Human Computation in Computer Vision

Slides adapted from Luis von Ahn, Manuel Blum, Nicholas Hopper, John Langford, James Hays, Brian O'Neil, and Alexander Sorokin

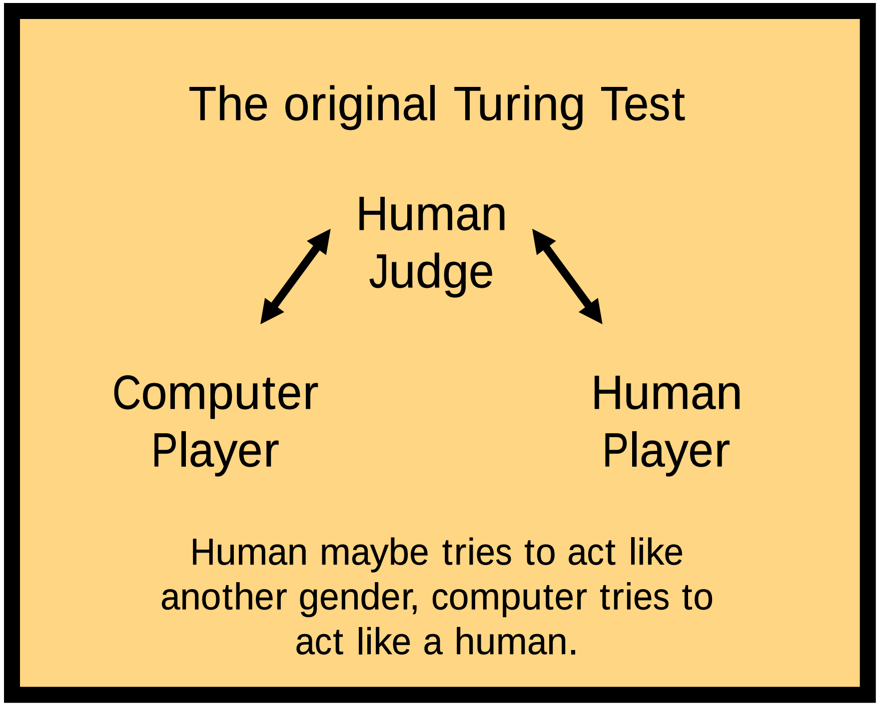
Turing Test

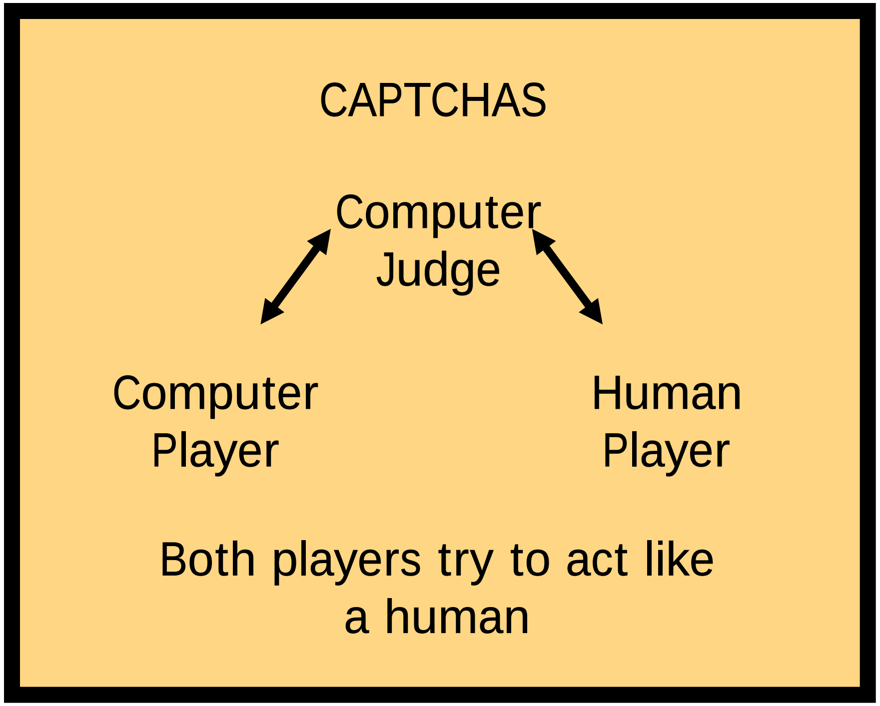
- Proposed to demonstrate machine intelligence
- Variation on a parlor game called the *imitation* game
- An *interrogator* asks questions (via teletype) of a *subject*
  - to guess their gender

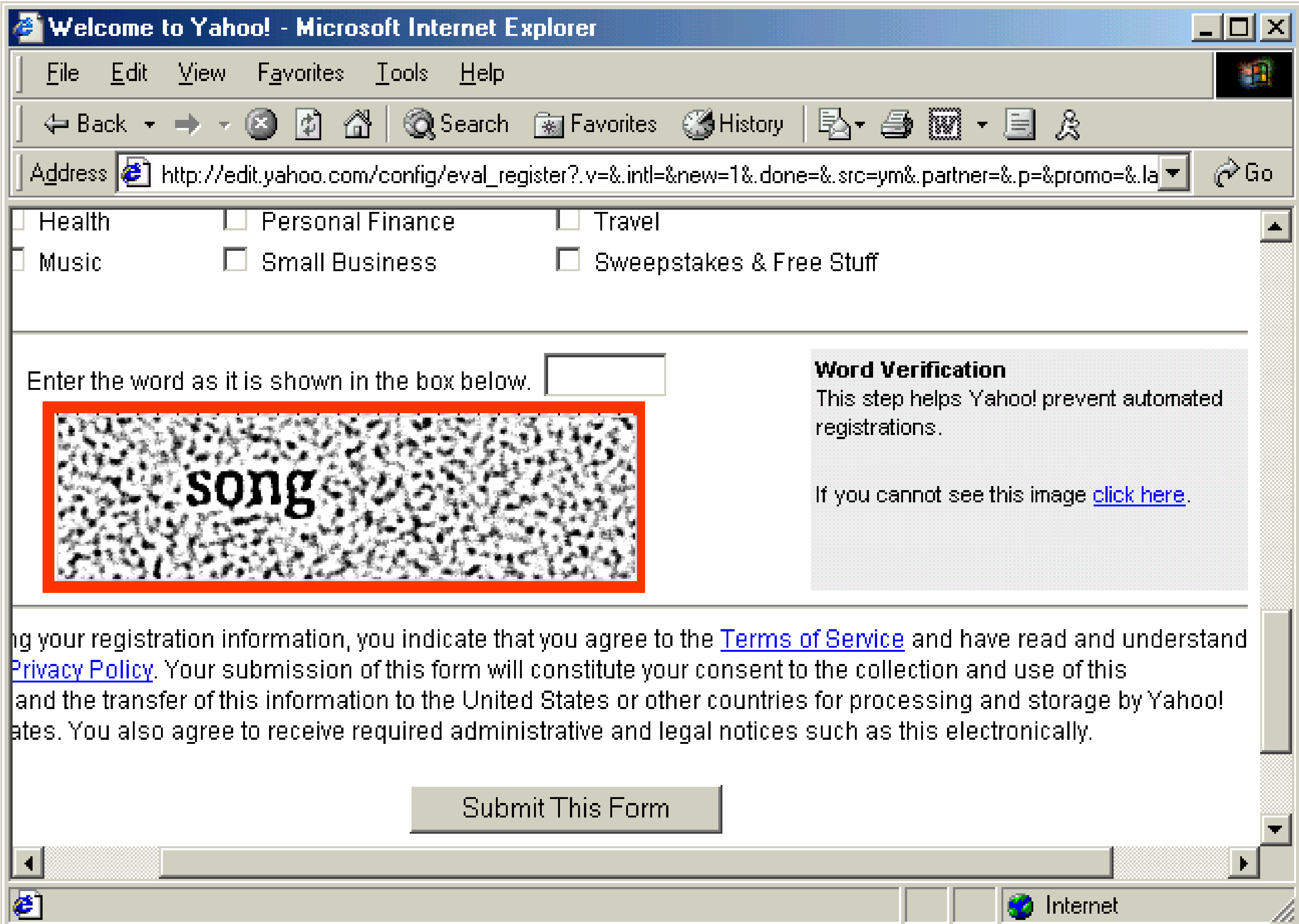
- Turing suggested replacing the subject with a computer
- If, after some agreed time, the interrogator cannot distinguish situations where a machine has been substituted for the man/woman, we should just agree to say the machine can think (says Turing)

Minor Modification

- Today, we are going to see if a *computer* can tell the difference between a person and a computer.






Welcome to Yahoo! - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites History
Address http://edit.yahoo.com/config/eval_register?.v=&.intl=&new=1&.done=&.src=ym&.partner=&.p=&promo=&.la
Go
Health Personal Finance Travel
Music Small Business Sweepstakes & Free Stuff
Enter the word as it is shown in the box below.
song
Word Verification
This step helps Yahoo! prevent automated registrations.
If you cannot see this image click here.
ng your registration information, you indicate that you agree to the Terms of Service and have read and understand
Privacy Policy. Your submission of this form will constitute your consent to the collection and use of this
and the transfer of this information to the United States or other countries for processing and storage by Yahoo!
ates. You also agree to receive required administrative and legal notices such as this electronically.
Submit This Form
Internet

CAPTCHA: "**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part"

A program that can *generate* and *grade* tests that:

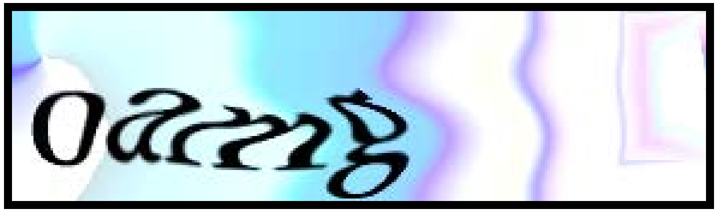
A. Most humans can pass
B. Current computer programs cannot pass

Example

Picks a random string of letters

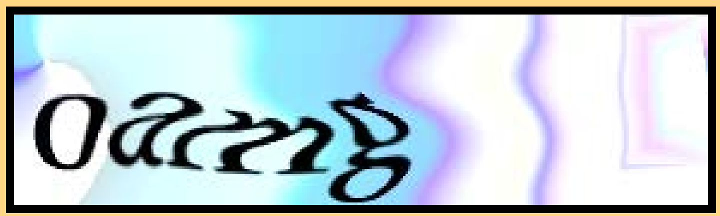
oamg

Renders the string into a randomly distorted image



Example

...and generates a test:



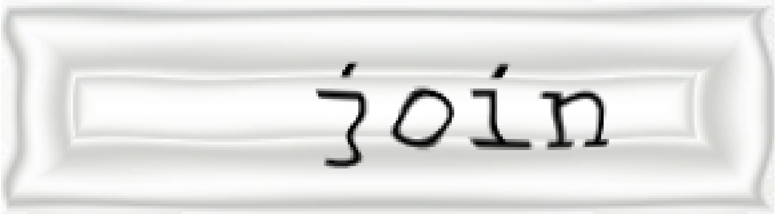
Type the characters that appear in the image

EZ-Gimpy Examples

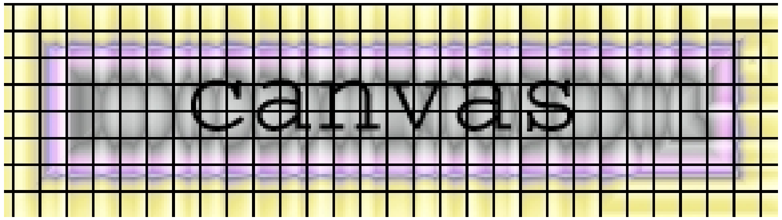
horse

spade

smile



join



canvas

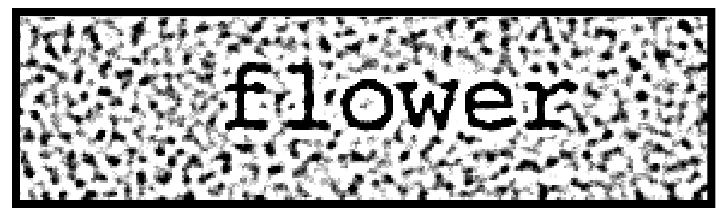
here

Advancing AI

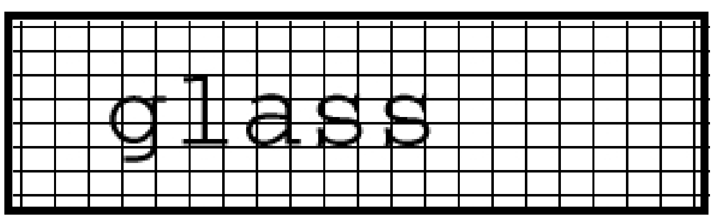
Mori and Malik, 2002: 92% accuracy against Yahoo! CAPTCHA (using Shape Context)

front

flower

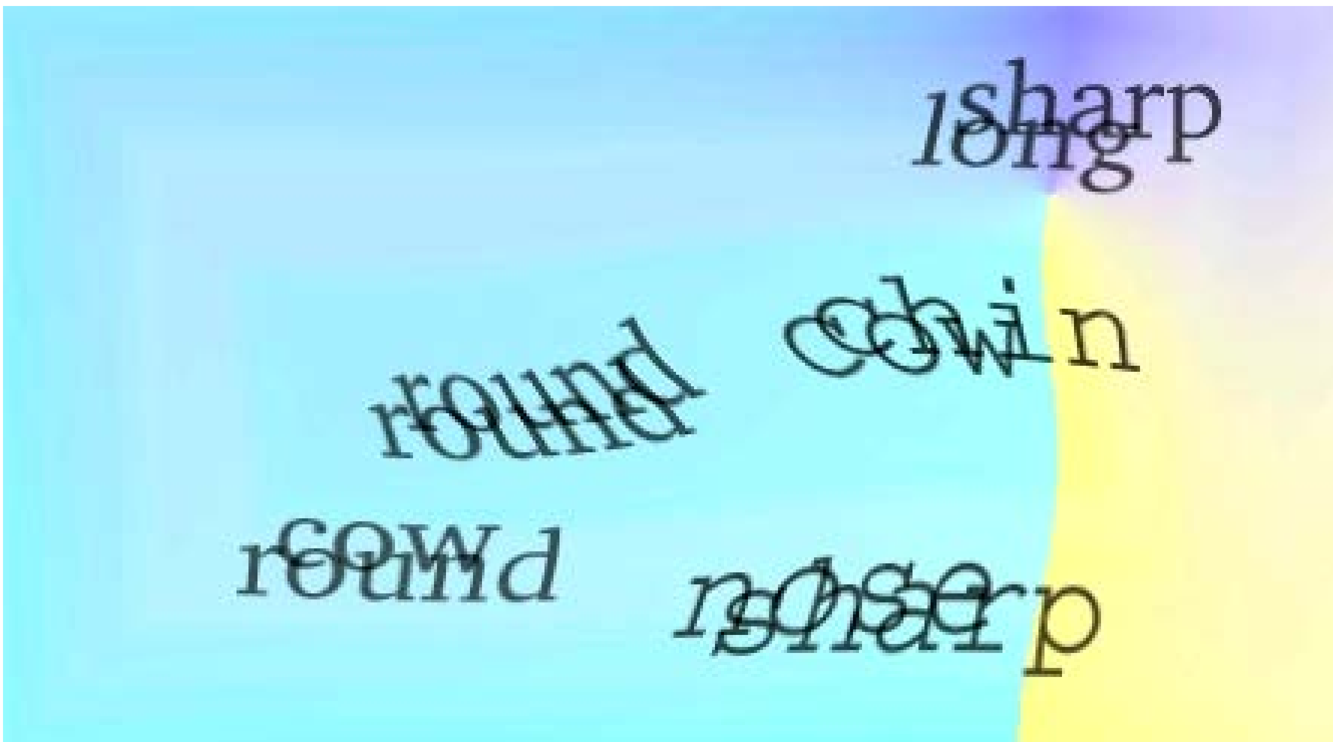


glass

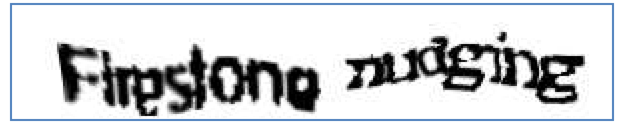


lock

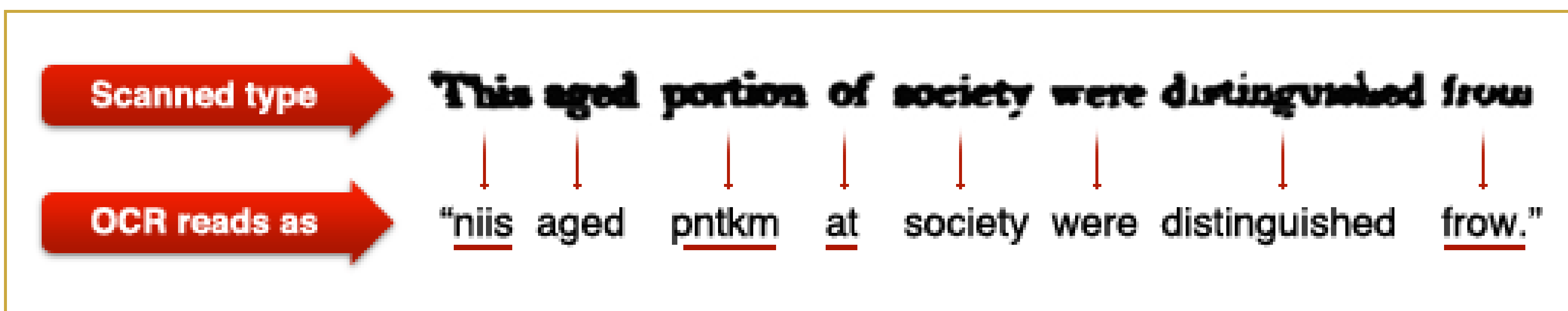
Make it harder (Gimpy)



ReCAPTCHA



- Similar to previous “word-based” CAPTCHAs
- Takes advantage of OCR failures
  - Presents 2 “hard” words: one known and one unknown
- Serves as CAPTCHA and helps to digitize books
- Designed by Luis von Ahn (CMU)
  - Awarded MacArthur Fellowship (“Genius Award”)



Other types of Captchas

What are these pictures of?

Pix

What are these pictures of?

The images need to be randomly distorted.
Why?

CAPTCHAs Are a Win-Win Situation

Either a CAPTCHA remains secure or
an open problem becomes solved

CAPTCHAs get malicious people to
work on AI problems!

Spin-off Idea

- Still in the realm of images…
- Instead of creating the test from words, use concepts from images

Labeling Images With Words

Donald Trump
Chair
Toupee

Until recently, a completely unsolved problem

Data Generator Goal

- We want a method for labeling images that:
  1. Actually looks at the images
  2. For any image gives several keywords that make sense
  3. Is very fast (Google indexes > 10B images)
     (as of July 2010)

Stealing Cycles From Humans

Over 50 million people in the United States play computer games on a regular basis!

The ESP Game (Now Google Image Labeler)

- Two-player online game
- Partners don't know each other and can't communicate
- Object of the game: type the same word
- The only thing in common is an image

The ESP Game

Player 1

Player 2

The ESP Game

Player 1

Guessing: car

Player 2

Guessing: boy

The ESP Game

Player 1

Guessing: car

Guessing: hat

Player 2

Guessing: boy

The ESP Game

Player 1

Guessing: car

Guessing: hat

Guessing: kid

Player 2

Guessing: boy

The ESP Game

Player 1

Guessing: car

Guessing: hat

Guessing: kid

Success!
You both agree on car

Player 2

Guessing: boy

Guessing: car

Success!
You both agree on car

Taboos guarantee that each image will get many different keywords

The ESP Game

Taboos guarantee that each image will get many different keywords

Preliminary studies suggest that people find the game fun

The ESP Game

Average labeling rate: 4 images per minute

5000 people simultaneously playing the game would label all the images on Google in 2 years! (in 2010)

$$\frac{5000}{2} \times 4 \times 60 \times 24 \times 365 \times 2 = 10{,}512{,}000{,}000$$

Individual games in Yahoo!, Pogo.com or MSN average well over 10,000 players at a time

Problems with the ESP Game

- Not so easy to get things to them to do useful things!
- This game devolves quickly, as people want to win
  - Images start being labeled by the primary color
    - Not informative
    - Also, a computer could figure that out
- Interesting question, perhaps, “how do people that can’t communicate develop strategies together?”

Locating Objects in Images

- The ESP game tells us if an image contains an object
    - It doesn’t say *where* in the image the object is

- Such information would be extremely useful for computer vision research

Other Games: Paintball

PLAYERS SHOOT AT OBJECTS ON THE IMAGE

SHOOT THE:
CAR

Give points and check accuracy by using images which we already know where the car is (similar to reCAPTCHA)

PAINTBALL GAME

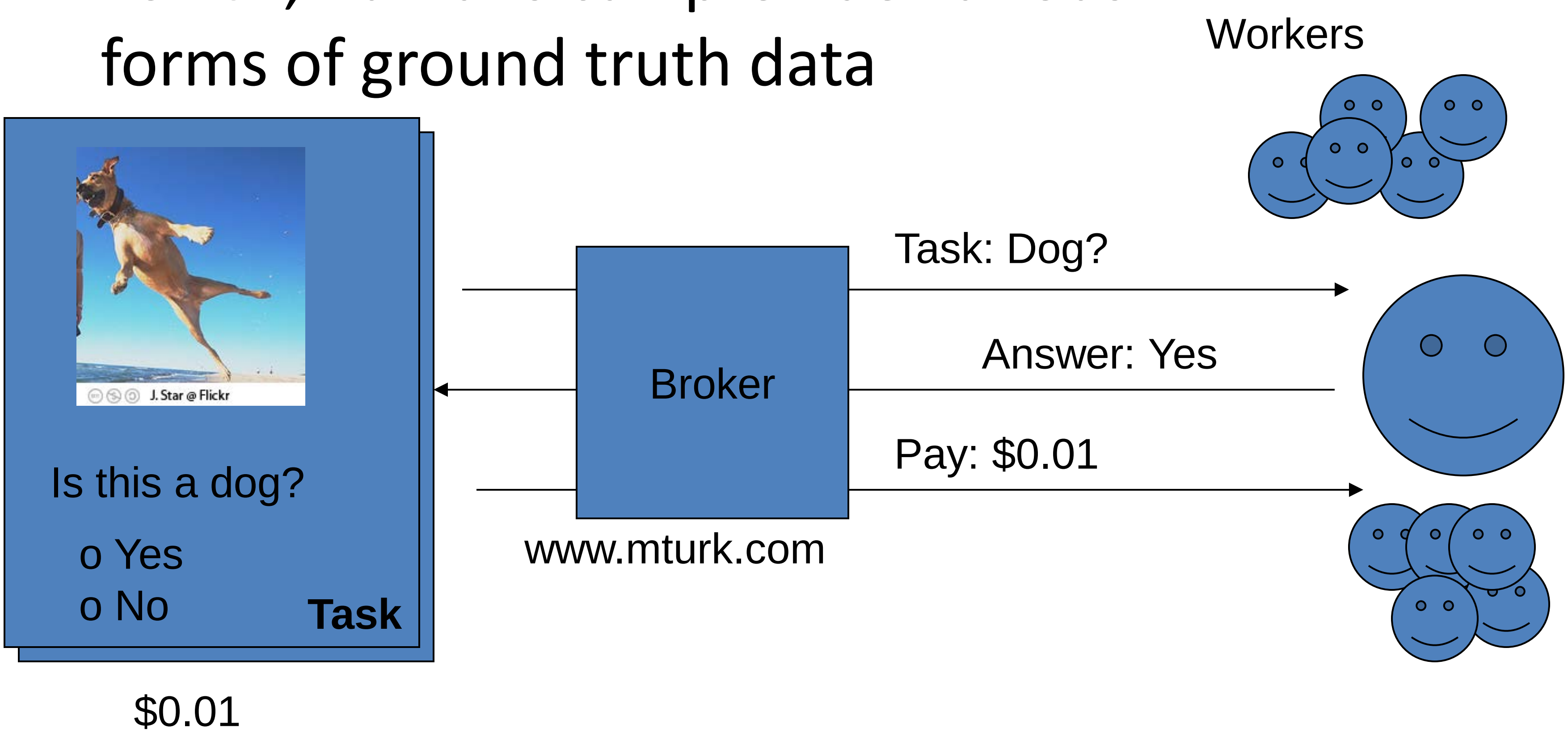
X% OF IMAGES

(100-X)% OF IMAGES

DON’T
KNOW

Beyond Games…

- Amazon Mechanical Turk (MTurk)
  - Human Intelligence Tasks (HITs)
- For CV, humans can provide various forms of ground truth data



Type keywords

Mechanical Turk Project

If you're using the turk, Be sure to copy the text back into the HIT page so that you can be credited.

- Photo should be rotated 90 degrees left (counter-clockwise)
- Photo should be rotated 90 degrees right (clockwise)
- Photo should be turned upside down
- Photo is oriented properly

Please describe the picture in the box using 10 words or more:

shells

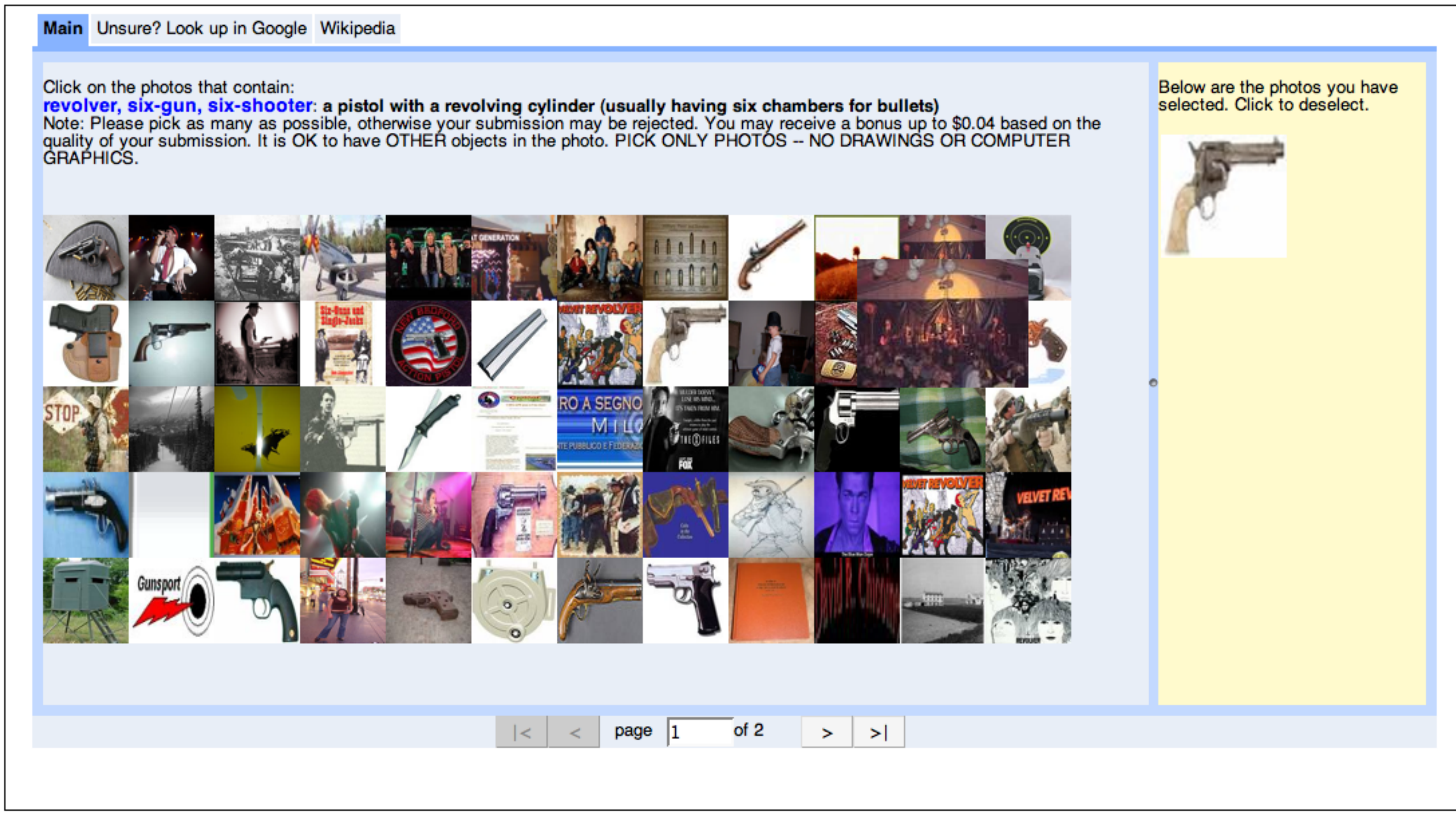
Submit Turk Skip / Load a different photo

The submit button MUST be clicked!

$0.01

http://austinsmoke.com/turk/.

Select examples



$0.02

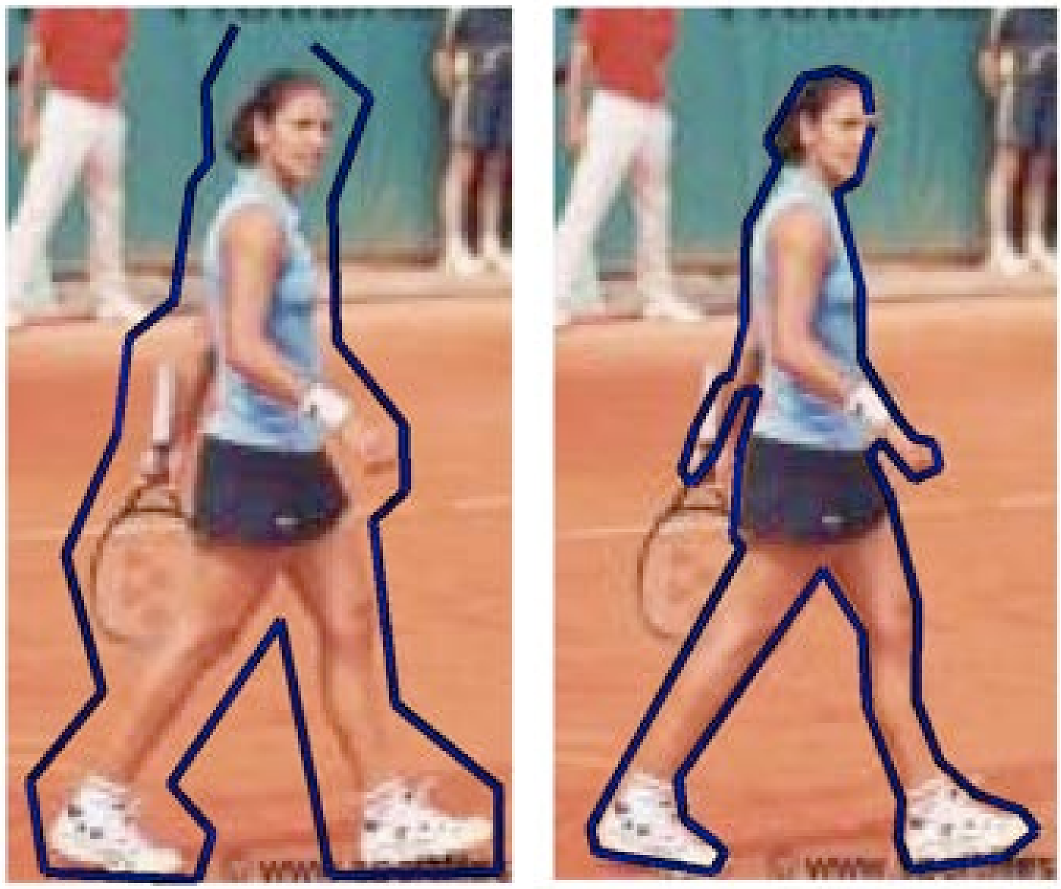
requester mtlabel

Click on landmarks

$0.01

http://vision-app1.cs.uiuc.edu/mt/results/people14-batch11/p7/

Outline something



$0.01 http://visionpc.cs.uiuc.edu/~largescale/results/production-3-2/results_page_013.html

Data from Ramanan NIPS06

Price? Quality?

X 100,000 = $5000

Custom annotations

Large scale

Is this a good deal?

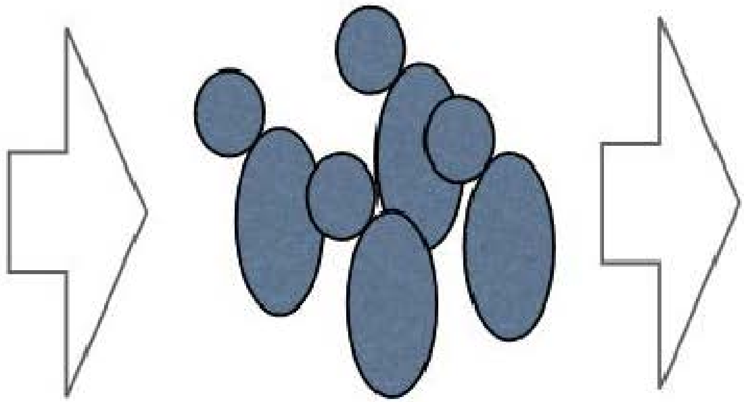
- Quality?
  - How good is it?
  - How to be sure?
- Price?
  - How to price it?

How do we get quality annotations?

Building datasets

6000 images from flickr.com

Annotators



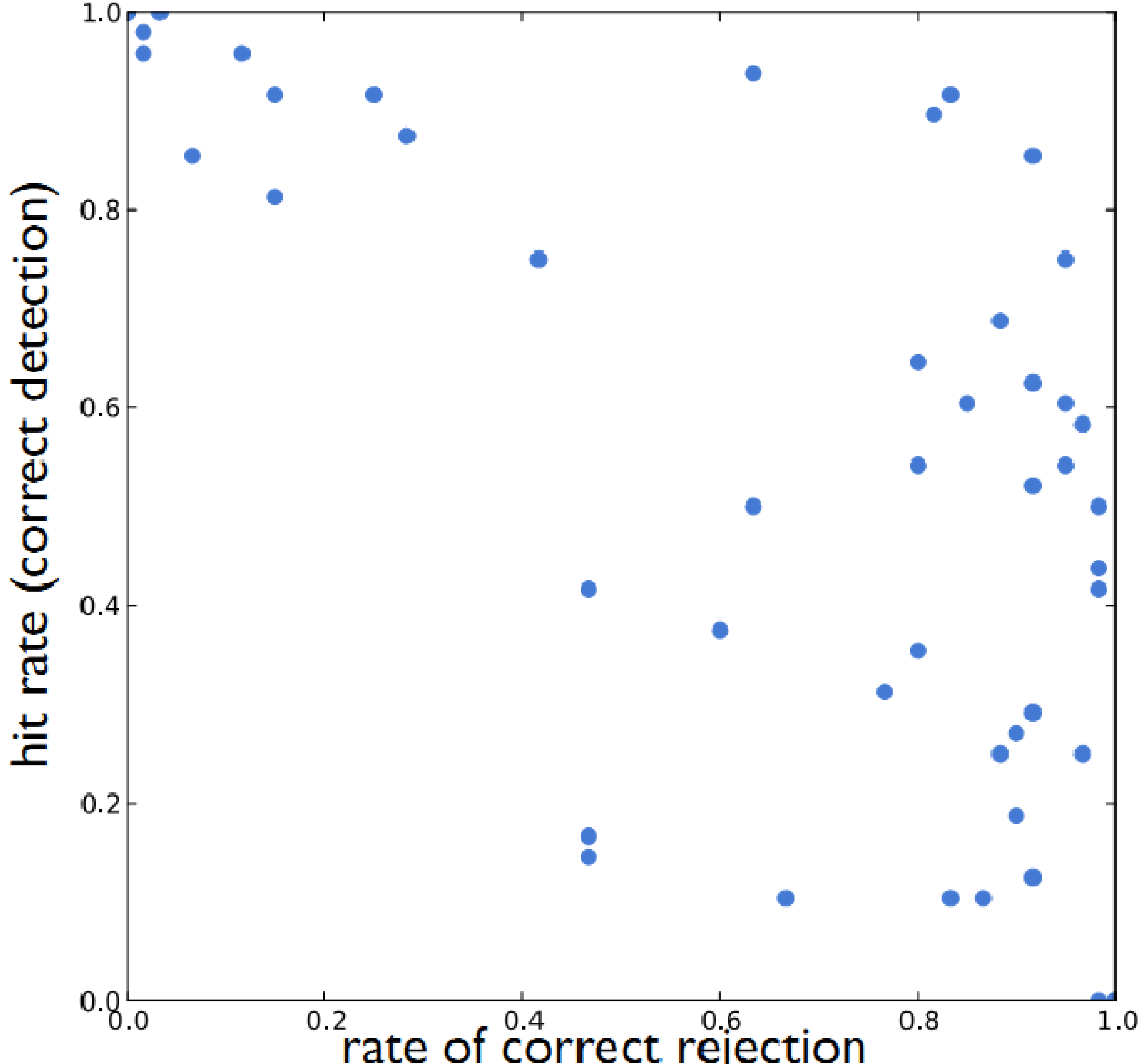
amazon mechanical turk
Artificial Artificial Intelligence

Is there an Indigo bunting in the image?

100s of training images

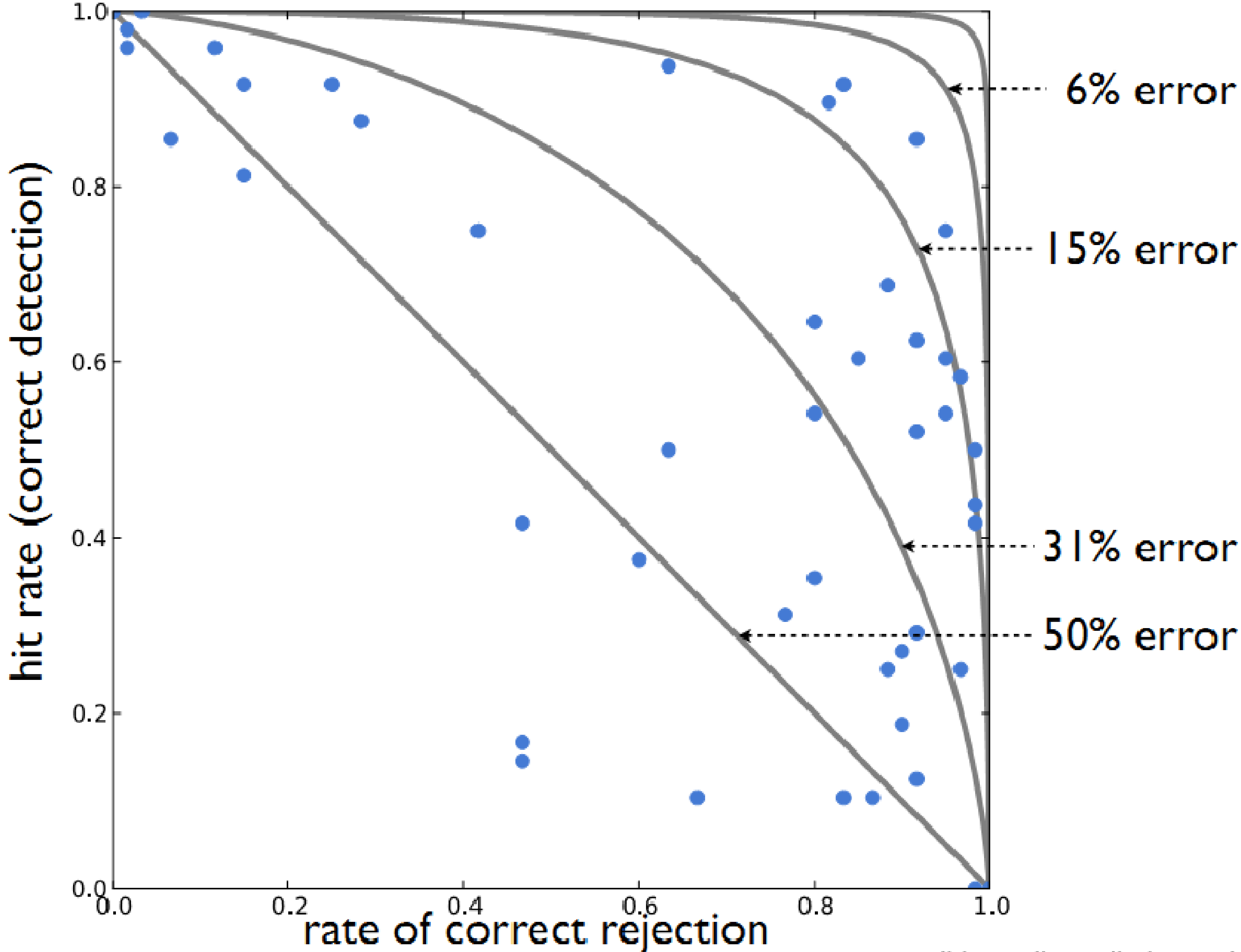
Slide credit: Welinder et al

Task: Find the Indigo Bunting



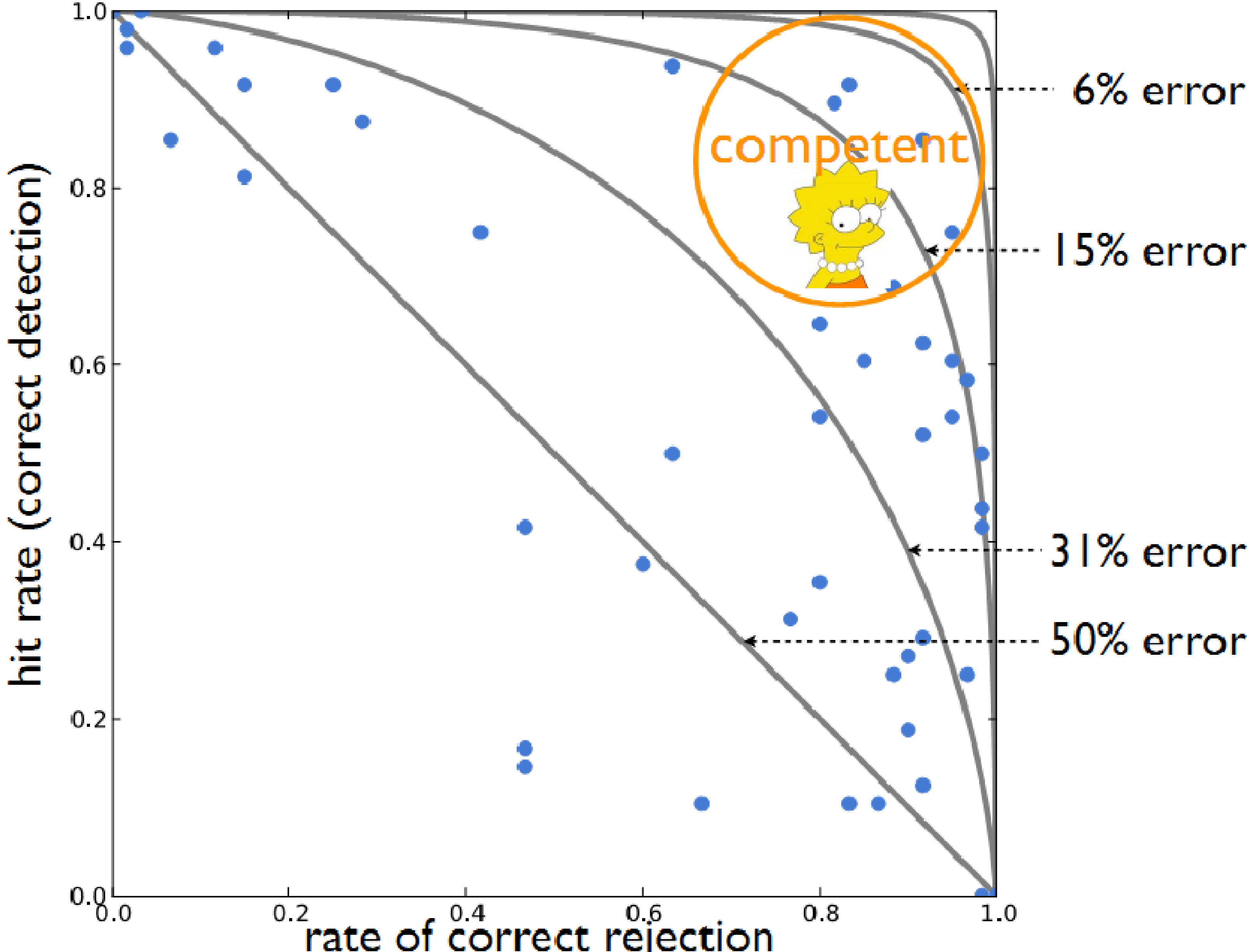
Slide credit: Welinder et al

Task: Find the Indigo Bunting



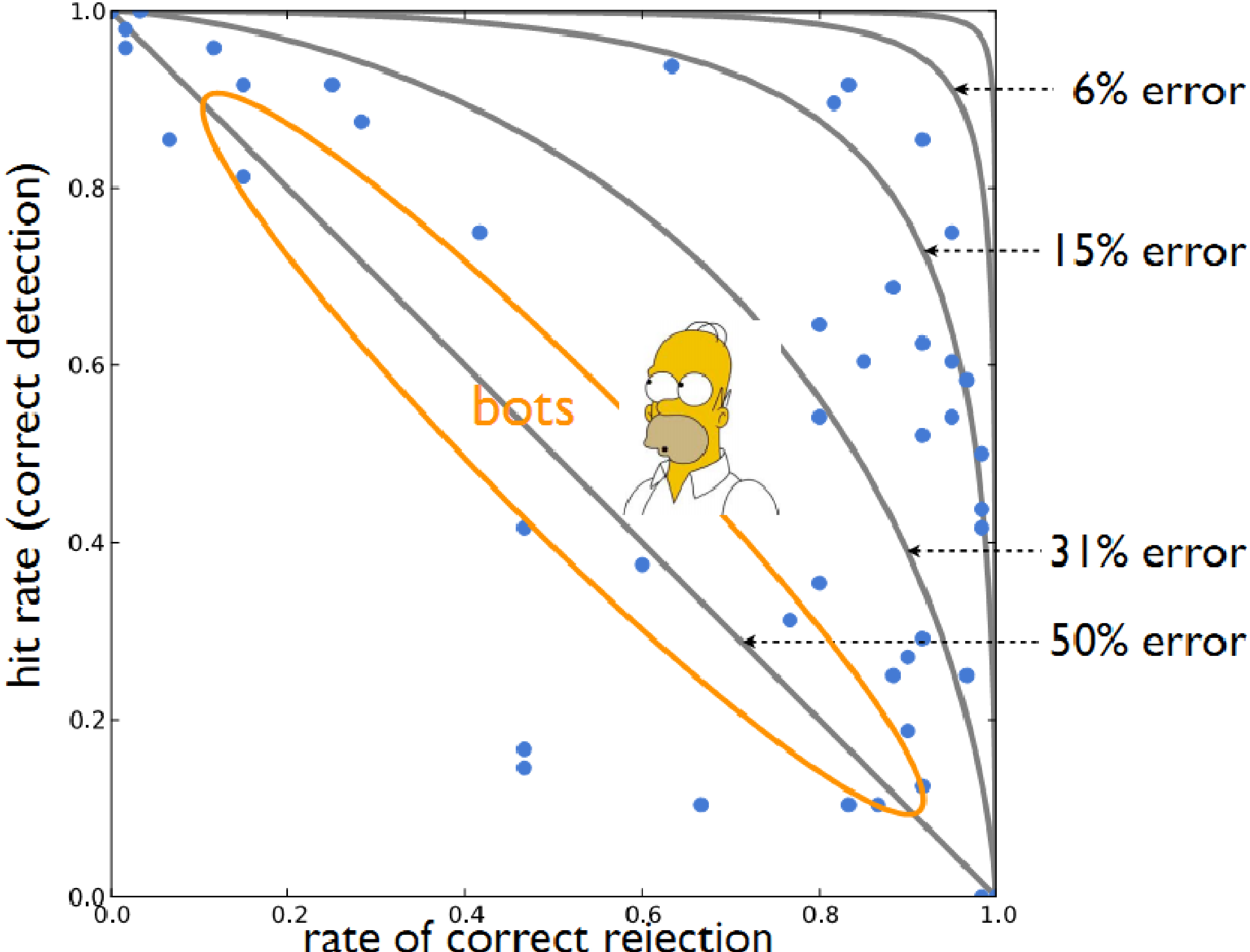
Slide credit: Welinder et al

Task: Find the Indigo Bunting



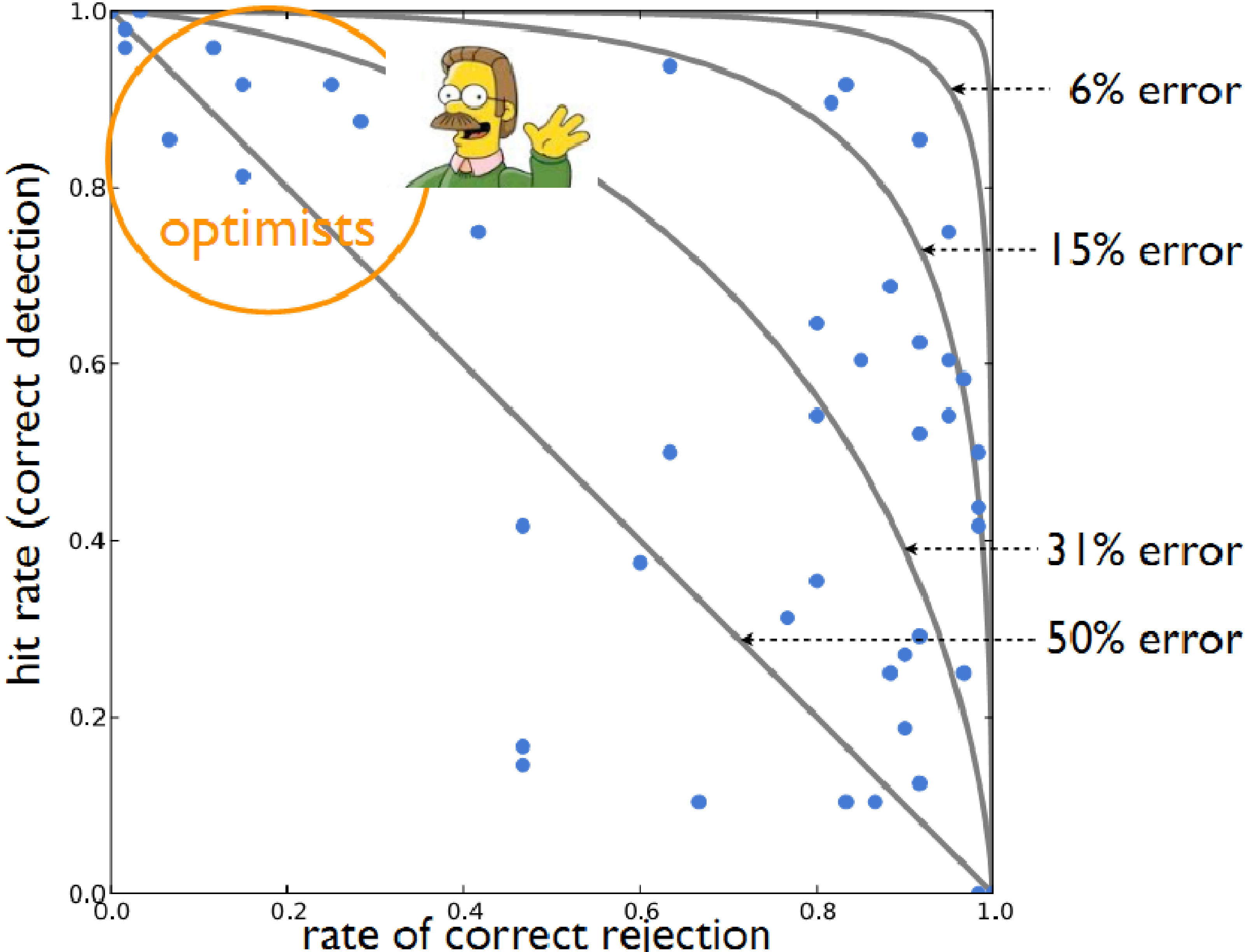
Slide credit: Welinder et al

Task: Find the Indigo Bunting



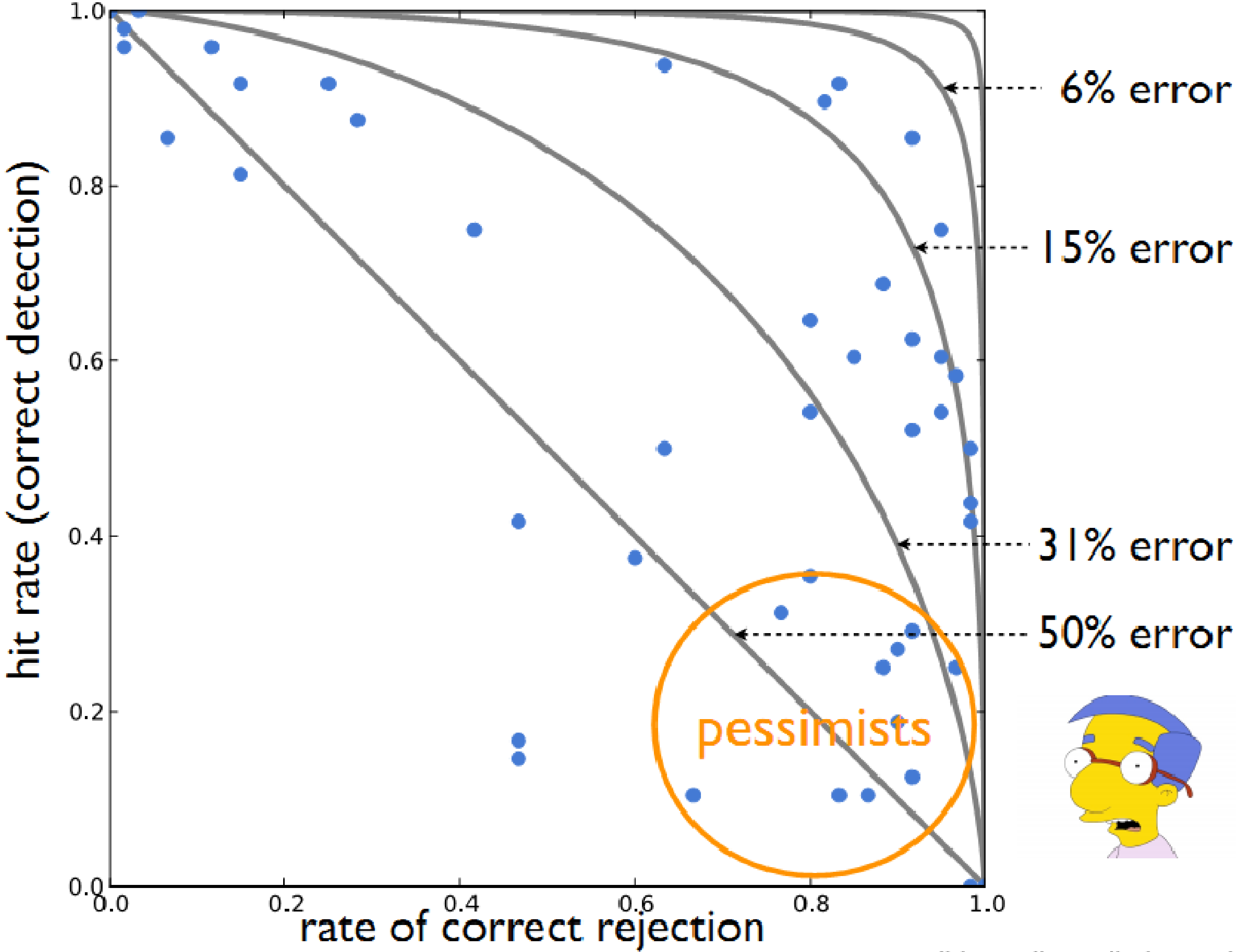
Slide credit: Welinder et al

Task: Find the Indigo Bunting



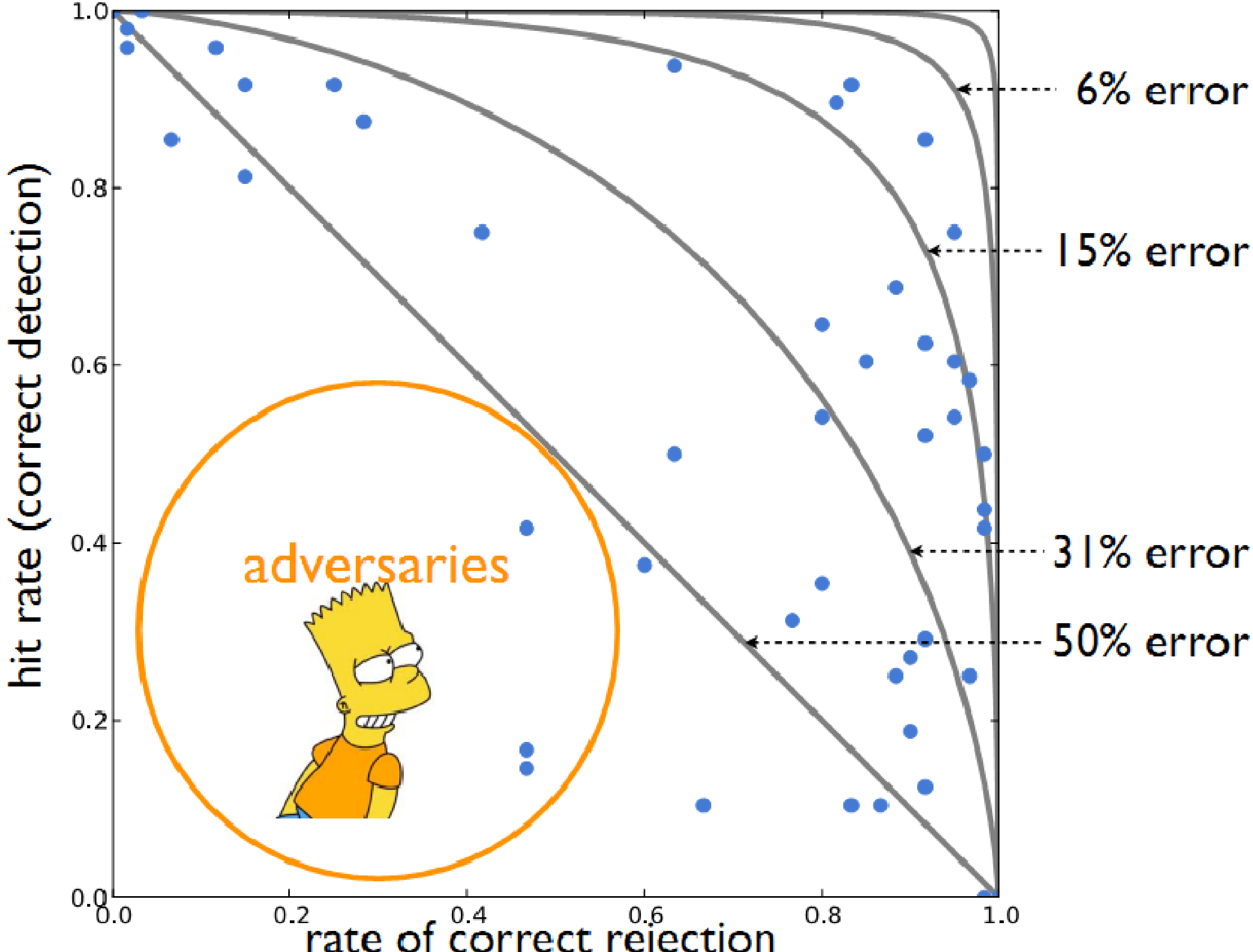
Slide credit: Welinder et al

Task: Find the Indigo Bunting



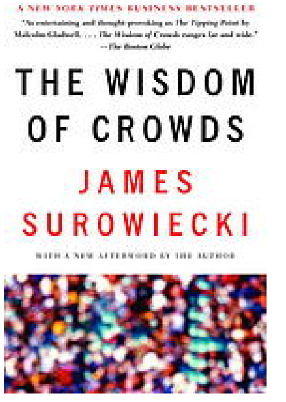
Slide credit: Welinder et al

Task: Find the Indigo Bunting



Slide credit: Welinder et al

Ensuring Annotation Quality



- Consensus / Multiple Annotation / "Wisdom of the Crowds"

- Gold Standard / Sentinel
  - Special case: qualification exam

- Grading Tasks
  - A second tier of workers who grade others

Pricing

- Trade off between throughput and cost
- Higher pay can actually attract scammers

Humans + Computers

(A) Easy for Humans

Chair? Airplane? …

Computers starting to get good at this.

(B) Hard for Humans

Finch? Bunting?…

If it's hard for humans, it's probably too hard for computers.

(C) Easy for Humans

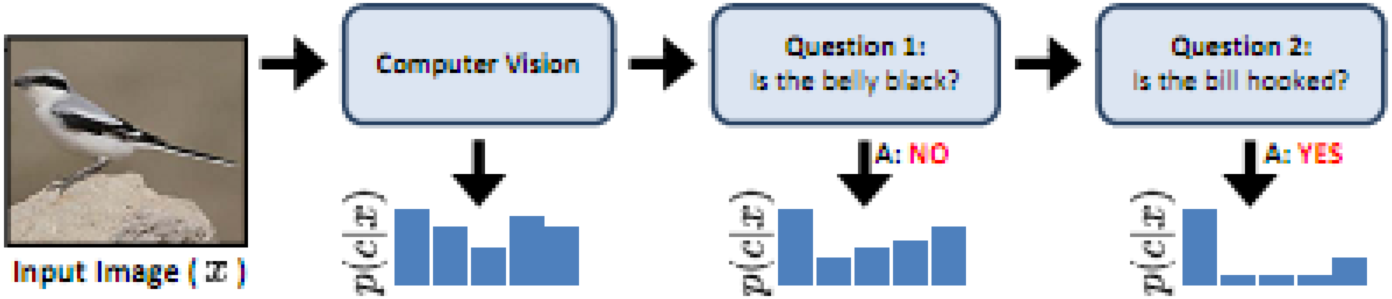
Yellow Belly? Blue Belly? …

Semantic feature extraction difficult for computers.

Combine strengths to solve this problem.

The Approach: 20 Questions

- Ask the user a series of discriminative visual questions to make the classification.



Which 20 questions?

- At each step, exploit the image itself and the user response history to select the most informative question to ask next

- Seek the question that gives the maximum information gain (entropy reduction) given the image and the set of previous user responses

Incorporating Computer Vision

- A visual recognition algorithm outputs a probability distribution across all classes that is used as the prior

- A posterior probability is then computed based on the probability of obtaining a particular response history given each class

The Dataset: Birds-200

- 6033 images of 200 species





Implementation

amazonmechanical turk

- Assembled 25 visual questions encompassing 288 visual attributes extracted from www.whatbird.com

- MTurkers asked to answer questions and provide confidence scores

User Responses

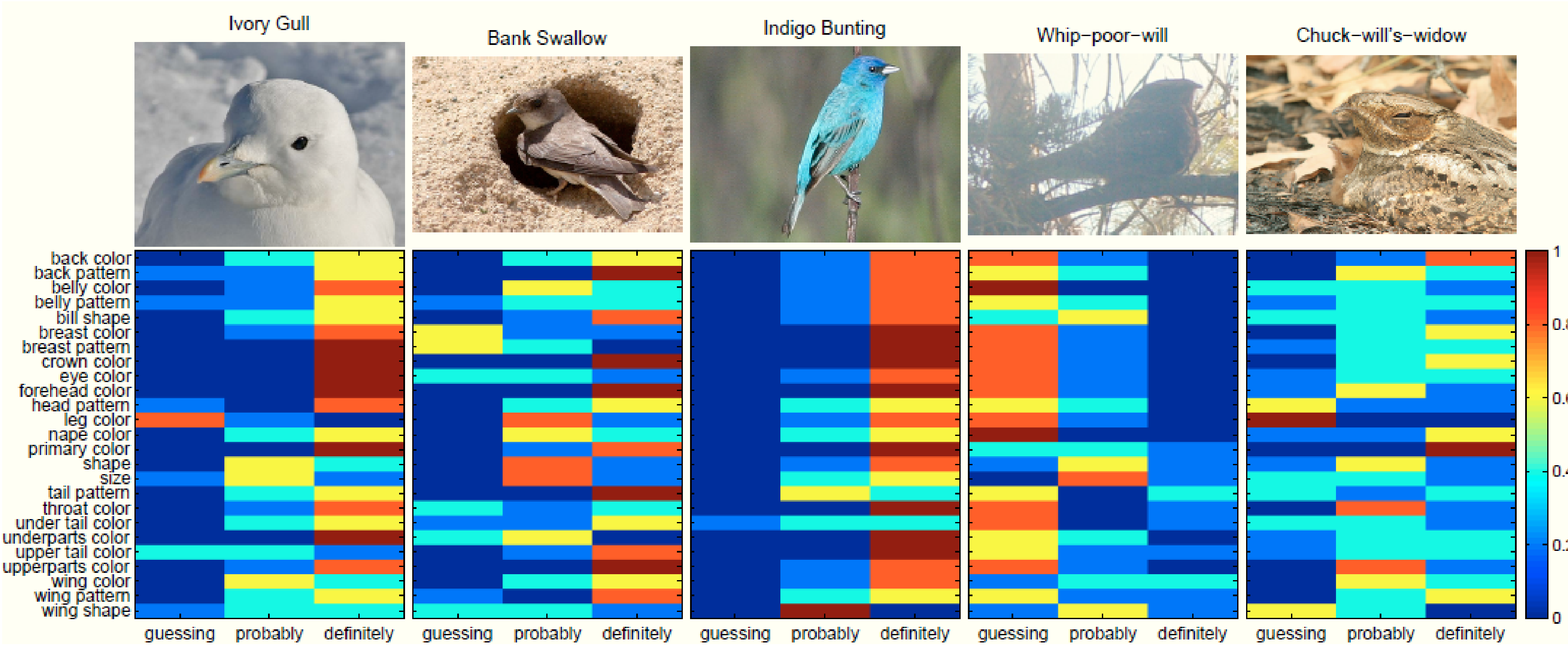
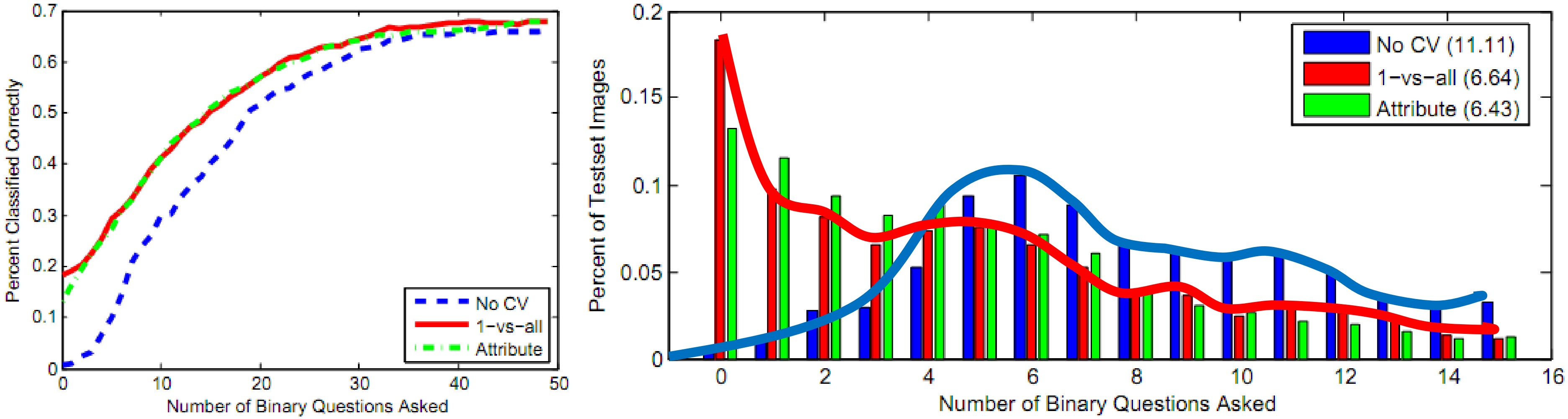


Fig. 4. Examples of user responses for each of the 25 attributes. The distribution over $\{Guessing, Probably, Definitely\}$ is color coded with blue denoting 0% and red denoting 100% of the five answers per image attribute pair.

Visual recognition

- Any vision system that can output a probability distribution across classes will work.
- Authors used Andrea Vedaldis's code.
  - Color/gray SIFT
  - VQ geometric blur
  - 1 v All SVM
- Authors added full image color histograms and VQ color histograms

Results



- Average number of questions to make ID reduced from 11.11 to 6.43
- Method allows CV to handle the easy cases, consulting with users only on the more difficult cases.

Key Observations

- Visual recognition reduces labor over a pure “20 Q” approach
- Visual recognition improves performance over a pure “20 Q” approach
  - 69% vs 66%
- User input dramatically improves recognition results
  - 66% vs 19%

Strengths and weaknesses

- Handles very difficult data and yields excellent results.
- Plug-and-play with many recognition algorithms.
- Requires significant user assistance
- Reported results assume humans are perfect verifiers
- Is the reduction from 11 questions to 6 really that significant?

Summary

- Most CAPTCHAs involve visual analysis
  - Strong comment on the state of computer vision
  - Provides interesting problems for computer vision
    - Both in designing and breaking CAPTCHAs
- Putting humans “in the loop” can simplify large scale problems
  - Amazon’s Mechanical Turk
- Combining strengths of computers and humans may be a good strategy