Crowdsourcing and Its Applications in Computer Vision

Catherine Wah

May 26, 2011

What is crowdsourcing?

"The act of taking a job traditionally performed by a designated agent … and outsourcing it to an undefined, generally **large network** of people in the form of an **open call**"

[Howe, 2009]

amazonmechanical turk
Artificial Artificial Intelligence

"Marketplace for work that requires human intelligence" (http://www.mturk.com)

- **Requesters**: post tasks called **HITs**
- **Workers ("Turkers")**: choose and select HITs to complete; payment is usually on the order of cents
- **Assignments**: total number of unique workers who can complete a specific HIT, e.g. for aggregation/ consensus

Questions as a requester

- What motivates workers to complete tasks?
- What is the best way to design a task?
- What is the best way to ensure high quality results?
- How can one use crowdsourcing in a cost effective manner?
- …

Goal: examine crowdsourcing strategies relevant to MTurk for computer vision applications

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Task incentives

- Workers may be motivated by:
  - Entertainment – games with a purpose
  - Altruism – citizen science
  - Financial reward – MTurk

Question: does increasing the rate of compensation yield better quality in results?

Higher pay ≠ better results

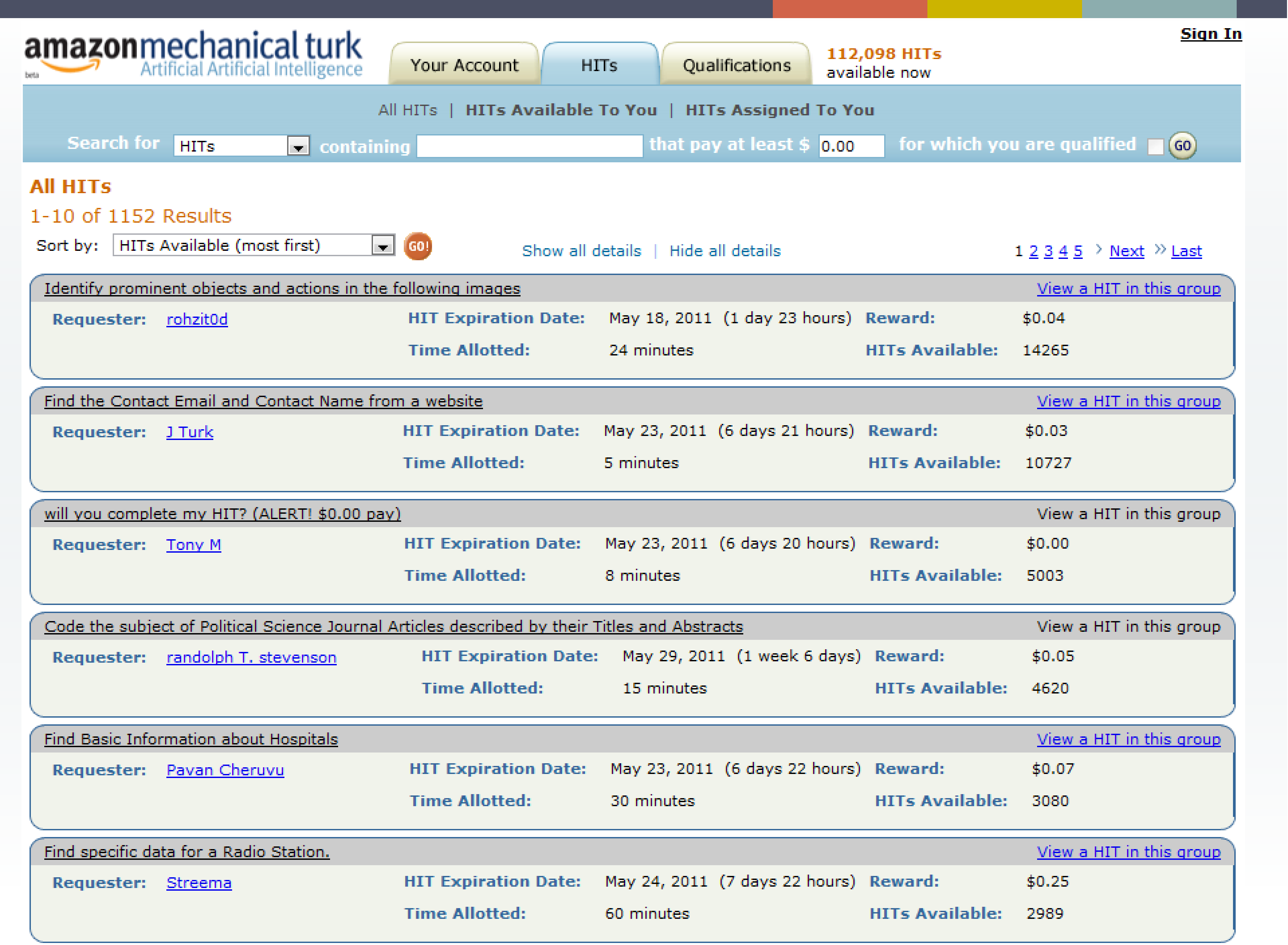
- Higher pay rate increases work *quantity* but has no effect on *quality*
  - Increase payment only in order to obtain results faster
- Workers consistently value their own work above the current pay rate
  - Paying *nothing* may yield results that are just as good

Quality of results cannot be adequately controlled using incentives.

[Mason & Watts, 2009]

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion


amazonmechanical turk
Artificial Artificial Intelligence
beta
Your Account
HITs
Qualifications
112,098 HITs
available now
Sign In
All HITs | HITs Available To You | HITs Assigned To You
Search for
HITs
containing
that pay at least $
0.00
for which you are qualified
GO
All HITs
1-10 of 1152 Results
Sort by:
HITs Available (most first)
GO!
Show all details | Hide all details
1 2 3 4 5 Next Last
Identify prominent objects and actions in the following images
View a HIT in this group
Requester: rohzit0d
HIT Expiration Date: May 18, 2011 (1 day 23 hours)
Reward: $0.04
Time Allotted: 24 minutes
HITs Available: 14265
Find the Contact Email and Contact Name from a website
View a HIT in this group
Requester: J Turk
HIT Expiration Date: May 23, 2011 (6 days 21 hours)
Reward: $0.03
Time Allotted: 5 minutes
HITs Available: 10727
will you complete my HIT? (ALERT! $0.00 pay)
View a HIT in this group
Requester: Tony M
HIT Expiration Date: May 23, 2011 (6 days 20 hours)
Reward: $0.00
Time Allotted: 8 minutes
HITs Available: 5003
Code the subject of Political Science Journal Articles described by their Titles and Abstracts
View a HIT in this group
Requester: randolph T. stevenson
HIT Expiration Date: May 29, 2011 (1 week 6 days)
Reward: $0.05
Time Allotted: 15 minutes
HITs Available: 4620
Find Basic Information about Hospitals
View a HIT in this group
Requester: Pavan Cheruvu
HIT Expiration Date: May 23, 2011 (6 days 22 hours)
Reward: $0.07
Time Allotted: 30 minutes
HITs Available: 3080
Find specific data for a Radio Station.
View a HIT in this group
Requester: Streema
HIT Expiration Date: May 24, 2011 (7 days 22 hours)
Reward: $0.25
Time Allotted: 60 minutes
HITs Available: 2989

amazon mechanical turk beta
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

114,544 HITs available now

Sign In

All HITs | **HITs Available To You** | **HITs Assigned To You**

Search for [HITs] containing [image labeling] that pay at least $ [0.00] for which you are qualified [ ] GO

HITs containing 'image labeling'

1-1 of 1 Results

Sort by: [HITs Available (most first)] GO!

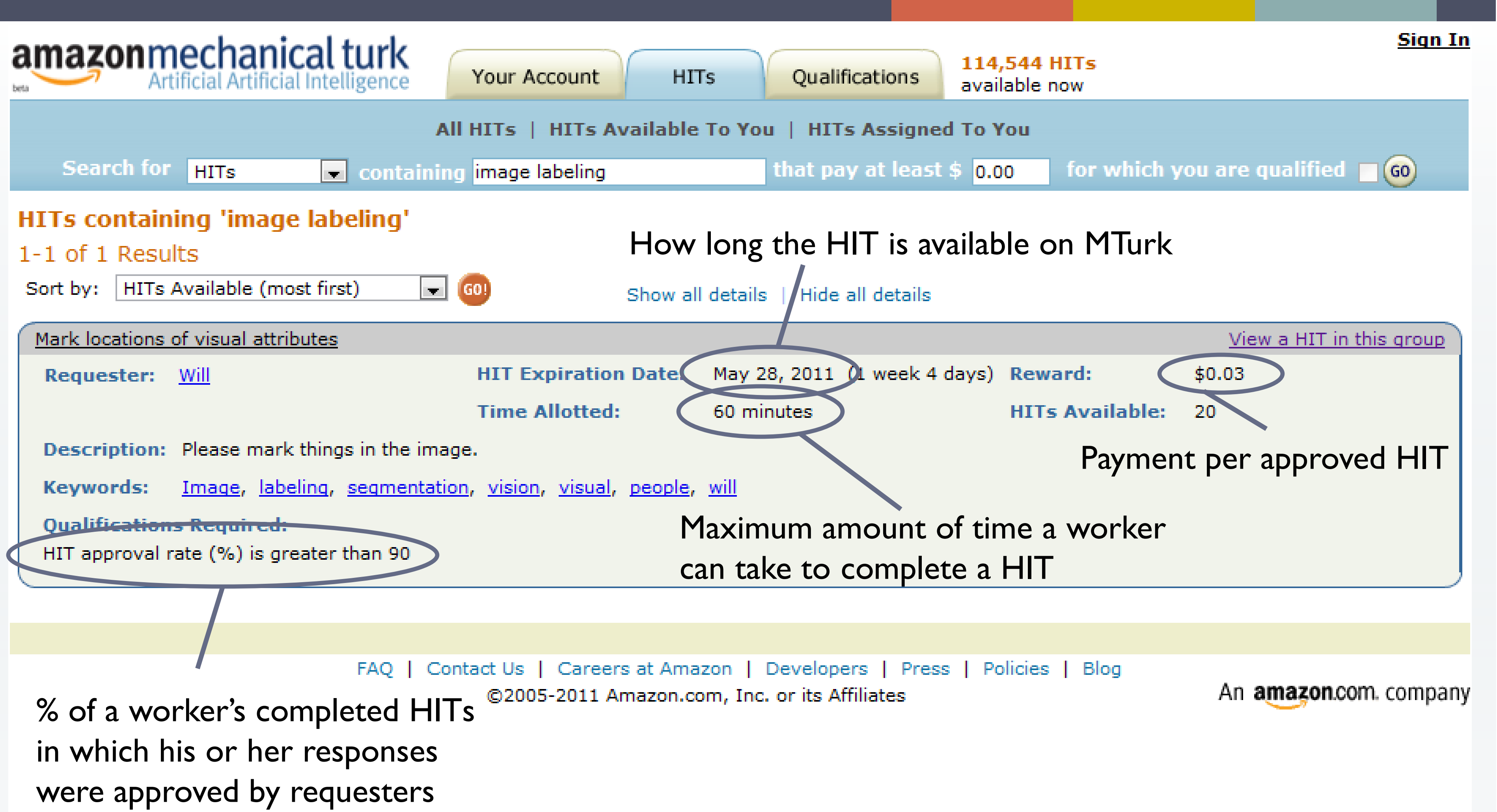
Show all details | Hide all details

Mark locations of visual attributes — View a HIT in this group

Requester:	Will	**HIT Expiration Date:**	May 28, 2011 (1 week 4 days)	**Reward:**	$0.03
		Time Allotted:	60 minutes	**HITs Available:**	20

FAQ | Contact Us | Careers at Amazon | Developers | Press | Policies | Blog

An amazon.com company


amazonmechanical turk
Artificial Artificial Intelligence
beta
Your Account
HITs
Qualifications
114,544 HITs
available now
Sign In
All HITs | HITs Available To You | HITs Assigned To You
Search for HITs containing image labeling that pay at least $ 0.00 for which you are qualified GO
HITs containing 'image labeling'
1-1 of 1 Results
Sort by: HITs Available (most first) GO!
How long the HIT is available on MTurk
Show all details | Hide all details
Mark locations of visual attributes
View a HIT in this group
Requester: Will
HIT Expiration Date: May 28, 2011 (1 week 4 days)
Reward: $0.03
Time Allotted: 60 minutes
HITs Available: 20
Payment per approved HIT
Description: Please mark things in the image.
Keywords: Image, labeling, segmentation, vision, visual, people, will
Qualifications Required:
HIT approval rate (%) is greater than 90
Maximum amount of time a worker
can take to complete a HIT
FAQ | Contact Us | Careers at Amazon | Developers | Press | Policies | Blog
©2005-2011 Amazon.com, Inc. or its Affiliates
An amazon.com. company
% of a worker's completed HITs
in which his or her responses
were approved by requesters

amazon mechanical turk
Artificial Artificial Intelligence

Your Account | HITs | Qualifications

114,977 HITs available now

Sign In

All HITs | HITs Available To You | HITs Assigned To You

Search for HITs containing that pay at least $ 0.00 for which you are qualified GO

Timer: 00:00:00 of 60 minutes

Want to work on this HIT? Accept HIT

Want to see other HITs? Skip HIT

Total Earned: Unavailable
Total HITs Submitted: 0

Mark locations of visual attributes

Requester: Will **Reward:** $0.03 per HIT **HITs Available:** 20 **Duration:** 60 minutes

Qualifications Required: HIT approval rate (%) is greater than 90

IMPORTANT: Read the instructions!!

(q) camcorder

IMPORTANT: Read the instructions!!

(q) camcorder

Image by: (loading) License: loading

[Sorokin and Forsyth, 2008]

Alexander Sorokin, University of Illinois, v0.1.8, http://code.google.com/p/cv-web-annotation-toolkit/

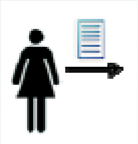
Task parameters

- Time restrictions
- Compensation – rewards and bonuses
- Worker qualifications
  - Geographic constraints
  - Reputation (i.e. , approval rating)
  - Custom qualification tests
- Task interface – visual appearance and layout

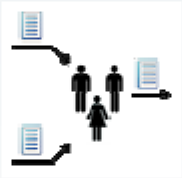
More complex tasks necessitate the division of labor into multiple sets or sequences of tasks.

Human computation processes

- Different steps in a computational process are outsourced to humans
- Little et al. categorize these into two types:



  - **Creation tasks** solicit new content



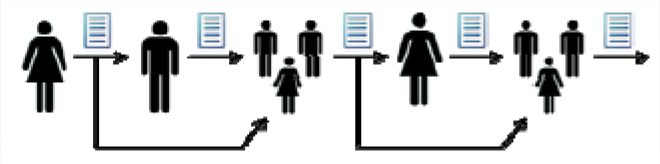
  - **Decision tasks** solicit opinions about existing content
    - Comparison
    - Rating

[Little et al., 2010]

Combining tasks

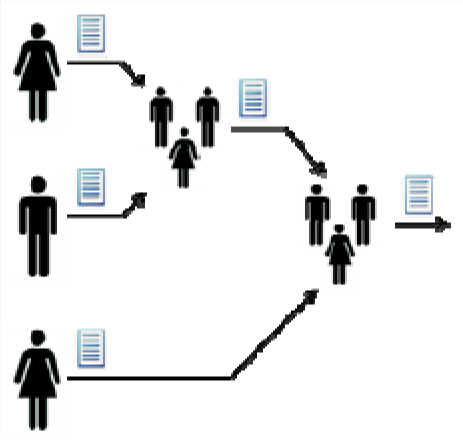
Iterative pattern:

sequence of creation tasks, with a comparison task in between



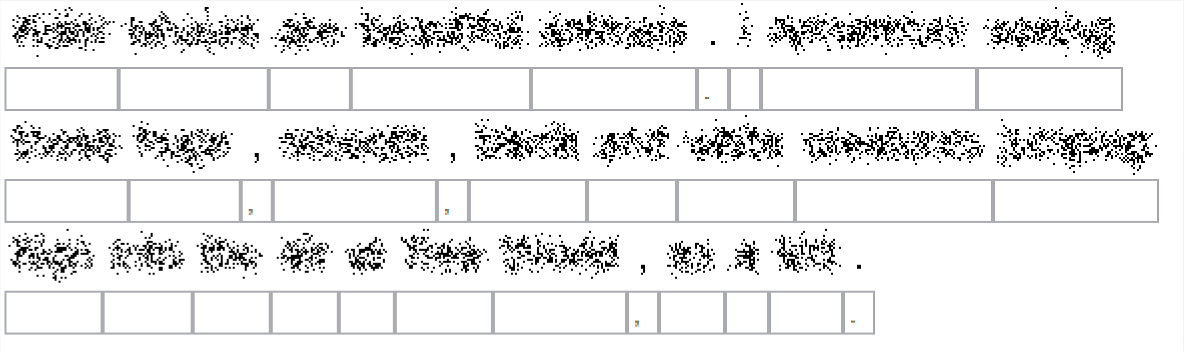
Parallel pattern:

set of creation tasks in parallel, merged with a sequence of comparison tasks



[Little et al., 2010]

Example: transcription



"Killer whales are beautiful animals. I remember seeing these huge, smooth, black and white creatures jumping high into the air at Sea World, as a kid."

[Little et al., 2010]

Iterative vs. parallel?

- Iterative processes tend to yield *higher quality* than parallel processes
- However, results have *lower variance*
  - Showing Turkers previous work can negatively affect quality by leading them down the wrong path

Other workflow strategies may be needed for certain tasks.

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Using heuristics

- Keeping track of workers ("**experts**") with consistently high quality results
- **"Gold standard"** – questions for which answers are known
- **Majority vote** – averaging multiple labels from non-experts

If labels are noisy, repeated labeling improves quality but requires more labels (higher cost).

The human annotation process

- Modeling various aspects of annotation:
  - *Worker competency* – accuracy in labeling
  - *Worker expertise* – better at labeling some things than others, based on their strengths
  - *Worker bias* – how one weighs errors
  - *Task difficulty* – ambiguous images are universally hard to label
  - *True label* – the ground truth label value
- We focus on the model described by Welinder et al.

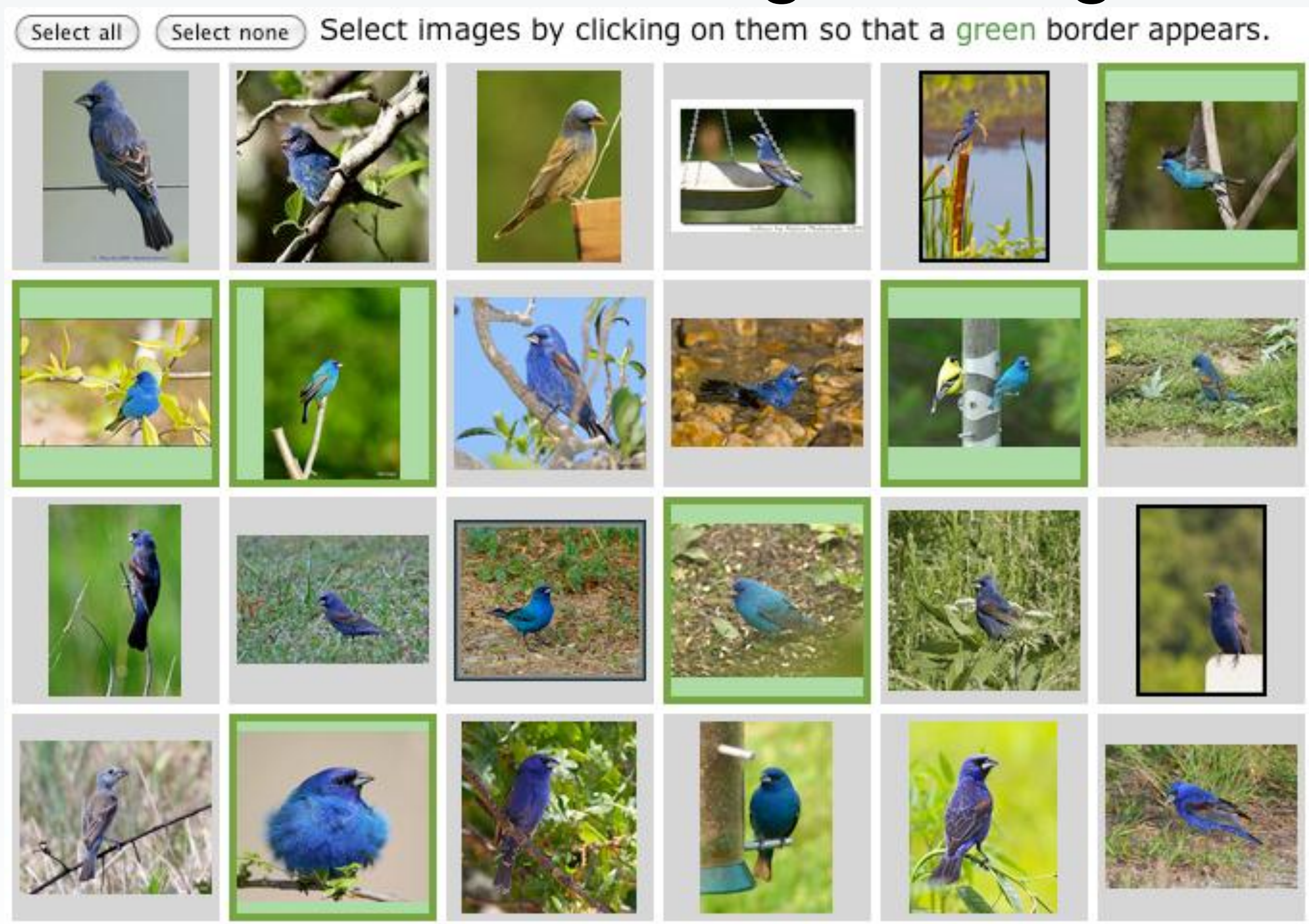
[Welinder et al., 2010]

Types of annotator errors

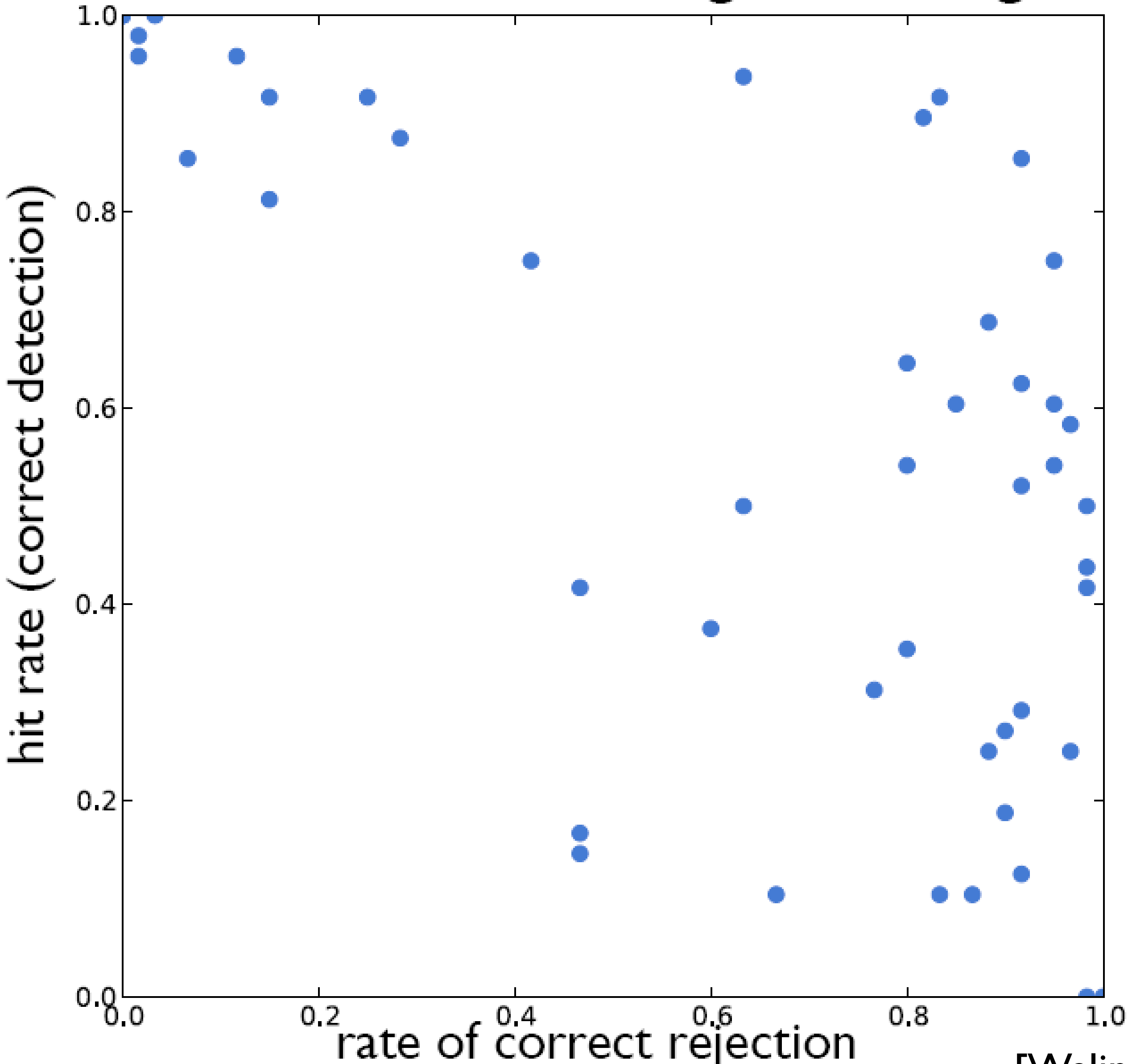
Task: Find the Indigo Bunting

Indigo Bunting

Blue Grosbeak

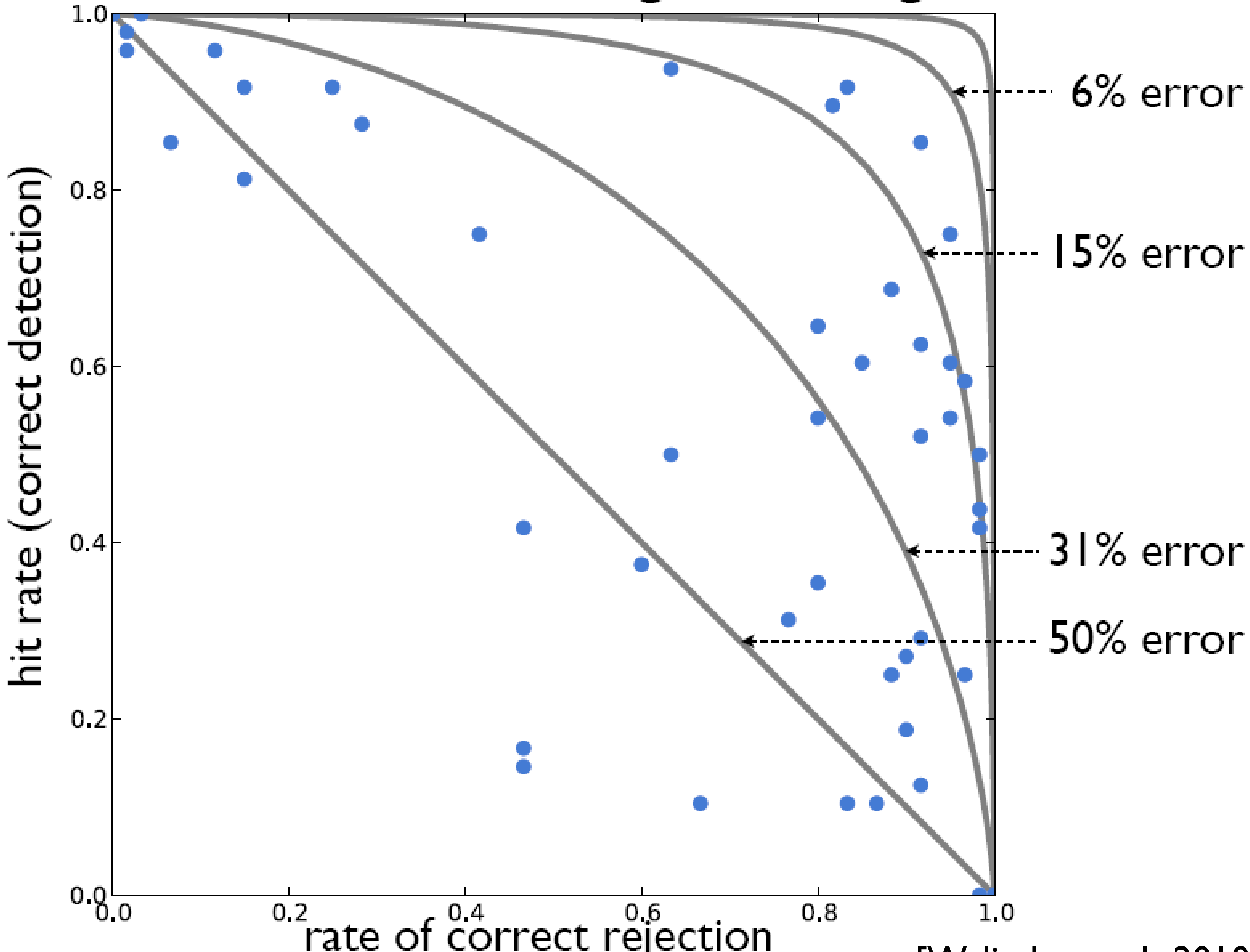


Task: Find the Indigo Bunting



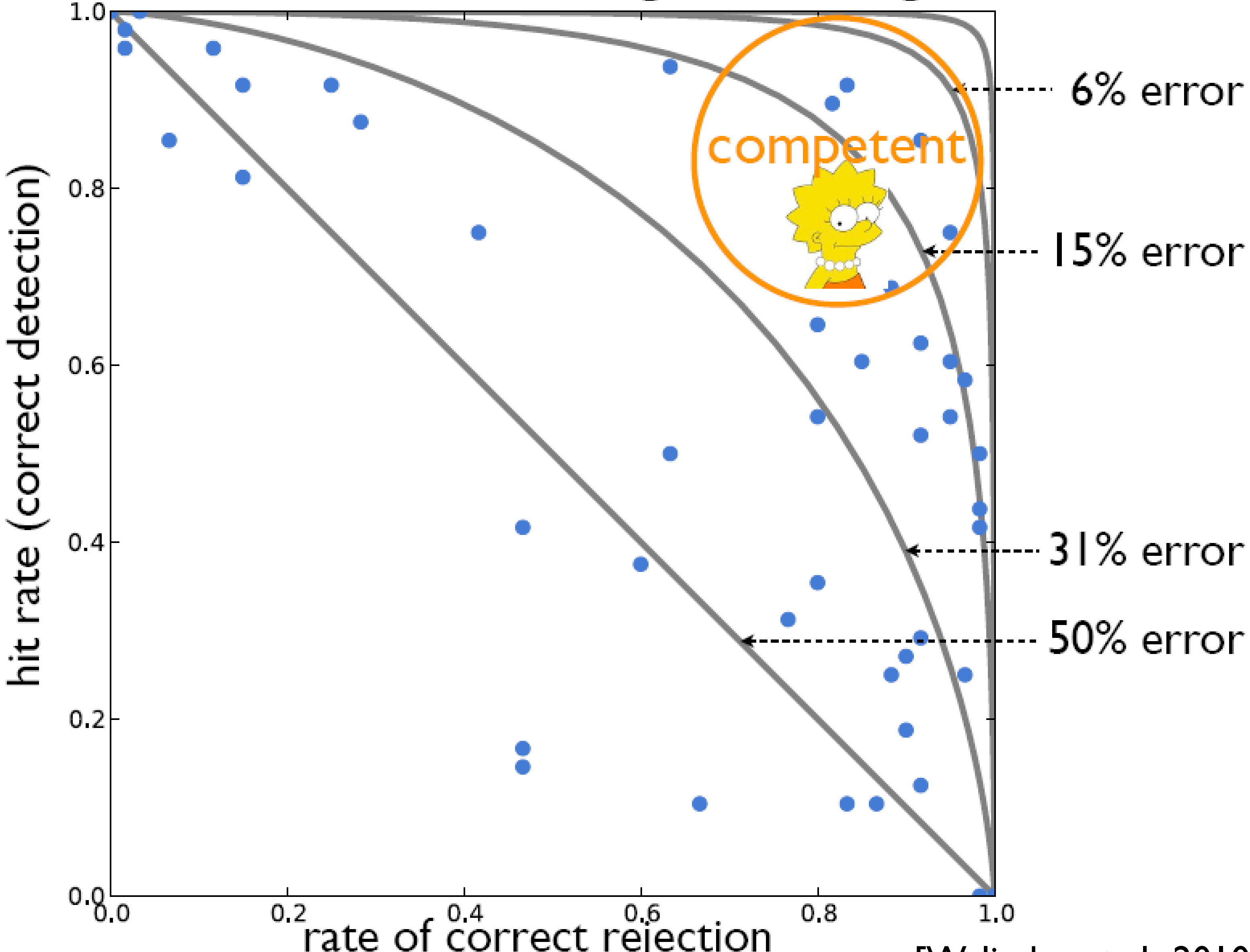
[Welinder et al., 2010]

Task: Find the Indigo Bunting



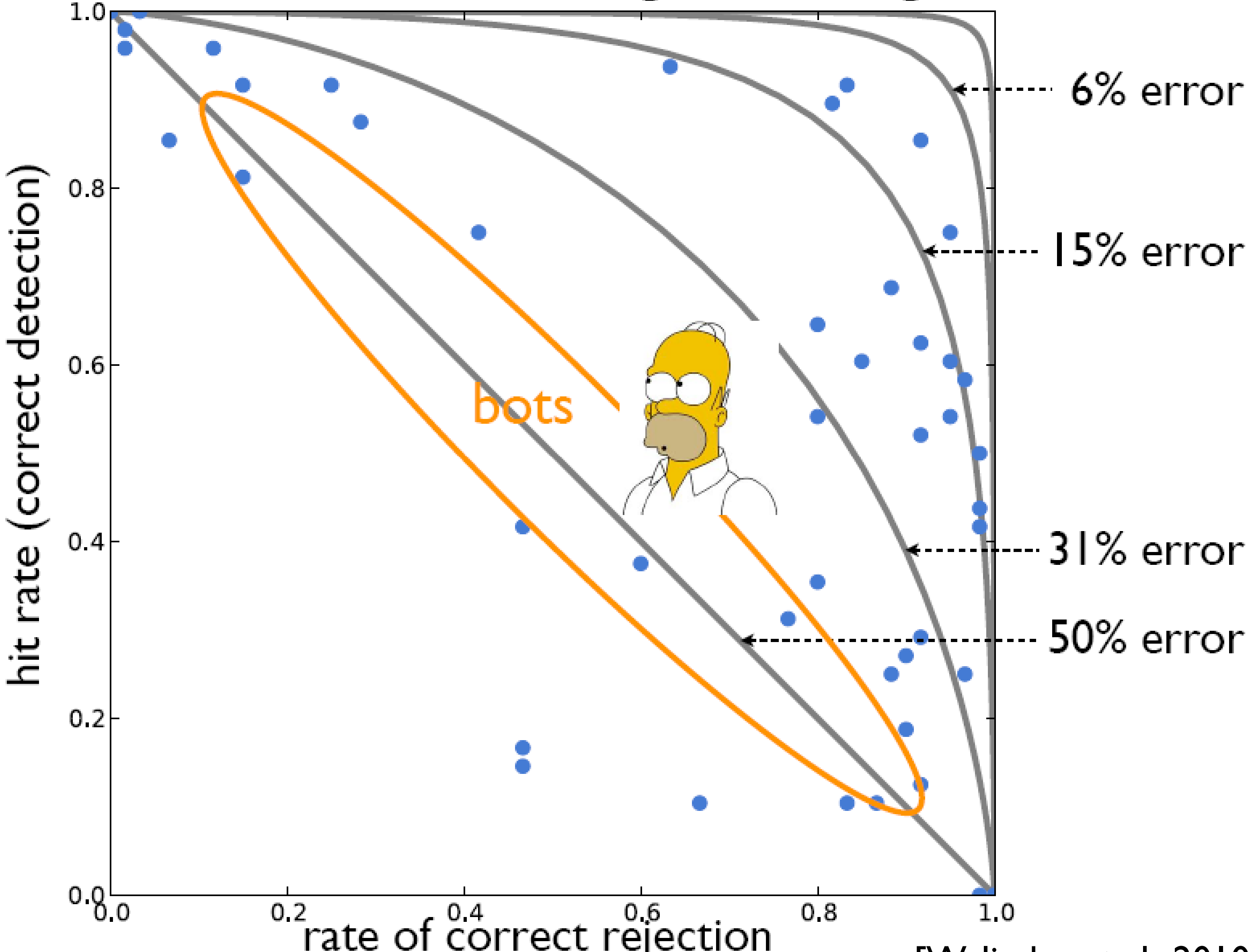
[Welinder et al., 2010]

Task: Find the Indigo Bunting



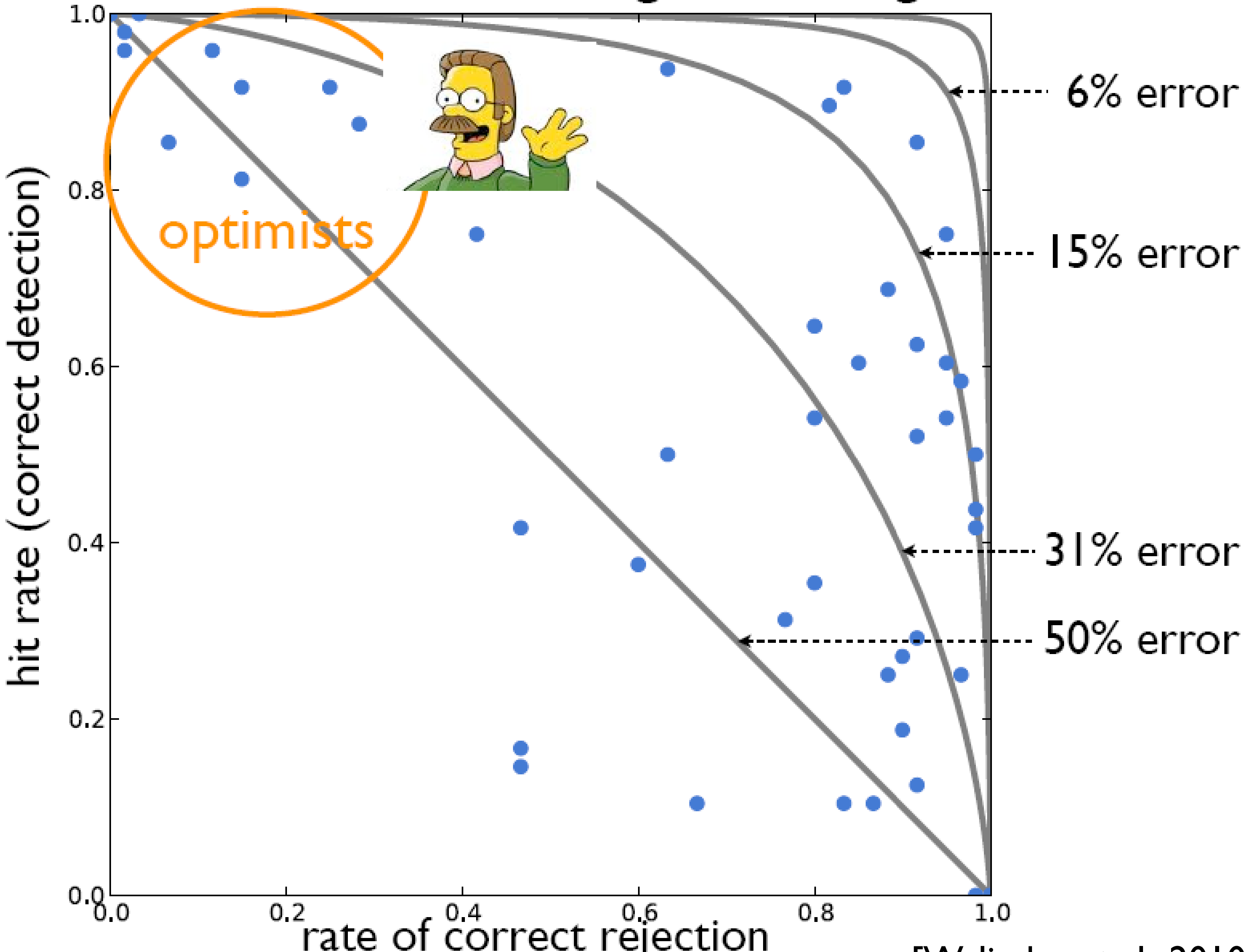
[Welinder et al., 2010]

Task: Find the Indigo Bunting



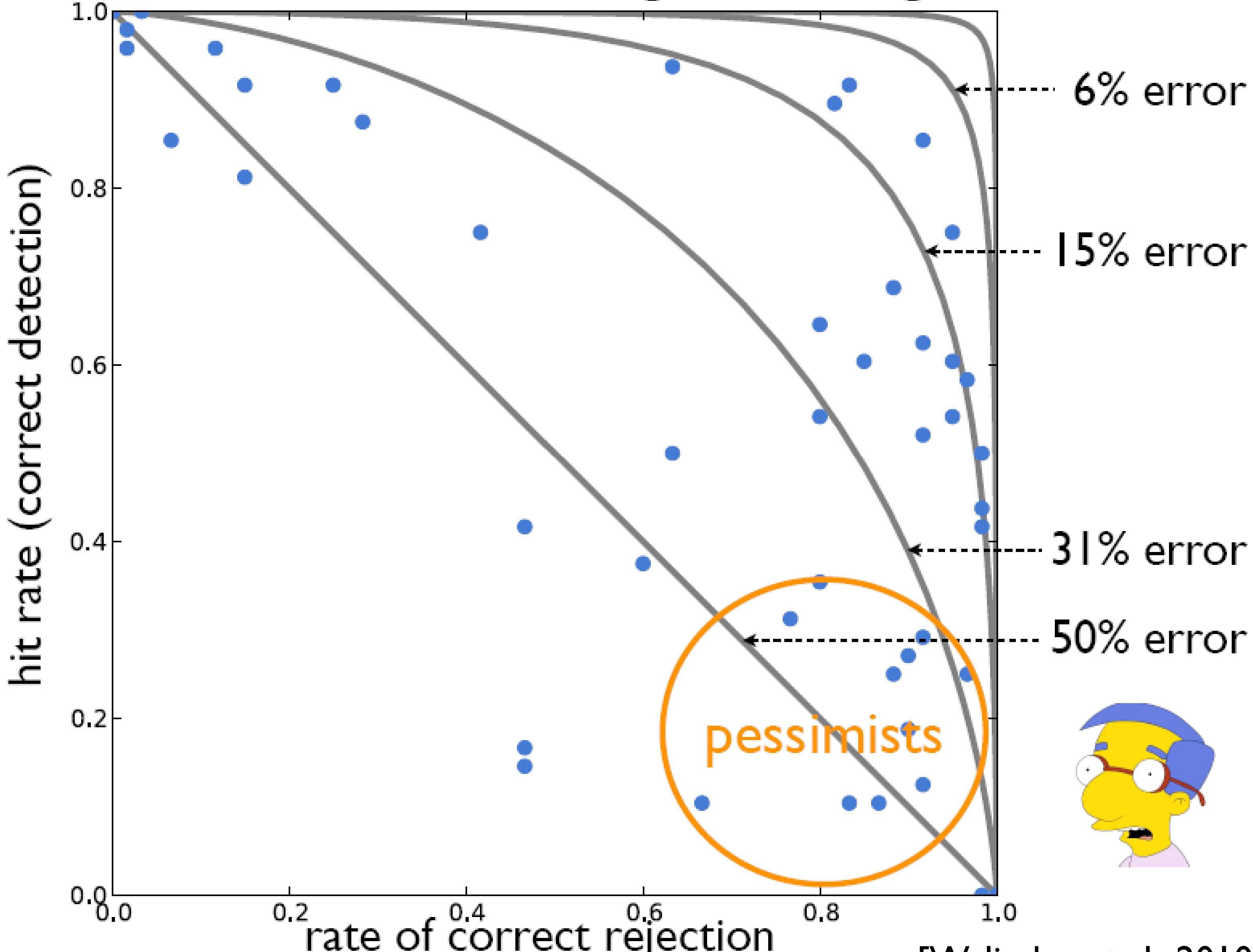
[Welinder et al., 2010]

Task: Find the Indigo Bunting



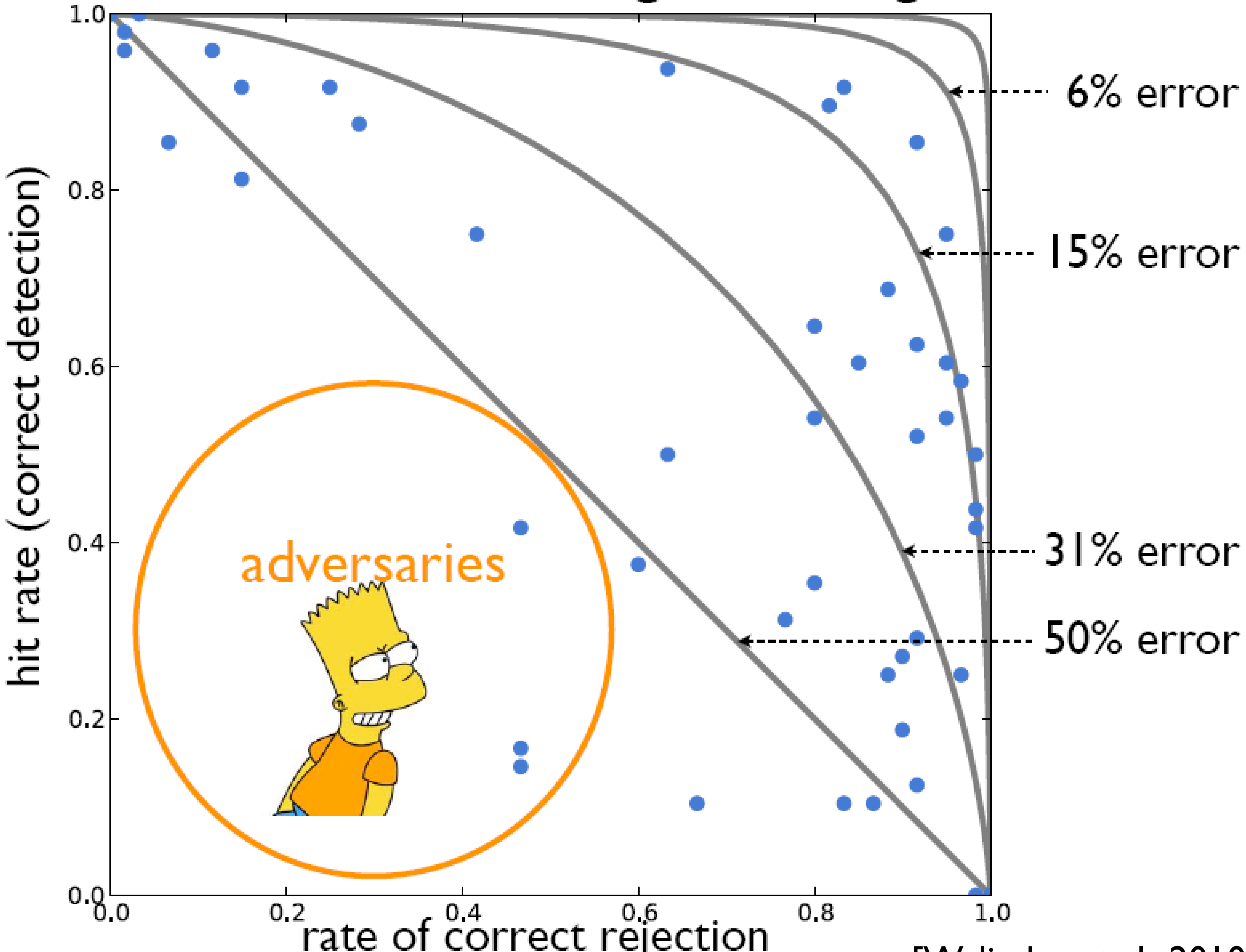
[Welinder et al., 2010]

Task: Find the Indigo Bunting



[Welinder et al., 2010]

Task: Find the Indigo Bunting

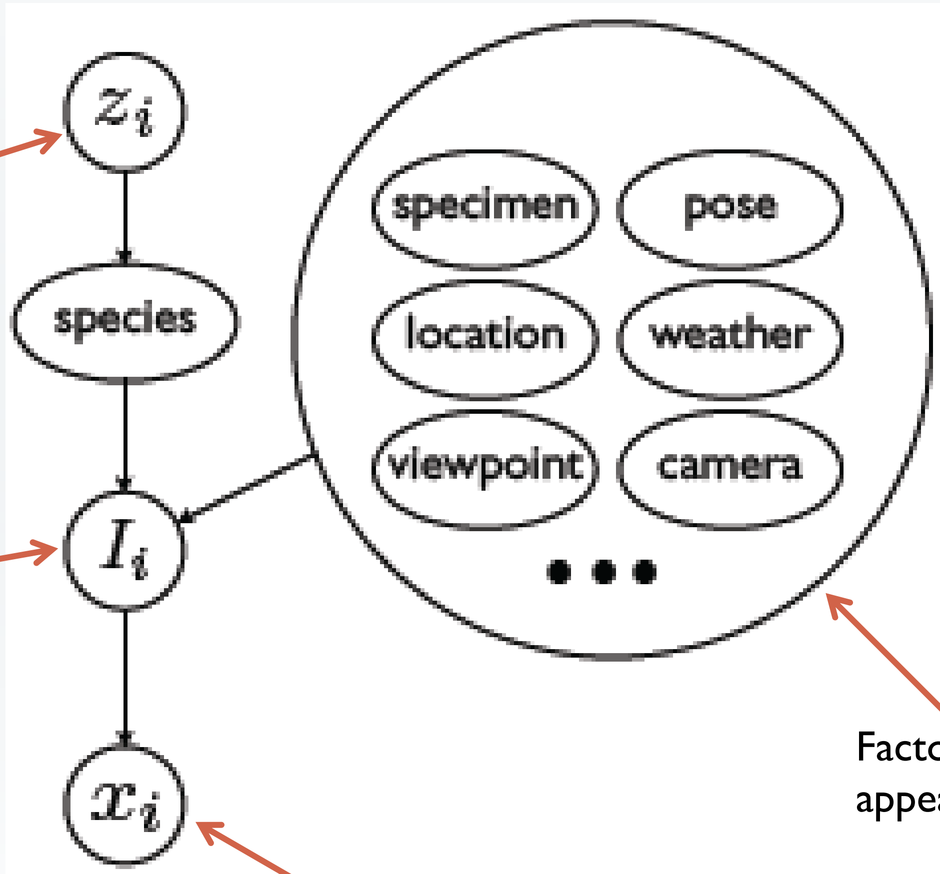


[Welinder et al., 2010]

Image formation process

Object presence or absence

Image i



Factors influencing appearance

Signal seen by ideal observer

Entire annotation process

$$p(\mathcal{L}, x, w, \tau) = \prod_{j=1}^{M} p(\tau_j|\gamma) p(w_j|\alpha) \prod_{i=1}^{N} \left(p(x_i|\theta_z, \beta) \prod_{j \in \mathcal{J}_i} p(l_{ij}|x_i, w_j, \tau_j) \right)$$

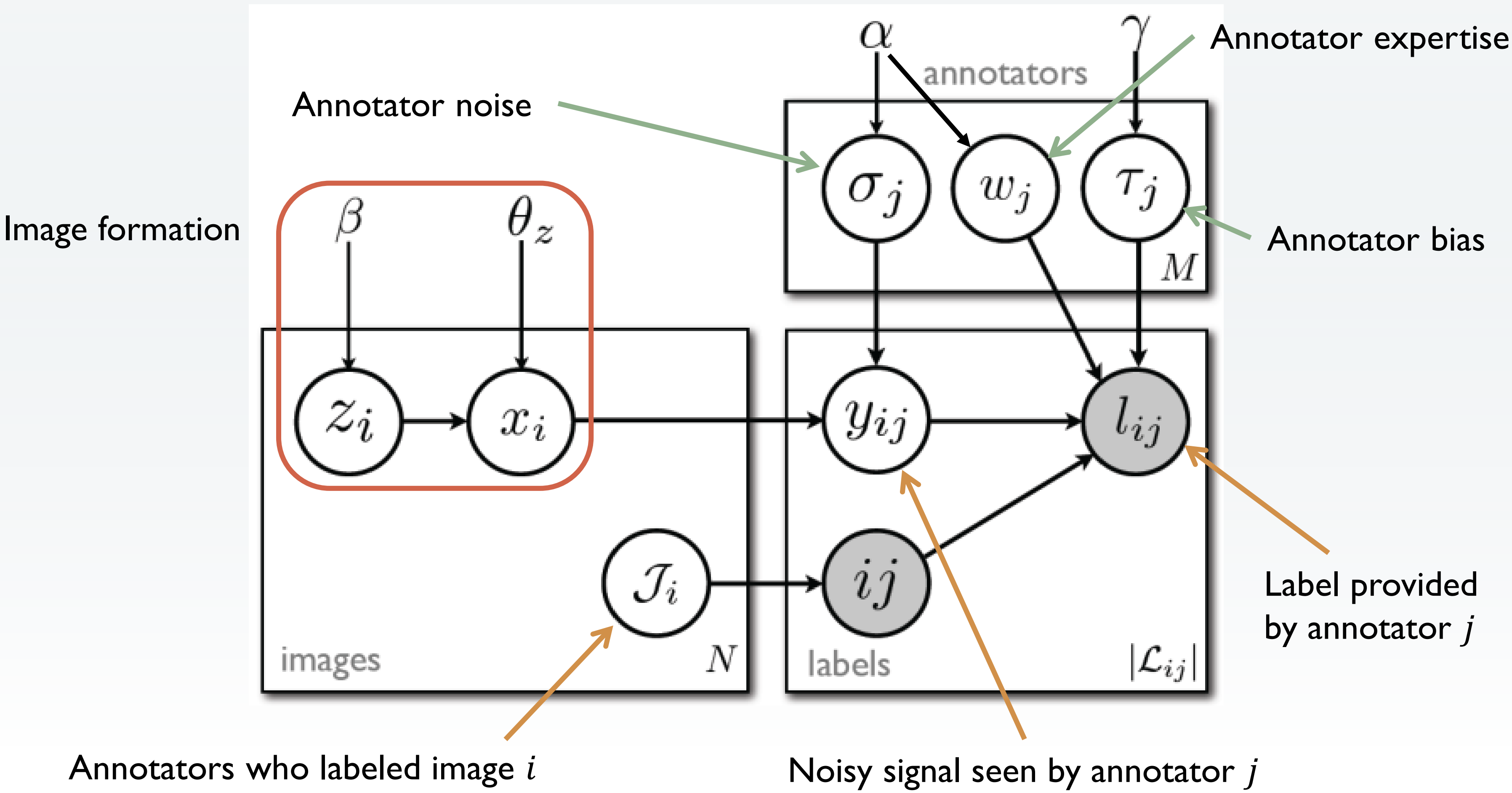
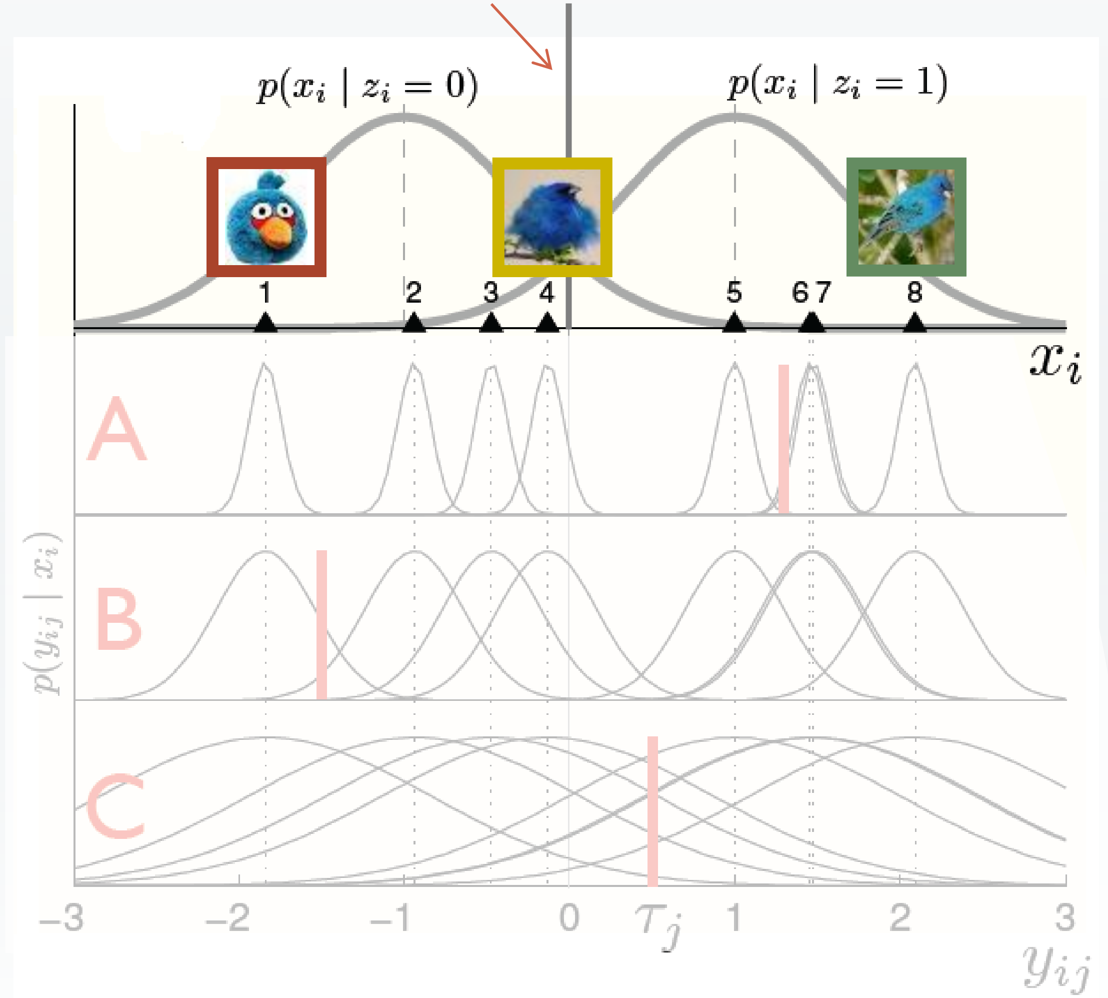
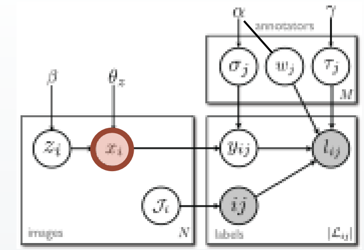


Image difficulty

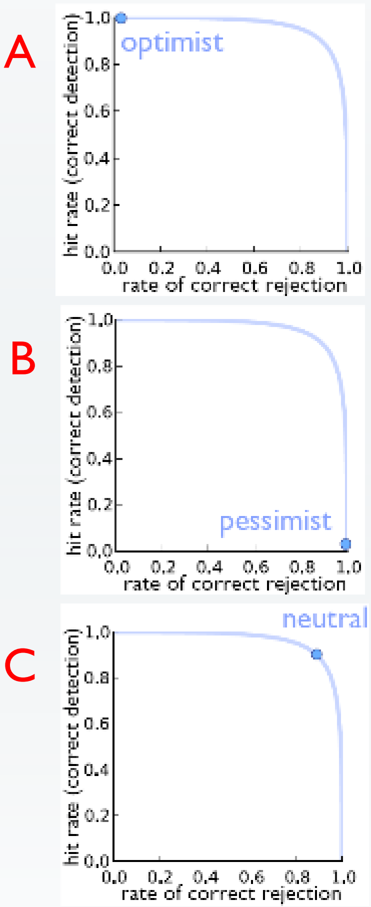
Ground truth decision plane at $x_i = 0$

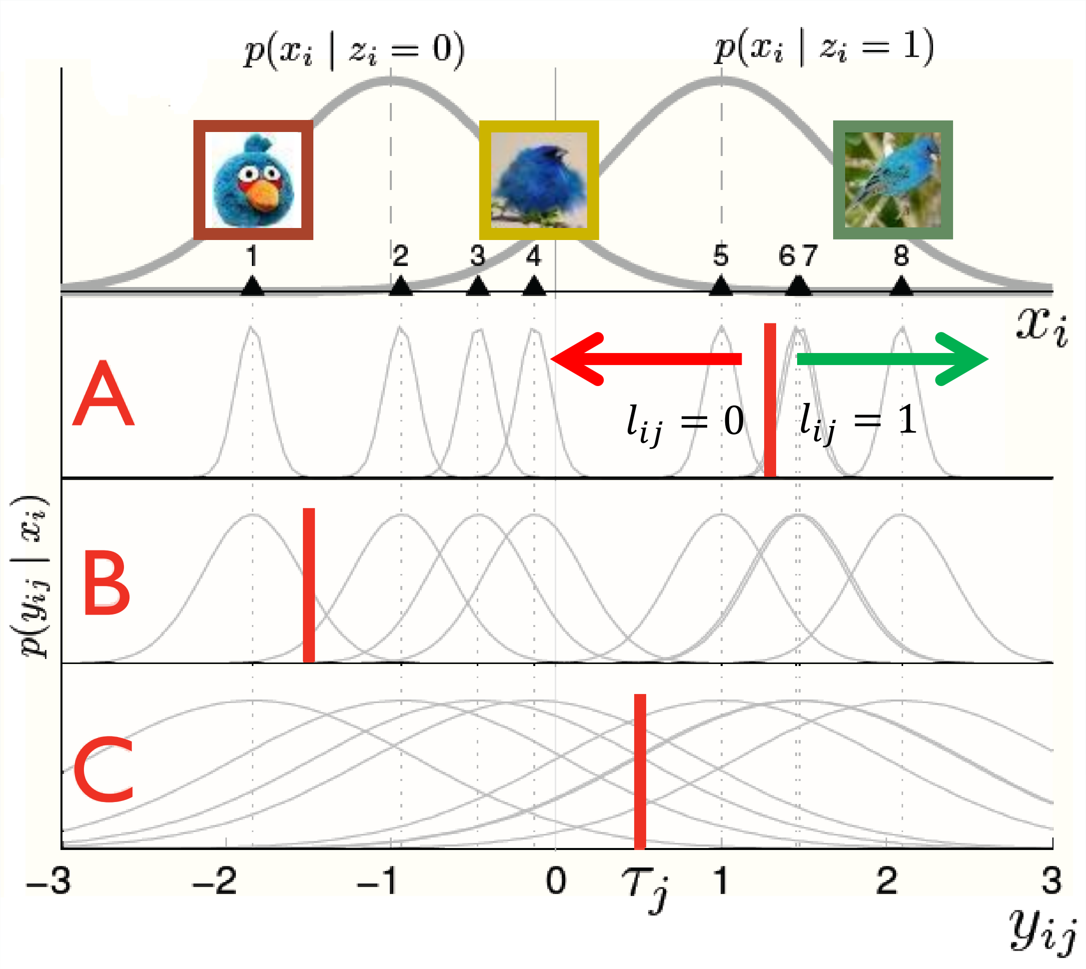


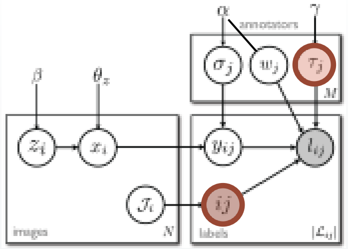


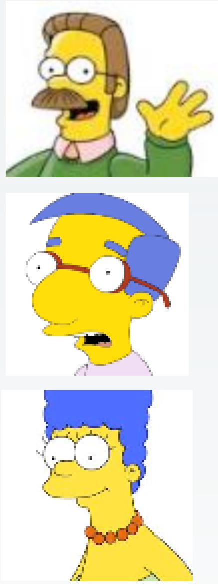
[Welinder et al., 2010]

Annotator bias



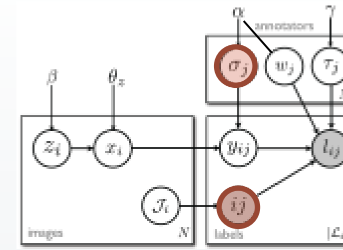


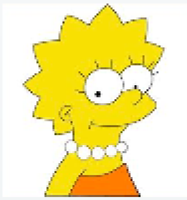




[Welinder et al., 2010]

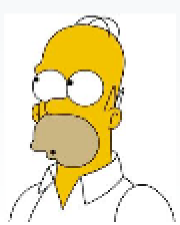
Annotator competence

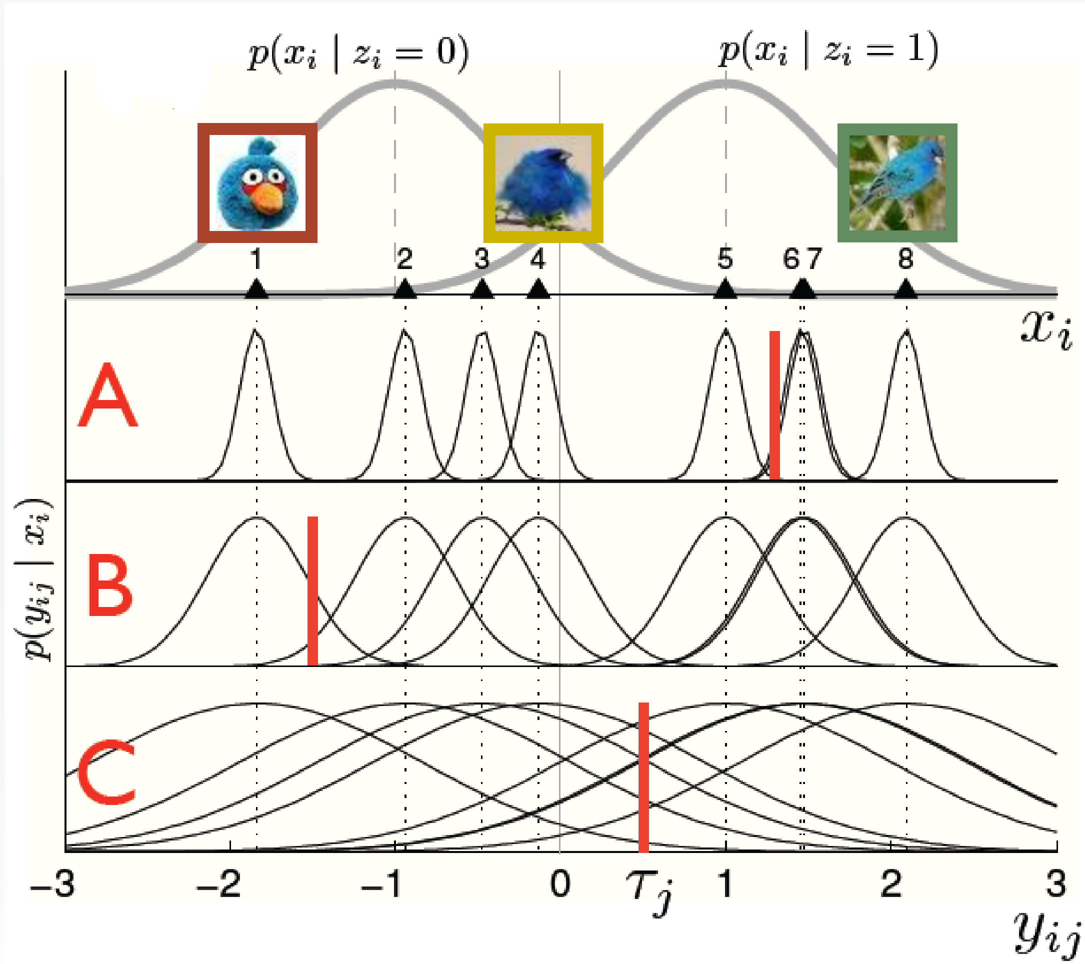




High competence

Low competence



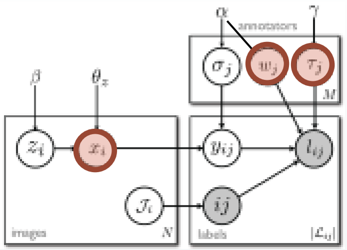


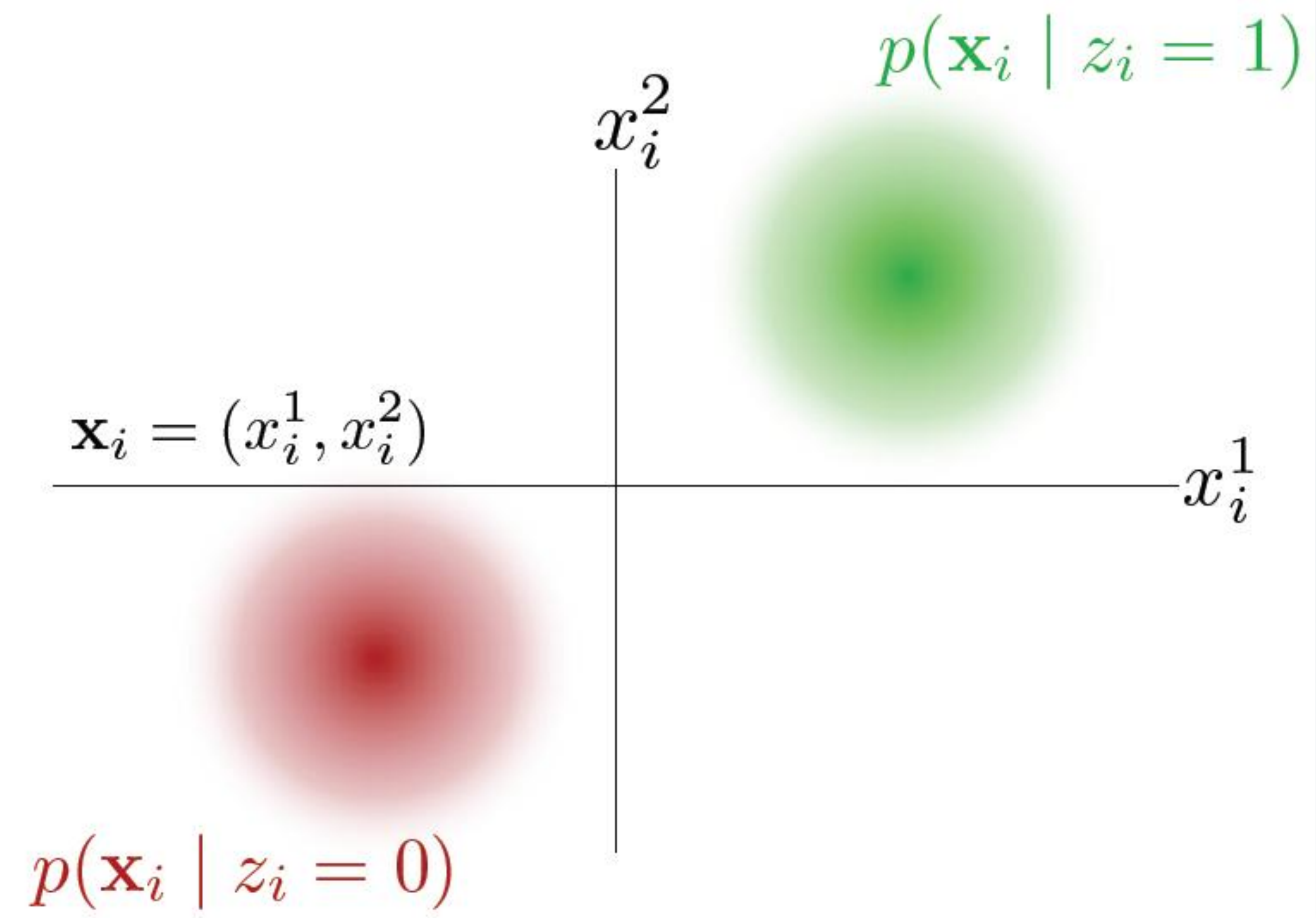
Low σ_j

High σ_j

[Welinder et al., 2010]

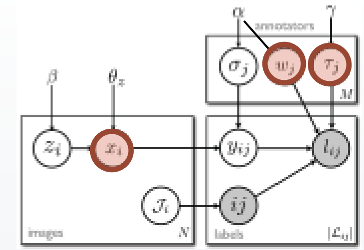
Multidimensional ability of annotators

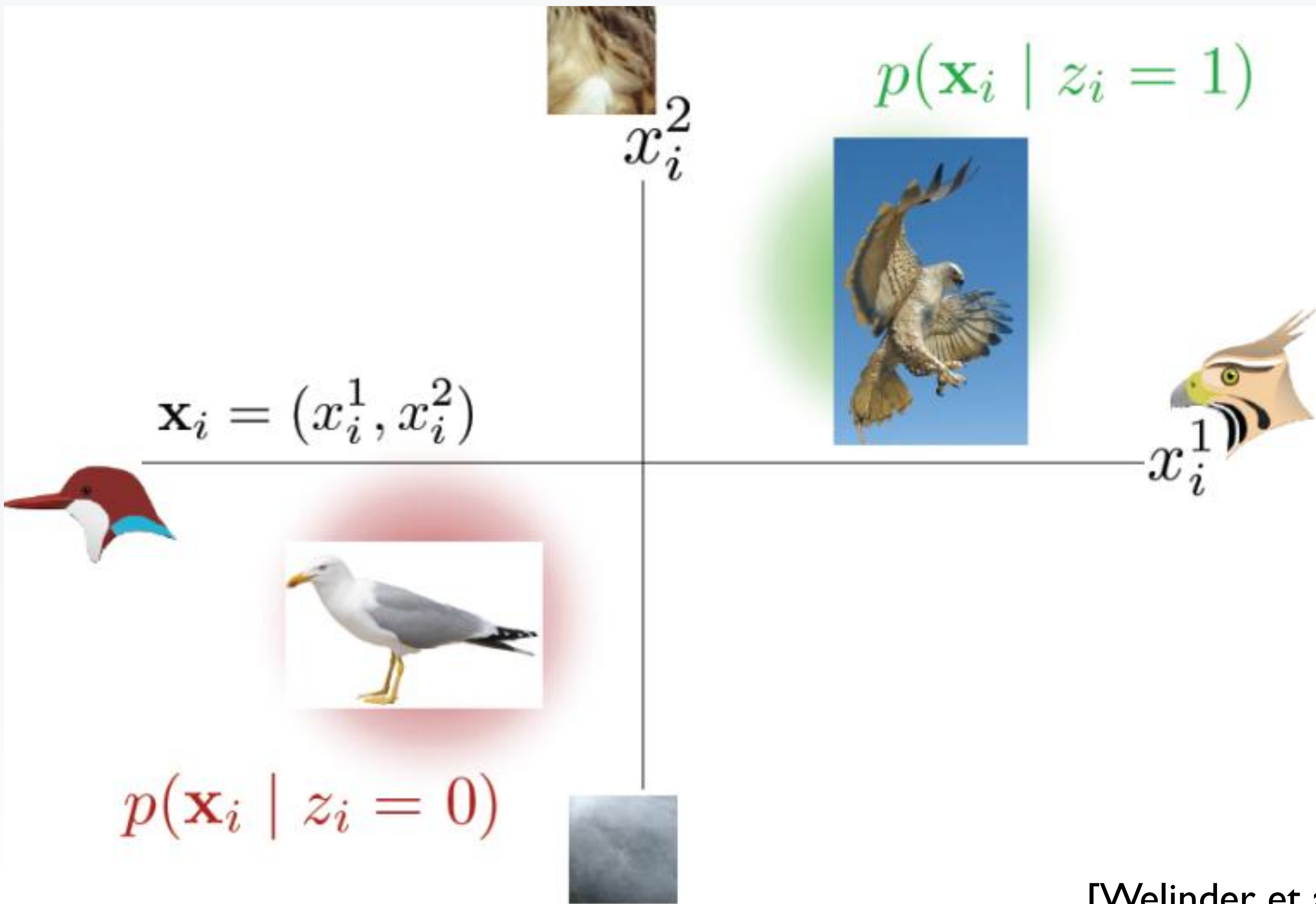




[Welinder et al., 2010]

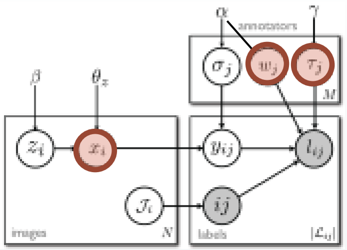
Multidimensional ability of annotators

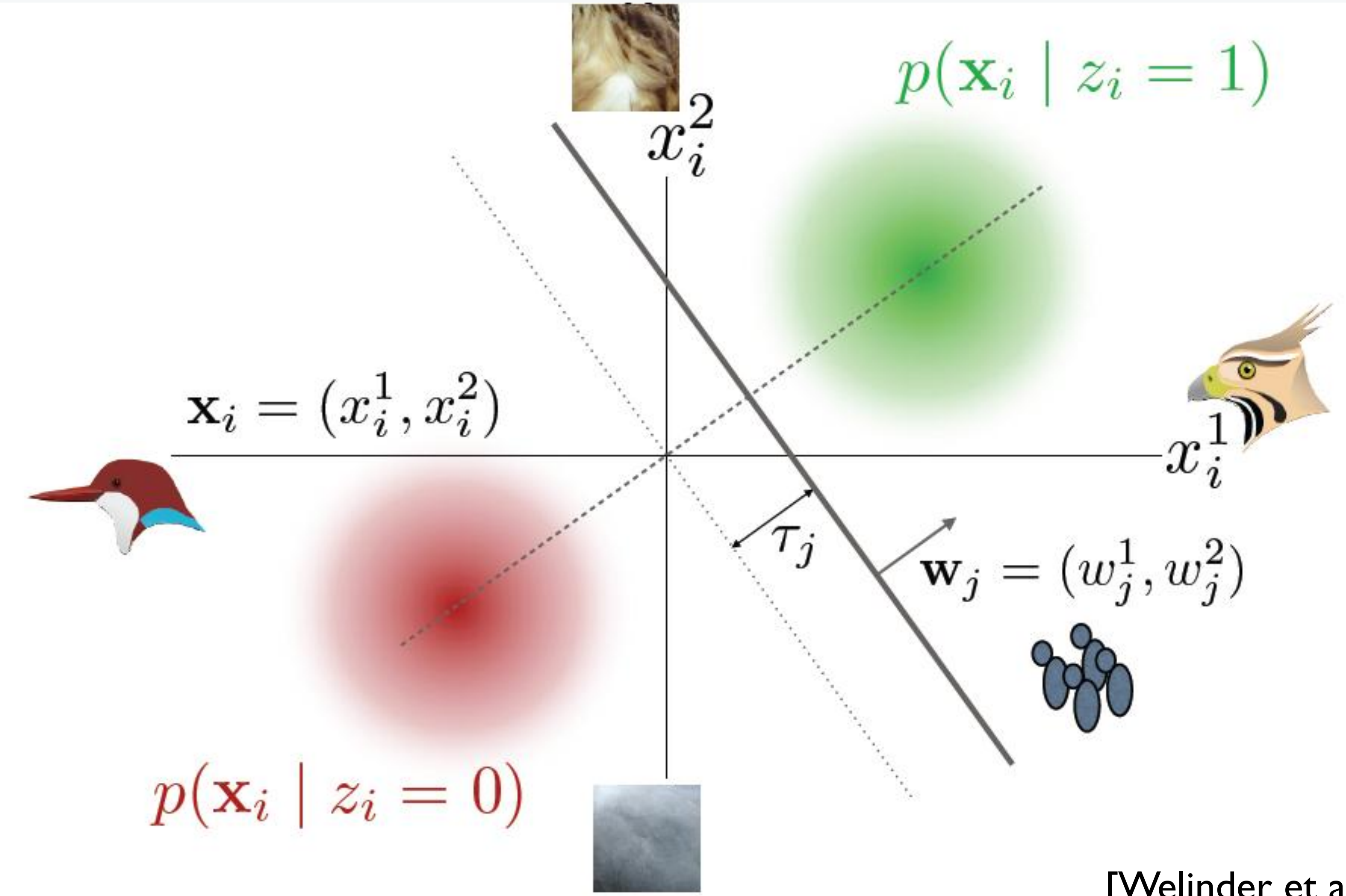




[Welinder et al., 2010]

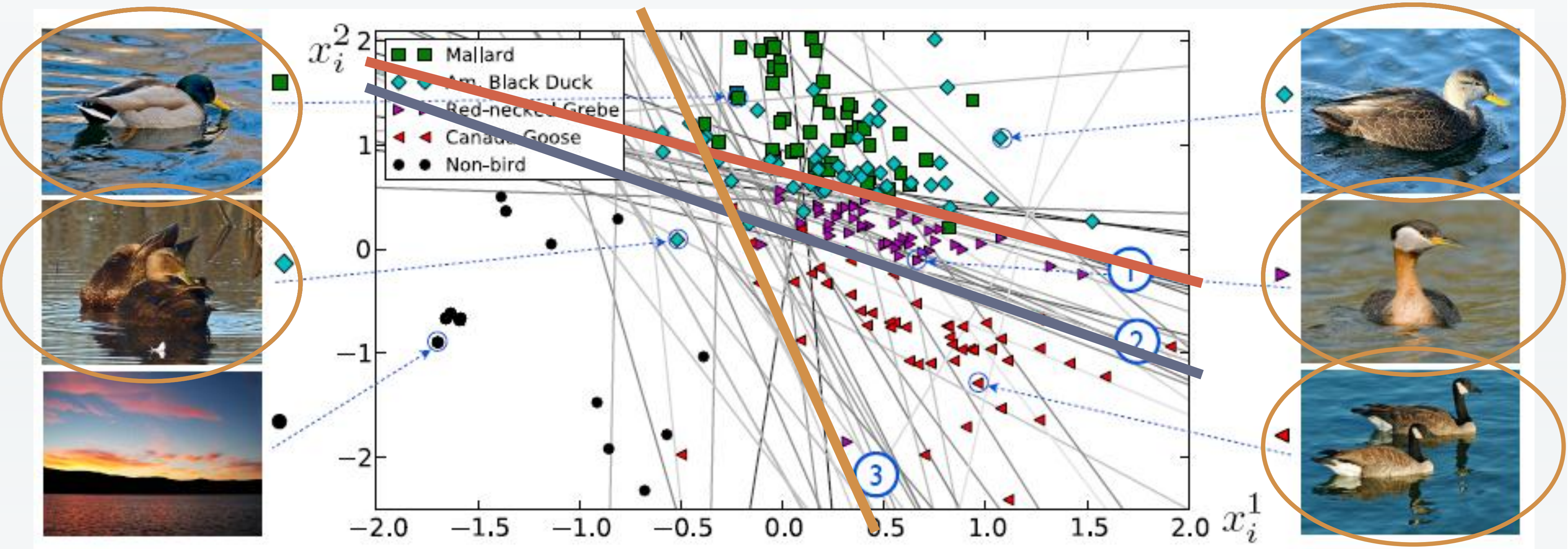
Multidimensional ability of annotators





[Welinder et al., 2010]

Worker "schools of thought"



Ducks

Ducks and grebes

Ducks, grebes, and geese

[Welinder et al., 2010]

Discussion: quality management

- Models can capture multidimensionality of annotation process
- How well does this generalize to continuous annotations?

Different tasks require different usages of reviewing strategies.

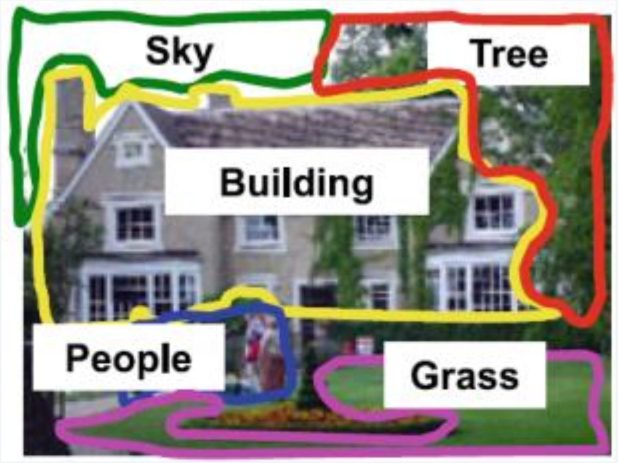
Predicting quality accurately can reduce the number of labels needed.

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Obtaining labels cost effectively

- Use only reliable Turkers
- Estimate number of assignments to request based on current labels



Question: What about different types of labels or multiple objects/image?

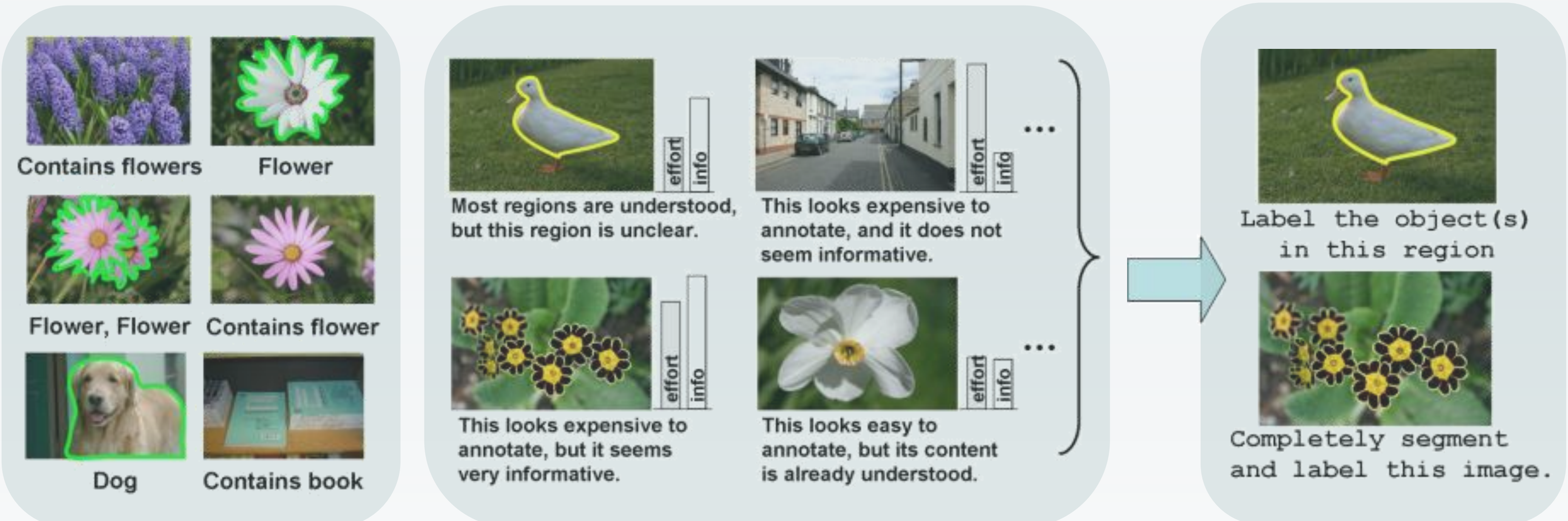
Objective: request the most promising annotations and use them to update a classifier

- Tradeoff between informativeness and manual effort

[Vijayanarasimhan & Grauman, 2010]

Objective: request the most promising annotations and use them to update a classifier

- Object categories learned from weak and strong labels
- Request annotations
- Update classifier

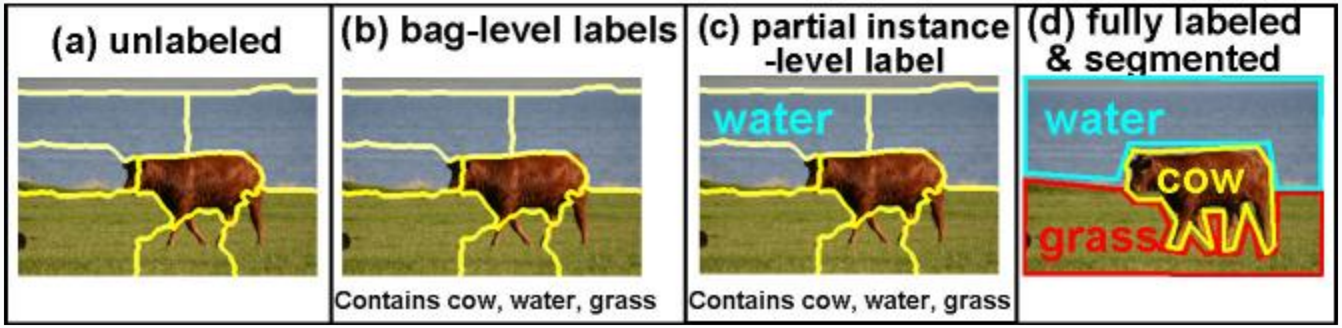


- Consider unlabeled and partially labeled images
- For each candidate annotation, predict tradeoff between informativeness and cost

[Vijayanarasimhan & Grauman, 2010]

MIML framework

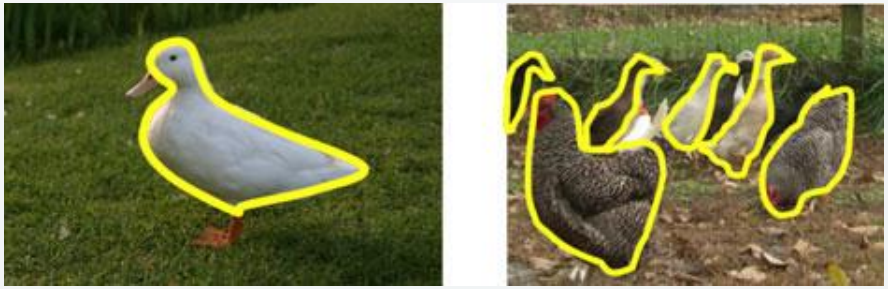
- Standard MIL: bags of instances
  - Positive bags – at least one positive instance
  - Negative bags – no positive instances
- **Multiple instance multi-label (MIML) learning:**
  - Each instance (**oversegmented region**) in a bag can have one of many class labels
  - Thus, a bag (**image**) can be associated with multiple labels



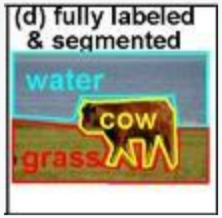
[Vijayanarasimhan & Grauman, 2010]

Predicting cost

- Cost = annotation time
- **Goal:** predict cost based solely on image content



1) Turkers asked to segment images and name objects in segmented regions
2) SVM trained to predict amount of manual effort required to annotate a particular image
   - Training label – average time to complete a full annotation



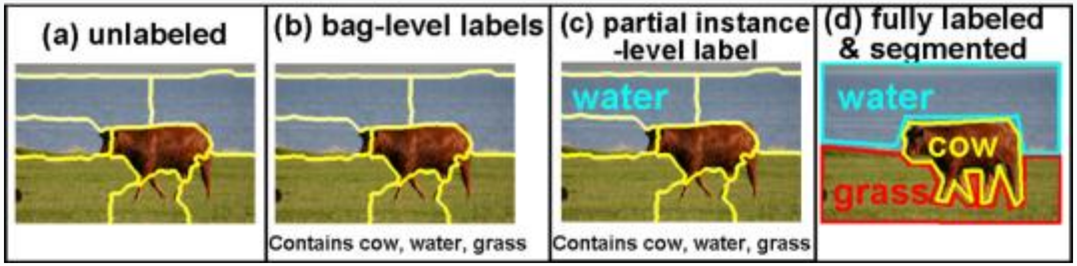
3) Build a cost function $\mathcal{C}(\boldsymbol{z})$
   - Input is candidate annotation z
   - Output is predicted time (sec)

[Vijayanarasimhan & Grauman, 2010]

Predicting informativeness

- **Goal:** predict image and annotation type combination that produces greatest decrease in risk for current classifier, while penalizing for manual effort required

- **Risk terms** $\mathcal{R}$ defined based on labeling completeness
  - $\mathcal{X}_U$: Set of unlabeled examples (bags and instances)
  - $\mathcal{X}_L$: Set of labeled examples
  - $\mathcal{X}_P$: Set of partially labeled examples



[Vijayanarasimhan & Grauman, 2010]

Value of information (VOI) measure

$$VOI(\boldsymbol{z}) = \underbrace{\mathcal{R}(\mathcal{X}_L) + \mathcal{R}(\mathcal{X}_U) + \mathcal{R}(\mathcal{X}_P)}_{\text{Total risk for current set of annotations}} - \left(\mathcal{R}(\widehat{\mathcal{X}}_L) + \mathcal{R}(\widehat{\mathcal{X}}_U) + \mathcal{R}(\widehat{\mathcal{X}}_P)\right) - \mathcal{C}(\boldsymbol{z})$$

Sets of data after obtaining annotation $\boldsymbol{z}$

Predicted cost of annotation $\boldsymbol{z}$

We want to maximize VOI: a high VOI indicates a decrease in total cost after adding an annotation.

- VOI depends on estimating risk for yet-to-be-labeled data
- Solution: estimate total risk with expected value

$$\mathcal{R}(\widehat{\mathcal{X}}_L) + \mathcal{R}(\widehat{\mathcal{X}}_U) + \mathcal{R}(\widehat{\mathcal{X}}_P) \approx \mathrm{E}\left[\mathcal{R}(\widehat{\mathcal{X}}_L) + \mathcal{R}(\widehat{\mathcal{X}}_U) + \mathcal{R}(\widehat{\mathcal{X}}_P)\right]$$

Full active learner pipeline

1) MIML classifier trained on initial set of labeled images
2) Active learner selects label and example with maximal VOI
3) Classifier is updated with added label
4) Repeat

[Vijayanarasimhan & Grauman, 2010]

Discussion: cost effective labeling

- MIML framework suggests further study
  - Other levels of supervision – scene layout, contextual cues, part labels, etc.
  - Domains outside visual categorization
- Alternative cost management approach: exploring quality within a budget

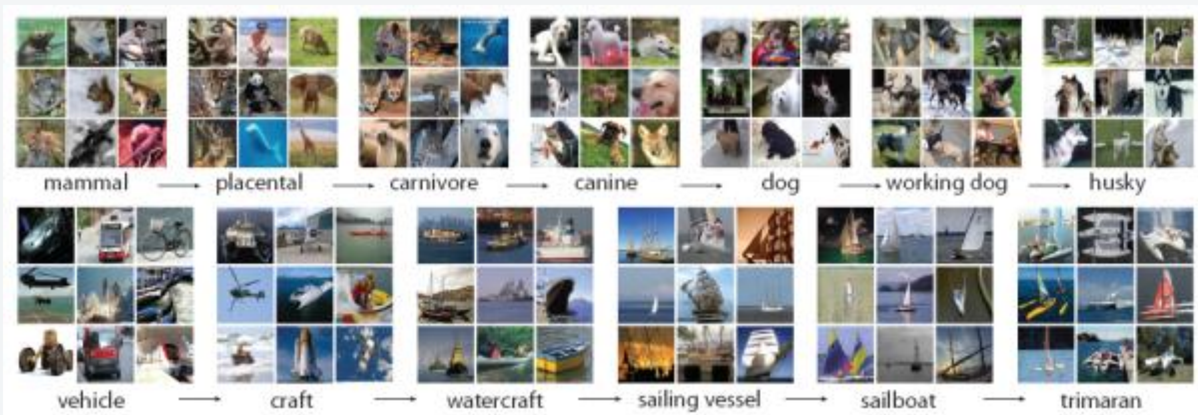
By predicting cost we can save in manual effort while improving classification accuracy.

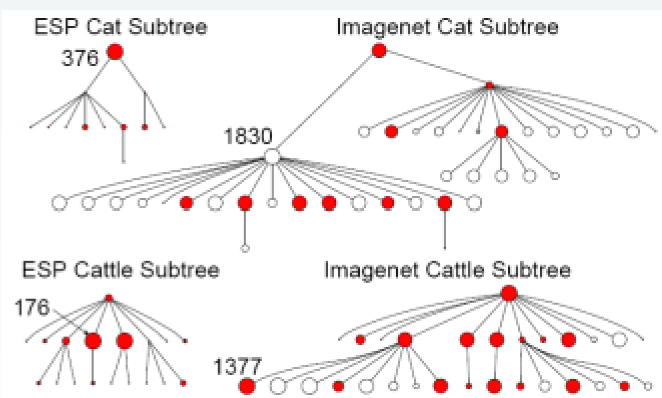
Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Large-scale data collection

- ImageNet





Is there a Burmese cat in the images?

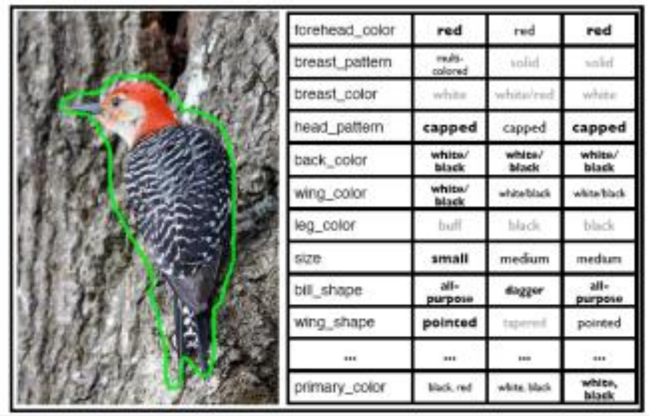
User 1	Y	Y	Y
User 2	N	Y	Y
User 3	N	Y	Y
User 4	Y	N	Y
User 5	Y	Y	Y
User 6	N	N	Y

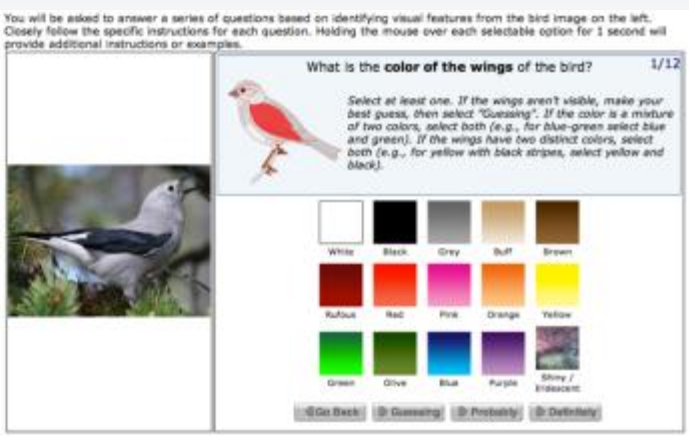
#Y	# N	Conf Cat	Conf BCat
0	1	0.07	0.23
1	0	0.85	0.69
1	1	0.46	0.49
2	0	0.97	0.83
0	2	0.02	0.12
3	0	0.99	0.90
2	1	0.85	0.68

[Deng et al., 2009]

Large-scale data collection

- CUB-200

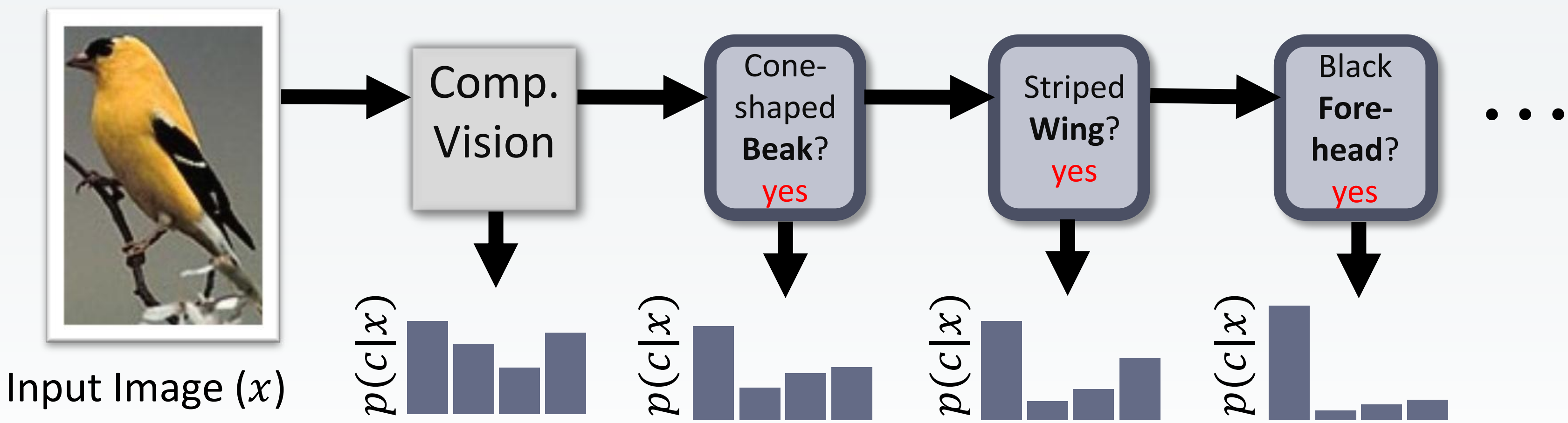




[Welinder et al., 2010]

Humans in the loop

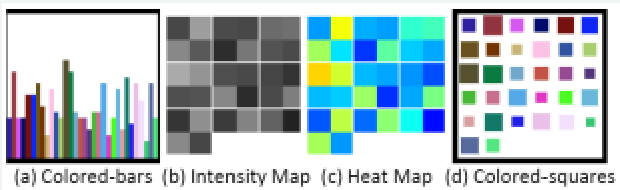
- Interactive field guides – visual 20 questions game

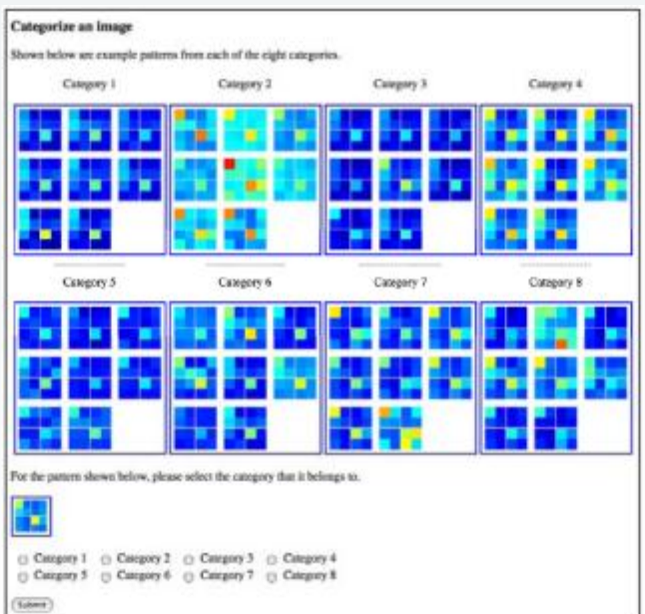


[Branson et al., 2010]

Humans in the loop

- Advancing knowledge about computer vision





"congested"

"spotted"

"green"

[Parikh et al., 2010, 2011]

Outline

- Task incentives
- Experimental design
  - Task parameter selection
  - Human computation process design
- Quality management
  - Heuristics
  - Modeling the human annotation process
- Cost effective strategies for obtaining labels
- Applications in computer vision
- Discussion

Discussion

- Potential lack of diversity in Turkers
  - Self-selection for tasks
- "Virtual sweatshop"
- MTurk spam estimated at over 40%

Even so, MTurk is a feasible, relatively affordable business model that is well-suited for tasks that are difficult for computers.

  - Methods exist for sufficient cost and quality management

Crowdsourcing in computer vision

- Questions:
  - What are the most suitable design patterns and labeling interfaces?
  - How do we integrate the crowd into a computer vision pipeline?
  - Which tasks should we focus on developing algorithms for and which are best suited for humans?

Future directions

- Hybrid approaches: when computers fall short, humans fill in
- Validate methods with crowd feedback
- Explore boundaries of computer vision algorithms

References

1. M.S. Bernstein, G. Little, R.C. Miller, B. Hartmann, M.S. Ackerman, D.R. Karger, D. Crowell, and K. Panovich, "Soylent: a word processor with a crowd inside," *Proceedings of the 23nd ACM UIST*, ACM, 2010, pp. 313-322.
2. S. Branson, C. Wah, F. Schroff, B. Babenko, P. Welinder, P. Perona, and S. Belongie, "Visual recognition with humans in the loop," *ECCV*, 2010, pp. 438-451.
3. J. Deng, W. Dong, R. Socher, L.J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," *CVPR*, Jun. 2009, pp. 248-255.
4. J. Howe, "The rise of crowdsourcing," *Wired*, vol. 14, 2006.
5. G. Little, "Exploring iterative and parallel human computation processes," *Proceedings of the ACM SIGKDD workshop on human computation*, 2010.
6. D. Parikh and K. Grauman, "Interactively Building a Discriminative Vocabulary of Nameable Attributes," *CVPR*, 2011.
7. D. Parikh and C.L. Zitnick, "The role of features, algorithms and data in visual recognition," *CVPR*, 2010.
8. V.S. Sheng, F. Provost, and P.G. Ipeirotis, "Get another label? improving data quality and data mining using multiple, noisy labelers," *Proceedings of the 14th ACM SIGKDD*, ACM, 2008, pp. 614–622.
9. A. Sorokin and D. Forsyth, "Utility data annotation with Amazon Mechanical Turk," *CVPR*, 2008.
10. S. Vijayanarasimhan and K. Grauman, "Cost-Sensitive Active Visual Category Learning," *International Journal of Computer Vision*, vol. 91, Jul. 2010, pp. 24-44.
11. P. Welinder, S. Branson, T. Mita, C. Wah, F. Schroff, S. Belongie, and P. Perona, *Caltech-UCSD birds 200*, 2010.
12. P. Welinder, S. Branson, S. Belongie, and P. Perona, "The Multidimensional Wisdom of Crowds," *NIPS*, 2010.
13. W. Mason and D.J. Watts, "Financial incentives and the 'performance of crowds'," *Proceedings of the ACM SIGKDD workshop on human computation*, 2009.