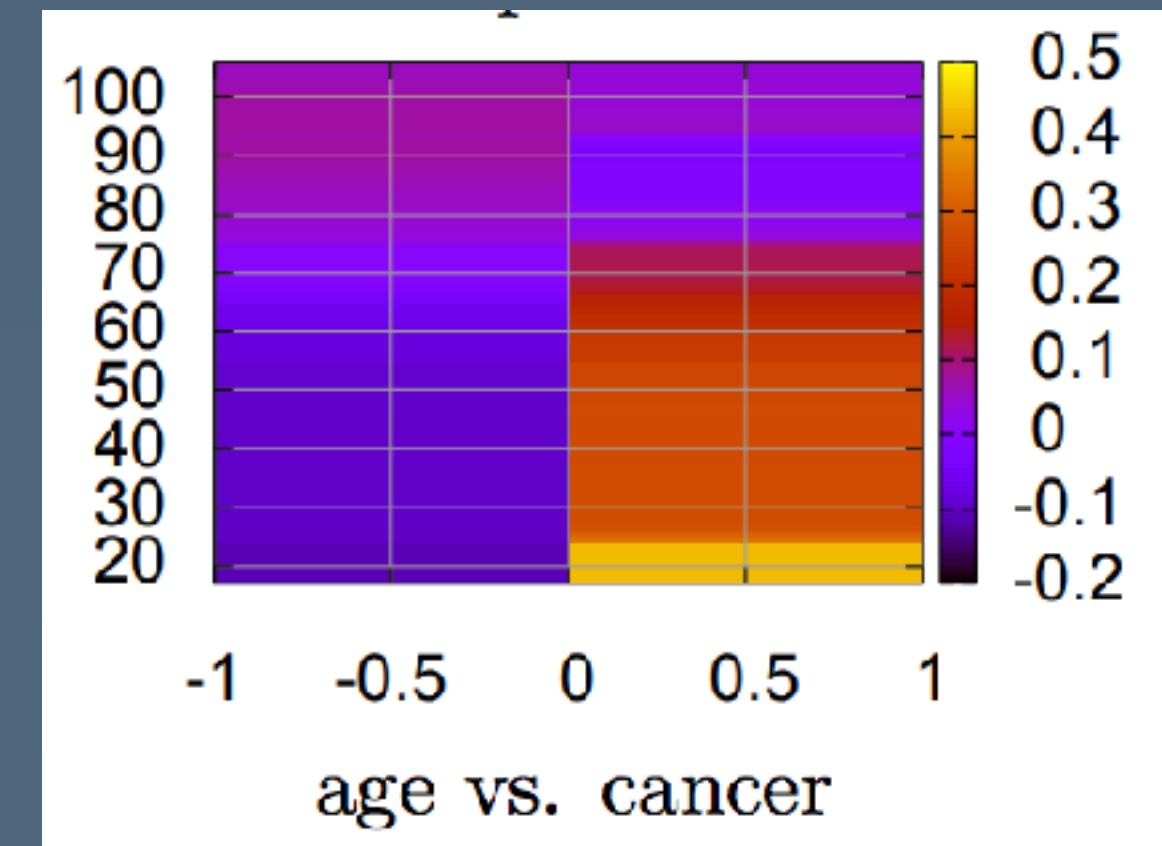
GA²M

[Lou et al. 2013, Caruana et al. 2015]

- A model that learns all features of the form: $\sum f_i(x_i) + \sum f_{ij}(x_i, x_j)$
- In other words: the system learns univariate and bivariate relationships between the input features and the outcome



Univariate features predicting pneumonia risk

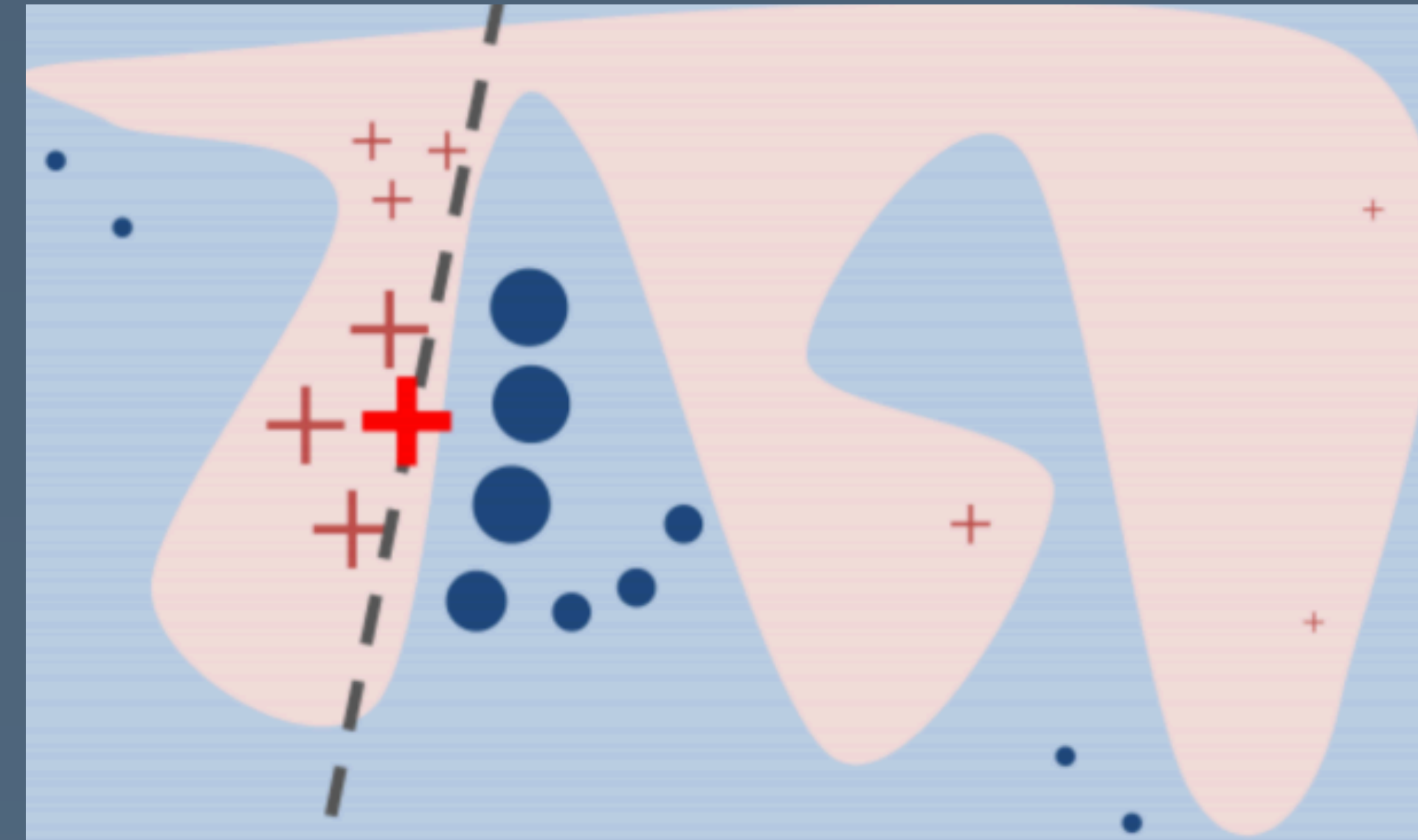


Bivariate interaction predicting pneumonia risk (note particular risk of young people with cancer)

LIME local explainers

[Ribeiro et al. 2016]

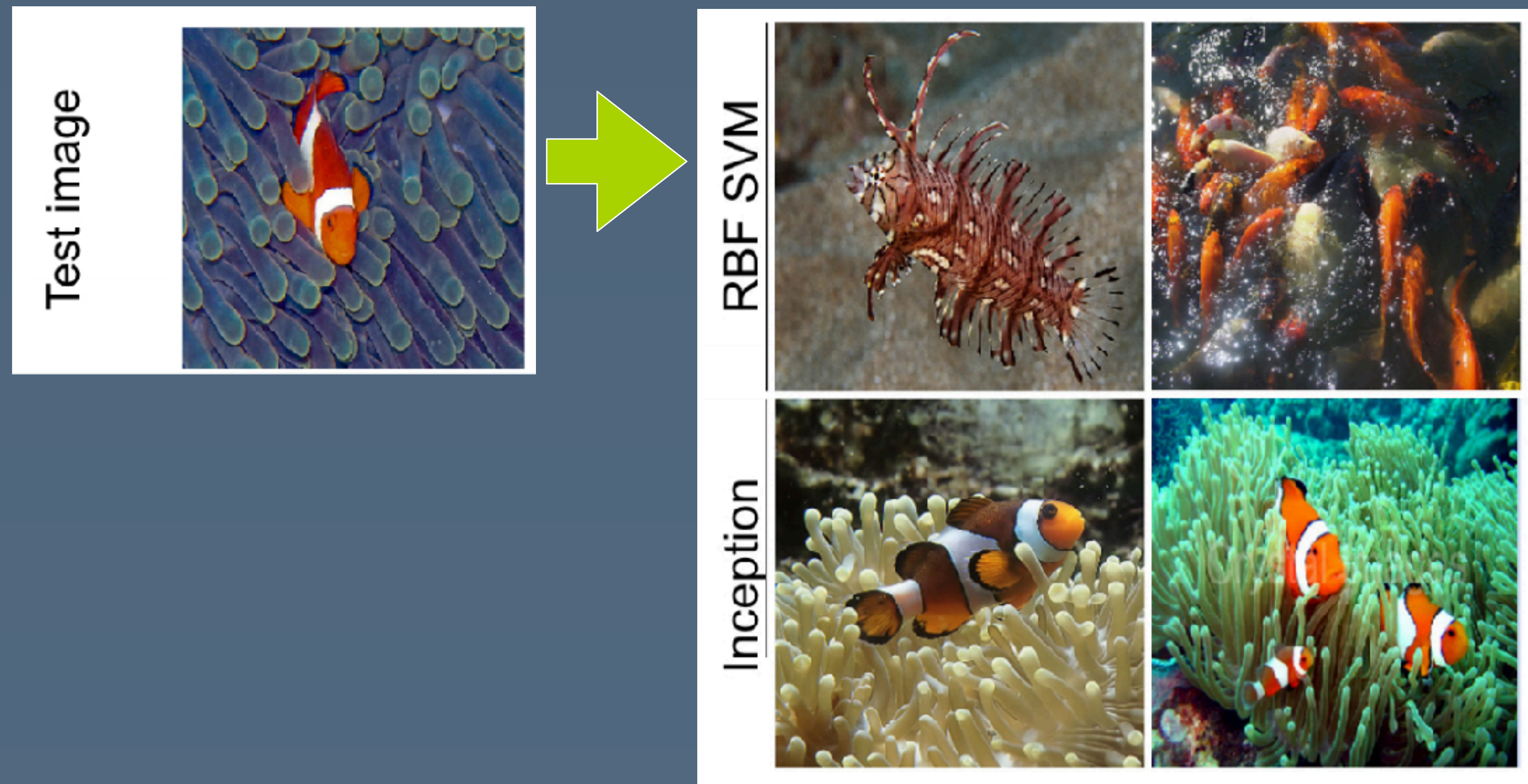
- Often the learned decision boundary (red vs. blue) is nonlinear →
- Suppose you were trying to explain the bright red cross example. What would you do?
- LIME: sample other points nearby (large red crosses and blue circles), weigh them in proportion to their proximity to the red cross, and learn a linear separator for them.
- This is *not* an accurate representation of the whole model! But still useful in local explanation.



Influence functions

[Koh and Liang et al. 2017]

- Mathematical approach that traces a prediction back to the most influential training points that produced the prediction



The dilemma

[Weld and Bansal 2018]

- Any model simplification is a lie!
- But any non-simplification is unintelligible.
- Recommendation: draw on psych research to guide explanation
 - Make explanations contrastive: “Why did you recommend Movie X?” (implicitly: “Why didn’t you recommend Movie Y?”)
 - Necessary causes are better than sufficient ones
 - Use few conjuncts

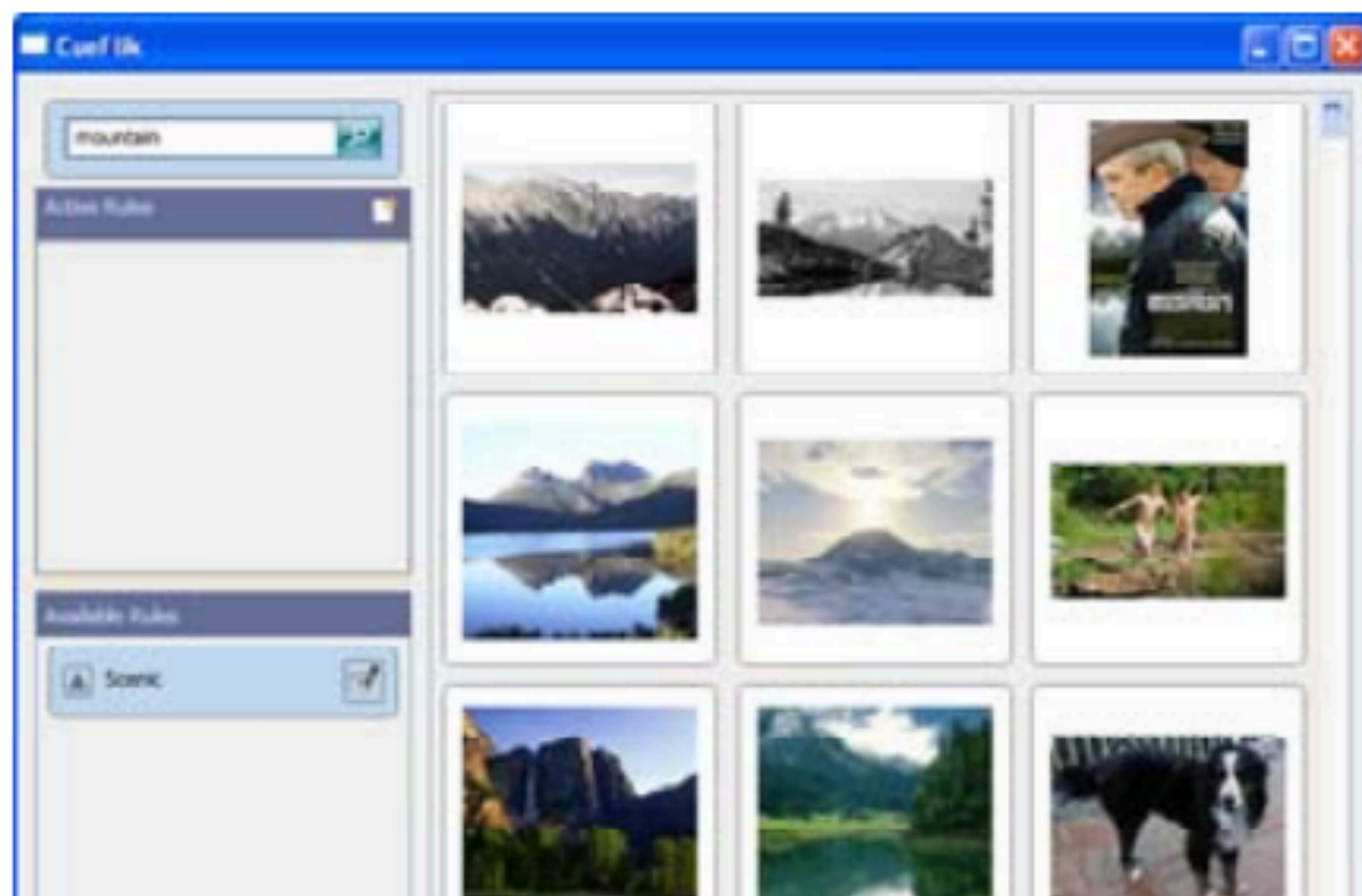
**Guiding users to
train effective
classifiers**

Interactive training

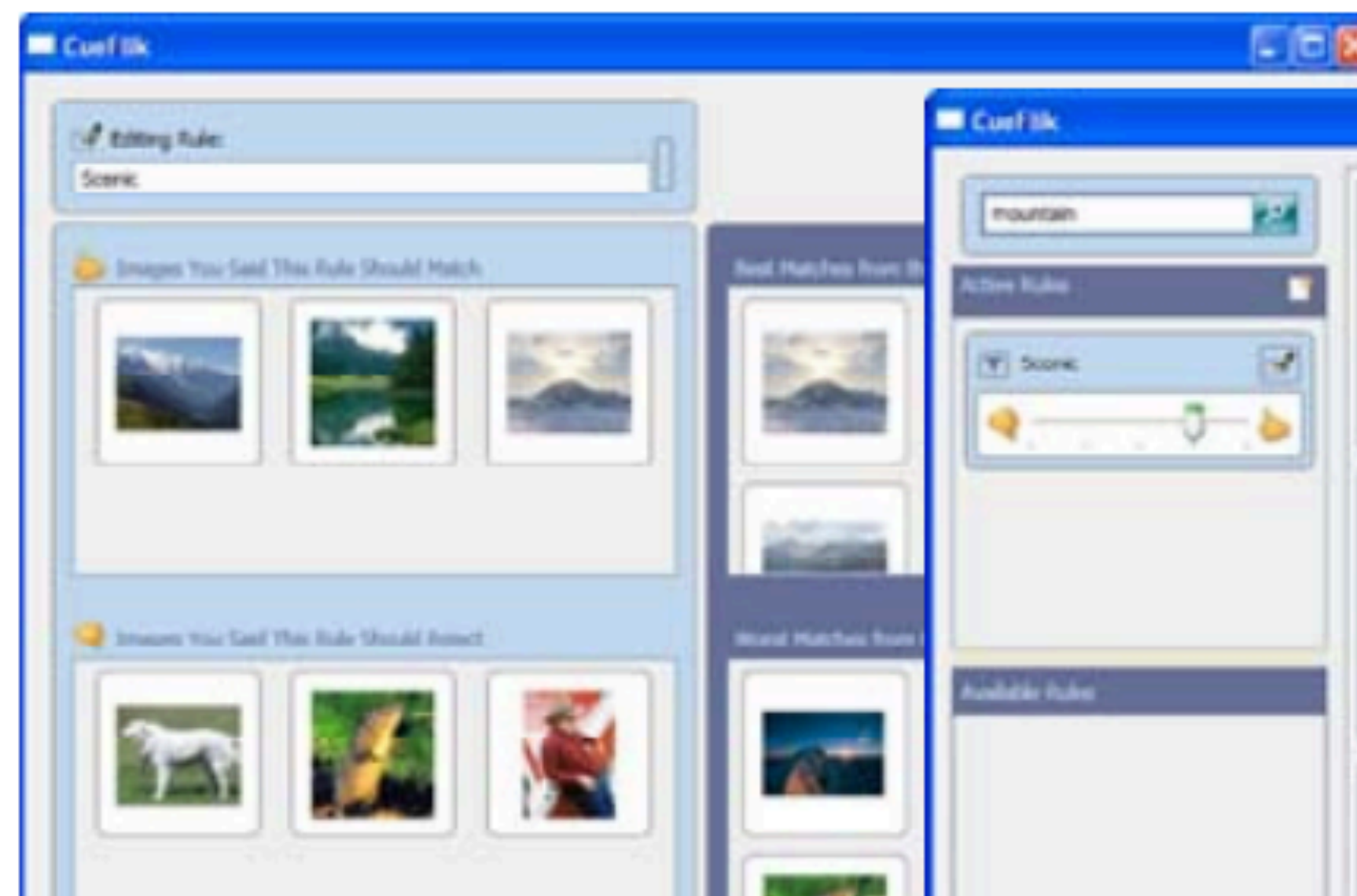
[Fogarty et al. 2008]

- Allow users to keep training and re-training by drag-dropping instances into positive and negative classes as they go

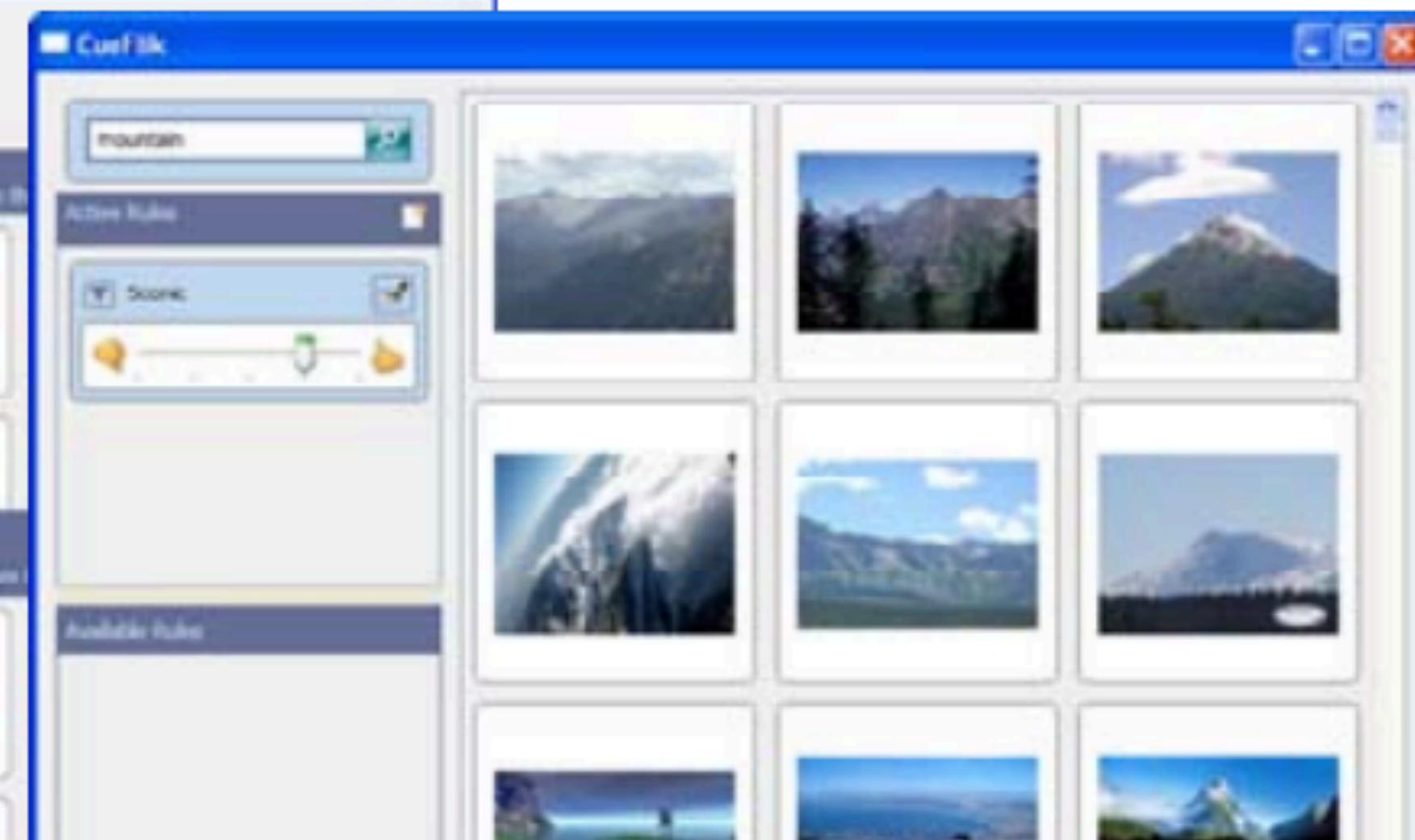
Image Search for
“Mountain”



Creating a “Scenic” Rule



Apply the “Scenic” Rule



Revising your training as you go

[Chang, Amershi and Kamar 2017]

- Facilitate concept evolution through a “could be” category that allows clustering into subcategories you can change labels for

The diagram illustrates the process of revising training through concept evolution. It shows three rows of images (tiger, real cat, cartoon cat) with their initial classification options. A green arrow points to the revised state where the tiger is grouped into 'big cats' and the cartoon cat is grouped into 'cartoon cats'.

Initial State (Left):

- Tiger: Cat, Not Cat, Maybe/NotSure
- Real Cat: Cat, Not Cat, Maybe/NotSure
- Cartoon Cat: Cat, Not Cat, Maybe/NotSure

Revised State (Right):

- Tiger: Create; big cats (selected), cartoon cats, cats with dogs
- Real Cat: Create
- Cartoon Cat: Create; big cats, cartoon cats (selected), cats with dogs

Play-along learning

[Fiebrink, Cook, and Trueman, ICMC '09]

- Create the output (sounds) you desire
- “Play along” and demonstrate the input that should generate that output



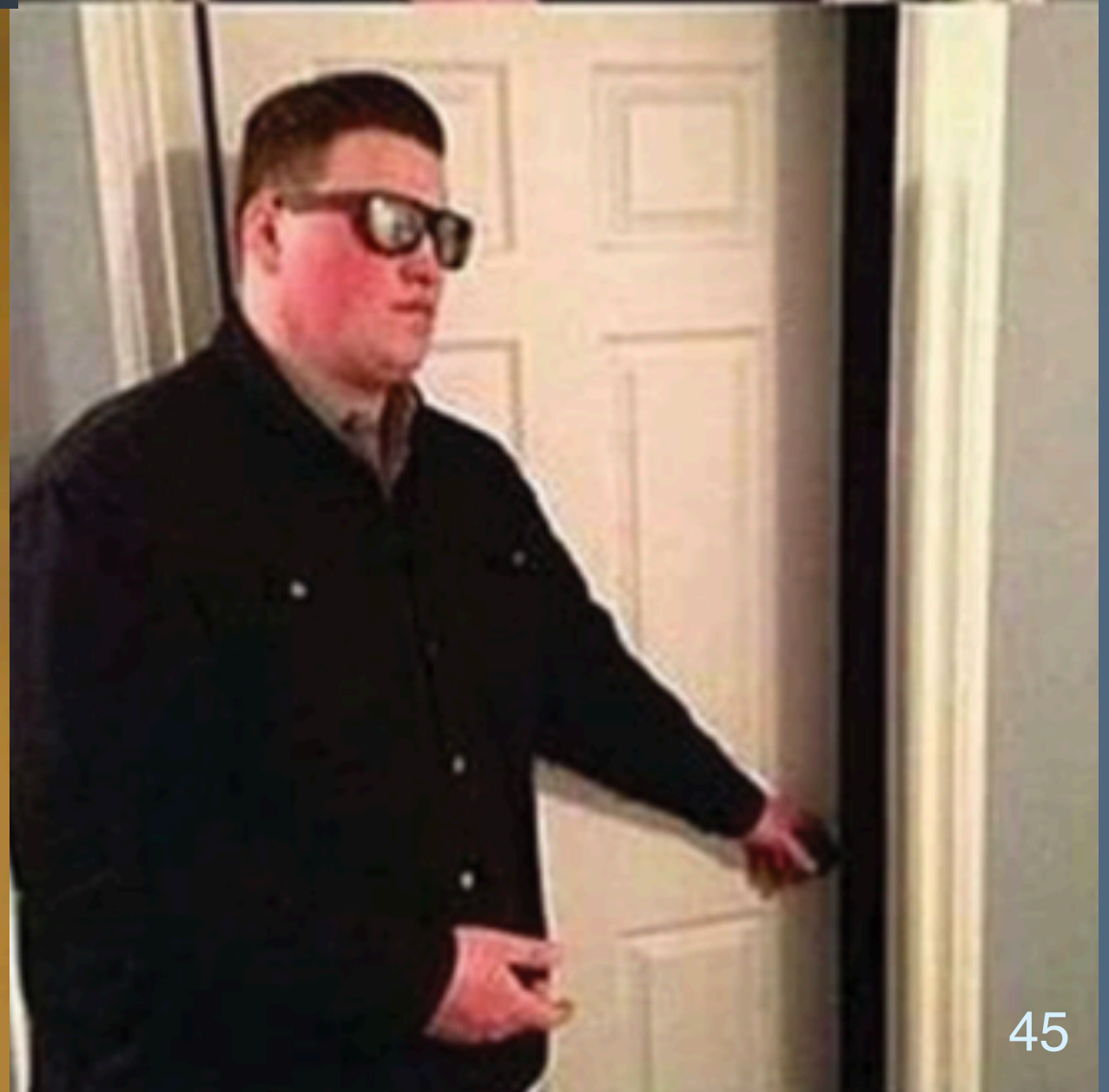
Developing intelligent software

If you wanted a smart doorbell...

- To automatically control entrance to your room
- To let in possible donors for your Stanford education



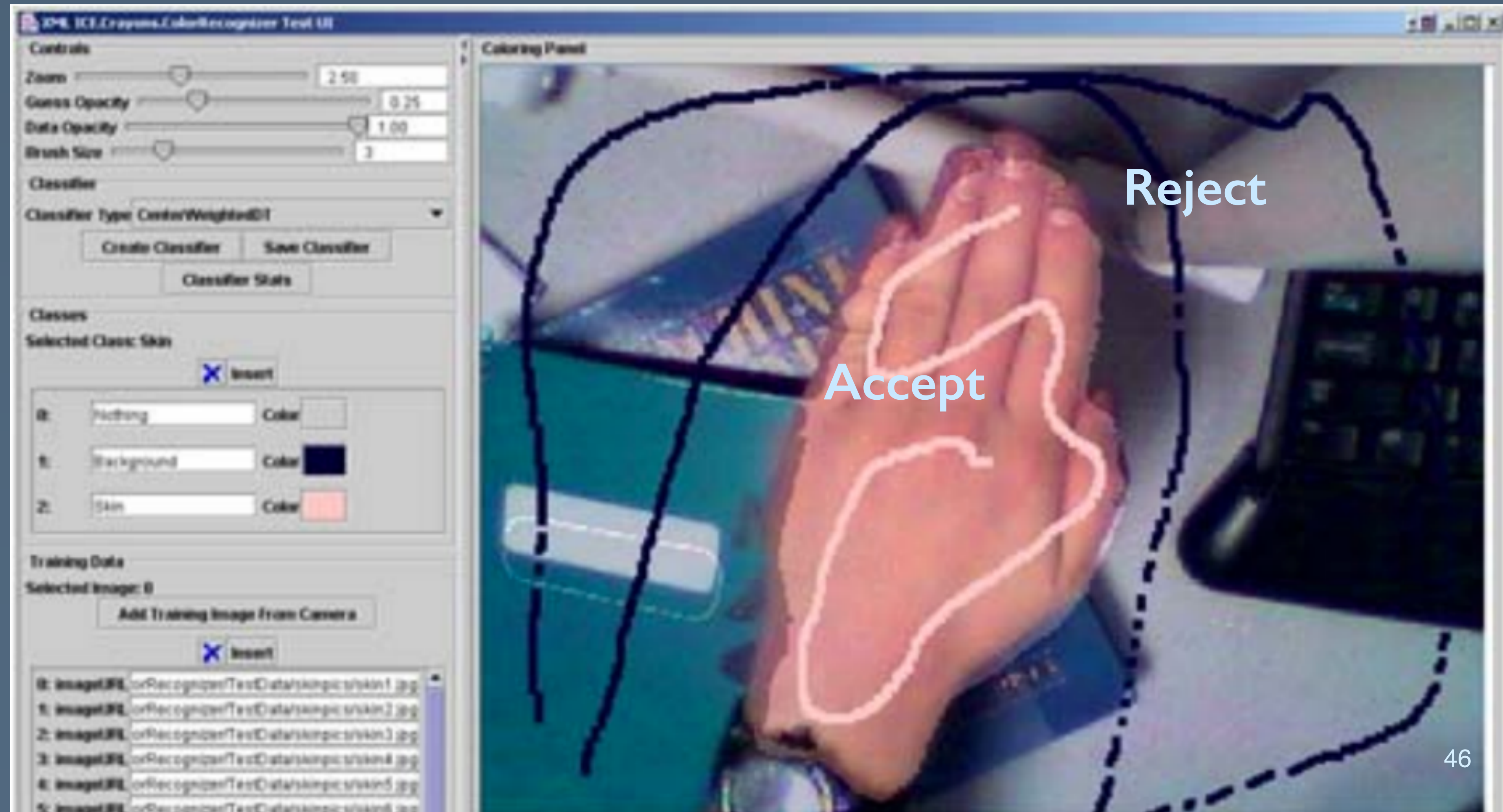
How would
you train the
system
quickly?



Crayons: camera-based interaction

[Fails and Olsen, CHI '03]

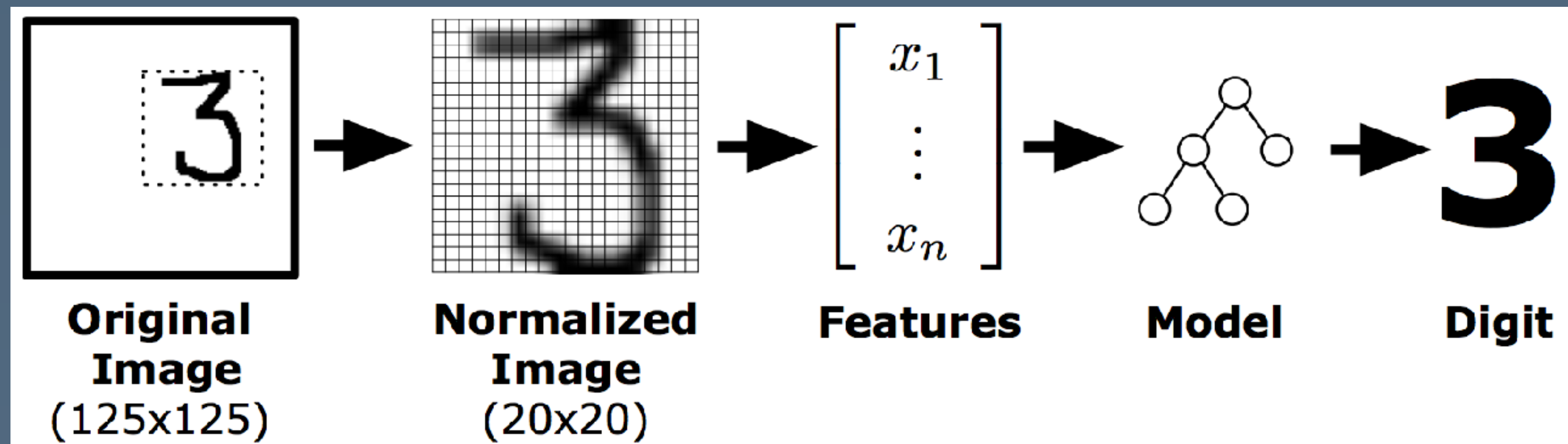
Direct-
manipulation
training



Development challenges with ML

[Patel et al., CHI '08]

- Software development benefits from modularity, but machine learning is iterative and nonlinear
- Difficulty understanding the statistical process underlying machine learning algorithms
- Evaluation of progress is difficult



Papier-Mâché: toolkit support for tangible input

[Klemmer et al., CHI '04]

- Monitoring window, wizard-of-oz input, of-oz input, listeners, designed and evaluated as a user interface

The screenshot displays the Papier-Mâché software interface, which is divided into several panels:





- Current Phobs:** A hierarchical tree view showing the system's state. It includes folders for RFIDReaders, VisionPhobGenerators, and BarcodePhobGenerators. Under VisionPhobGenerators, there is a sub-entry for "VisionPhobGenerator: (Video Camera)" with five colored circular markers representing detected objects. Each marker is associated with a set of bounding box coordinates, size, hue, and saturation values. For example, a red marker has "Bounds[148, 7, 219, 124]; Size[1753]; Hue[196], Sat[94]".
- Vision input (Video Camera):** A video window showing a real-time feed of a wooden surface with a green sticky note, a red pen, a barcode sticker, and a pair of scissors. A green dot on the scissors indicates a detected feature. A timestamp "00:00:46.65" is visible at the bottom of the video frame.
- Classifier Map:** A panel on the right side of the interface. It contains a list of classifiers, including "TagPhob" and "VisionPhobGenerator: (Video Cam)". A "New classifier" button is located at the bottom of this panel.
- Association Map:** A panel at the bottom right showing a list of detected objects with their bounding box coordinates and sizes. For example, "Bounds[22, 48, 63, 127]; Size[335]".
- Vision Processing:** A panel at the bottom center showing "Edge Detection Thresholds" with three sliders. Each slider has a range from 0 to 255. Below the sliders, there are three small preview windows showing the results of the edge detection process.

IDE support for ML development

[Patel et al., UIST '10]

- Explicit support for each step: feature extraction, model generation, training and testing

The screenshot displays an IDE interface for ML development. On the left is a 'Legend' window listing various classes with corresponding colored squares: arrow (330), caret (330), check (330), circle (330), delete_mark (330), left_curly_brace (330), left_sq_bracket (330), pigtail (330), question_mark (330), rectangle (330), right_curly_brace (330), right_sq_bracket (330), star (330), triangle (330), v (330), and x (330). The main area shows 'Classification Results' in 'Table View' mode. The table has columns: Raw Object, Class, RowNumber, Number, Milliseconds, Name, Subject, AppName, AppVer, TimeOfDay, and Date. Two rows of data are visible, each showing a 'Stroke' and its 'Normalized Stroke'.

Raw Object	Class	RowNumber	Number	Milliseconds	Name	Subject	AppName	AppVer	TimeOfDay	Date
 Stroke	arrow	0	1	547	arrow01	1	Gestures	3.5.0.0	5:05:00 PM	Monda
 Normalized Stroke										
 Stroke	arrow	1	2	557	arrow02	1	Gestures	3.5.0.0	5:05:01 PM	Monda
 Normalized Stroke										

AI-driven design

Does your design look a bit like this?

Video Sonic Labs Inc
Electronics Repair & Installation Authority
1 877 900 HDTV (4388)

[On-Line Store](#) [Contact](#) [WHAT'S NEW](#) [Search](#) [About Us](#) 6:28:45PM

We Repair & Service All Make Model HDTV Experts

PLASMA/LCD

LCD, LCOS, DLP Projections TV

Projectors

VISIT OUR NEW LOCATION IN FABULOUS LAS VEGAS

[Click Here To Visit Our ON-LINE STORE](#)

[Need Installation? Click Here INSTALLATION](#)

Factory Authorized Service Center For

DC DAEWOO ELECTRONICS

Address & Hours of Operations

LOCATION#1

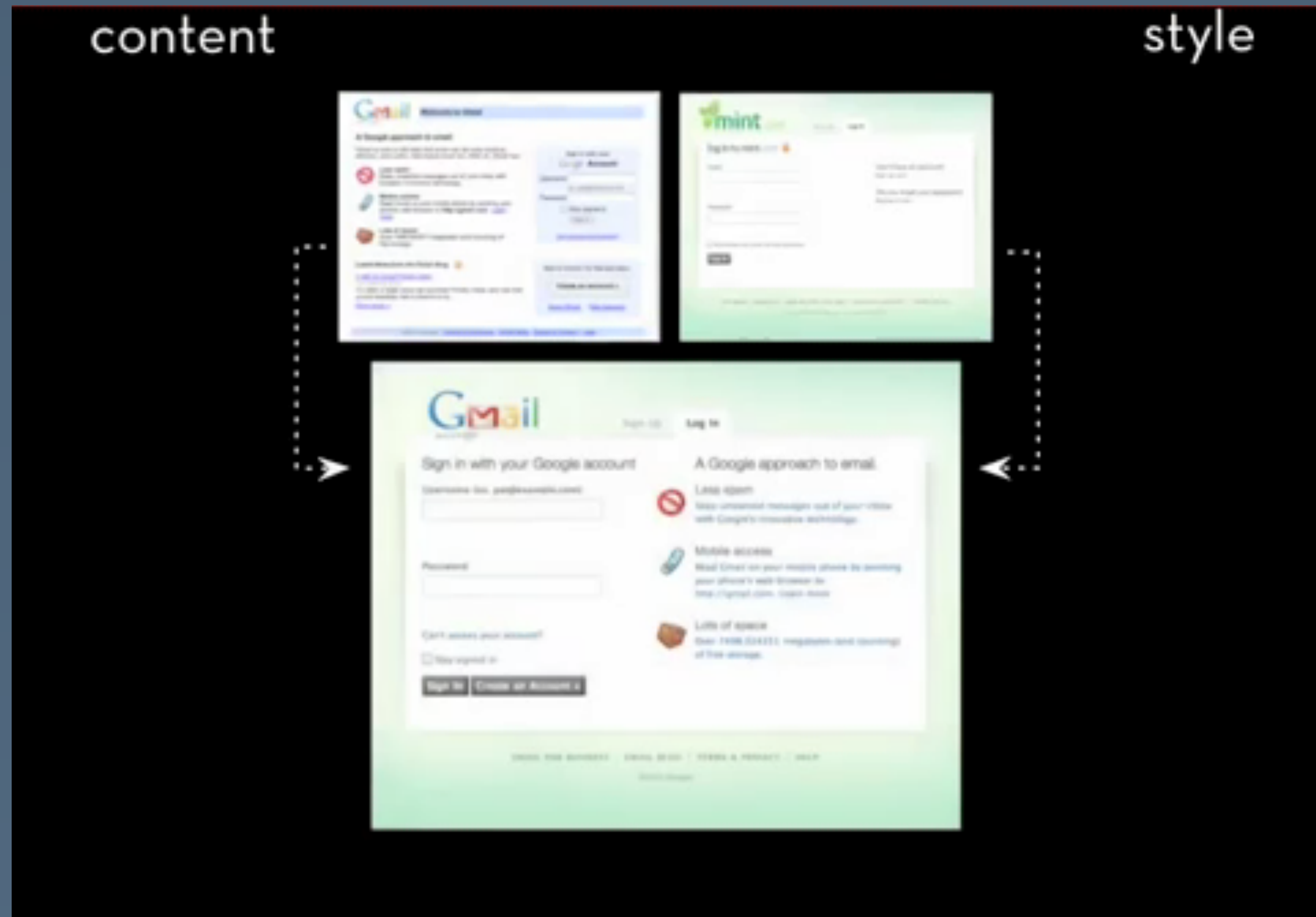
Monday To Friday
10:00AM to 6:00PM
Saturday

AI-driven design

- Learn design patterns from high-quality designs, and make it straightforward to apply those design patterns to your own designs

Retargeting designs

- “Can I borrow your design?” for the web
- Structured tree mapping algorithm
 - Roughly: costs associated with violating ancestry and sibling relationships in creating a mapping



Webzeitgeist

[Kumar et al., CHI '13]

- Crawl the web and index large-scale design elements
- Main idea: what happens if we start data mining designs, rather than user behavior?





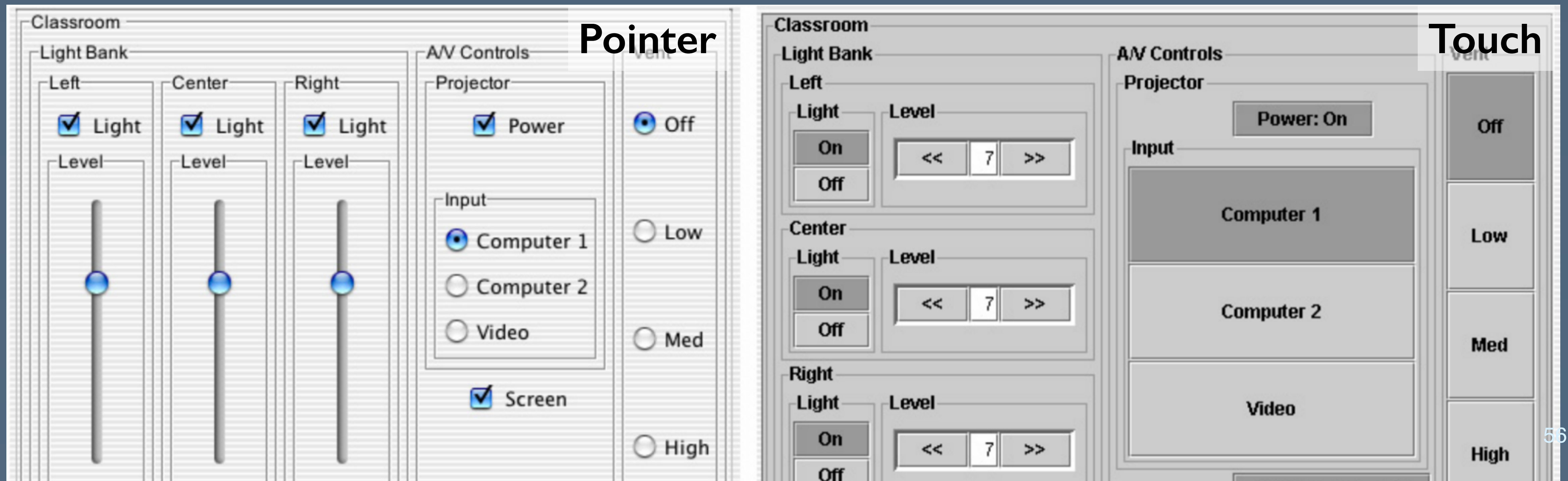
Human-robot interaction: enforcing social norms

[Porfirio et al., UIST '17]

Adaptive interfaces

[Gajos and Weld, IUI '04]

- Reactive design: remaps to output affordances
- Minimize a cost function derived from navigating between widgets in user traces

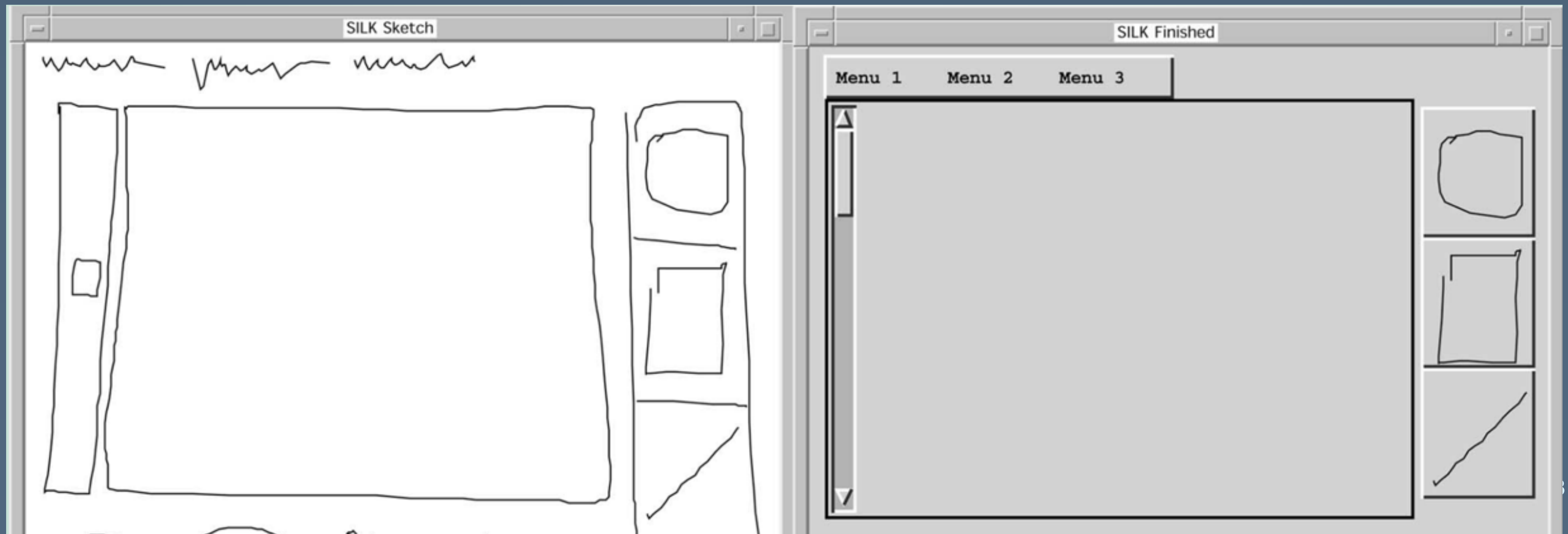


We see this all over...

- ...as in papers we already covered!

Sketch the interaction to produce working systems

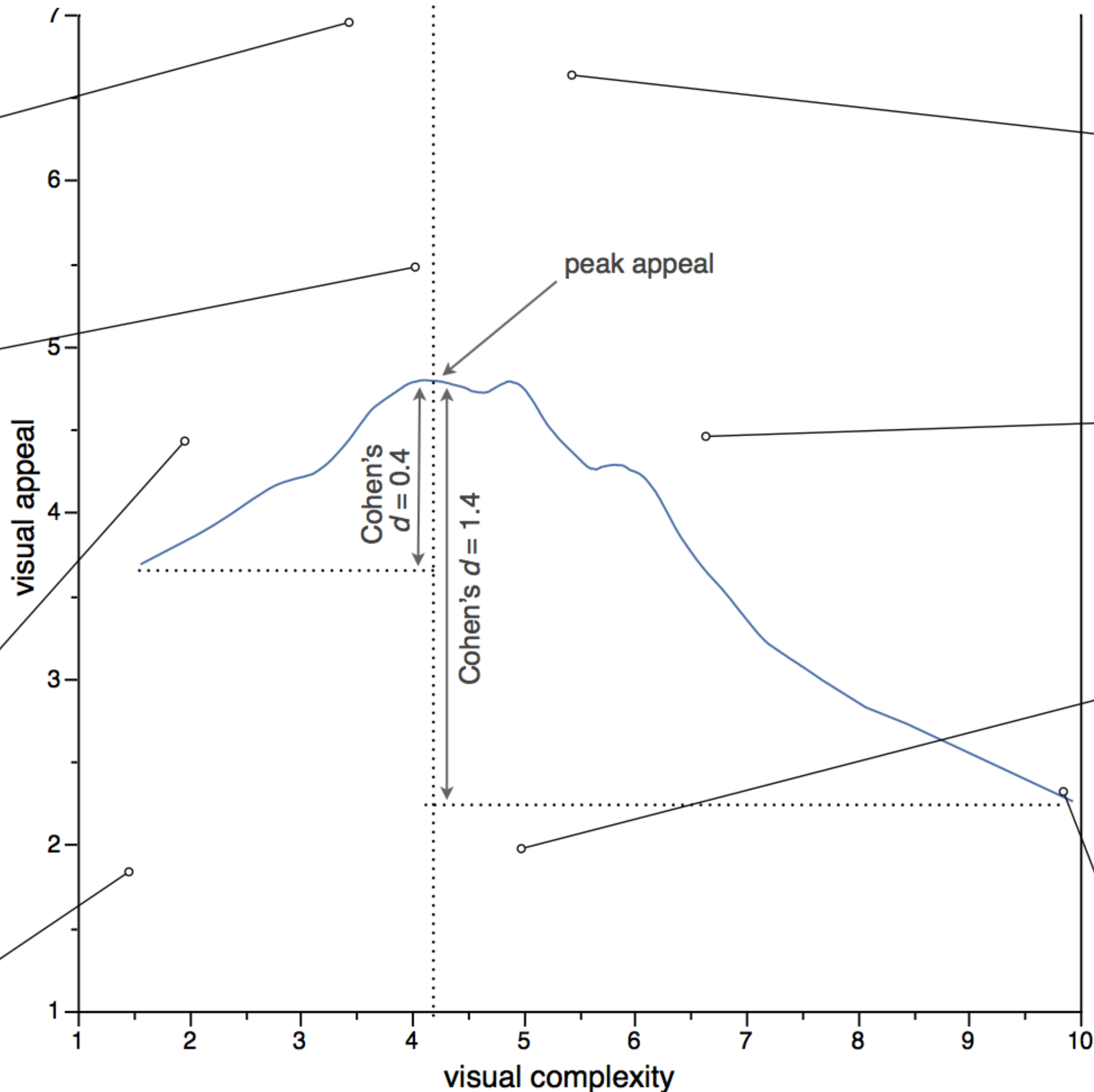
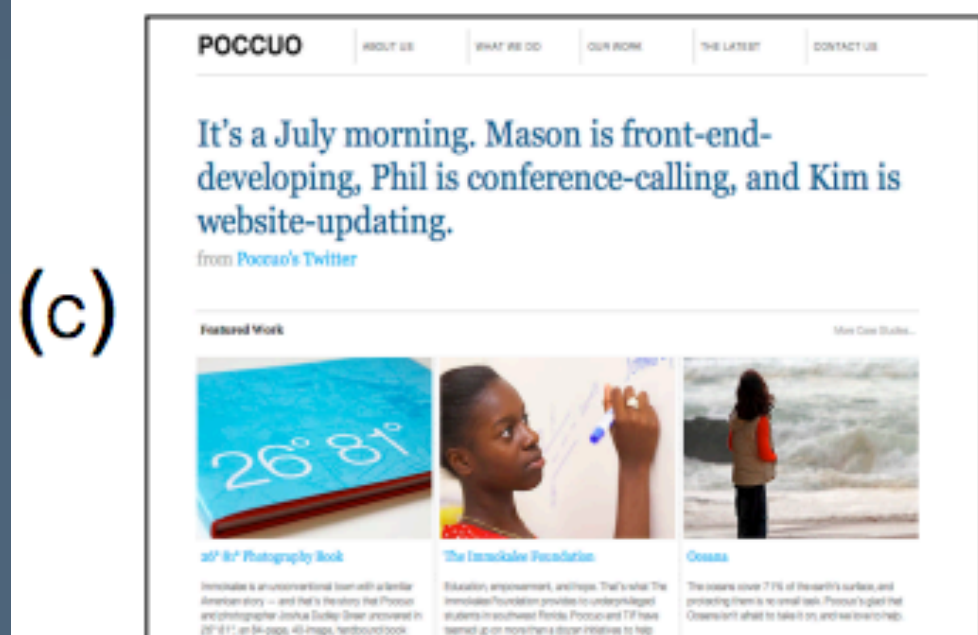
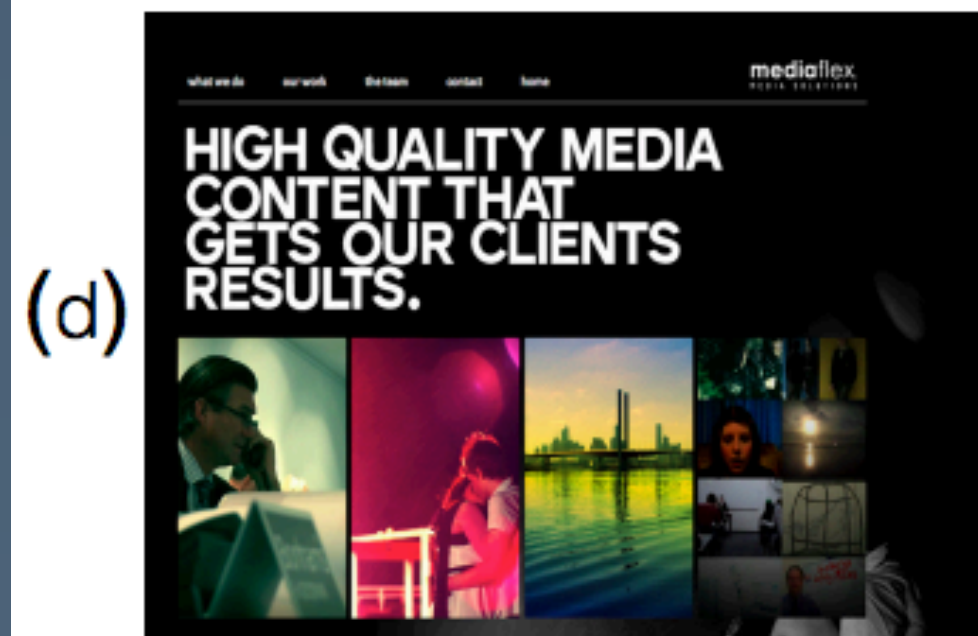
- SILK [Landay, CHI '96]



Quantifying Visual Preferences

[Reinecke and Gajos CHI 2014]

- LabInTheWild data via a quiz about which web sites you like



(d)

(c)

(b)

(a)

(e)

(f)

(g)

(h)

DesignScape: interactive layout

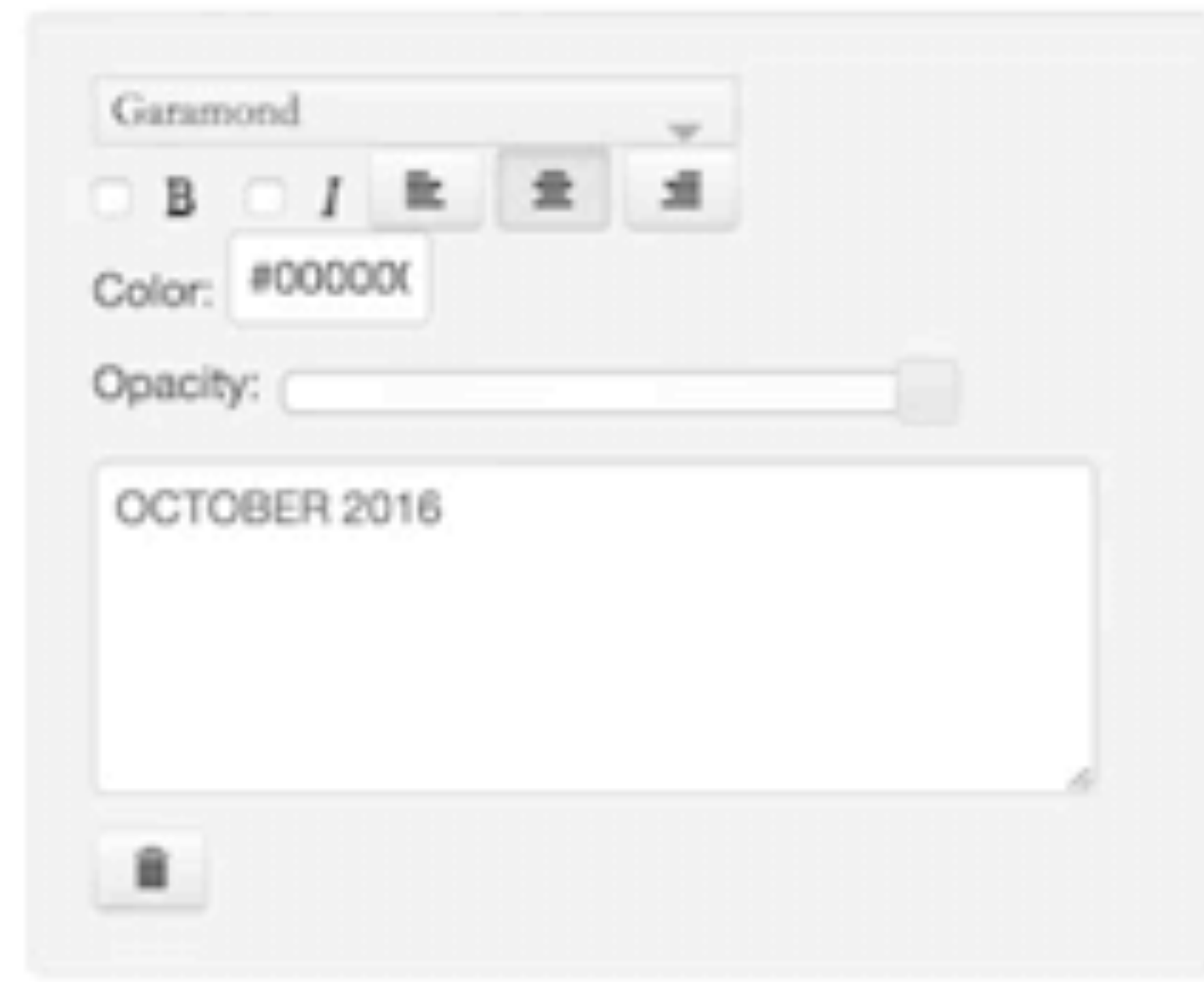
[O'Donovan, Agarwala, and Hertzmann CHI '15]

The image shows a software interface for designing a flyer for a "High School Chemistry Tutor". The interface is divided into three main sections: "Tweak Your Design", a central workspace, and "Brainstorm New Designs".

Tweak Your Design: On the left, there are three panels, each showing a different layout of icons (microscope, calculator, flask) and text for the flyer. The top panel is selected.

Central Workspace: The main area shows the selected design. At the top, there is a toolbar with buttons for "File", "Add", "Undo", "Redo", a lock icon, a zoom level of "12", a "Randomize" button, and a "Save" button. The design itself features three large icons: a blue microscope, a red calculator, and a blue flask with red liquid. Below the icons, the text reads: "HIGH SCHOOL CHEMISTRY TUTOR", "Available after 6pm weekday evenings, 10am-5pm on weekends", "555-555-5555", "help@tutor.ca", and a list of skills: "Learn how to: - identify types of chemical reactions - balance chemical equations - balance redox reactions - convert grams to moles - write in scientific notation".

Brainstorm New Designs: On the right, there are three panels showing alternative design variations. Each panel shows a different arrangement of the icons and text, such as "Available after 6pm weekday evenings, 10am-5pm on weekends" in a different position or font.



Learning Visual Importance

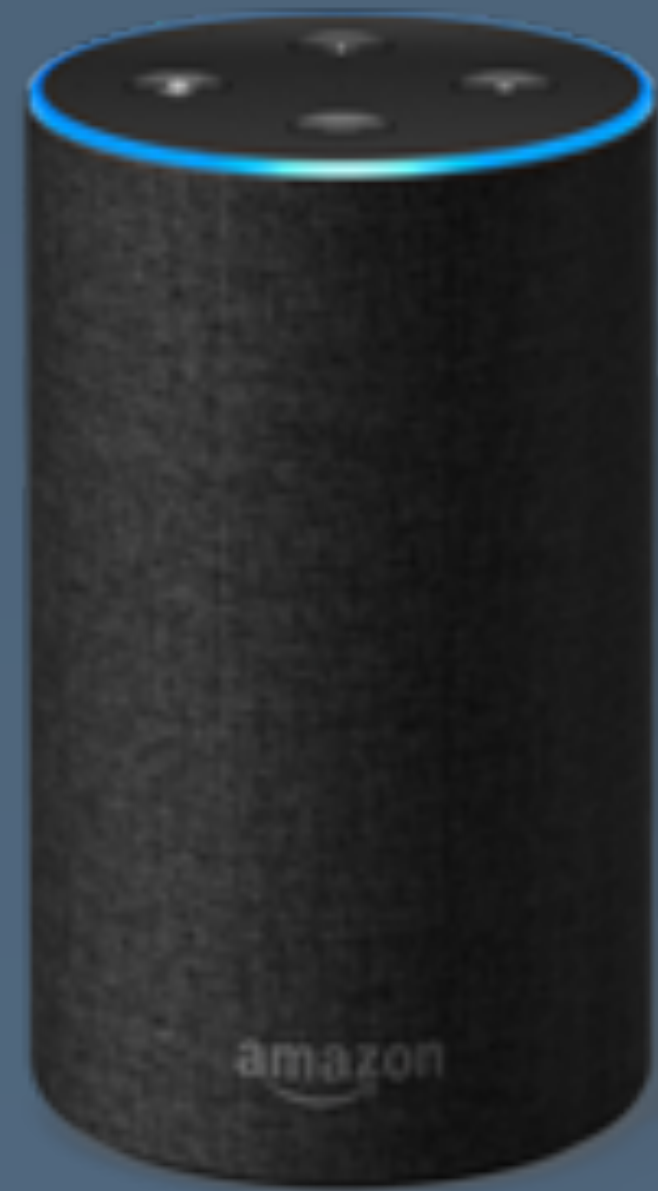
[Bylinskii et al., UIST '17]

What to take away?

- AI can...
 - Help identify effective designs
 - Help map your design onto a target design
 - Predict how people will react to your design

Voice, gesture,
and vision-based
interaction

Interaction off the desktop requires intelligence



Amazon Echo
+ Siri



FitBit + Apple Watch



Nest thermostat

AI is now a component of many sensing pipelines

- ...as in papers we already covered!

Activity recognition

- Sense the user's physical state by using minimally invasive sensors
- For example, wearing five 2d accelerometers and predicting tasks like walking, watching TV, reading, eating...

Activity Recognition from User-Annotated Acceleration Data

Ling Bao and Stephen S. Intille

Massachusetts Institute of Technology

1 Cambridge Center, 4FL

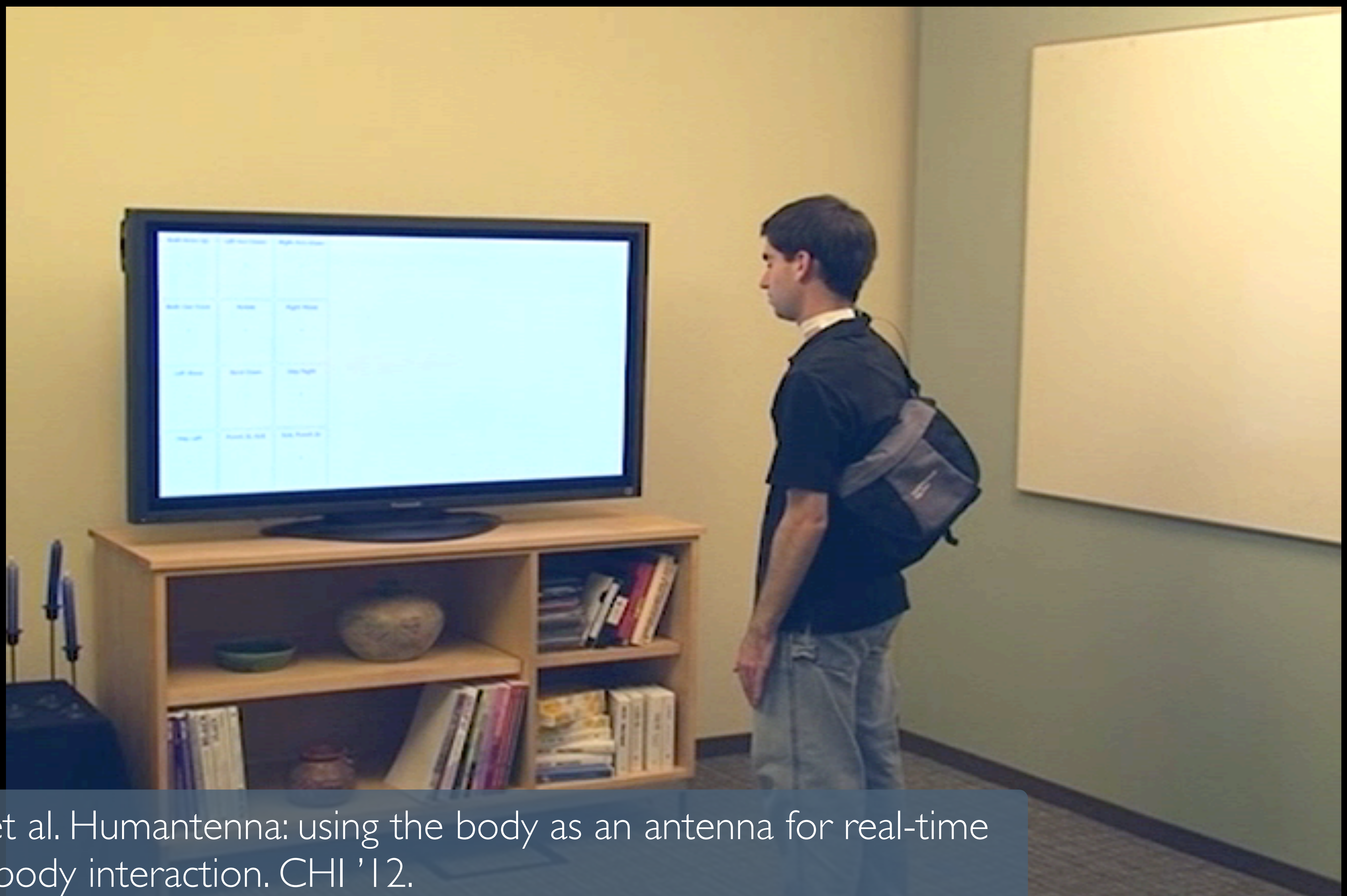


Custom
Powerline
Interface

USB Data
Acquisition/
Oscilloscope

PC

Patel et al. At the Flick of a Switch: Detecting and Classifying Unique Electrical Events on the Residential Power Line. UbiComp '07.



Cohn et al. Humantenna: using the body as an antenna for real-time whole-body interaction. CHI '12.



Harrison, Morris, Tan. Skinput: Appropriating the Body as an Input Surface. CHI '10.



Harrison, Benko, Wilson. Omnitouch: Wearable Multitouch Interaction Everywhere. UIST '11.

EM-Sense

Touch Recognition of Uninstrumented,
Electrical and Electromechanical Objects

Gierad Laput

Chouchang Yang

Robert Xiao

Alanson Sample

Chris Harrison

**Carnegie
Mellon
University**

Disney Research

Laput, G. et al. 2015. EM-Sense: Touch Recognition of Uninstrumented, Electrical and Electromechanical Objects. UIST '15.

Acoustics

Laput et al. Acoustruments: Passive, Acoustically-Driven Interactive Controls for Hand Held Devices. UIST '15.

Multimodal interaction

Using simultaneous inputs

- Sensor fusion can help disambiguate multiple noisy signals



Put That There

Speech N-best	Gesture N-best	Multimodal N-
Zoom in	Checkmark	Zoom out
Show info		
Show all		
Zoom out		

Quickset

[Oviatt, CHI '99]

Reflections

- AI is a powerful tool, but brings massive user interaction problems as a result of the uncertainty it introduces for the user
- Smart interaction design can hide or manage that uncertainty.
- “Don’t let your AI write a check that your UI can’t cash; Don’t let your UI write a check that your AI can’t cash.”
 - Eytan Adar, University of Michigan