



Crowdsourcing
CSCI 4849 Spring 2019

Crowdsourcing

- What?
- Why?
- How?
- Notable examples

What is crowdsourcing?

One definition:

- “Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call.”
 - [Jeff Howe, Wired](#)

Crowdsourcing experiences?

Amazon Mechanical Turk

- Primary platform for conducting paid crowdsourcing
- Amazon manages payment accounts
- Requesters create tasks and provide a cash value
- Workers can search tasks, and complete them
- Requesters can verify work and provide payment

Mechanical Turk demo

- As requester
- As worker

Integrating MTurk into code

```
# Import CrowdLib and your settings file with your account key and HIT defaults.
import crowdlib as cl, crowdlib_settings, time

# Create a HIT type, with the title and description for this group of HITs.
hit_type = cl.create_hit_type("Who invented this?", "Answer a simple question.")

# Post a HIT.
hit = hit_type.create_hit([cl.text_field("Who invented the umbrella?", "q1")])

# Wait until the HIT has been completed by a worker.
while hit.is_available:
    time.sleep(10) # poll every 10 seconds

# Fetch and print the result.
print( tuple(hit.assignments)[0]["q1"] )
```

Crowdsourcing for accessibility

- People can perform tasks that some people can't do (and AI can't yet do) - or faster, more accurately, etc.
- Reading text, annotating images, generating speech...
- Prototype “intelligent”, AI-like systems quickly and without tons of training data
- Supports many domains with minimal customization

Trade-offs

Trade-offs

- Cost per use
- Requires internet connectivity
- Latency
- Potential privacy concerns

Why design for the crowd is interesting

- Need to design for unskilled workers
 - Or train them on the job
- Design for tasks that can be dropped and picked up at any time
- Accounting for bad input (intentional or not)

Crowdsourcing research

Key challenges

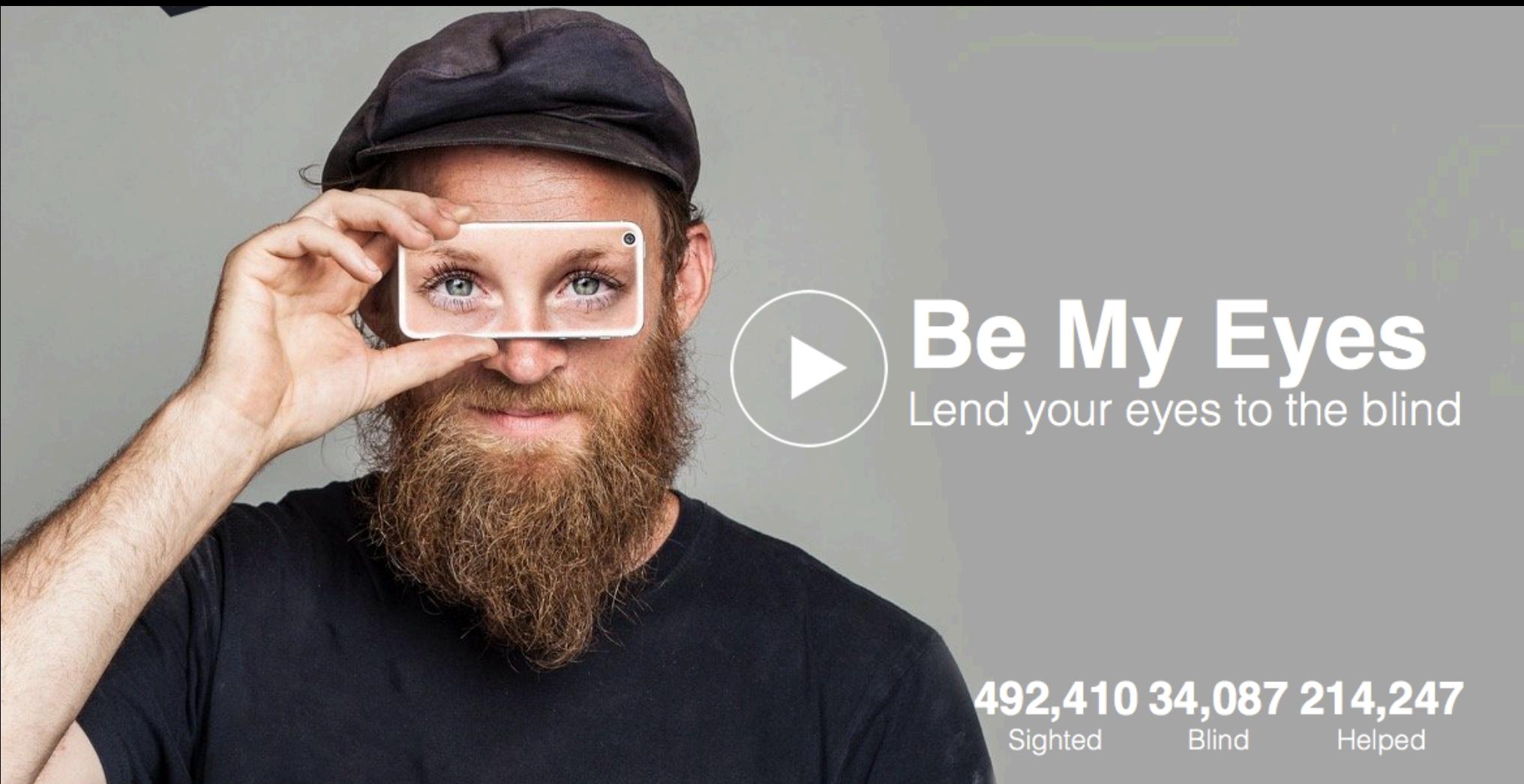
- Breaking down a complex task into smaller parts
- Ensuring quality output from (potentially noisy) data
 - → Crowd algorithms
- Minimizing response time
- Providing meaningful output to user
- Managing privacy for end users and workers

How to do it

- Choose a problem that can be supported by crowd
 - Or break an existing problem into very small steps
- What skills must crowd workers have?
- What contextual info does the crowd need?
- What is worker's incentive?
- How to ensure quality/fast results?

Incentivizing crowd workers

- Volunteering - [Be My Eyes](#)
- Paid
 - Experts - [UpWork](#)
 - Non-experts - [Mechanical Turk](#)
- Gamification - [The ESP Game](#)



Examples from research

- The ESP Game
- Soylent
- VizWiz
- Legion
- LegionScribe
- RegionSpeak
- Crowdsourced map generation

The ESP Game

([von Ahn and Dabbish 2004](#))



Figure 2. The ESP Game. Players try to “agree” on as many images as they can in 2.5 minutes. The thermometer at the bottom measures how many images partners have agreed on.

How to play the ESP Game

- 2 players are matched at random
- Both players are shown the same image
- Players receive points for guessing the same words as others
- As long as they are not “taboo” words



Figure 2. The ESP Game. Players try to “agree” on as many images as they can in 2.5 minutes. The thermometer at the bottom measures how many images partners have agreed on.

What's the deal?

- Why is the ESP Game designed this way?
- Why are there taboo words?



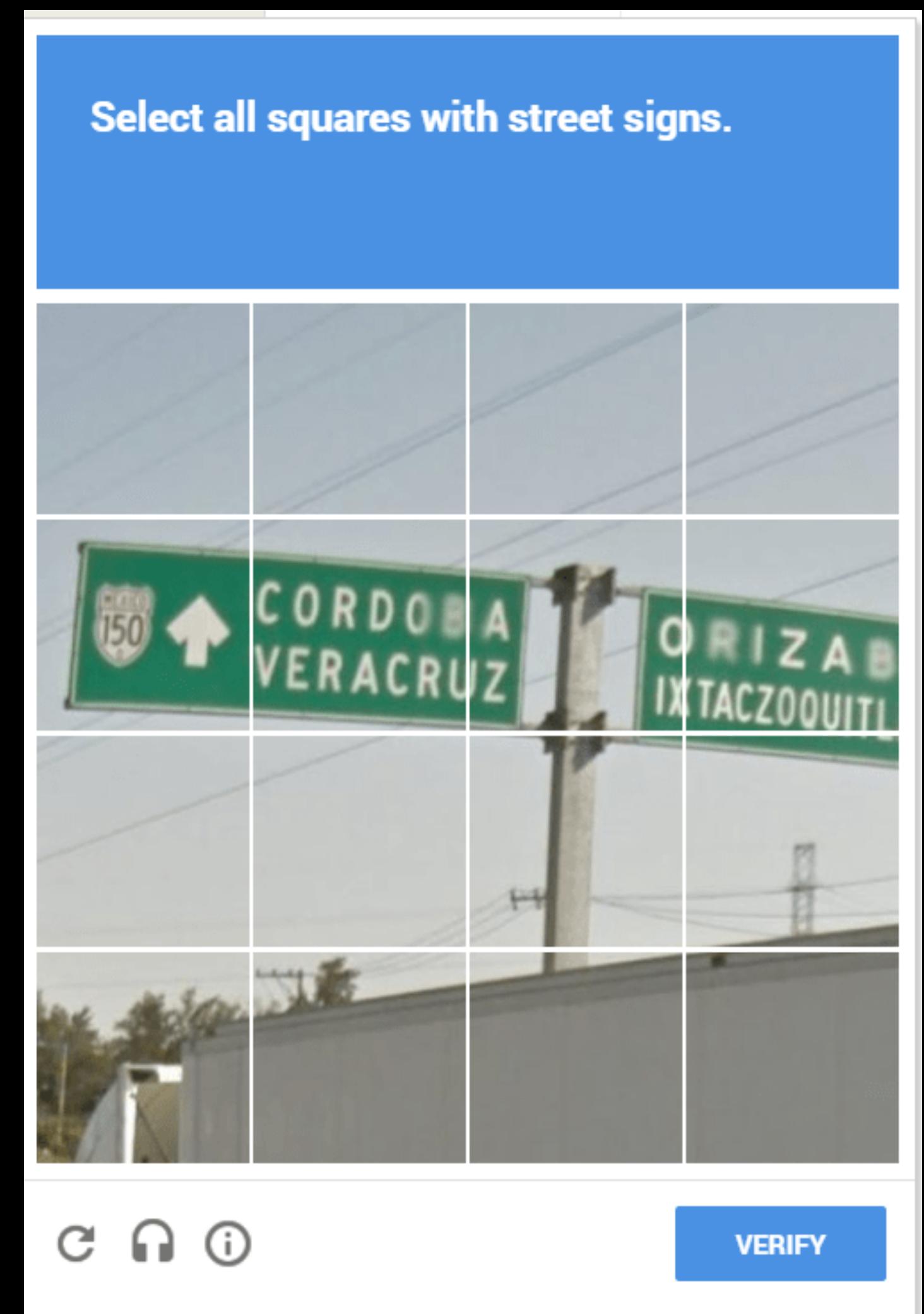
Figure 2. The ESP Game. Players try to “agree” on as many images as they can in 2.5 minutes. The thermometer at the bottom measures how many images partners have agreed on.

The ESP Game: Behind the scenes

- Players are incentivized to label images, solving a critical accessibility problem
- “Taboo” words prevent people from coming up with the most common word only, encourages new words

The beauty of the ESP game

- Generates useful accessibility data for free
- Right combination of fun task + useful output
- van Ahn also founded reCAPTCHA
- Received the MacArthur “Genius” Award for this work

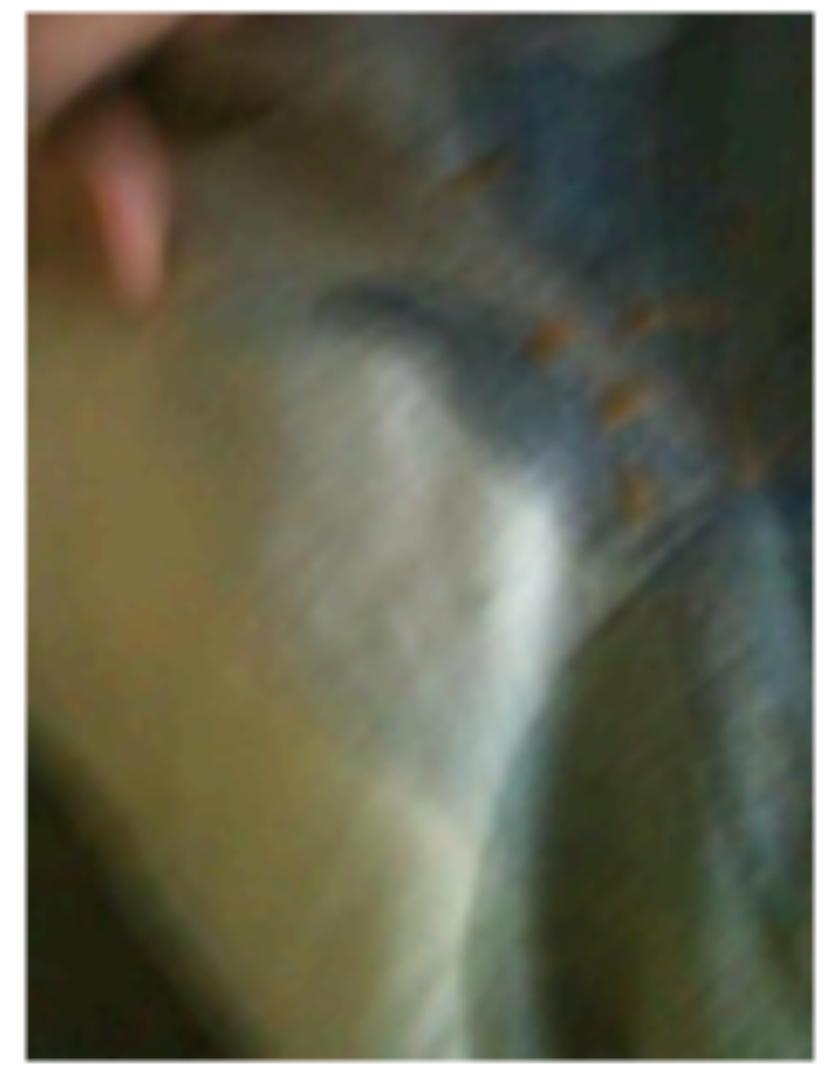


VizWiz: Nearly real-time answers to visual questions ([Bigham et al. 2010](#))

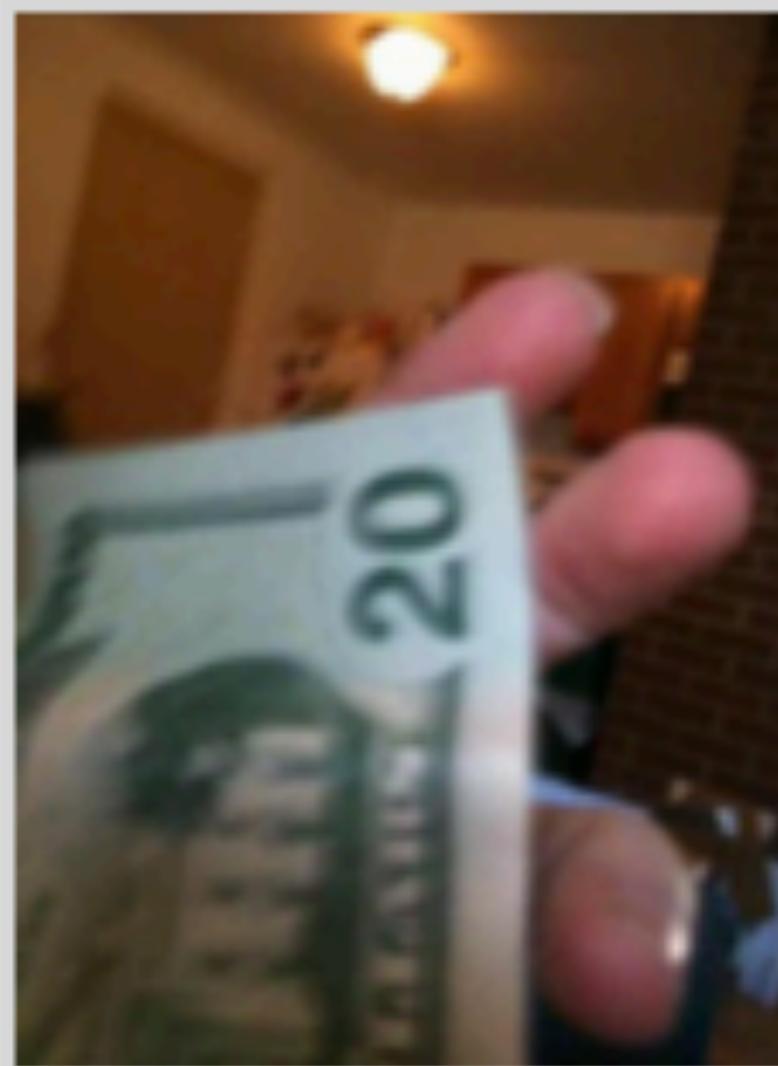
- A blind person takes a photo with a smartphone and speaks a question
- Text is transcribed and sent to crowd workers
- QuikTurkit allows for fast responses; recruits workers to complete filler tasks until the user's request is ready

VizWiz questions

What color is this pillow?



What denomination is this bill?



Do you see picnic tables across the parking lot?



What temperature is my oven set to?



Can you please tell me what this can is?



What kind of drink does this can hold?



(89s) I can't tell.
(105s) multiple shades of soft green, blue and gold

(24s) 20
(29s) 20

(13s) no
(46s) no

(69s) it looks like 425 degrees but the image is difficult to see.
(84s) 400
(122s) 450

(183s) chickpeas.
(514s) beans
(552s) Goya Beans

(91s) Energy
(99s) no can in the picture
(247s) energy drink

Figure 2: Six questions asked by participants, the photographs they took, and answers received with latency in seconds.

What's cool about VizWiz?

- Solved a seemingly impossible AI task – answering arbitrary questions about images
- Addressed the main limitation of crowd markets – latency for recruiting workers
- A real app that you can download from the App Store – although not currently available :(

VizWiz for fashion

(Burton et al., ASSETS 2012)

- Use human assistants for their subjective fashion sense
- Important to match raters with end users
- “Yes, I would like to know if the volunteer preferred ‘big city’ or ‘small town’ style of dress. Also, I would be interested to know if the volunteer had exposure or experience with various other cultures. Like color preferences and accent pieces for a complete look.”

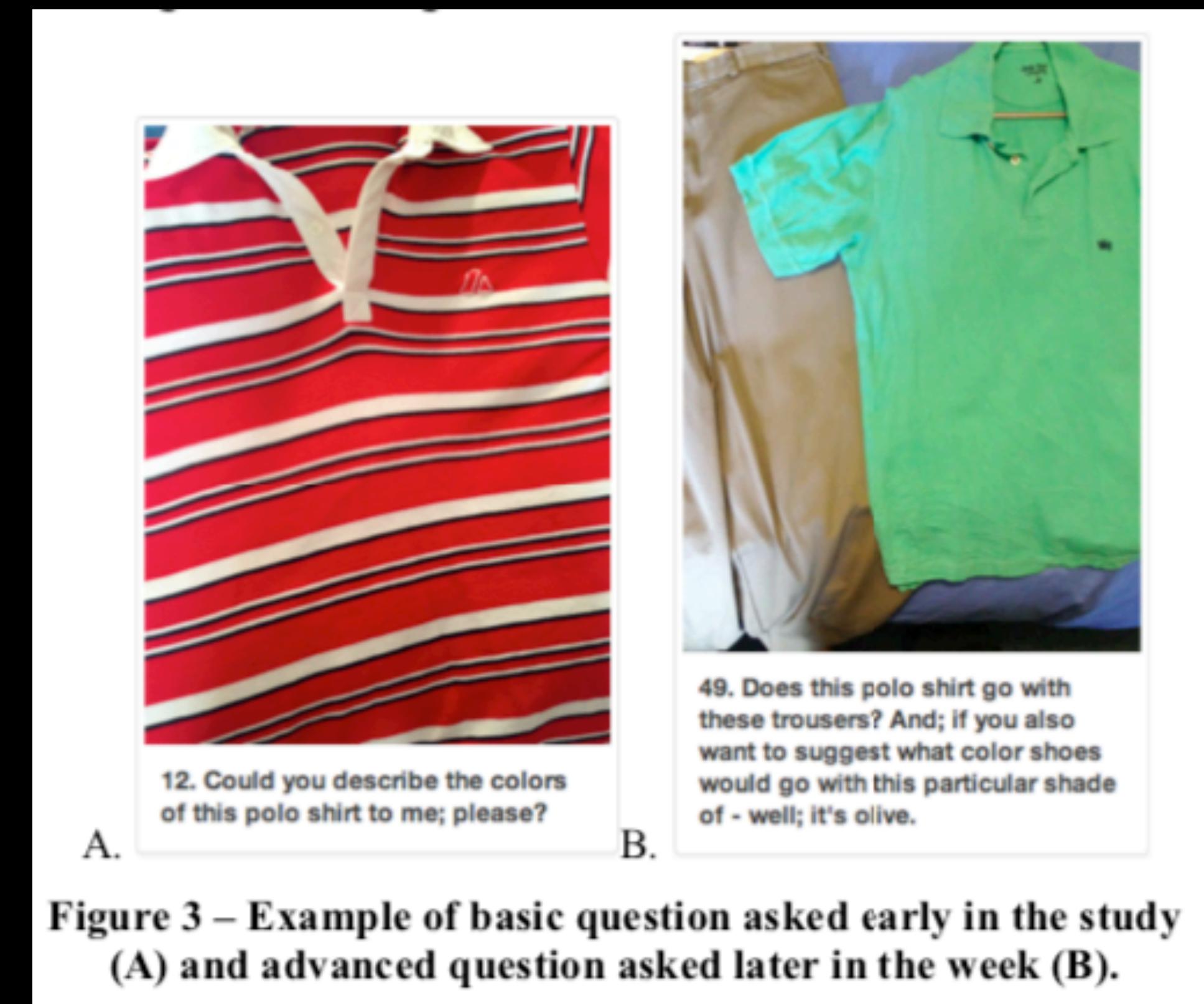


Figure 3 – Example of basic question asked early in the study (A) and advanced question asked later in the week (B).

Legion Scribe

([Lasecki et al. 2014](#))

- Support audio transcriptions using non-expert workers
- Provides support for capturing and merging partial transcripts

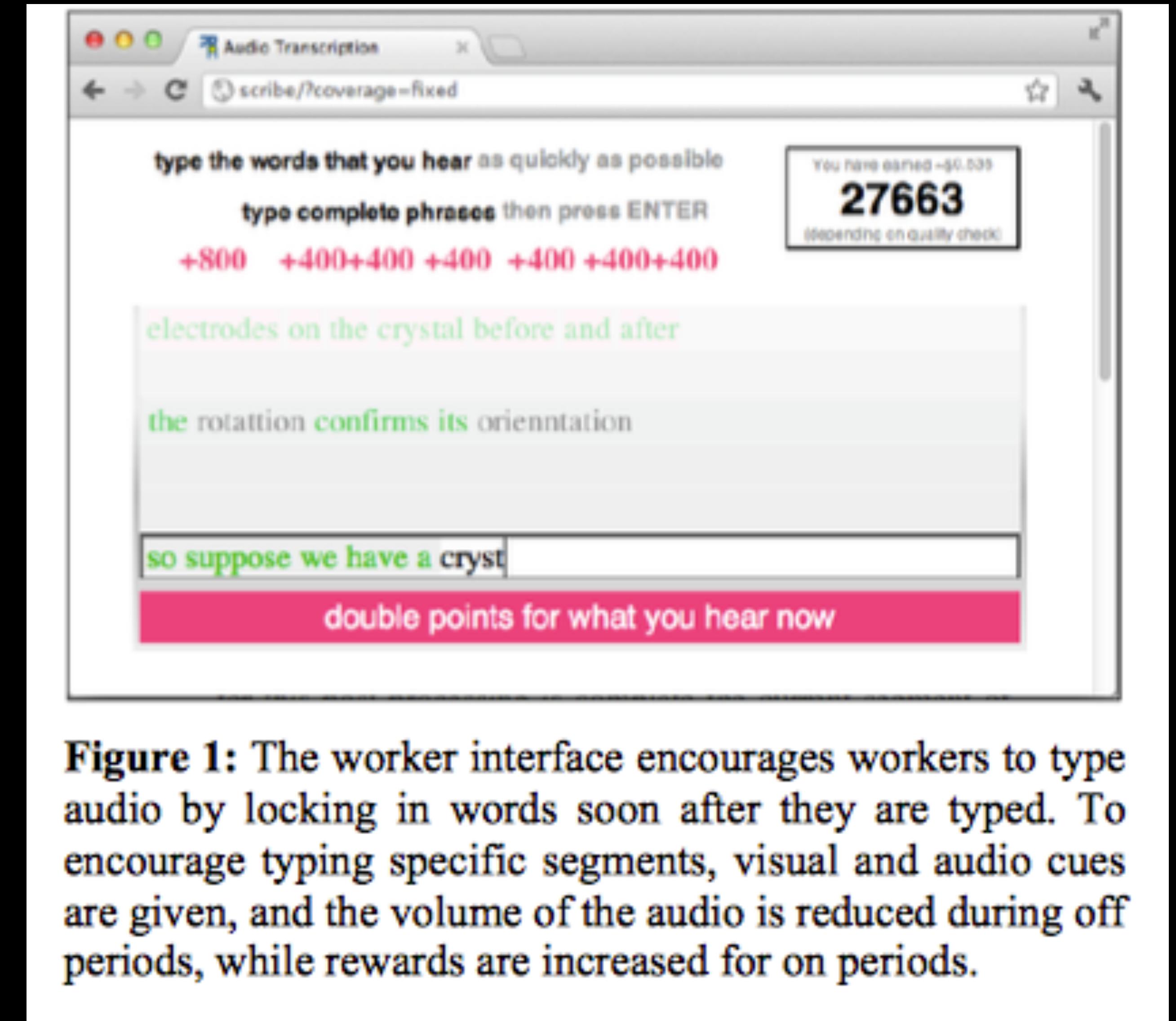
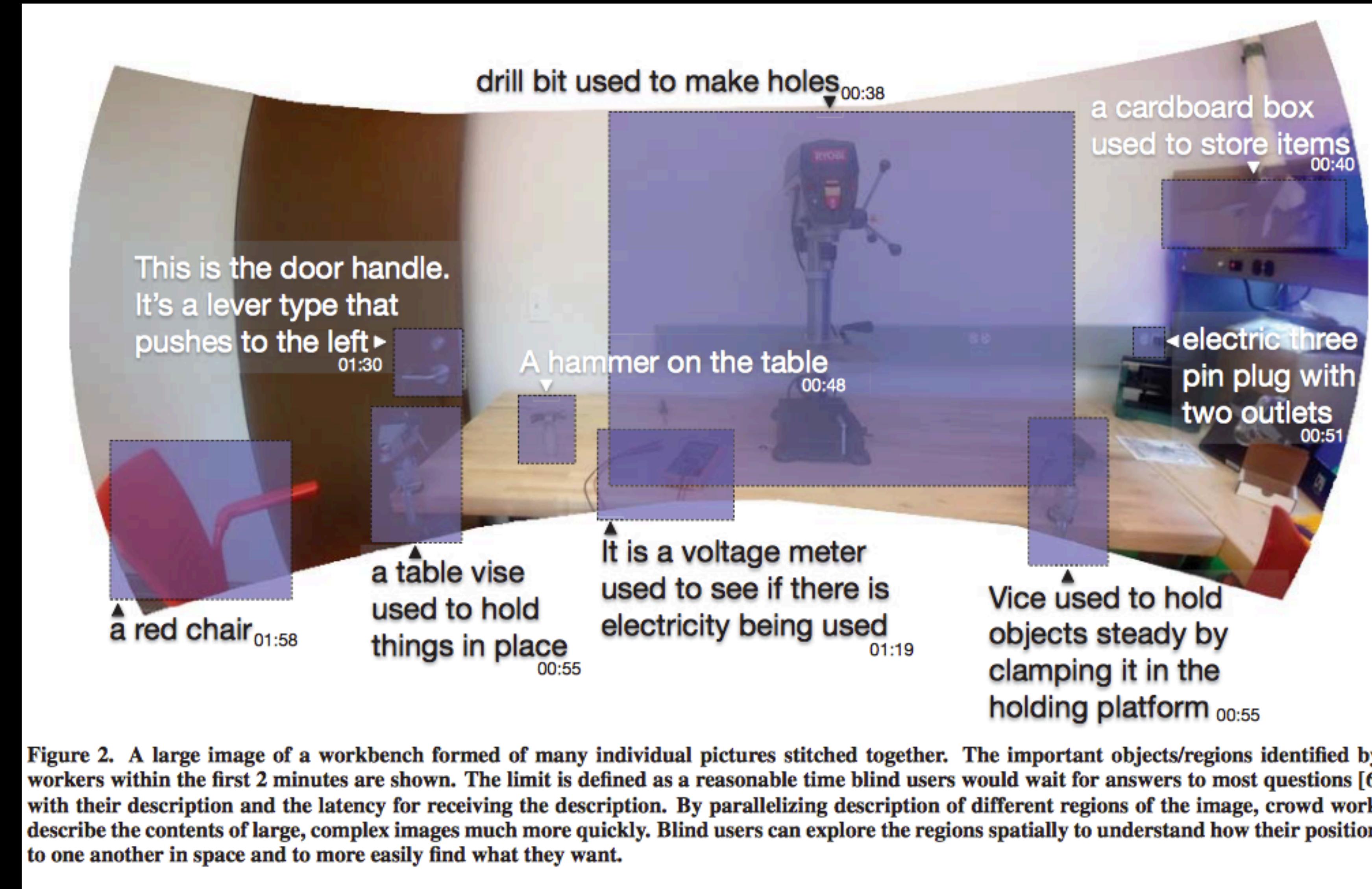


Figure 1: The worker interface encourages workers to type audio by locking in words soon after they are typed. To encourage typing specific segments, visual and audio cues are given, and the volume of the audio is reduced during off periods, while rewards are increased for on periods.

RegionSpeak ([Zhong et al. 2015](#))

- Combines crowd-based image tagging with computer vision-based tracking
- Crowd labels parts of an image
- User can explore image interactively by touching screen



Limitations of these approaches?

Some limitations

- Do crowd workers give good answers?
- Image quality
- Privacy
- Worker rights

Ensuring quality answers

- Have workers complete a **gold standard** task for which the answer is known
- Recruit multiple workers to complete the same task; correct answers should be common across multiple workers
- Can use more complex crowd algorithms, like **Find-Fix-Verify**

Soylent: A word processor with a crowd inside

([Bernstein et al., UIST 2010](#))

- How do we enable paid crowd workers to edit a document?
- Keeping in mind that crowd workers:
 - May make mistakes
 - May input junk data
 - May quit halfway through
 - **Benefits:** smart find-and-replace (e.g. Michelle to Michael), natural language queries, human intelligence

The solution: crowd algorithms!

- Break down tasks in a way that enables error checking, smaller sub-tasks
- Soylent introduces **find, fix, verify algorithm**
- Each sub-step handled by different workers
- Sent to multiple workers for agreement

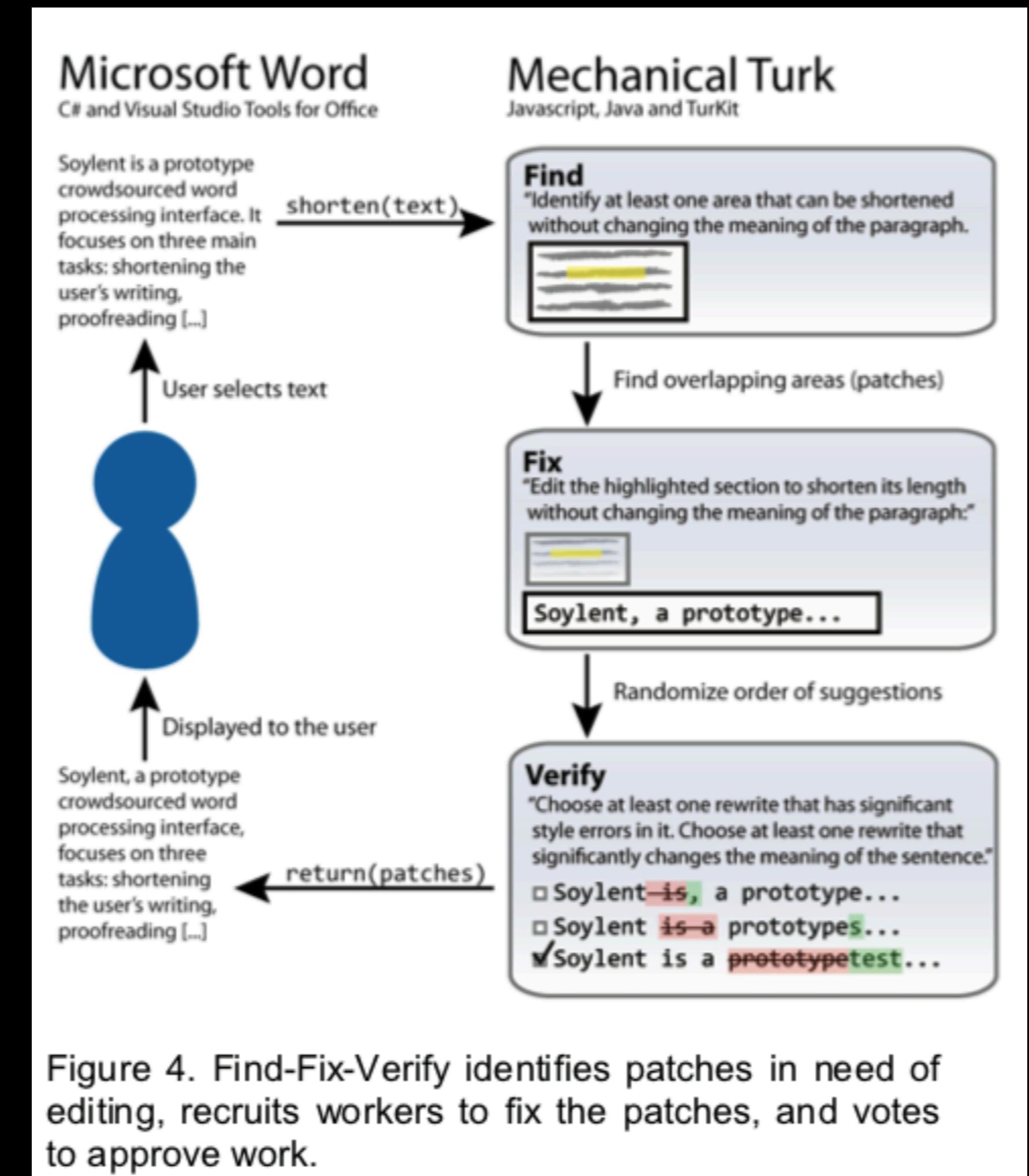
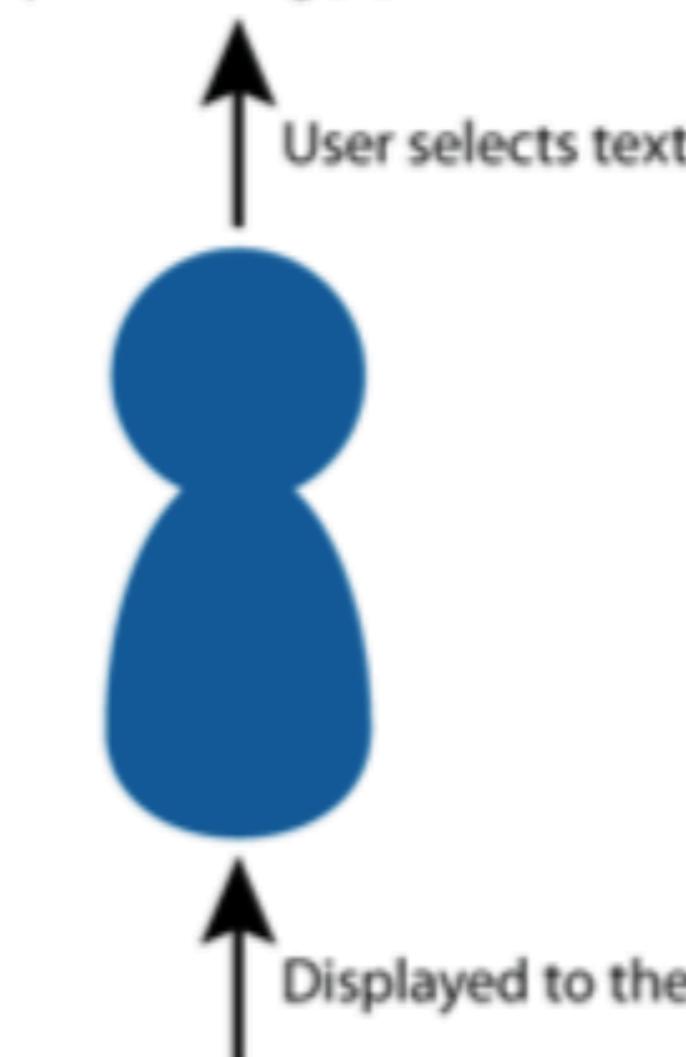


Figure 4. Find-Fix-Verify identifies patches in need of editing, recruits workers to fix the patches, and votes to approve work.

Microsoft Word

C# and Visual Studio Tools for Office

Soylent is a prototype crowdsourced word processing interface. It focuses on three main tasks: shortening the user's writing, proofreading [...]



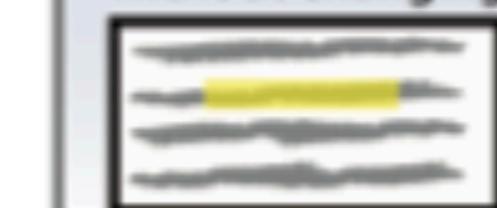
Soylent, a prototype crowdsourced word processing interface, focuses on three tasks: shortening the user's writing, proofreading [...]

Mechanical Turk

Javascript, Java and TurKit

Find

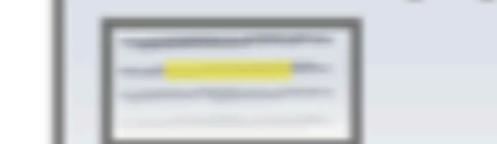
"Identify at least one area that can be shortened without changing the meaning of the paragraph."



Find overlapping areas (patches)

Fix

"Edit the highlighted section to shorten its length without changing the meaning of the paragraph."



Soylent, a prototype...

Randomize order of suggestions

Verify

"Choose at least one rewrite that has significant style errors in it. Choose at least one rewrite that significantly changes the meaning of the sentence."

- Soylent ~~is~~, a prototype...
- Soylent ~~is-a~~ prototypes~~s~~...
- Soylent is a ~~prototype~~test...

return(patches)

Figure 4. Find-Fix-Verify identifies patches in need of editing, recruits workers to fix the patches, and votes to approve work.

Structuring input to improve quality

(Morash et al. 2015)

- Providing a template for describing an image results in higher-quality descriptions

This is a line graph titled *title*. A caption reads: “*caption*.” The horizontal axis is labeled *x-axis label* and ranges from *x-axis min* to *x-axis max* in increments of *x-axis increments* *x-axis units*. The vertical axis is labeled *y-axis label* and ranges from *y-axis min* to *y-axis max* in increments of *y-axis increments* *y-axis units*. There *is / are* *number* *line / lines* on the graph. Line 1 is labeled *line 1 label* and appears *line 1 trend*.

Line *number* is labeled *line label* and appears *line trend*.

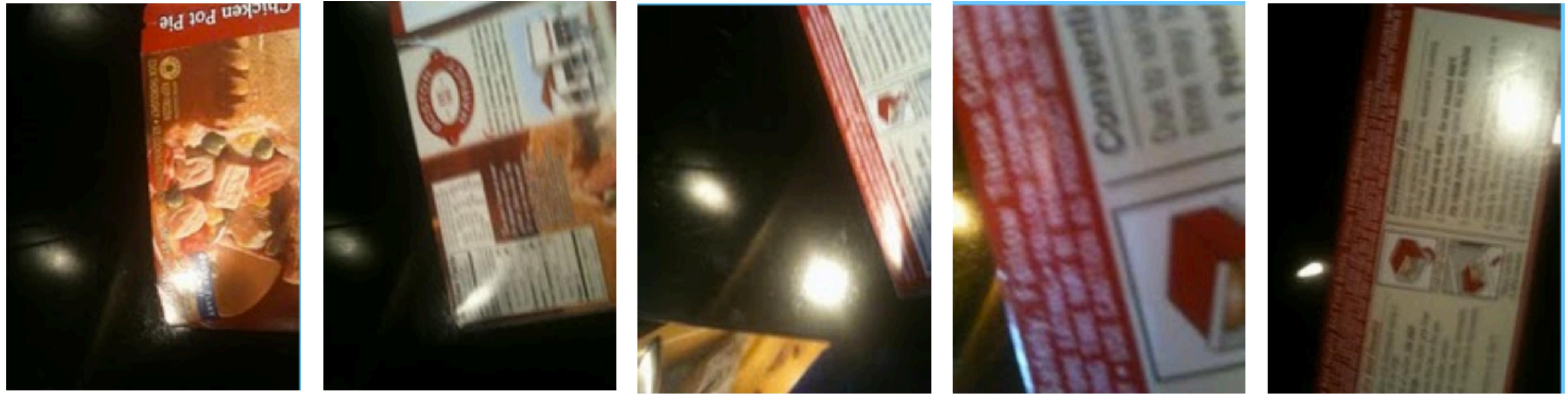
repeat to number of lines

Privacy

- Data may unintentionally contain private information
 - “What is my credit card number?”
 - Inadvertently including private information in image
- Important for users to know that humans have access to their data
- Detect sensitive information when possible

Image quality ([Brady et al. 2013](#))

- Capturing images to use in crowdsourcing can be a real challenge for blind and visually impaired people
- Issues: lighting, blur, framing
- Sometimes take picture of wrong thing entirely
- Feedback can help; this can be an iterative process



Instructions are likely on
the other side of the box.

■ ■ ■

Box is mostly out of
frame to the right.

■ ■ ■

Preheat to 400, remove from
box, cover edges with foil and
put on tray, cook 70 minutes

Figure 2. Sequence of pictures taken to answer, “How do I cook this?” with VizWiz Social. Each question was answered quickly, but cooking instructions took more than 10 minutes to receive as crowd workers helped the user frame the right part of the box.

Crowd work for PWD?

- Crowd work could be a viable employment for people with disabilities (flexible hours, can work from home)
- But existing crowd systems present accessibility challenges
 - Time limits
 - Inaccessible user interfaces (often inaccessible halfway through a task)
 - No way to discover task accessibility
- For more info, see [Zyskowski et al. 2015](#)

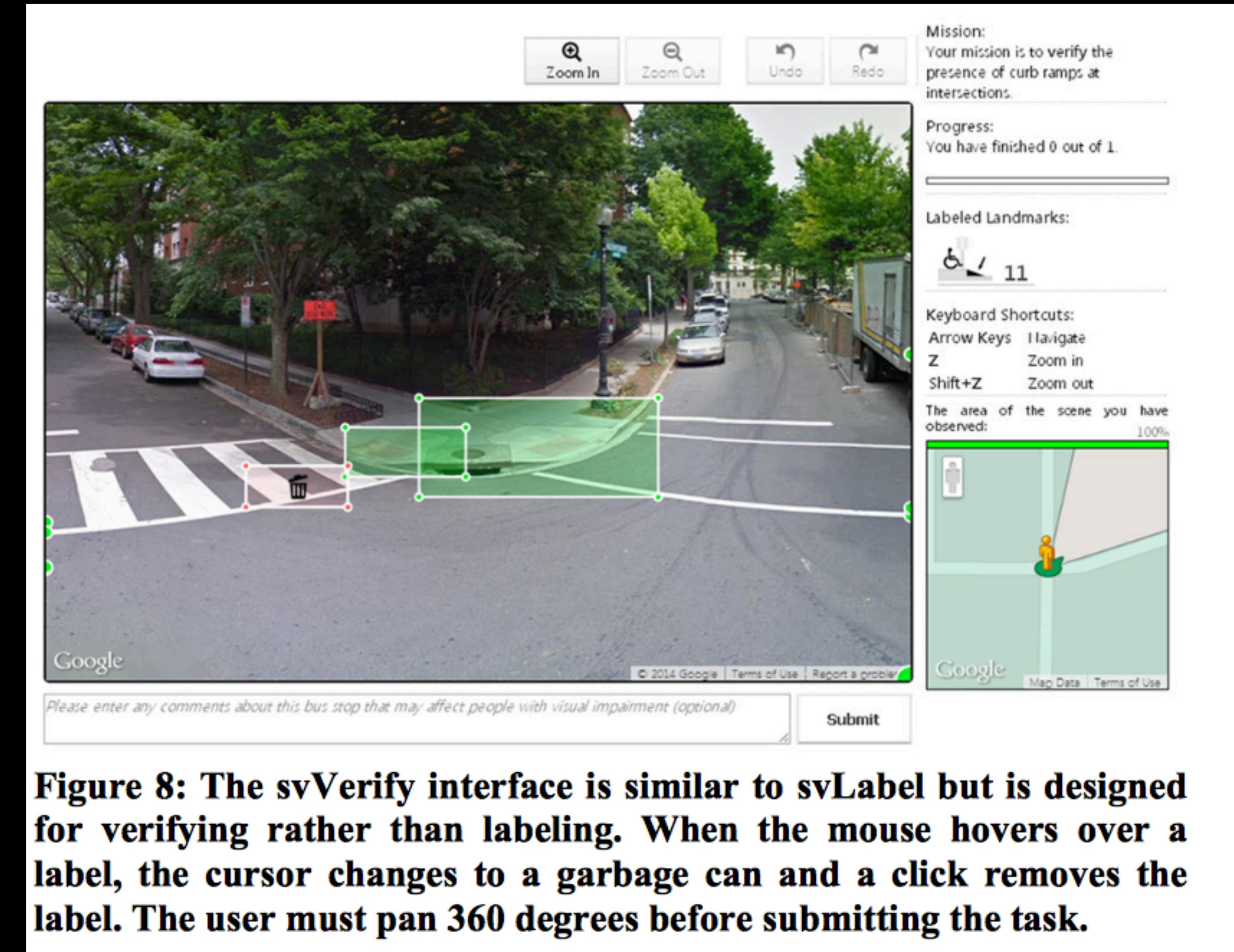
Fair pay & worker rights

- Hourly rates vary wildly
- Be clear what work is expected for a task
- One option: pay workers at least minimum wage where you are
 - But costs of living vary widely across the world
- New crowdsourcing platforms like [Daemo](#) seek to provide workers with more control

Other crowdsourcing systems

Tohme ([Hara et al. 2014](#))

- Computer vision system to detect pedestrian navigation features
- Use crowd workers to remove false positives/negatives; train ML system
- Human workers can often **verify** results quickly and accurately



SceneTalk

(Kane et al. 2017)

- Use crowd to generate likely phrases for AAC
- Crowd identifies regions in images (still difficult for ML methods)
- Can be automated later

Create phrases for this image

This photo has been taken by an individual with a communication disability. We are asking you to highlight important items in the image to help this individual with their communication. Given the highlighted item, generate some phrases that a person might wish to say about this item.



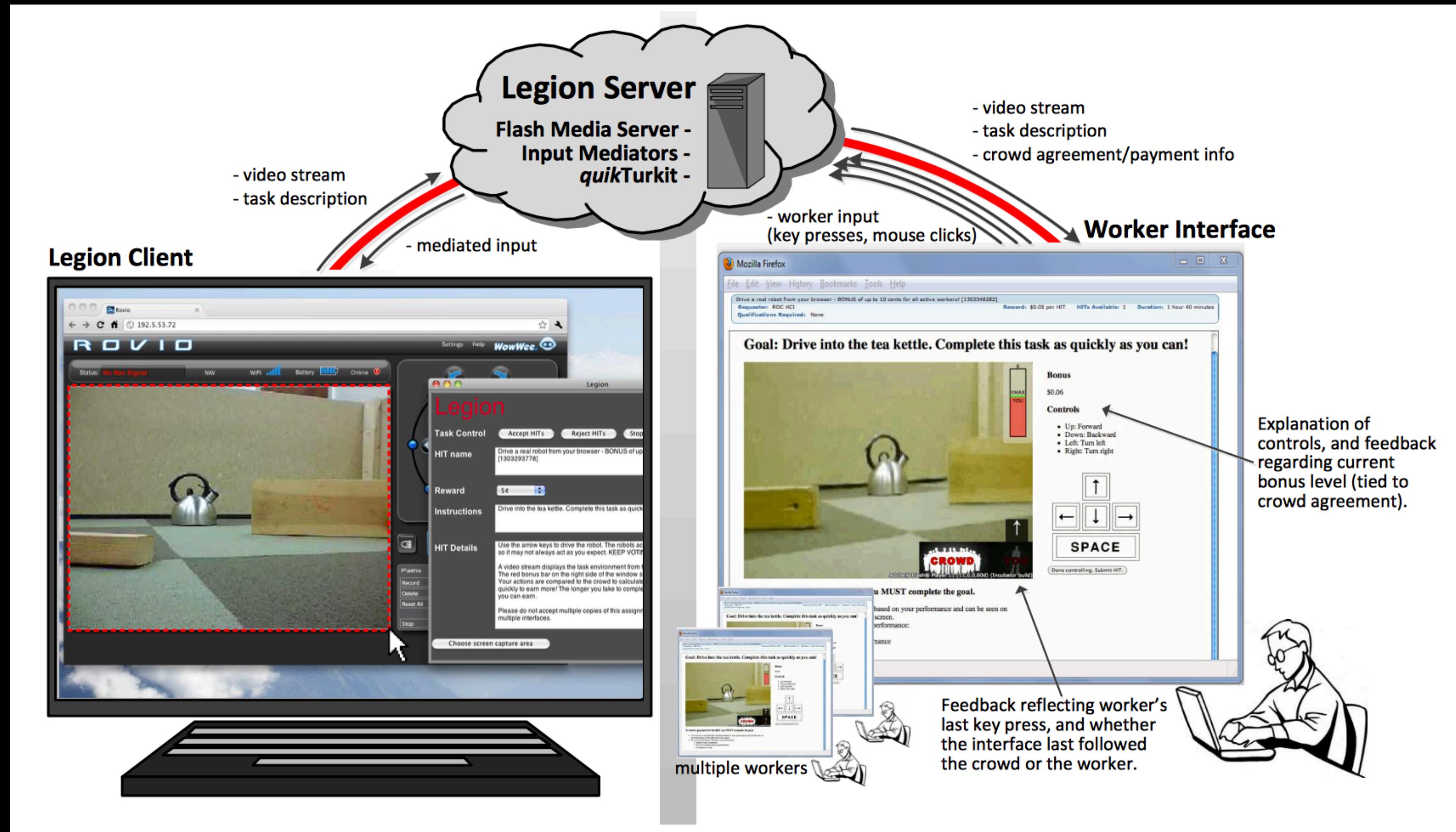
green peppers

Phrase 1

Can you get me some green

Legion ([Lasecki et al. 2011](#))

- Use crowd to control a complex user interface (e.g. a robot)
- Combine crowd input to choose best path



Making crowd input usable (Legion)

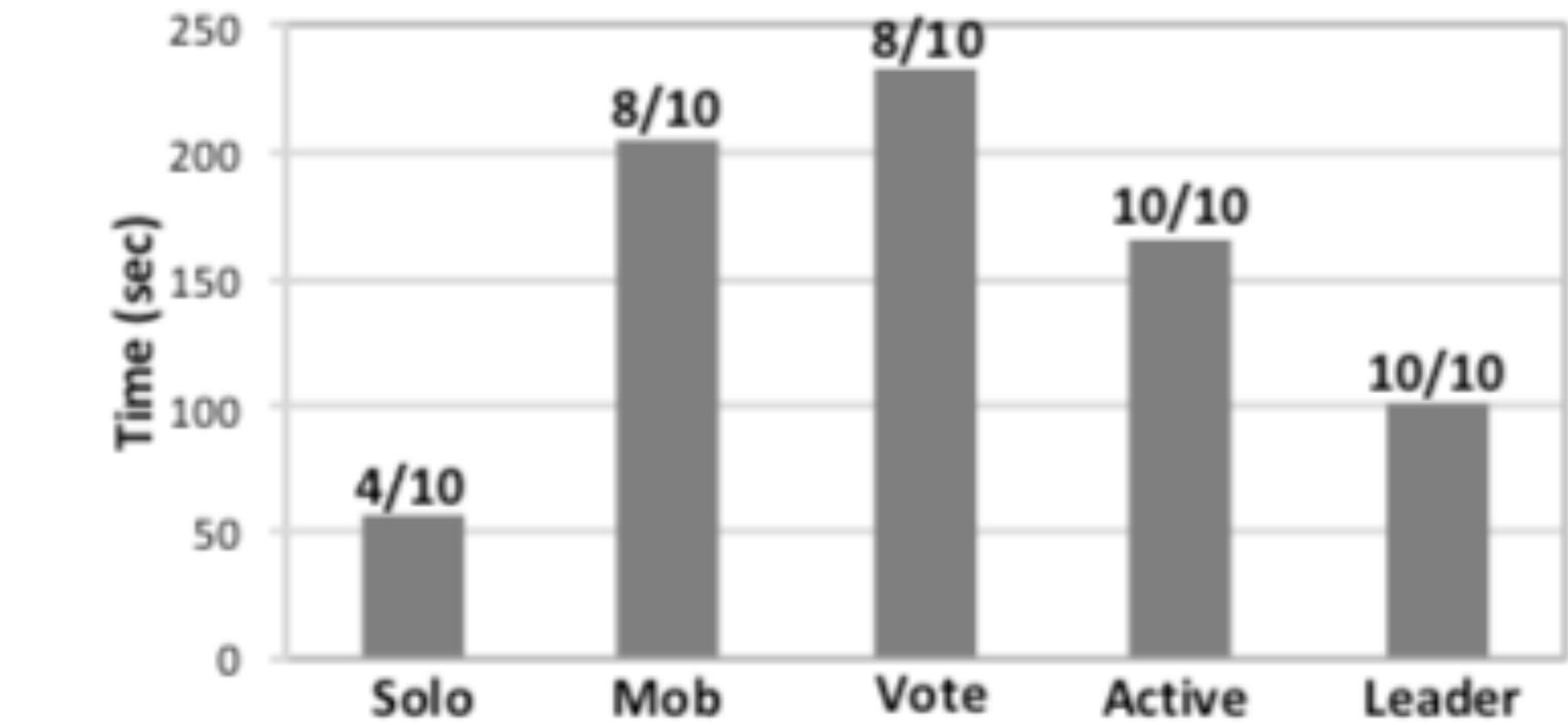
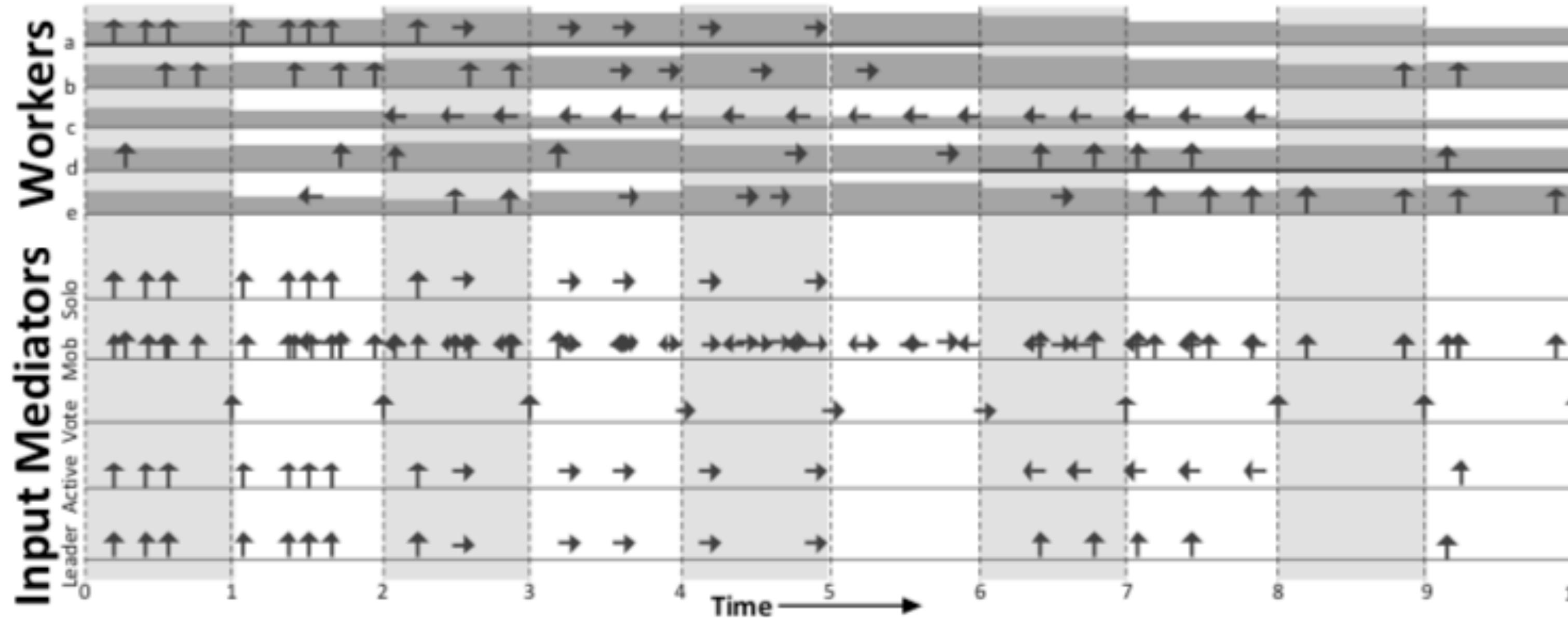
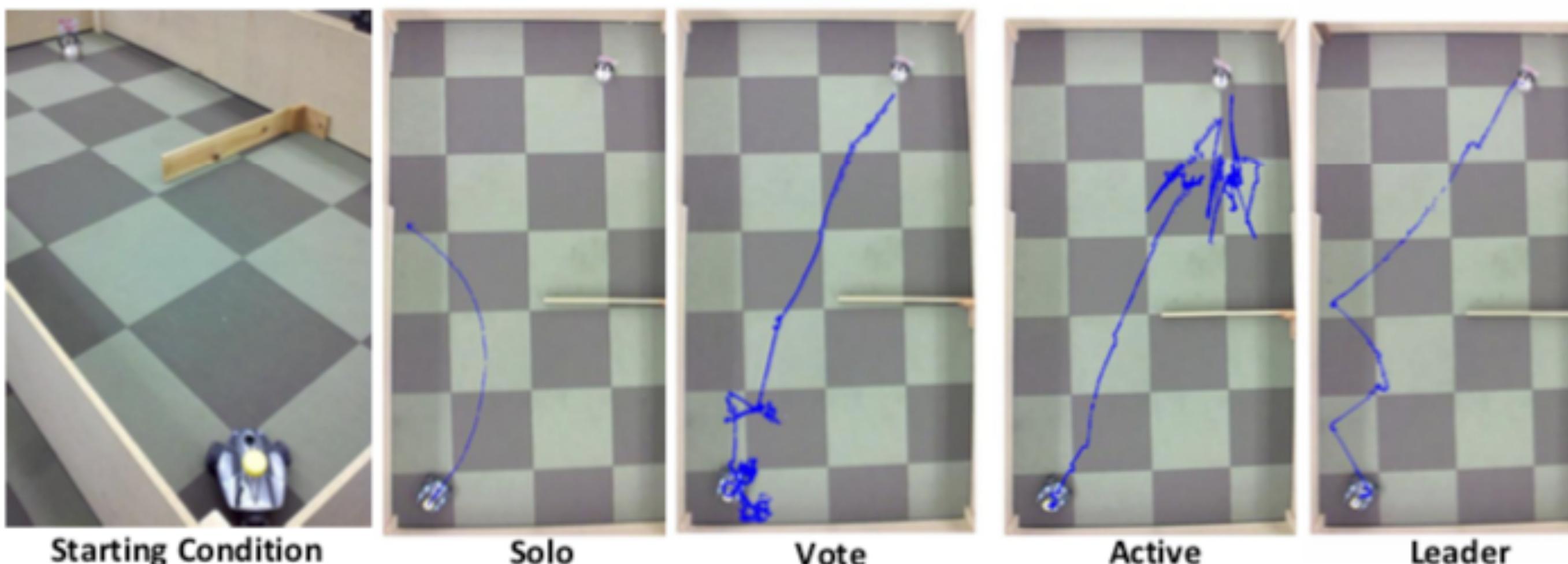
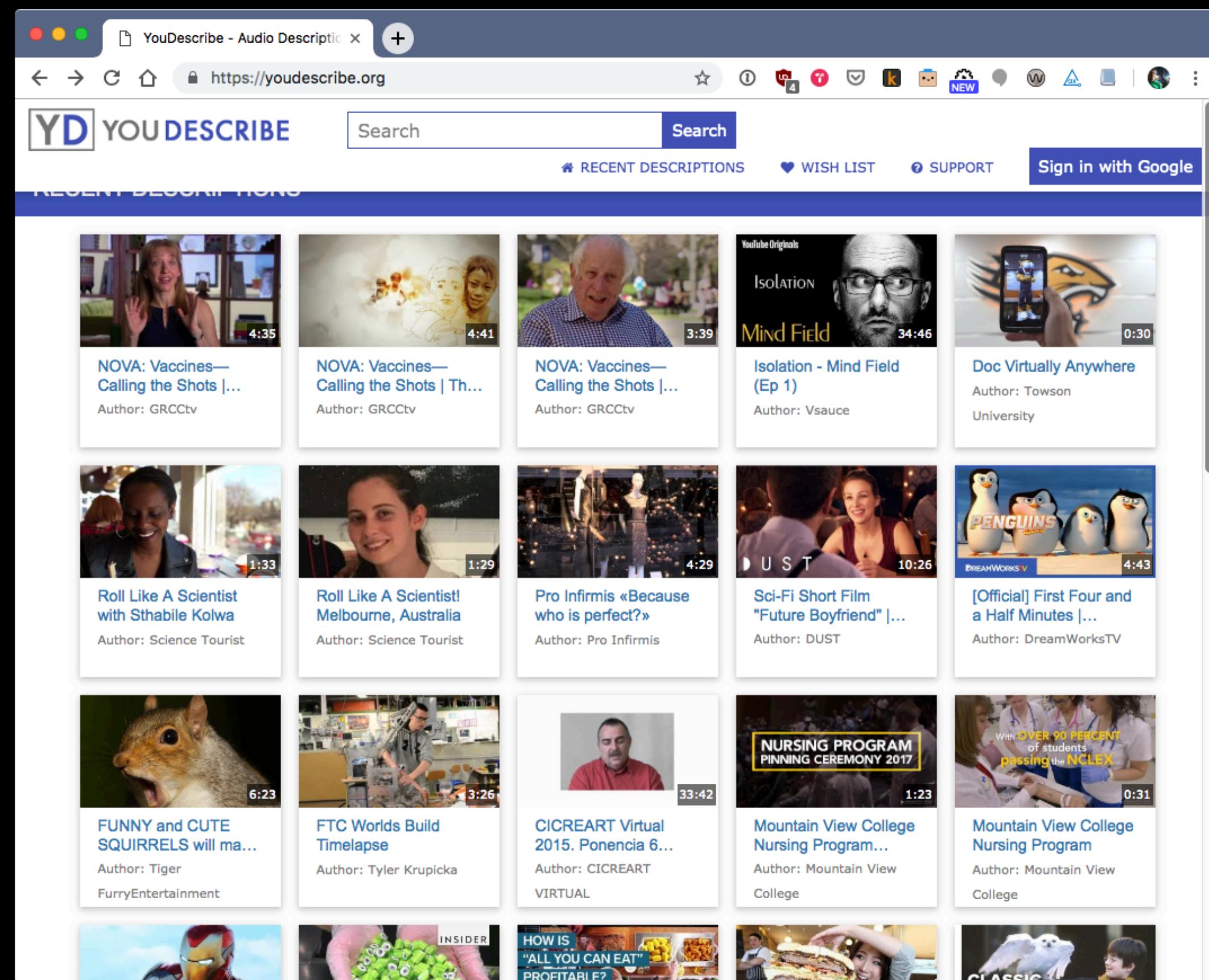


Figure 6: The average time required to complete the robot navigation task for each of the input mediators, along with the fraction of successful runs. *solo* was the fastest but only completed 4 of 10 runs, whereas *leader* was close in time and completed all runs. Choosing leaders based in crowd agreement in *leader* outperformed *active*, which chooses leaders randomly.



Fansourcing

- Where else can we find willing volunteers? **Fan communities.**
- Take advantage of fans' excitement to share their favorite thing



Design exercise

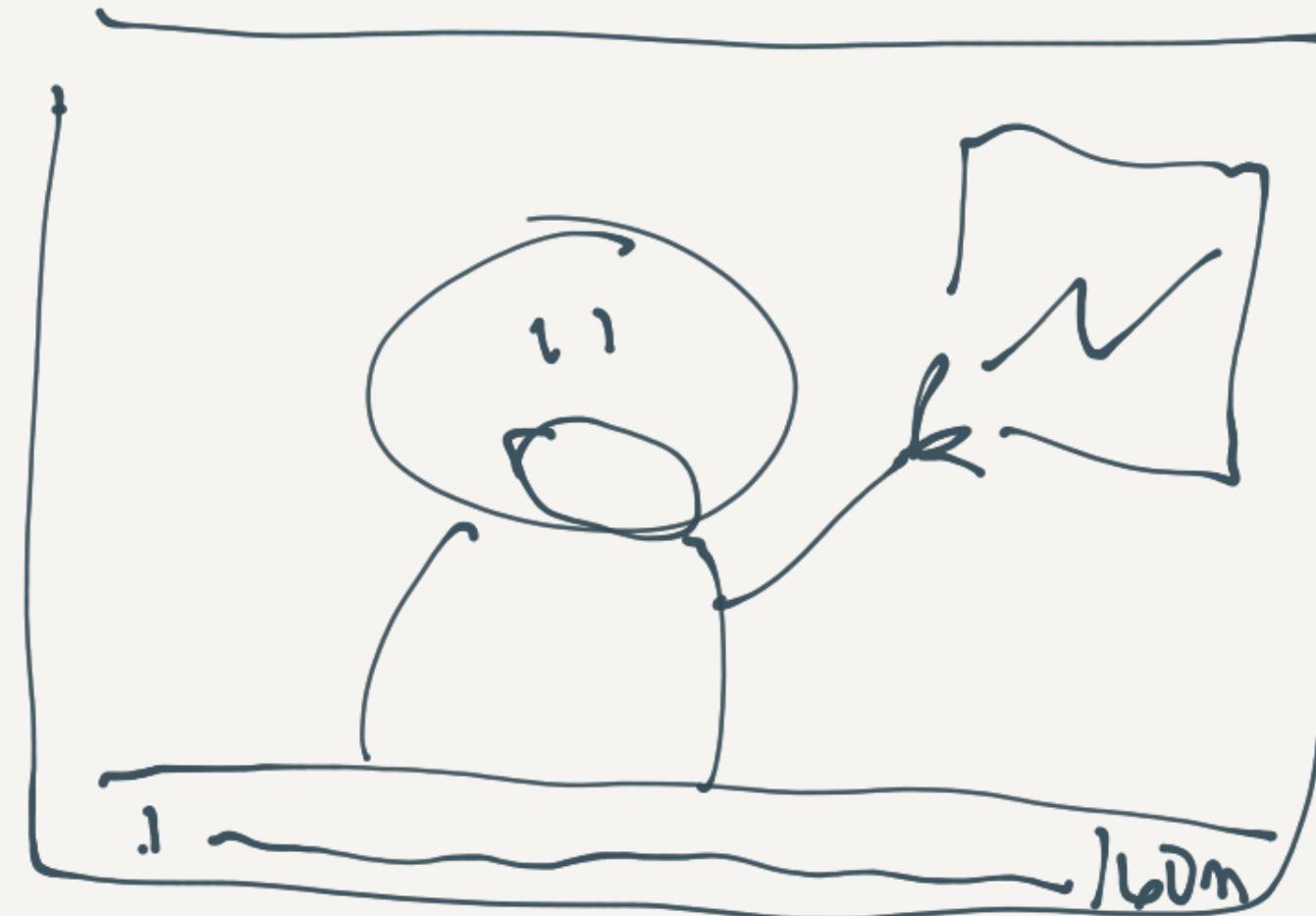
Working with crowd workers...

- They are untrained (unless we filter or train them)
- They may have trouble following directions (especially if directions are confusing)
- They may not always do a good job
- They may quit if a task is boring
- They take time to recruit, and to complete the task

How to do it well

- A practice example: use the crowd to transcribe computer science talks
- We'll start with a bad version, and make a better one

Please transcribe
this video.



Submit

Solutions

- Break into 1 minute segments, Give a sentence at a time
- Auto transcribe first and correct
- Competition! (between friends, high score list)
- Give correspondence between text and video
- Controls for video speed
- Gold standard test video
- Overlap between segments

Creating tasks in MTurk

- Use APIs to generate human intelligence tasks (HITs), retrieve results
 - [Official MTurk API](#)
 - [Boto](#) (a pretty good Python API)
- MTurk tasks: can be a web form created on mturk.com, or can create a custom web app (that handles specific MTurk query parameters)