

Lecture 14

Non-Independent Work

Instructor: Chien-Ju (CJ) Ho

Logistics: Project

- Milestone 1: Due this Friday
 - Brief literature survey
 - A plan (with a timeline)
- Plan:
 - You should have a relatively clear idea on how “exactly” you are going to do things. You would then only need to really do it (or figure out why it can’t be done).
 - Theory/simulation project: what your model looks like
 - Data analysis projects: where is your data (source of dataset)
 - Experimental project: Prototype design (even just on paper without implementation)
 - Literature survey project: A high-level category of directions you plan to survey

Logistics: Project

- Milestone 2: Due Nov 20 (Friday)
- Midterm Project Pitch and Discussion: Nov 19 (Thu) in Lecture
 - A chance to get feedback or collect data from other students
 - Think about whether/how you can best utilize this time

Lecture Today

Human Computation



“Crowd”



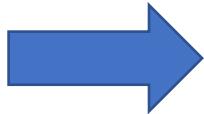
Flower
Dog
Cute
...

Input

Black Box

Output

Open the Black Box



Flower
Dog
Cute

...

Human as data sources:
Label aggregation
Probabilistic reasoning to
aggregate noisy human data

Humans are “Humans”:
Incentive design
Game theoretical modeling of
humans and incentive design

The Crowd is Connected



Input

A Connected Crowd

Flower
Dog
Cute
...

Output

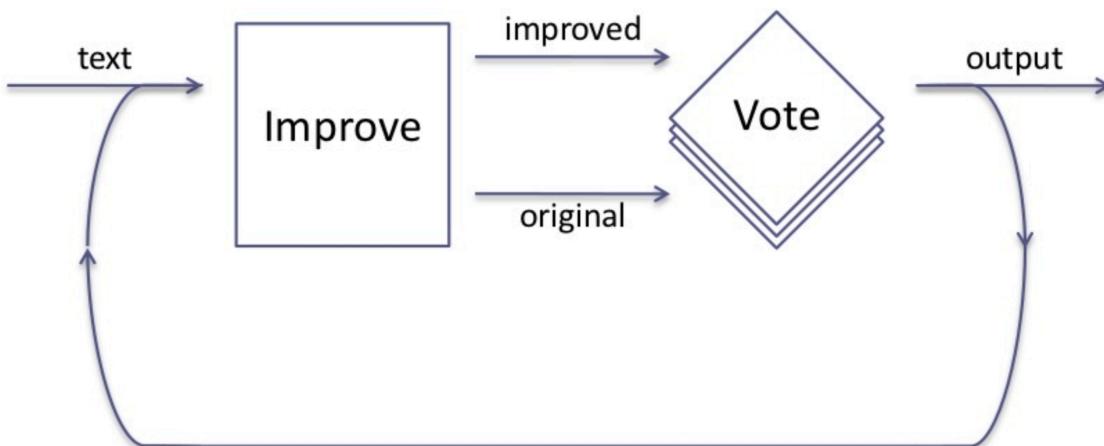
Beyond the Independent Workers

Consider this handwriting transcription task

You wrote us several words. Also you have some words written
in other which are few grammatical mistakes. Overall your writing style
is a bit too flowing. You do not have good friends,
but they get lost amidst the writing

Jyoti

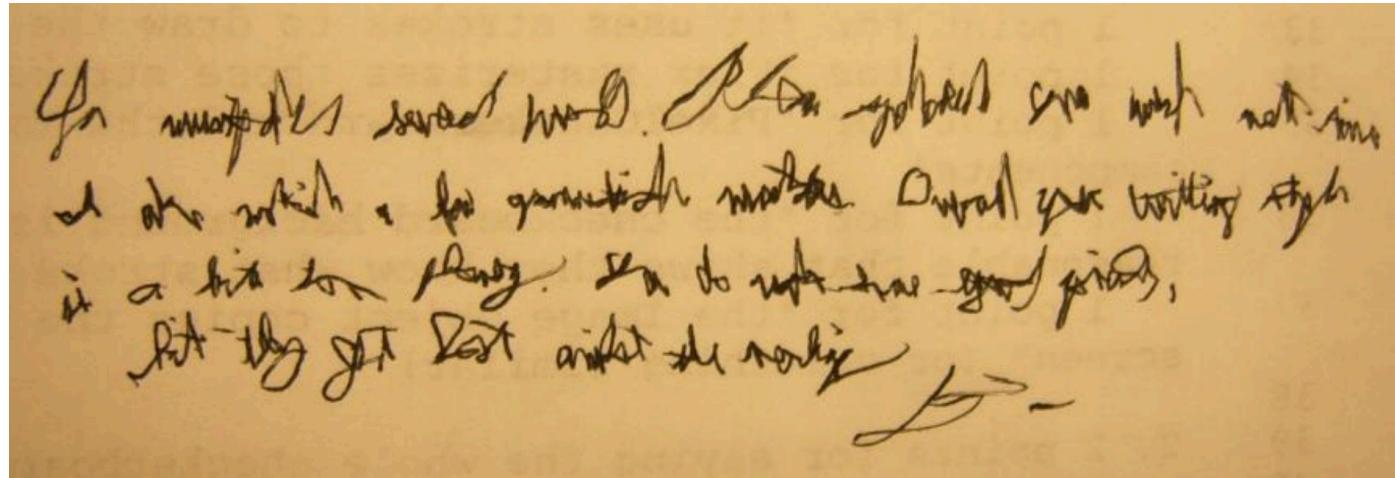
Designing Workflows [Lecture on Nov 3]



- Some workers are asked to perform **improvement** tasks
- Some workers are asked to **vote** on whether the improvement is good

Designing Workflows [Lecture on Nov 3]

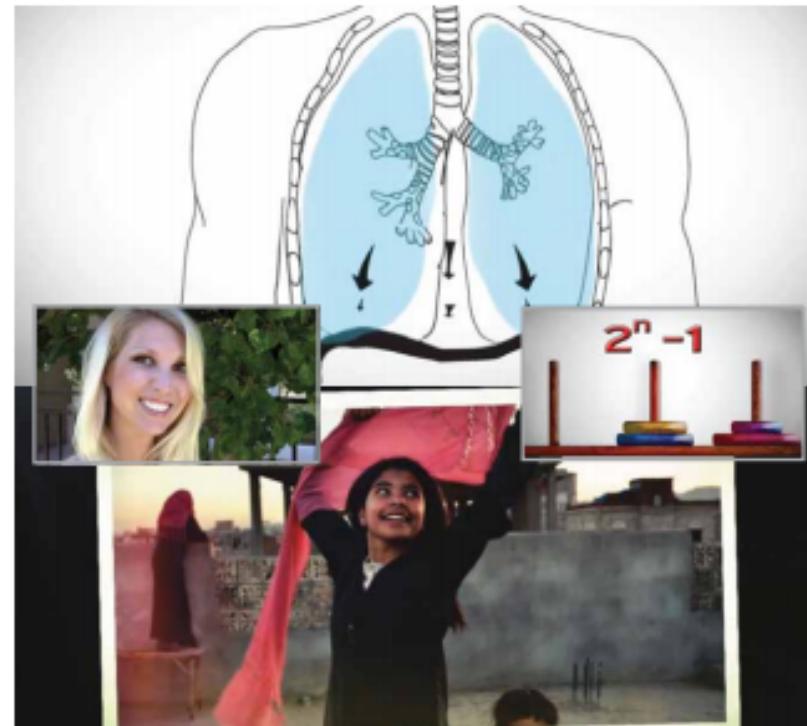
Version 6



You (misspelled) (several) (words). Please spellcheck your work next time. I also notice a few grammatical mistakes. Overall your writing style is a bit too phoney. You do make some good (points), but they got lost amidst the (writing). (signature)

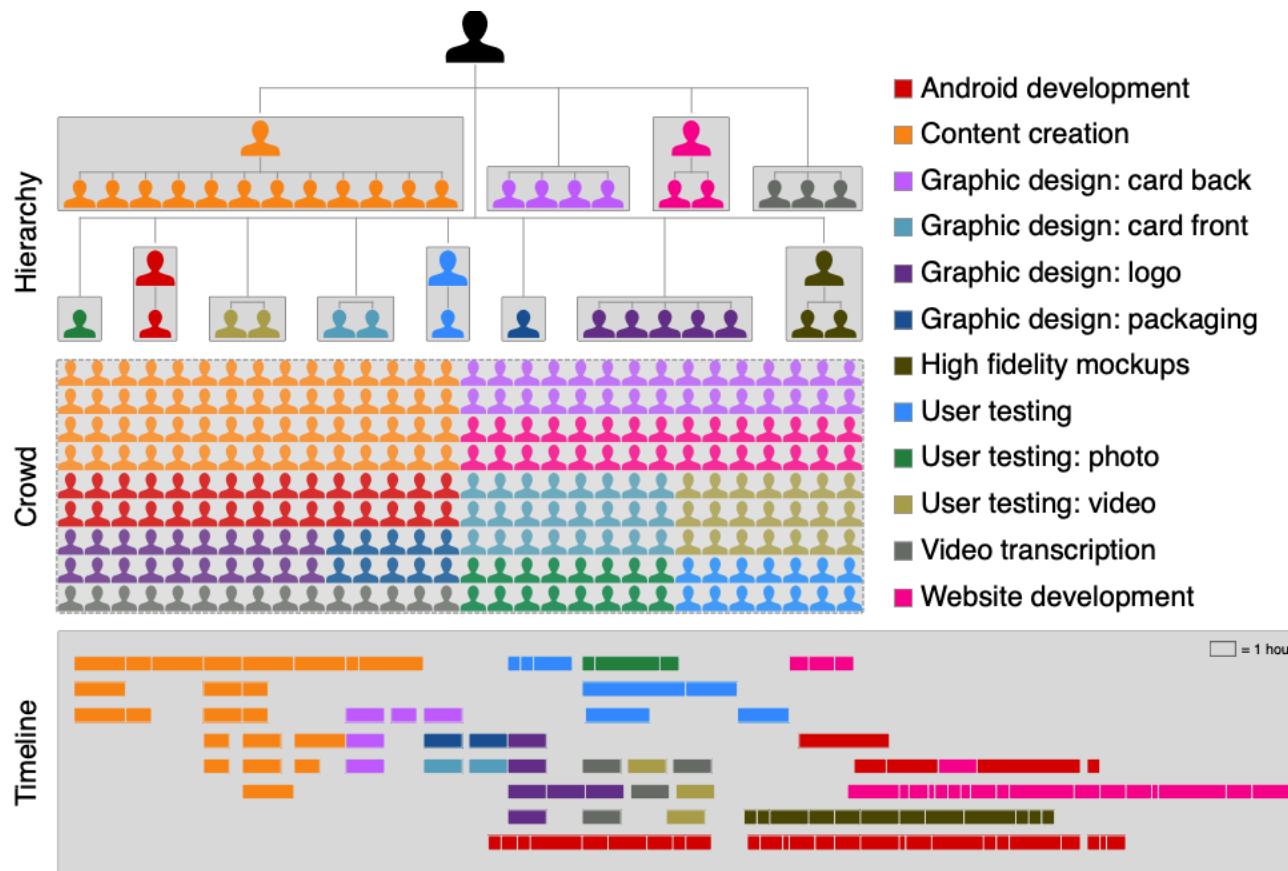
Expert Sourcing and Teams [Nov 5 Lecture]

- Task: Making an animation; crafting a course website; etc



Expert Sourcing and Teams [Nov 5 Lecture]

- Dynamically assembling a team of experts from crowd



MicroTalk: Using Argumentation to Improve Crowdsourcing Accuracy

Ryan Drapeau, Lydia B. Chilton, Jonathan Bragg, Daniel S. Weld. HCOMP 2016.

Complex vs Hard Tasks

- Complex tasks
 - Complexity of the task is high
 - Might decompose the tasks into reasonable small chunks
- Hard tasks
 - Hard to be correctly solved
 - hard to decompose



Claim: **Nicolas Sarkozy** "lived in" **France**

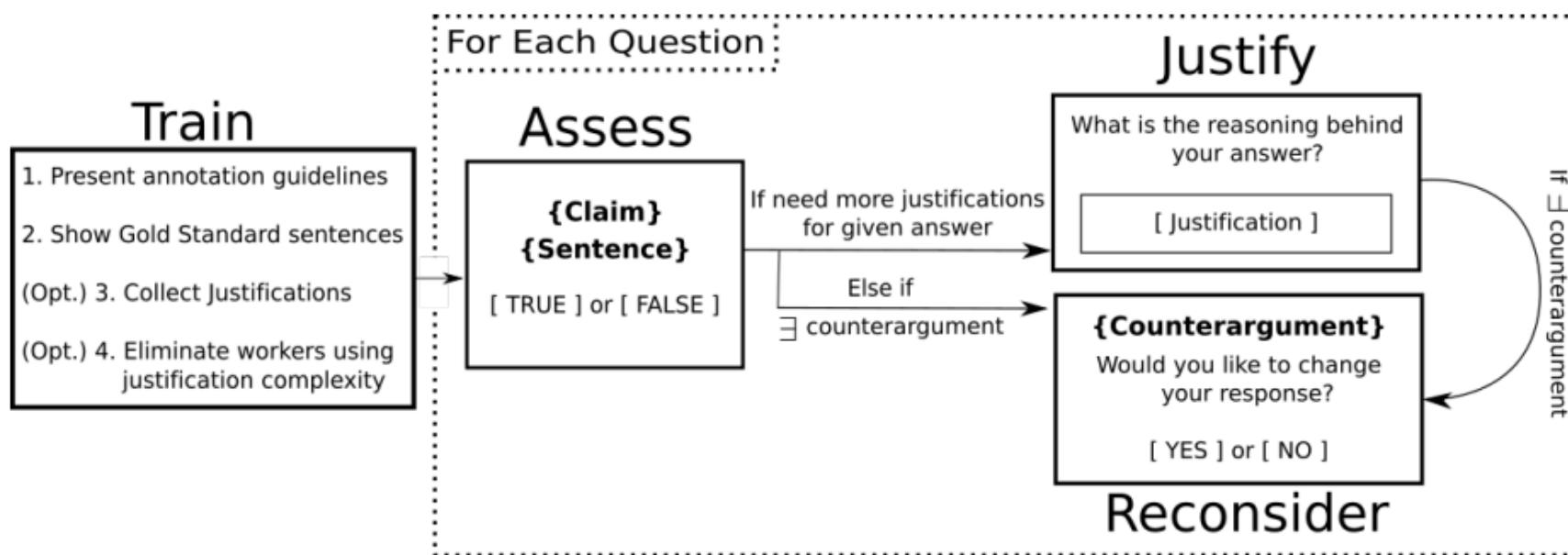
France is a Country.

Sentence: **Nicolas Sarkozy** leads Bastille Day celebrations, his first after being elected as **France**'s president.

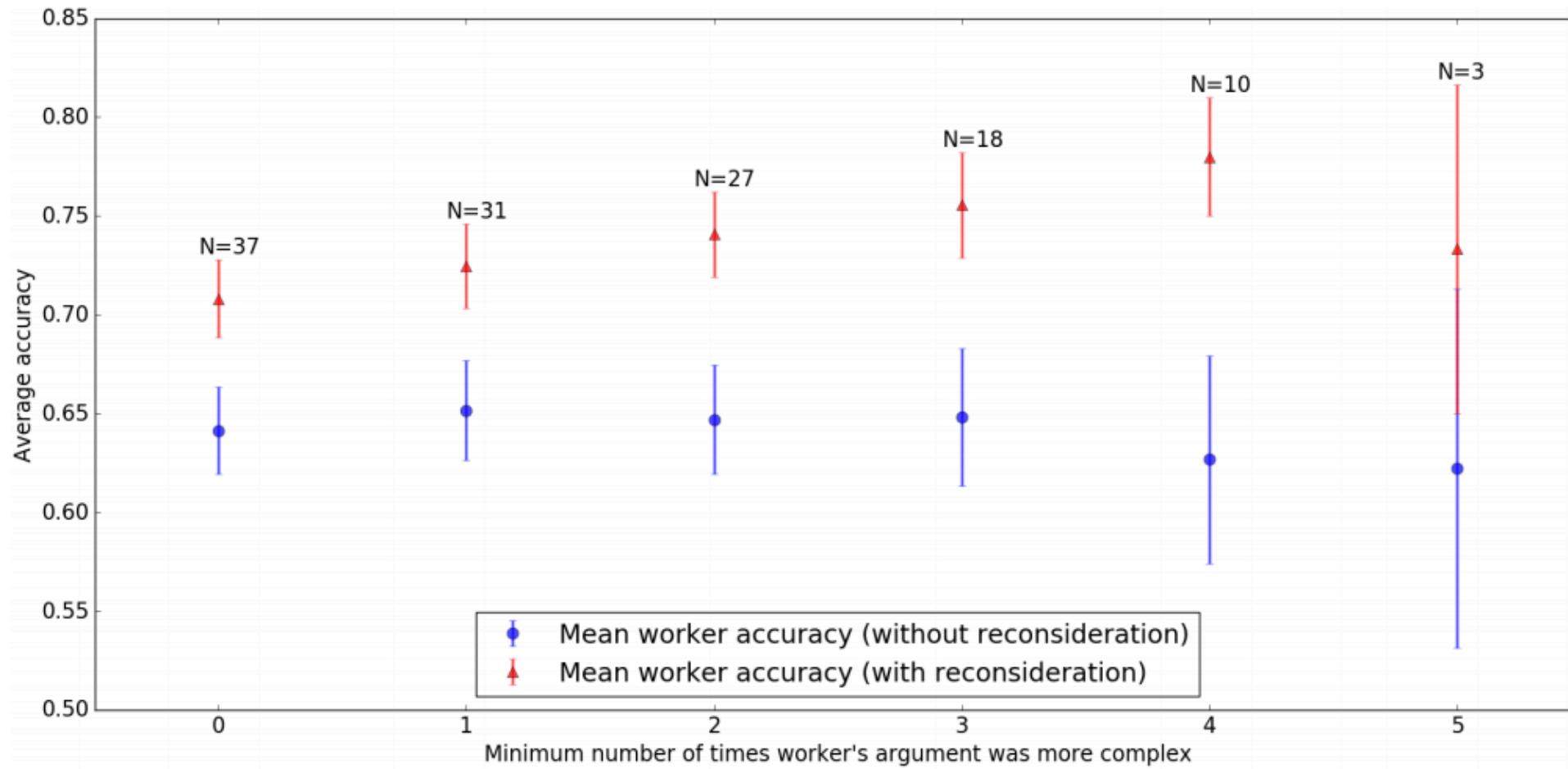
- True
- False

Dealing with Hard Tasks

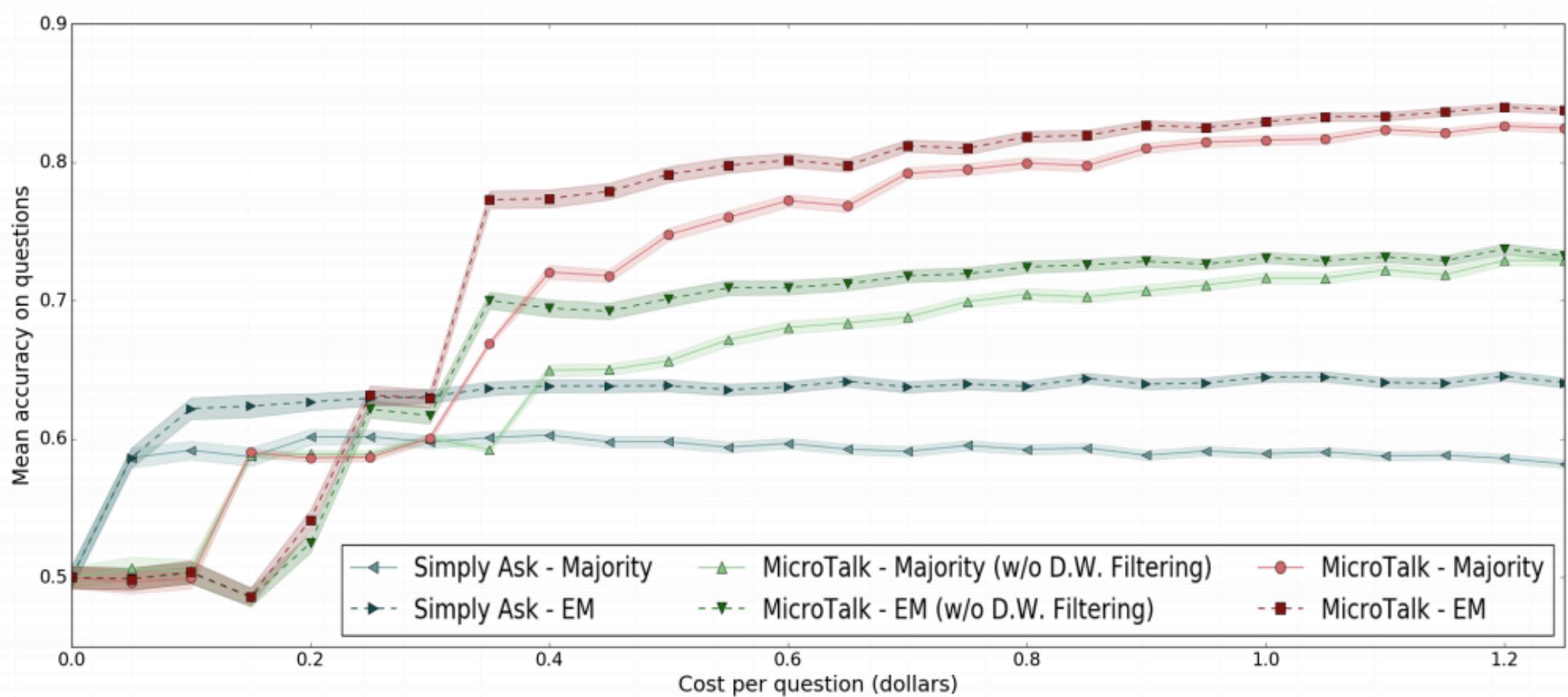
- Justification
 - Asking workers to provide rationale of their answers improve accuracy
- Reconsideration with counter-arguments
 - Enabling wisdom of crowd by presenting alternative explanations



Average Quality of Work



Aggregation Results Improves w.r.t. Cost

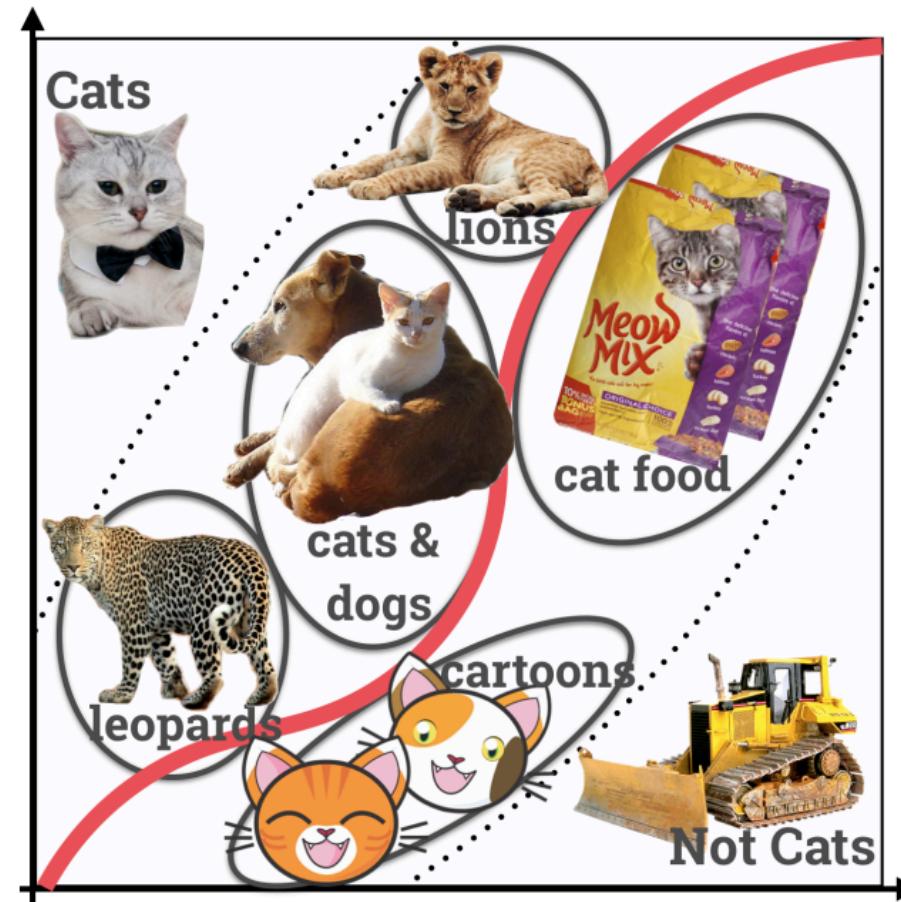


Side-Product of Justification

Revolt: Collaborative Crowdsourcing for Labeling Machine Learning Datasets.
Chang et al. CHI 2017.

Labels Might Have Different Interpretations

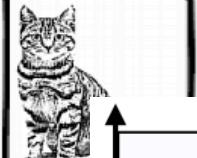
- Example: Label whether the image theme is "cat"



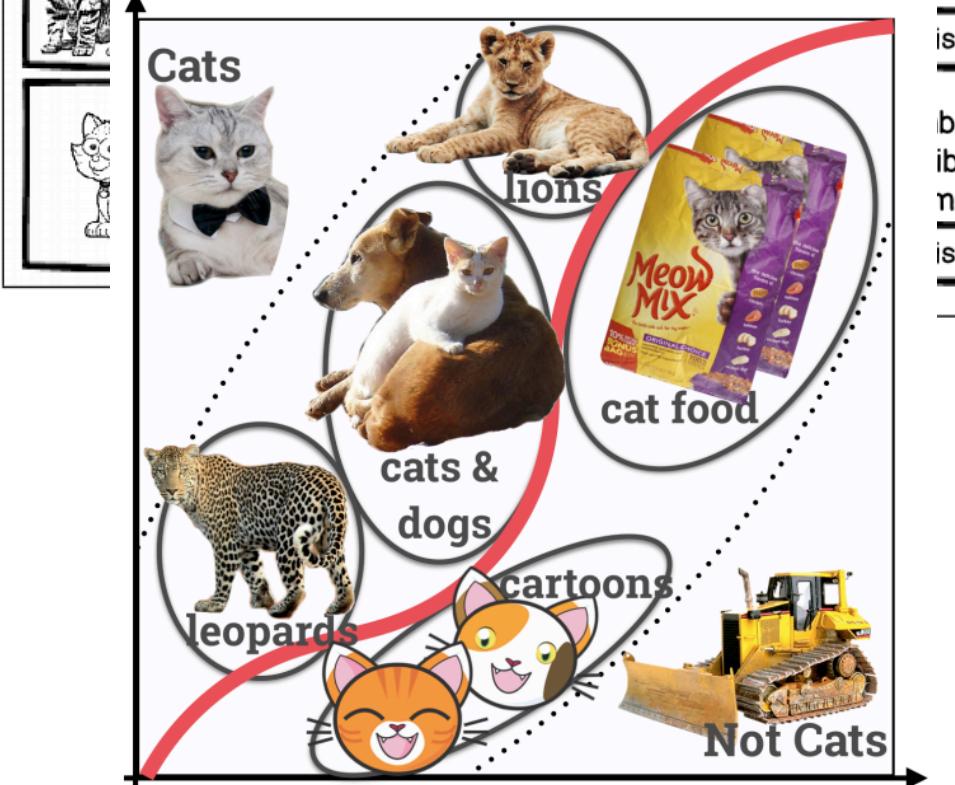
We want to know if the main theme of the items below are "Cats". Label "Cat" if you think the main theme of the item is Cats, otherwise label "Not Cat". Label "Maybe/Not Sure" for items that you are uncertain about or if you think other workers might pick different labels.



- Cat
- Not Cat
- Maybe/Not Sure



- Cat
- Not Cat



Assess

Justify

The other workers have also finished labeling the same items you just labeled. The following items received different labels. Please provide an explanation for each of your labels below.



You labeled "Not Cat". Please focus on describing things about the item that could have made it difficult or ambiguous for others.

is a tiger.

Save

beled "May
ibing things
made it diff
is a cartoon

You labeled differently on the following items. Please review all the explanations provided by other workers and pick or come up with good category names so the requesters can make an informed decision afterwards.



Create

- big cats
- cartoon cats
- cats with dogs

worker1: This is a tiger.

worker2: This is a big cat.

worker3: Do lions and other big cats



Create

- big cats
- cartoon cats
- cats with dogs

worker1: This is a cartoon drawing of a cat.

worker2: Cat drawing.

worker3: Do cartoon cats count?

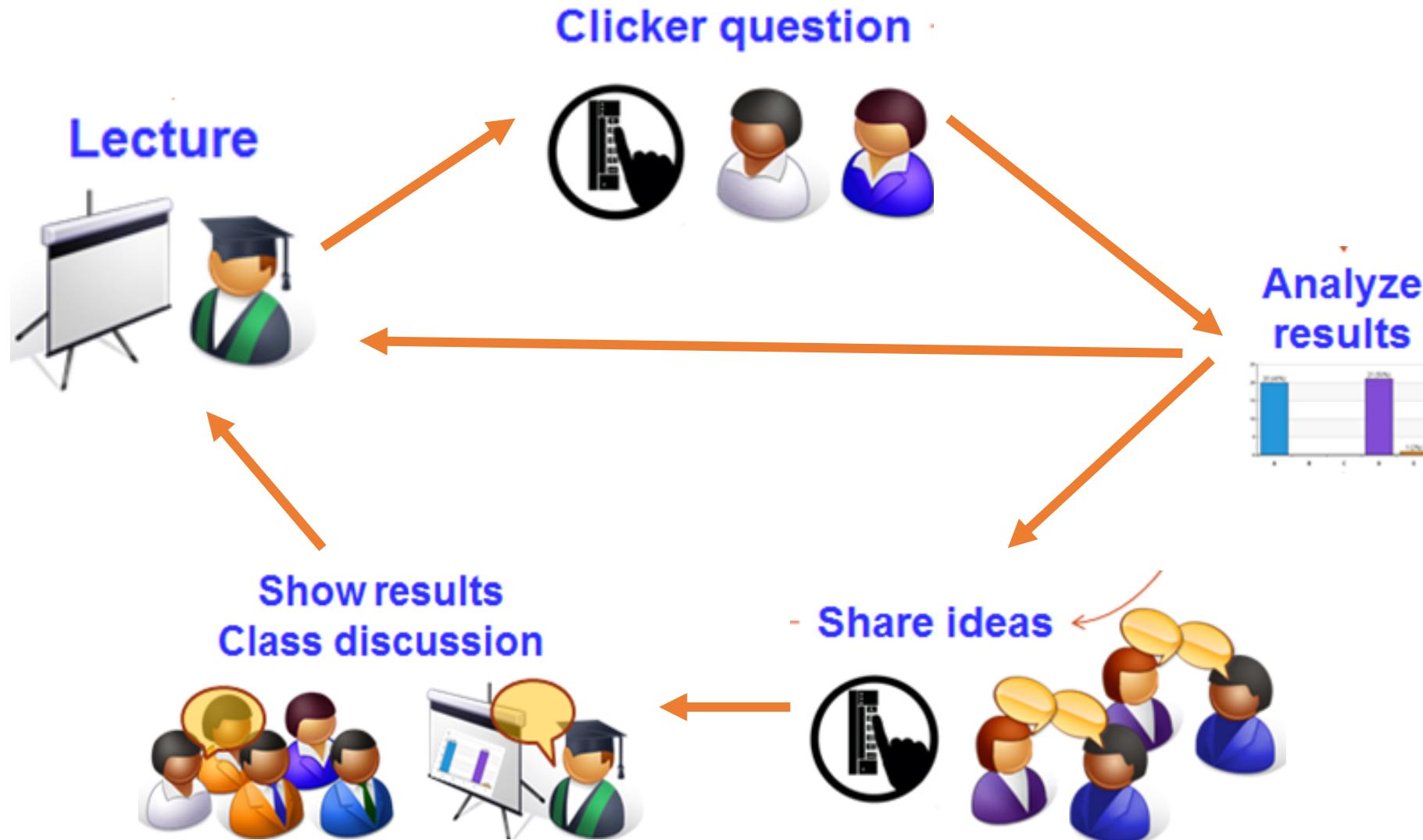
Categorize justifications => Future guidelines

Discussions

- What are the other types of **communications between workers** you think we can utilize to improve the quality of crowdwork?
- What are the benefits of the method you mentioned? What are the technical challenges? How do you implement it?
- Is enabling communication always good? What are the situations things might go south? Anyway to address that?

Peer Instruction

How Peer Instruction Works



Year	Method	FCI pre	FCI post	Absolute gain (post-pre)	Normalized gain $\langle g \rangle$	MBT	MBT quant. questions	<i>N</i>
Calculus-based								
1990	Traditional	(70%)	78%	8%	0.25	66%	62%	121
1991	PI	71%	85%	14%	0.49	72%	66%	177
1993	PI	70%	86%	16%	0.55	71%	68%	158
1994	PI	70%	88%	18%	0.59	76%	73%	216
1995	PI	67%	88%	21%	0.64	76%	71%	181
1996	PI	67%	89%	22%	0.68	74%	66%	153
1997	PI	67%	92%	25%	0.74	79%	73%	117
Algebra-based								
1998	PI	50%	83%	33%	0.65	68%	59%	246
1999	Traditional	(48%)	69%	21%	0.40	129
2000	PI	47%	80%	33%	0.63	66%	69%	126

^aThe FCI pretest was administered on the first day of class; in 1990 no pretest was given, so the average of the 1991–1994 pretest is listed. In 1995 the 30-question revised version was introduced (Ref. 15). In 1999 no pretest was given so the average of the 1998 and 2000 pretest is listed. The FCI posttest was administered after two months of instruction, except in 1998 and 1999, when it was administered the first week of the following semester to all students enrolled in the second-semester course (electricity and magnetism). The MBT was administered during the last week of the semester after all mechanics instruction had been completed. For years other than 1990 and 1999, scores are reported for matched samples for FCI pre- and posttest and MBT. No data are available for 1992 (EM was on sabbatical) and no MBT data are available for 1999.

Can we apply similar ideas in
crowdsourcing?

Leveraging Peer Communication to Enhance Crowdsourcing.
Wei Tang, Chien-Ju Ho, and Ming Yin. WWW 2019

Peer Instruction => Peer Communication

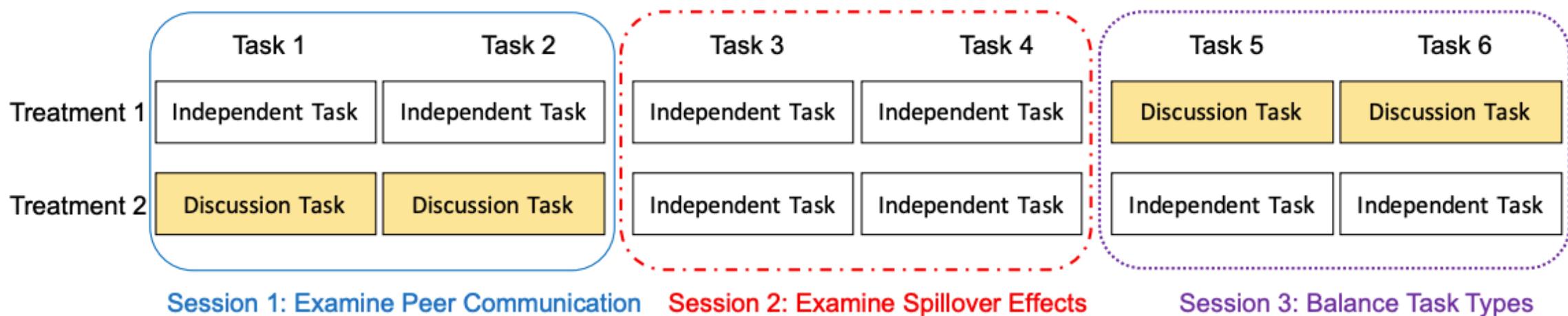
- Apply the idea in crowdsourcing
 - Recruit a pair of workers
 - Each complete and submit the task on his/her own
 - Discuss with another worker for 2 min
 - Decide whether to revise the answer
- Research Questions:
 - Does peer communication lead to higher quality of work?
 - Are there spill-over effects?

Experiment Design

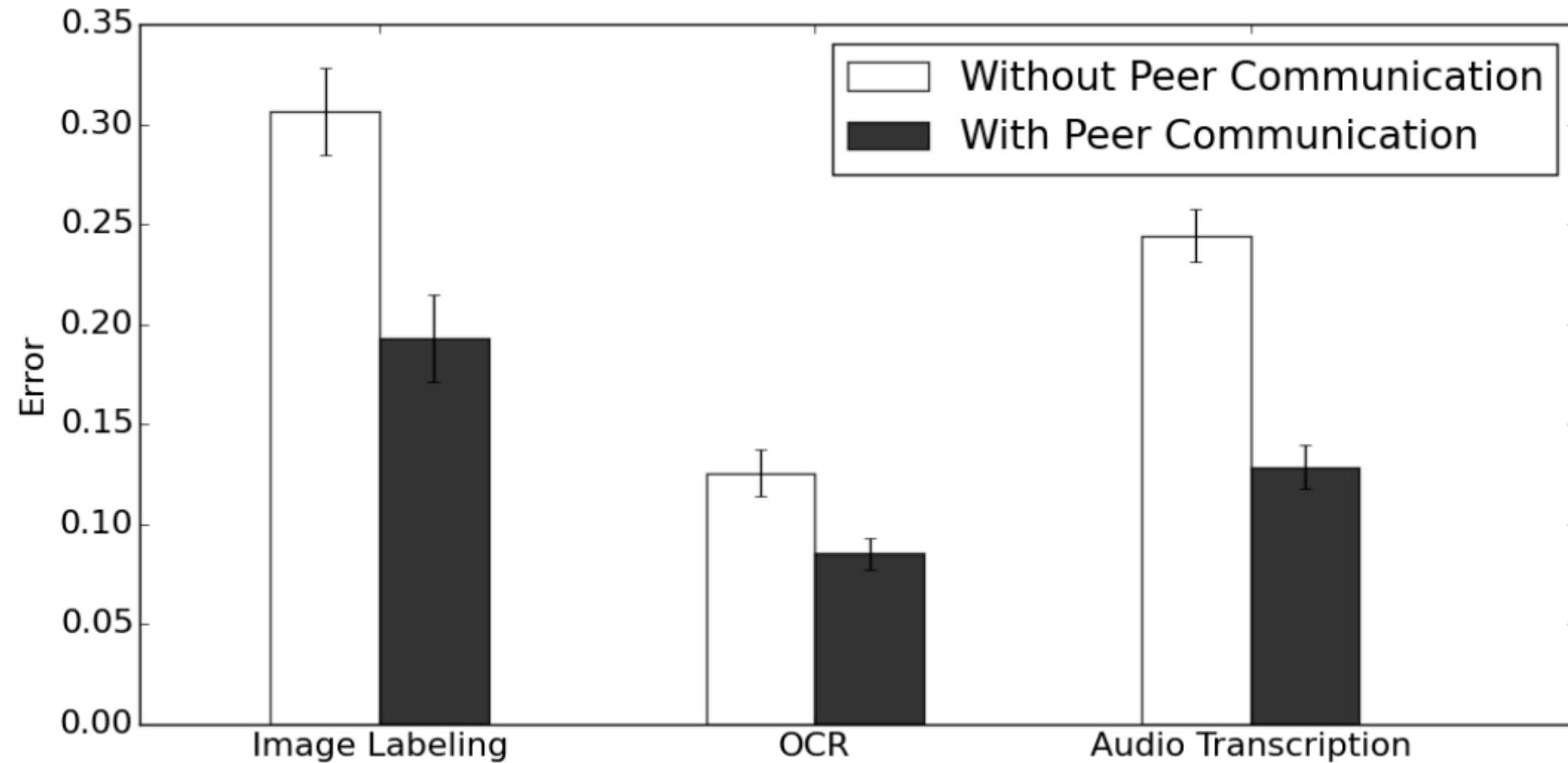
- Task Types:
 - Independent tasks: work individually
 - Discussion tasks: work with peer communication
- A naïve experiment design
 - Post two HITs
 - One with independent tasks
 - One with discussion tasks
 - Measure the end results
- What might go wrong?
 - Self selection (the worker population won't be the same across treatments)

Experiment Design

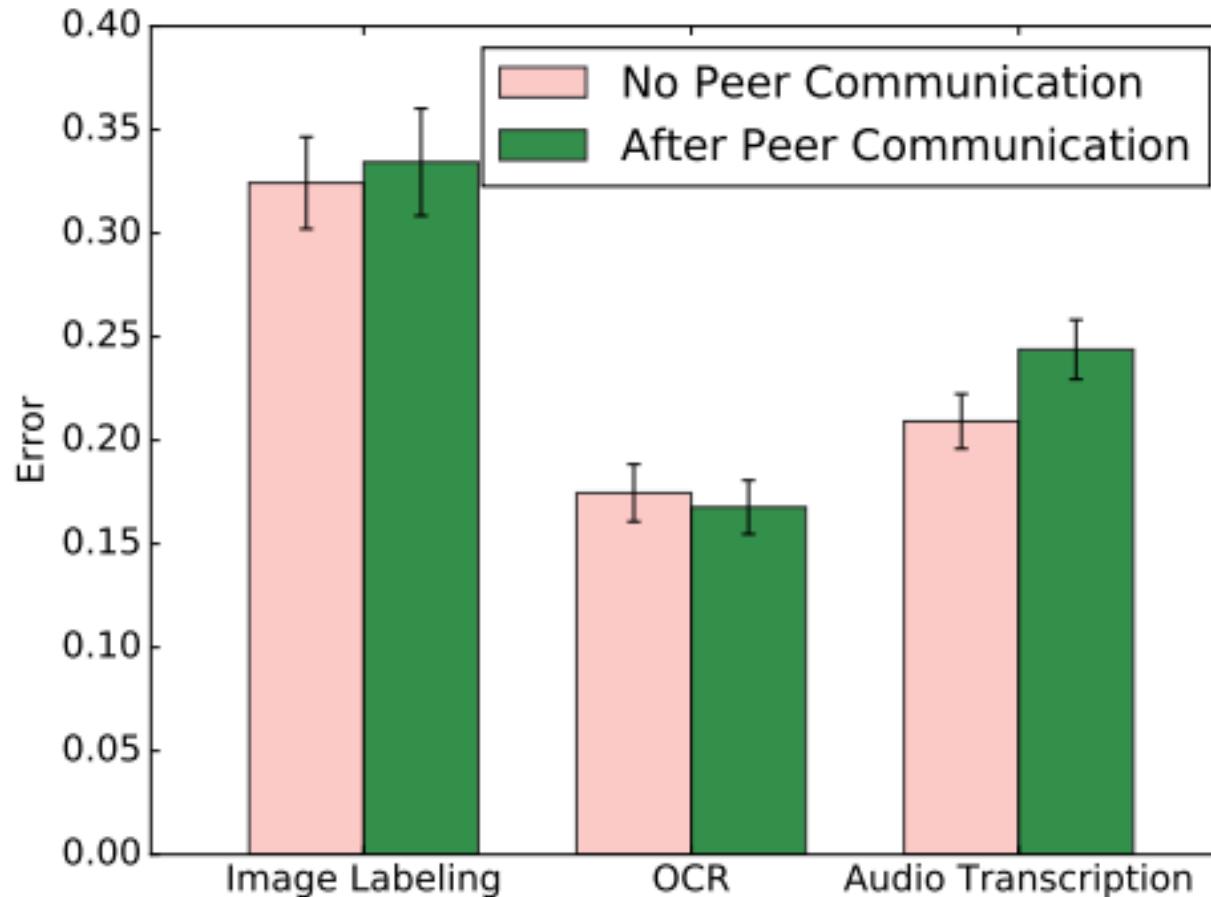
- Goal of our experiment design
 - Ensure there won't be self selection concerns
 - Be able to answer the two research questions
 - Q1: Does peer communication lead to higher quality of work?
 - Q2: Are there spill-over effects?



Peer Communication Leads to Better Work



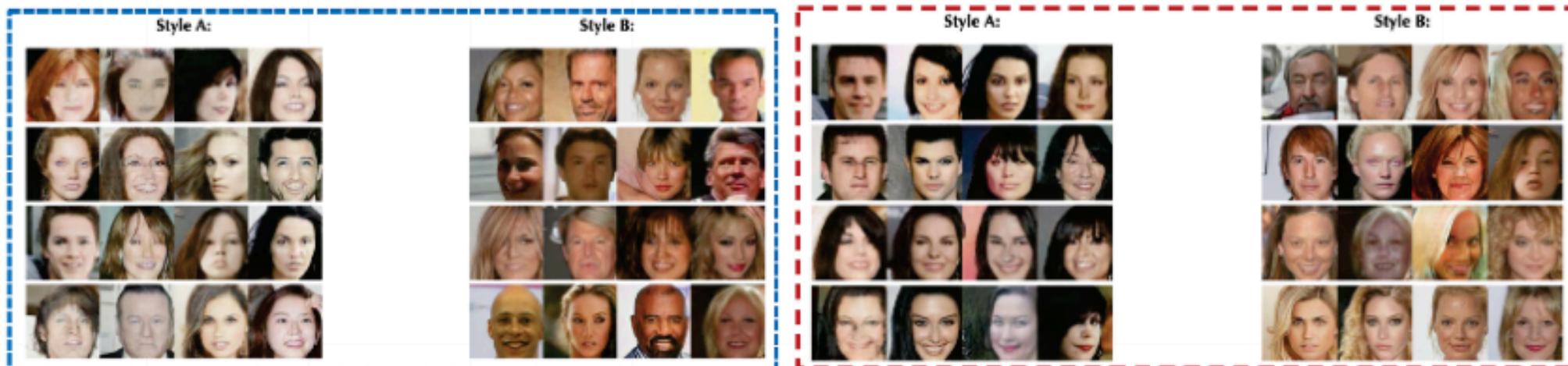
Long-term Performance Might Not Improve



- Might be due to the design
 - Need longer communication time?
 - Need feedback?

Does Diversity in Communication Help?

- Not that trivial [Duan et al. HCOMP 2020]
- Example:
 - Diversity in (artificial) skills



Training Type 1: Skin color salient

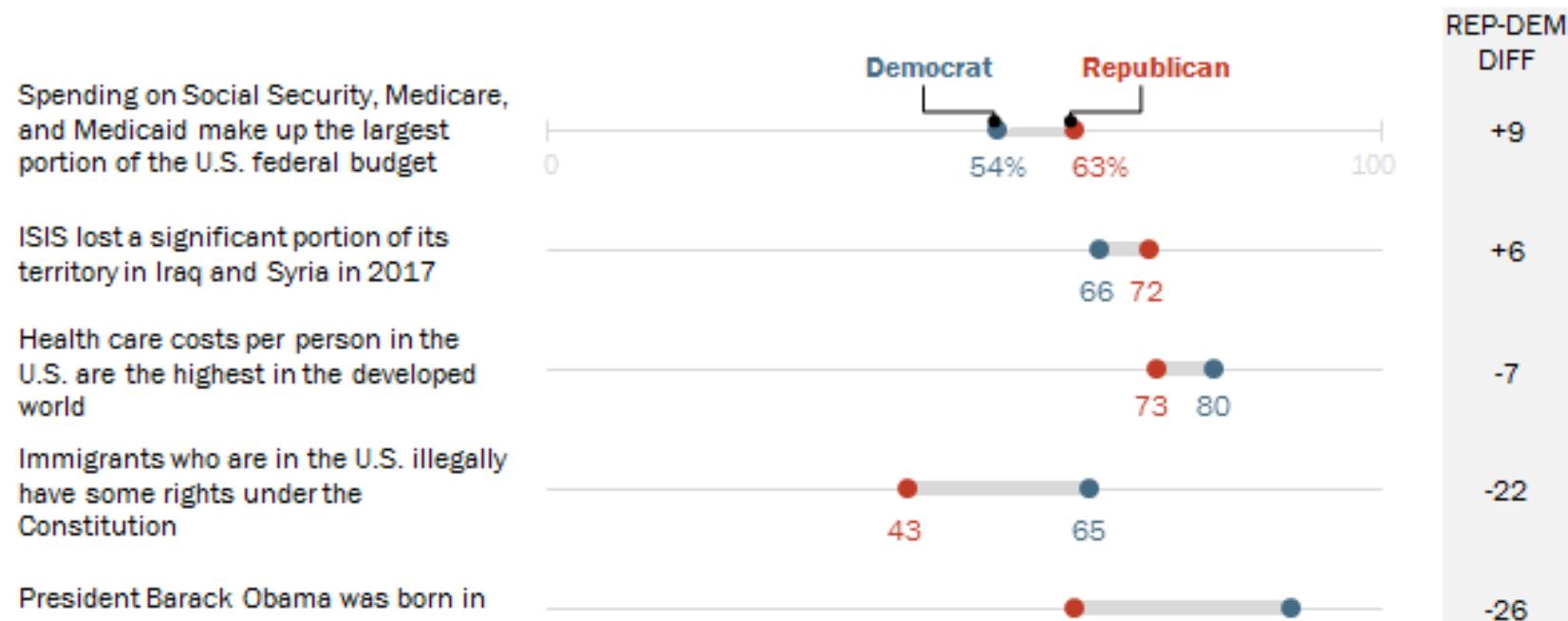
Training Type 2: Hair color salient

Does Diversity in Communication Help?

- Diversity in perspectives (Republicans/Democrats)
 - Opinions or factual statements

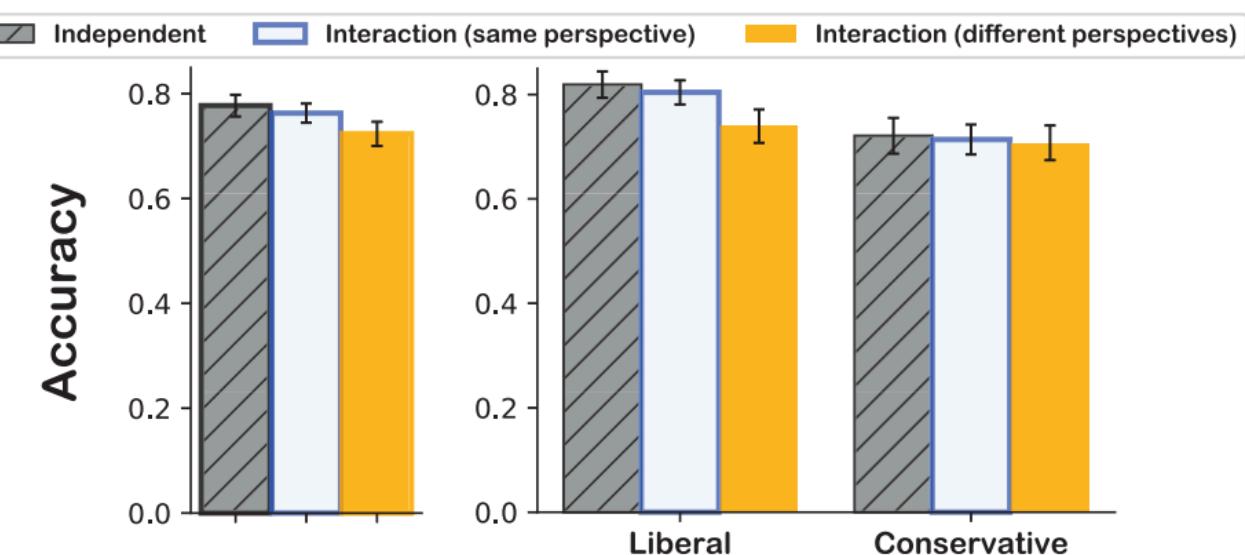
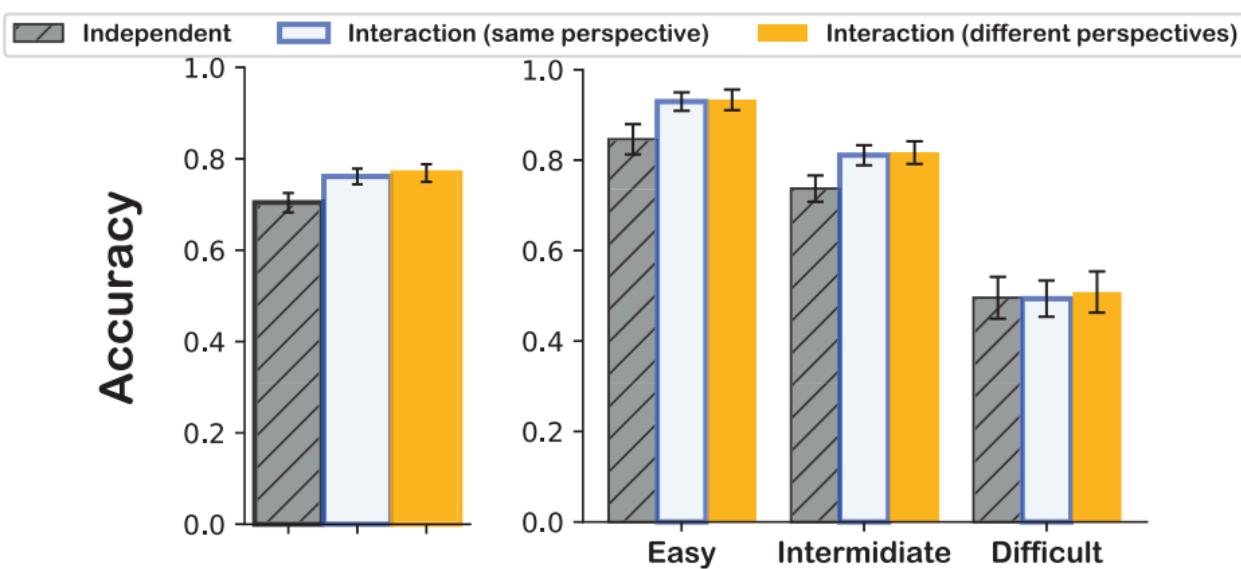
Republicans and Democrats more likely to see factual and opinion news statements as factual when they favor their side

% who classified each factual statement as factual



Does Diversity in Communication Help?

- Not that trivial [Duan et al. HCOMP 2020]
 - A couple of designs are tested; none lead to significant improvements
 - Could be that the designs happen to not work
 - Or could be due to that the “micro” properties of tasks in crowdsourcing



Peer Communication and Network



- Broader question
 - Peer communication is an edge (point to point) of a network
 - Design better network topology?
 - What do we mean by better
- What if we can't design network
 - Networks are already there
 - The influences are also there
 - What'll happen?

Discussion

- What can you do to **utilize** the network/communication to improve crowdsourcing or to achieve social good in general?
- What are the potential drawbacks given that humans are in the network and are influenced/biased by others (e.g., echo chambers)? What might we do to address that?

Hard but Important Questions

RESEARCH ARTICLE

Social learning and partisan bias in the interpretation of climate trends

Douglas Guilbeault,  Joshua Becker, and  Damon Centola

PNAS September 25, 2018 115 (39) 9714-9719; first published September 4, 2018;
<https://doi.org/10.1073/pnas.1722664115>

Edited by Matthew O. Jackson, Stanford University, Stanford, CA, and approved August 3, 2018 (received for review January 4, 2018)



RESEARCH ARTICLE

Exposure to opposing views on social media can increase political polarization

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky

PNAS September 11, 2018 115 (37) 9216-9221; first published August 28, 2018;
<https://doi.org/10.1073/pnas.1804840115>

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved August 9, 2018 (received for review March 20, 2018)



"we find that belief exchange in structured bipartisan networks **can significantly improve** the ability of both conservatives and liberals to interpret climate data, eliminating belief polarization."

"Republican participants expressed substantially more conservative views after following a liberal Twitter bot, whereas Democrats' attitudes became slightly more liberal after following a conservative Twitter bot"